TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC



PROCEEDINGS

www.tiptekno.net

TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC

Index

Foreword	**3**
Committees	**4**
Scientific Program	**5**
Papers	**12**

TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC

2025 Medical Technologies Congress, TIPTEKNO'25, was organized with the technical support of IEEE Turkey EMB Branch, Biomedical and Clinical Engineering Association (www.biyoklinikder.org), and Izmir University of Economics, Electrical and Electronics Engineering Department (eee.ieu.edu.tr) between October 26-28, 2025 in, Kaya Artemis Hotel, Gazi Magusa, TRNC, with international participation.

In our congress, users, manufacturers, researchers, managers and public representatives working in the field of medical technologies were brought together; scientific research results in the fields of biomedical engineering and clinical engineering were shared and common solutions to the problems experienced in the sector were discussed. As in the previous TIPTEKNO congresses, which were held with great success, special sessions were organized on current research topics, and problems in the development, production, use and management of medical technologies.

In addition, invited talks by leading scientists were presented on latest developments in Biomedical and Clinical Engineering areas.

Thank you for your scientific contribution and support to TIPTEKNO'25 Congress.

TIPTEKNO'25 Congress Organizing Committee
www.biyoklinikder.org

COMMITTEES

Congress Chair

Aydın AKAN

Organization Committee

Aydın AKAN
Levent EREN
Devrim ÜNAY
Mehmet TÜRKAN
Esra SAATÇI
Ozlem KARABIBER CURA

Technical Program Committee*

Abdurrahman Gumus / *Izmir Institute of Technology*
Ahmet Mert / *Bursa Technical University*
Ali Umit Keskin / *Yeditepe University*
Andac Hamamci / *Yeditepe University*
Atilla Özmen / *Kadir Has University*
Aydin Akan / *Izmir University of Economics*
Aysegul Guven / *Erciyes University*
Aysel Ersoy / *Istanbul University-Cerrahpasa*
Bülent Bolat / *Yildiz Technical University*
Burçak Alp / *İzmir University of Economics*
Cengiz Polat Uzunoglu / *Istanbul University-Cerrahpasa*
Cevriye Kalkandelen / *Istanbul University-Cerrahpasa*
Damla Gürkan Kuntalp / *Dokuz Eylul University*
Devrim Unay / *Izmir Democracy University*
Didem Şen Karaman / *Izmir Katip Celebi University*
Dilek Çökeliler / *Baskent University*
Dilek Goksel-Duru / *Turkish-German University*
Ertugrul Saatci / *Istanbul Kultur University*
Esin Öztürk Işık / *Boğaziçi University*
Esra Saatci / *Istanbul Kultur University*
Faezeh Yeganli / *İzmir University of Economics*
Fatih Uysal / *Kafkas University*
Fatma Latifoğlu / *Erciyes University*
Fatma Patlar Akbulut / *Istanbul Kültür University*
Feride Utku / *Yeditepe University*
Fırat Kacar / *Istanbul University*
Gizem Kaleli-Can / *Izmir Democracy University*
Gokhan Bilgin / *Yildiz Technical University*
Gulay Tohumoglu / *Dokuz Eylul University*
Gülizar Çalışkan Bilgin / *Izmir University of Economics*
Günnur Güler / *Izmir Institute of Technology*
Haşim Özgür Tabakoğlu / *Arel University*
Hazim Kemal Ekenel / *Istanbul Technical University*
Imran Goker / *Istanbul Arel University*
Kemal Polat / *Bolu Abant Izzet Baysal University*
Kemal Şerbetçi / *Izmir Democracy University*
Mahmut Ozturk / *Istanbul University-Cerrahpasa*
Mana Sezdi / *Istanbul University-Cerrahpasa*
Mehmet Aksahin / *Gazi University*
Mehmet Zubeyir Unlu / *Izmir Institute of Technology*
Mehmet Soylu / *Erciyes University*
Mehmet Akif Özçoban / *Gaziantep University*
Merih Palandoken / *Izmir Katip Celebi University*
Mesut Cevik / *Altinbas University*
Murat Okatan / *Istanbul Technical University*
Mustafa Sen / *Izmir Katip Celebi University*
Nermin Topaloglu / *Izmir Katip Celebi University*
Nizamettin Aydin / *Yildiz Technical University*
Olcay Akay / *Dokuz Eylul University*
Onan Guren / *Izmir Katip Celebi University*
Osman Doluca / *Izmir University of Economics*
Osman Erogul / *TOBB University of Economics and Technology*
Ovunc Polat / *Akdeniz University*
Ozan Karaman / *Izmir Katip Celebi University*
Ozgun Baser / *Izmir Katip Celebi University*
Pinar Özel / *Nevsehir Haci Bektas Veli University*
Savas Sahin / *Izmir Katip Celebi University*
Semra Icer / *Erciyes university*
Senay Mihcin / *Izmir Institute of Technology*
Tulay Yildirim / *Yildiz Technical University*
Umut Ceyhan / *Izmir Katip Celebi University*
Utku Kursat Ercan / *Izmir Katip Celebi University*
Yalcin Cekic / *Bahcesehir University*
Yalçın Albayrak / *Akdeniz University*
Ziya Telatar / *Ankara University*

**Names are sorted alphabetically by name.*

TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC

SCIENTIFIC PROGRAM

Scientific Program

Day 1 – Sunday, October 26, 2025

12.00-13.00	**Registration**
13.00-13.15	**Opening Speech** Prof. Dr. Aydın AKAN *(Congress Chair)* Doç.Dr. Esra SAATÇI *(Biyomedikal ve Klinik Müh. Derneği Başkanı)*
13.15-14.00	**Invited Talk – 1 Prof. Lotfi Chaari (Toulouse INP)** **"Image Denoising, From Inverse Problems to Deep Learning"**

	Hall-1	Hall-2
14.00-15.30	**Computer-Aided Diagnosis Systems-1** **Chair: Berrin Yanıkoğlu**	**Special Session: Advances in Assistive Technologies and Prosthetic Systems for Human-Centered Rehabilitation** **Chair: Elif Hocaoğlu**
1	**3-**Secure and Interpretable Dyslexia Detection Using Homomorphic Encryption and SHAP-Based Explanations **Mhd Raja Abou Harb, Barış Çeliktaş, and Günet Eroğlu**	**14-**Scolifix: Integrated Wearable Solution for Remote Scoliosis Rehabilitation **Fırat Matur, Gözde Hilal Öçalan, and Bora Büyüksaraç**
2	**4-**Simplifying Hydration Monitoring: Wearable Sensor for Accessible Dehydration Detection in Vulnerable Populations **Dilay Memiş Zeynep İlke Akbay, and Gülizar Çalışkan Bilgin**	**69-**Konuşma Engelliler için Yapay Zeka Tabanlı Duygusal Ses Üretim Sistemi **Efe Eren Yağcı, Eren Kalkan, and Güray Yılmaz**
3	**36-**Comparison of Leading Deep Learning Architectures on Skin Lesion Classification Using the ISIC 2019 Dataset **Anil Aktan, Alaa Almouradi, Erchan Aptoula and Berrin Yanikoglu**	**82-**Preliminary Evaluation of the Passive Rehabilitation Phase of the REACT System **Habibe Asran, Zikra Alfaris, and Elif Hocaoğlu**
4	**40-**PCA-Enhanced Deep Learning Models for Breast Cancer Detection: A Comparative Study **Ece Akaydın, Ayse Rüya Arslan, and Gülay Çiçek**	**85-**Development of Proprioceptive and Vibrotactile Training Modes for BALANSens Platform **Emad-ud-din Sheikh, Hafız Muhammad Bilal, Pınar Kaya, and Elif Hocaoğlu**
5	**64-**Determination Health Condition of Newborns Using Apgar Score and Thermal Camera Imaging **Farhad Nassehi, Doğa Kaya, Pelin Koçak, Emine Bilun Haskaraman, Seray Akbağ, and Osman Eroğul**	**88-**Wearable Soft Actuator for Restoring Finger Adduction–Abduction **Yazan Kanjo, Abed Alrahman Al Sayyed, and Elif Hocaoğlu**
6		**94-**A Hybrid Feature-Based CNN-BiLSTM Model for Real-Time Classification of Motor and Non-Motor Seizures from EEG **Kübra Kayan, Özgür Koray Şahingöz, and Ramin Rasi**
15.30-16.00	**Coffee Break**	**Coffee Break**
16.00-16.45	**Invited Talk – 2 Prof. Nihal Sarier (Istanbul Kultur Univ.)** **"Nanotechnology Shaping the Future of Medicine"**	

Scientific Program

	Hall-1	Hall-2
16.00-17.30	**Panel 1: Tıbbi Cihaz Üretiminde CE Belgelendirme Süreçleri** **Chair: Dr. Öğr. Üyesi Onur Koçak** CE MDR Regülasyonlarına Bakış (panel açılış konuşması & bilgilendirme)	**Biomedical Image Processing – 1** **Chair: Faezeh Yeganli**
1	**1.** Sınıf IIA – IIB – III cihazlarda Teknik Dosya Hazırlama ve Risk Yönetimi **Selen Koçak (Marla Teknoloji Müh. Kalite Güvence Uzmanı)**	**41-**Enhancing Fundus Image Quality for Improved Age-Related Macular Degeneration Detection using Deep Learning **Erdem Okur, Ceren Yılmaz, Sinem Yeşil, and Elif İlkay Özkan**
2	**2.** Tıbbi Cihaz Üretiminde Klinik Araştırma Tarafları ve Sorumlulukları **Ertan Köseoğlu (CRM / CRO Firma Sahibi)**	**56-** Integrating CLAHE and Image Enhancement with Deep Learning for Multi-Class Brain CT Analysis **Arda Şenyüz, Begüm Gençer, Miray Solmaz, Yakup Atıcı, Erdem Okur, Ali Salbaş, Ali Murat Koç**
3	**3.** İyi Klinik Uygulamaları **Dr. Evrim Köseoğlu (CRM / CRO Planlama Müdürü)**	**27-**A Multi-Chatbot Evaluation Framework for Knee MRI Diagnosis Assistance **Serhat Uzunbayir, Beyza Altuner, Deniz Eren Terziler, and Cinar Gedizlioglu**
4	**4.** Sağlık Hizmet Sunucularda Klinik Mühendislik Faaliyetleri **Uğur Bayraktar (PNN Eğitim ve Danışmanlık Hizmetleri Ltd. Şti. – EMO İstanbul Şubesi Biyomedikal MDK Üyesi)**	**26-**Olfactory Emotion Recognition Using EEG Spectral Topographic Heatmaps and CNNs **Faezeh Yeganli, Aydin Akan, and Rıza Sadıkzade**
5		**32-**Real-time 3D Human Pose Estimation Model for Physiotherapy Exercises on Mobile Devices **Hamdi Yalın Yalıç, Alaettin Uçan, Ali Yaşar Yiğit, and Adem Ali Yılmaz**
6		**74-** IoT-Enabled Camera-Based Eye Tracking and Data Analysis for Real-Time **Saadet Kılıç, Aytaç Durmaz, İrem Karagöz, Taha Çırpıcı, Berkay Uçar, Gökçer Eskikurt, Ezgi Ildırım, İrem Köçer, Cenk Efe Bıkmaz, and Aşkın Miraçhan Ergen**
7		**13-** Attention Augmented U-Net for Robust Thyroid Ultrasound Segmentation **Özgür Ömer Gülsoy, Mohammad Qusay Mawlood Mawlood, Mahsa Mikaeili, and Sina Alp**

	POSTER AREA
16.00-17.30	**Interactive Oral Presentations-1** **Chair: Kaan Kurtel**
	34- An AI-Supported Hospital Laundry Processing System (AIHLPS) **İbrahim Zincir, Kaan Kurtel, and Ufuk Celikkan**
	39- Navigating the Healthcare Maze through Natural Language: HAAN An AI-Based Health Tool **Tareq M F AlHammoodi, Kaan Özer, Hulki Enes Uysal, Erol, and İlker Korkmaz**
	43- Design of a Digital Content Framework with Gamification and Progress-Tracking for Mobile-Assisted Diabetes Self-Education **Göksel Okandan, A. Burak İnner, Ayaz Aktaş, Adnan Kavak, Alparslan Bozkurt, Sajjad Khan, Yeliz Demirhan, and Özgür Çakır**
	48- Portable Plasmonic Biosensor for Simultaneously Detection of Neisseria gonorrhoeae and Chlamydia trachomatis in Urine Samples **Furkan Kocer, Meryem Beyza Avcı, and Arif E. Çetin**
	76-A Comprehensive Decision Support System for Family Physicians **Emre Şimşek, Fatma Ezgi Erkat, İlker Korkmaz, and Ahmet Sermet Anagün**
	102- Comparative Error Analysis of 3d- Modeling, -Printing And -Scanning Methods Used in Orthodontics **Rabia Simal Selamet and Senol Piskin**
	108-Hemodynamic Effects on Acute Mesenteric Ischemia of Superior Mesenteric Artery **Hazal Azazi**
	111- From Arousal to Acceptance: A Review of EDA/GSR and BVP Applications in Food Consumer Research **Rıza Sadıkzade and K. Nazan Turhan**
	35-A Low-Cost Portable Vital Sign Monitoring System and Mobile Application Customized for Medical Device Designs **Arif Kocoglu, Halid Akkus, Sacid Erdogan, Cansel Ficici, Onur Kocak, and Ziya Telatar**
	37-Design of a Smart Infusion Pump with Decision Support **Melahat Kızıl, Berkay Tunakan, Nursun Üstünkarlı, and Aydın Akan**
	52-Design of an Intelligent Triage System Using Wearable Technologies **Nursun Üstünkarlı, Aydın Akan, Cansu Cozar, and Kadir Furkan Cozar**

Scientific Program

Day 2 – Monday, October 27, 2025

	Hall-1	Hall-2
09.00-10.15	**Graduate Student Workshop** **Chair: Ayça Kumluca Topallı**	**Biomaterials** **Chair: K. Nazan Turhan**
1	**21-** Smartphone Platform for Image-Based Cell Confluency Analysis **Meryem Beyza Avcı, Arif E. Çetin, and Mehmet Türkan**	**18-** Antibacterial Activity of Green Synthesized Silver Nanoparticles from Aloe Vera Plant **Imene Belhout, Hakan Solmaz, and Canan Bagci**
2	**46-** Tumor-Targeted Nanoparticle Delivery Prediction Using Machine Learning **Oguzhan Panatli, Açelya Yilmazer, and Gazi Erkan Bostancı**	**86-** Extraction and Characterization of Chitin and Chitosan from Shrimp Shell and Squid Pen Waste: Application in Biofilm Production **Duygu Geçkin, Gizem Ayna Duran, and Yaraslau Krotau**
3	**75-** Investigation of Effectiveness of Aspiration Technique in Simplified Artery with Clot Using Artificial Neural Network **Ali Bahadır Olcay, Arda Ataç, Burak Kalayoğlu, and Ezgi Işık**	**23-** Self-assembling Peptide Hydrogel Design for Enhanced Osseointegration and Vascularization of Dental Implants **Elif Deniz Toprakçı, Buse Sezer, and Günnur Pulat**
4	**81-**Evaluation of Thermal Damage Thresholds and Coagulation Behavior of 809 nm Diode Laser in Liver Tissue **Oğuzhan Efe Sapmaz and Ayşe Sena Sarp**	**92-** HbA1c Estimation with Machine Learning and Implementation on Diabetes Self-Management Platform Based on CGM Measurements **Ş. Can Tiryaki, Adnan Kavak, and A. Burak İnner**
5	**114-** Evaluation of No-Reference Image Quality Measurements for Medical Imaging **Dilek Özer**	
10.15-10.30	**Coffee Break**	**Coffee Break**
	Hall-1	
10.30-11.15	**Invited Talk – 3 Alex Tarchini (The MathWorks, Inc.)** **"Impacting Higher Education to Drive Societal Change: Leveraging MATLAB and AI for Enhanced Teaching"**	
	Hall-1	**Hall-2**
11.15-12.45	**Biomedical Signal Processing – 1** **Chair: Ziya Telatar**	**Biomedical Image Processing – 2** **Chair: Fatma Patlar Akbulut**
1	**30-** Tumor Classification via Subregion Combinations of Cortical L5PC6 Pyramidal Neurons **Zübeyir Özcan, İlknur Kayıkçıoğlu Bozkır, and Temel Kayıkçıoğlu**	**20-**Assessment of Oxidative Effects of Cold Atmospheric Plasma via Image Processing Techniques **Elçin Erkal, Mehmet Akif Ozdemir , Gizem Dilara Ozdemir, Utku Kursat Ercan, and Onan Guren**
2	**31-**A System Design for Scoliosis Detection **Cansel Fıçıcı, Kutay Yıldırım, İsmail Can Yılmaz, Arif Koçoğlu, Onur Koçak, and Ziya Telatar**	**66-** Asymmetric Attention-Enhanced Diffusion Models for Cardiac MRI Synthesis **Mertcan Özdemir and Osman Eroğul**
3	**44-**Transfer Learning Strategies for Leveraging Multi-Session and Multi-Subject Datasets to Improve sEMG Recognition **Nicolas Sicard and François Nougarou**	**83-** Benchmarking Deep Neural Networks for Lung Nodule Classification in LUNA25 **Muhammet Gökcan and Songül Varlı**
4	**65-**Giyilebilir Sistem Uygulamaları için Elektrokardiyogram İşaretinden Solunum Hızı Kestirim **Doğan Yörük, Surhan Bozkurt, Güleycan Dedecengiz Varol, and Umut Engin Ayten**	**90-** Classifying Age-Specific Stages of Cervical Vertebral Maturation Using Deep Feature Analysis **Banu Akkuş, Shireen Al-Ofary, Hamza Osman İlhan, and Yasin Yaşa**
5	**68-**Enhanced External Counterpulsation (EECP) Therapy System Supported by Synchronous Vital Monitoring **Onur Kocak, Cansel Ficici, and Ziya Telatar**	**91-** Patch-Based Deep Learning Approaches for Automated 3D Skull Segmentation in CT Imaging **Fatih Ekrem Onat, Yiğit Kahraman, Berk Aridağ, Melih Fırat Budak, Caner Bulut, Berke Apaydinli, Abdulkadir Günay, Yunus Emre Çakmaklı, Nuri Serdar Baş, Hamza Osman İlhan, Görkem Serbes, and Mihrigül Ekşi Altan**
6	**70-** Few-Shot Medical Image Segmentation with Meta-Learning: An Empirical Study on Organ Segmentation **Ismail EL JAMIY, Amel Aissaoui, and Lotfi Chaari**	**96-**Ensemble and Transfer Learning Baselines for Dermoscopic Skin Cancer Classification: A Comparative Study **Yiğit Aydoğan and Arda Canbaş**
12.45-14.00	**Lunch Break**	**Lunch Break**
	Hall-1	
14.00-14.45	**Invited Talk – 4** **Mauro Costagli (University of Genoa)** **"Quantitative Susceptibility Mapping: A Tool to Investigate Brain Anatomy and Function with Magnetic Resonance Imaging"**	

Scientific Program

	Hall-1	Hall-2
15.00-16.15	**Panel 2: Günümüz Teknolojileri Çerçevesinde Biyomedikal Mühendisliği Eğitimine Bakış** **Chair: Prof. Dr. Aydın Akan** (İzmir Ekonomi Üniversitesi – Elektrik & Elektronik Mühendisliği Böl. Bşk.) Teknolojik Gelişmeler Işığında Mühendislik Eğitimi Bakış(Panel Açılış Konuşması & Bilgilendirme)	**Diagnosis and Treatment Systems** **Chair: Kutluhan Erol**
1	**1.** Uygulamalı Biyomedikal Mühendisliği Eğitim Modeli **Prof. Dr. Ziya Telatar - (Başkent Üniversitesi Biyomedikal Müh. Böl. Bşk.)**	**33-**Enhanced Tumor Classification via Radiomics and Deep Feature Generation using Neural Network **Dilek Betül Sarıdede, Rukiye Nur Kaçmaz, and Sevim Cengiz**
2	**2.** Günümüz Teknolojilerinin Biyomedikal Mühendisliği Eğitimine Konum ve Katkıları **Prof. Dr. Rasime Uyguroğlu - (Doğu Akdeniz Üniversitesi Elektrik Elektronik Müh. Böl. Bşk.)**	**45-**A Focused Survey on Patient Simulation with Large Language Models **Selin Katarcı and Ayça Kumluca Topallı**
3	**3.** Üniversite & Sanayi İşbirliği Uygulamalarında Biyomedikal Mühendisliği **Dr. Öğr. Üyesi Aytaç Durmaz (İstinye Üniversitesi Biyomedikal Müh. Böl. . – PieVision Teknoloji Yazılım A.Ş.)**	**49-**Knowledge Graph Augmented Retrieval Applications in Healthcare **Anıl Çoğalan and Ayça Kumluca Topallı**
4	**4.** Yapay Zeka ve Biyomedikal **Dr. Öğr. Üyesi Meliz Yuvalı - (Yakın Doğu Üniversitesi Tıp Fakültesi - Bioistatistik Bölümü Öğr. Üyesi)**	**67-**Breast Cancer Stage Classification with Pyramid Vision Transformer and StyleGAN3-Based Synthetic Data Augmentation **Begüm Şener**
5		**103-** A Unified Pipeline for Consistent MultiLabel Mask Generation in Pediatric Cardiac Segmentation **Omer Faruk Sahin, Beyza Nur Yazıcı, Ruqaiyah Mirza, Musab Ahmed Khan Umair, Mohamad Humam Alzaeim, Irem Yilmaz, Fatmanur Aydogdi, Doruk Günveren, Fatma Zeynep Gungoren, Cansaran Tanıdır, Okan Yildiz, Kevser Banu Köse, JS Femilda Josephin, Ibrahim Faress, Şenol Pişkin and Sakip Önder**
6		**112-**Privacy-Aware Mobile Emergency Response System Design **Melahat Kızıl, Zeynep Şişli, Çınar Gedizlioğlu, and Kutluhan Erol**
16.15-16.30	**Coffee Break**	

	Hall-1	Hall-2
16.30-18.00	**Panel 3: Tıbbi Cihaz Sektörü İhtiyaçlarına Bakış** **Chair: Prof. Dr. Ziya Telatar** (Başkent Üniversitesi Biyomedikal Müh. Böl. Bşk.) Tıbbi Cihaz Sektöründe Uzun Dönem Staj ve Bitirme Projesi Uygulamaları (panel açılış konuşması & bilgilendirme)	**Biomedical Signal Processing – 2** **Chair: Ertuğrul Saatçı**
1	**1.** KKTC'de Elektrik Mühendisleri Odası Faaliyetleri Modeli **Prof. Dr. Mehmet Toycan (Bahçeşehir Kıbrıs Üniversitesi Rektörü & KKTC EMO Başkanı)**	**71-** Interaction-Aware Interpersonal Graph Neural Network (IA-IGNN) for Stress Detection with Biosignals **İrem Aksoy and Fatma Patlar Akbulut**
2	**2.** Elektrik Mühendisleri Odası'nda Biyomedikal Mühendisliği Meslek Alanı Modeli **Ramazan Pektaş (EMO 49. Dönem Y.K. Üyesi)**	**72-** Evaluation of a Telemonitoring Interface for Neurodegenerative Disorders through Wearable Biosensors **Vael Kokach, Büşra Kocaçınar, Çilen Tavşanoğlu, Uğur Ovacık, Feray Güngör, Kübra Nur Menengiç, Fatma Patlar Akbulut, Ela Tarakcı, and Akhan Akbulut**
3	**3.** Nano Çağda Biyomedikal Mühendisliği Eğitimi"; Geleceğin Sağlık Teknolojilerini Yetiştirmek: Nanoteknolojinin Eğitimdeki Rolü Modeli **Prof. Dr. Doğa Kavaz (Kuzey Kıbrıs Türk Cumhuriyeti Sağlık Bakanlığı Devlet Laboratuvarı Dairesi)**	**84-** Design and Implementation of a Phonocardiography System Combining ECG and Stethoscope Features **Ziya Telatar, Onur Kocak, and Cansel Fıçıcı**
4	**4.** Tıbbi Cihaz Üretimine Akademik & Sektörel Bakış Modeli **Dr. Öğr. Üyesi Onur Koçak (Başkent Üniversitesi Biyomedikal Müh. Böl. Bşk. Yrd. – İBÜTEM ArGe Ltd. Şti. – EMO MEDAK Üyesi)**	**89-** Detection of Sleep Apnea from Single-Lead Electrocardiogram (ECG) Signals **Sona Semsari Parapari and Esra Saatçi**
5	**5.** K. Kıbrıs'da Biyomedikal Mühendislik Hizmetleri Modeli **Dr. Öğr. Üyesi Ali Işın - Kuzey Kıbrıs Türk Cumhuriyeti Sağlık Bakanlığı Biyomedikal Birim Sorumlusu & Yakın Doğu Üniversitesi Yarı Zamanlı Öğr. Gör.)**	**104-** Volume Control in a Hybrid Mock Circuit of the Cardiovascular System using Command Filtered Backstepping Approach **Ahmed Alhajyounis, Yaren Kuzucu, Fatih Adıgüzel, and Kamuran Kadıpaşaoğlu**
6		**115-** Altered Salience-Network Functional Connectivity in Depression **Mehmet Akif Özçoban**
7		**116-** Optimum Yük Kapasitesinin Yüzey Elektromiyografisi ve Borg Ölçeği'ne Bağlı Olarak Modellenmesi **Ayşe Kesgin and Alpaslan Koç**

Scientific Program

POSTER AREA
Interactive Oral Presentations-2 **Chair: Erdem Okur**
15-Functional Brain Network Analysis in Dyslexia Using Graph Theory Metrics **Özlem Karabiber Cura, Sevde Dervişoğlu, and Günet Eroğlu**
16-Determination of the Effect of Anticancer Laser Therapies on the Cancer Cell Migration and Invasion via Artificial Intelligence Methods **Sümeyra Arslan, Özlem Karabiber Cura, Büsra Sirek, and Nermin Topaloğlu Avşar**
24-Classification of Emotions Under Multiple Olfactory Stimuli Using EEG Signals and Machine Learning **Özge Ada Bozbaş, Burak Akbuğday, Sude Pehlivan Akbugday, and Aydın Akan**
29-Ion-Dependent Behavior of the Hodgkin-Huxley Neuron Model under Spatial Constraints **Zübeyir ÖZCAN, İlknur Kayıkçıoğlu Bozkır, and Temel Kayıkçıoğlu**
42-Analysis of Dementia EEG Signals Using Empirical Mode Decomposition Variants and Deep Learning **Yahya Şenol, Özlem Karabiber Cura, and Aydin Akan**
51-Artificial Intelligence Based Post-Disaster Trauma Analysis and Prioritization System **Nimet Seyrek, Miray Solmaz, Ceren Biçici, Rümeysa Tuntaş, Levent Arı, and Aydin Akan**
106-Portable Smartphone-Integrated Incubator for Real-Time Cell Culture Monitoring **Ali Avcı, Bilge Can, Meryem Beyza Avci, and Arif E. Cetin**

Scientific Program

Day 3 – Tuesday, October 28, 2025

	Hall-1	Hall-2
09.00-10.00	**Biomedical Image Processing – 3** **Chair:** Aytaç Durmaz	**Other Topics in Biomedical Engineering – 1** **Chair:** Surhan Bozkurt
1	**98-** D U-Net-Based Deep Learning Approach for Pediatric Cardiovascular MRI Segmentation **Musab Ahmed Khan Umair, Omer Faruk Sahin, Irem Yilmaz, Mohamad Humam Alzaeim, Ahmed Boray, Ruqaiyah Mirza, Beyza Nur Yazıcı, Ayham Pezri, Ibrahim Faress, Nedim Yildirim, Cansaran Tanıdır, Fatma Zeynep Gungoren, Kevser Banu Köse, JS Femilda Josephin, and Senol Pişkin**	**10-** Optimized Radix Sort for Efficient Health Data Sorting in Biomedical Backend Systems **Tugay Direk**
2	**99-** Structure-Aware V-Net Framework for 3D Multi-Class Segmentation of Pediatric Cardiovascular Structures **Irem Yilmaz, Omer Faruk Sahin, Musab Ahmed Khan Umair, Beyza Nur Yazıcı, Mohamad Humam Alzaeim, Ruqaiyah Mirza, Doruk Gunveren, Fatmanur Aydogdi, Kevser Banu Kose, JS Femilda Josephin, Fatma Zeynep Gungoren, Cansaran Tanıdır, Okan Yildiz, Senol Pişkin and Sakip Önder**	**63-** Autononym: Multimodal Anonymization of Health Data using Named Entity Recognition and Structured Medical Data Processing **Hamdi Yalın Yalıç, Murat Dörterler, Alaettin Uçan, Ali Yaşar Yiğit, and Adem Ali Yılmaz**
3	**101-** Deep Learning Framework for B-Mode Ultrasound Image Reconstruction **Mahsa Mikaeili and Hasan Şakir Bilge**	**87-** Biyomedikal Mühendislikte Üretim ve Regülasyonlar: Mevzuat, Kalite ve Gelecek Perspektifleri **Selen Koçak, Onur Koçak, Arif Koçoğlu, and Ziya Telatar**
4	**113-**AI Assisted Polyp Detection and Segmentation in Colonoscopy **Değer Ayata**	**97-** Multi-modal Biosignal and Media Content Fusion for Human Interest Prediction Using HGNNs **Teoman Berkay Ayaz and Akhan Akbulut**
5	**80-**Kalp Üfürümlerinin Erken Tespiti Için Yenilikçi Yapay Zeka Çözümleri – Innovative AI Solutions For Early Detection Of Heart Murmurs **Asma ALTAIB, Hayati Türe, and Muhammed Aslan**	**62-**Graph Neural Networks-Based Digital Twin Modeling of WBC and ANC Dynamics for Personalized Time-Series Prediction **Utku Köse, Gözde Özkan Tukel, Emine Betul Surucu, Muhammet Enes Karatas, and Ilhan Uysal**
10.00-10.15	**Coffee Break**	
	Hall-1	**Hall-2**
10.15-12.30	**Biomechanics** **Chair:** Serhat Uzunbayır	**Medical Device Design and Implementation** **Chair:** Lale Canan Dülger
1	**1-** Doku Takip İşlemlerinde Formaldehit Fiksatifinin Otomatik Üretim Sisteminde Tamponlanarak Kullanılmasının Atık Üretimi Üzerine Etkilerinin İncelenmesi **Habib Serdar Aday**	**47-** Actuation and Sensor Issues: Portable CPR Machine **Lale Canan Dülger, Melahat Kızıl, Begüm Karaaslan, Tahsinefe Gülek, Doğa Eroğlu, Mustafa Güven Güneş, and Anıl Uzşen**
2	**2-** Tıbbi Patoloji Laboratuvarında Otomatik Takip Sistemlerinde Etanol ve Ksilen Geri Kazanım Uygulamasında Verim Değerlendirilmesi **Habib Serdar Aday**	**73-**Standardized Surgical Guide for Femoral Osteotomy **Manar Edris, Fuat Bilgili, and Hayati Durmaz**
3	**8-** Evaluating Respiratory Effects of 3D-Printed Scoliosis Bracing in Different Physiological States **Büşra Sağlam and Samet Çıklaçandır**	**77-**An Extensible Cloud Platform for Integrating Heterogeneous IoT Devices in Physiological and Behavioral Research **Berat Bakı and Fevzi Aytaç Durmaz**
4	**9-** Force Mapping of 3D-Printed Scoliosis Braces at Varying Tightness Levels Using Calibrated FSR Sensors **Büşra Sağlam and Samet Çıklaçandır**	**100-**Aortic-Flow Reference Shaping via Q-Learning for CF-LVAD Support **Mahmud Karno, Mert Yiğit, Ahmed Alhajyounis and A. Kamuran Kadıpaşaoğlu**
5	**11-** The Effects of Different Knee Braces on Gait-Related Muscle Activity **Serra Gedik, Fatma Günseli Yaşar Çıklaçandır, Özlem Karabiber Cura, and Samet Çıklaçandır**	**105-**Design and Virtual Performance Testing of the Accessory Equipment for a Left Ventricular Assist Device **Sinemis Çelik, and Kamuran Kadıpaşaoğlu**
6	**12-** Designing a Novel Implant for Preventing the Risk of Postoperative Pancreatic Fistula in Whipple Surgery **Ahmet Emir Öğüt and Samet Çıklaçandır**	
7	**54-**Fixed-Gain PID Controller Performance and Brittleness for a Simulated Knee Joint Under Varying Loads: A Simulink Study **Elif Umut, Ayman Bouamama, Fırat Matur, and Bora Büyüksaraç**	

TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC

Papers

Doku Takip İşlemlerinde Formaldehitin Tamponlanarak Kullanılmasının Fiksasyona Etkilerinin İncelenmesi

Investigation of the Effects of Buffered Formaldehyde in an Automatic Production System on Fixation in Tissue Processing Procedures

Habib Serdar Aday https://orcid.org/0009-0004-1069-199X
İstanbul Tıp Fakültesi
İstanbul Üniversitesi
Çapa(İstanbul, Türkiye)
saday@istanbul.edu.tr

Özetçe— Tıbbi Patoloji Anabilim Dalı rutin histopatoloji laboratuvarında gerçekleşen analitik süreçte yer alan doku takip işlemlerinde kullanılmakta olan formaldehit temini, muhafazası, tamponlanması, dozajlanması ve kullanıma hazır hale getirilerek doku takip sistemlerinde kullanılması ve analitik süreçte doku takip sonrası faaliyetler kayıt altın alınarak genel fiksatif olarak kullanılan % 37-40 saflıkta formaldehitin kullanıma hazır hale getirilmesi ve fiksasyon etkinliği bu çalışmamızda farklı yönleri ele alınarak değerlendirilmiştir.

Formaldehit, doku takip işlemlerinde henüz yerine efektif bir unsur ikame edilememiş doku fiksasyon unsuru olarak kullanılan kimyasal bir unsurdur. Formaldehit; üstün fiksasyon yeteneklerini ideal şartlarına kavuşması halinde gösterebilmektedir. Bir doku takip laboratuvarında doku fiksasyonunu etkinlik ve hızının ideal şartlarda oluşması için öncesinde bu şartların sağlanması esastır. Formaldehit solüsyonu laboratuvar yönetiminin tercihine göre hazır tamponlanmış ve tamponsuz halde ticari halde pazarlanmaktadır. Her iki formda olsa da solüsyonun etkinliği için öncelikle üretim ve tavsiye edilen son tüketim tarihi arasında ideal zamanlamada çalışılması önemlidir. Formaldehit fiksatifinin ideal şartlarda hazırlanması ile doku takip işlemleri ve sonrasında gerçekleşecek mekanizların uygulanmasına olumlu etki ettiği, tanı süresi ve işlem maliyetini azaltarak hasta ve kamu menfaatine olumlu etkisinin gösterilmesi amaçlanmıştır.

Anahtar Kelimeler —Doku takip, formaldehit, sodyum fosfat tamponu..

Abstract— In this study, different aspects of formaldehyde supply, storage, buffering, dosing and use in tissue tracking systems by making it ready for use and fixation efficiency of 37-40% pure formaldehyde used as a general fixative by recording the activities after tissue tracking in the analytical process carried out in the routine histopathology laboratory of the Department of Medical Pathology were evaluated.Formaldehyde is a chemical element used as a tissue fixation element that has not yet been effectively replaced in tissue tracking procedures.

Formaldehyde can show its superior fixation capabilities when it reaches ideal conditions. In a tissue tracking laboratory, it is essential to provide these conditions beforehand for the effectiveness and speed of tissue fixation to occur under ideal conditions. Formaldehyde solution is commercially marketed as ready-buffered and unbuffered according to the preference of the laboratory management. Although it is in both forms, it is important to work at the ideal timing between production and the recommended expiration date for the effectiveness of the solution. It is aimed to show that the preparation of formaldehyde fixative under ideal conditions has a positive effect on tissue tracking procedures and the application of the mechanisms that will occur afterwards, and to show its positive effect on the patient and public interest by reducing the diagnosis time and process cost.

Keywords — Tissue fixation, formaldehyde, sodium phosphate buffer.

I. GİRİŞ

Tıbbi Patoloji laboratuvarında doku fiksasyonu için uzun süredir kullanılan unsur formaldehit olagelmiştir. Formaldehit; doku fiksasyonunda etkisini kısa sürede oldukça etkin biçimde göstermesi, maliyetinin uygun olması, kolaylıkla temin edilerek kullanılabilir olması avantajlarına sahip vazgeçilmesi zor bir fiksatif olarak varlığını koruyabilmiştir. Bu avantajlarının etkin halde sürebilmesi, formaldehit asitliğinin korunabilmesi ile doğrudan bağlantılıdır, ayrıca konsantrasyon oranı, aldehit yüzdesi ve dağılımı, aldehit üretim ve kullanım tarihi, hazırlanma şekli, kullanılan su kalitesi, ortam ve fiksatif sıcaklığı, su ve ürün asitliği, doku veya organın daha önce farklı bir kimyasal ile etkileşiminin olup olmaması, ortam hava kalitesi

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

ve hava ile temas ve süresi formaldehit etkinliğini belirleyen etkenlerdir. Ayrıca formaldehitin asidik ortamda; pikrik asit ve formik asite parçalanması ile preparatlarda siyah-kahverengi pigment oluşumu, dokuda formik asitten kaynaklanan harabiyet olasığının artması, moleküler analizlerde formik asitin genetik materyali fragmante edebilme riski farklı sıklıkla kullanıcılar tarafından geri bildirim olarak sunulabilmektedir.

Formaldehit, doku takip işlemlerinde henüz yerine efektif bir unsur ikame edilememiş doku fiksasyon unsuru olarak kullanılan kimyasal bir unsurdur. Formaldehit; üstün fiksasyon yeteneklerini ideal şartlarına kavuşması halinde gösterebilmektedir. Bir doku takip laboratuvarında doku fiksasyonunu etkinlik ve hızının ideal şartlarda oluşması için öncesinde bu şartların sağlanması esastır. Formaldehit solüsyonu laboratuvar yönetiminin tercihine göre hazır tamponlanmış ve tamponsuz halde ticari halde pazarlanmaktadır. Her iki formda olsa da solüsyonun etkinliği için öncelikle üretim ve tavsiye edilen son tüketim tarihi arasında ideal zamanlamada çalışılması önemlidir. Formaldehit fiksatifinin ideal şartlarda hazırlanması ile doku takip işlemleri ve sonrasında gerçekleşecek mekanizların uygulanmasına olumlu etki ettiği, tanı süresi ve işlem maliyetini azaltarak hasta ve kamu menfaatine olumlu etkisinin gösterilmesi amaçlanmıştır.

II. MATERYAL METOD

Taze halde olan formaldehit sıklıkla %37-40 veya kullanıma hazır %10 halde pazarlanmaktadır. Optimum şartlarda temin edilen ürünün teknik şartlarını karşılayan bir laboratuvar ortamında çalışılması 6331 sayılı iş güvenliği ve çalışan sağlığı cari mevzuatı gereğince yasal zorunluluktur. İdeal termal şartları sağlanmış (sıcaklık 19-23^{0}C %30-60 bağıl nem) bir laboratuvarda ortam ve fiksasyon sıcaklığının kontrolü gerek dokunun ideal sıcaklıkta fiksatifin dokuya nüfuz edebilme hızını gerekse dokunun fiksatiften korunma zarara uğramama şartlarını optimize etmesi açısından gözetilmelidir.

Zira ortam ve fiksatif sıcaklığının gerek takip sistemlerinde gerek ortamda ideal şartların üzerine çıkması halinde; fiksatifin doku içerisinde yayılma hızı beklenen hızın üzerine çıkacaktır. Bu durum fiksatif ile dokunun daha hızlı ve uzun süre etkileşim halinde kalmasına ve dolayısı ile fiksasyon kimyasal etkinliğinin artmasına neden olacaktır. Bu durum, doku takip işlemi sonrasında gerek boyama gerekse de histokimyasal, immunhistokimyasal ve moleküler çalışmalarda doku etkinliğinin azalmasına neden olabilir.

Yüksek sıcaklıkta fiksayon ile menfi etkinin azaltılabilmesi için dokunun öncelikle oda sıcaklığında (298.15 K (25 °C, 77 °F) fiksasyonun başlatılması ardından 40-450C sıcaklığa çıkartılarak uzun süreli fiksasyonun dokuda etkinliği artırması beklenecektir. Formaldehiti %10 formda, ideal şartlarda hazırlanmış dokuya tüm yüzeylerinden uzun süreli gece takibi (4 Saat-6 saat) eşdeğere şeklilde etkin fiksasyonun oluşmasına katkı sağlayacaktır.(Tablo 1, Tablo 2, Tablo 3, Tablo 4)

Tablo.1 Penetrasyon Hızı

Formulasyon	d = k. √ t	
%10 Formaldehit		
Değer	Ölçüm	Birim
k	0,88	mm

Tablo.3 Penetrasyon Hızı

Formulasyon	d = k. √ t	
%10 Formaldehit		
Değer	Ölçüm	Birim
k	0,88	mm

Tablo.2
Tıbbi Patoloji A.D. Fiksasyon Süresi(saat)

Formaldehit	Etanol
4	6

Tablo.4
Tıbbi Patoloji A.D. Penetrasyon Hızı(mm)

Formaldehit	Etanol
0.7x √4=1,4	1x√6=2,3

Fiksatifin doku içerisine nüfuz etme işlemine fiksatifin pentrasyonu olarak adlandırılmaktadır. Penetrasyon; farklı bir sisteme, ağa konumuz gereği dokuya izin alınmaksızın yetkisiz bir şekilde giriş yapma işlemidir. Bilgisayar sistemine veya ağına bilişim virüsü girmesi ve bakteriyofaj penetrasyona örnek olarak verilebilir. Fiksasyon işlemlerinde; işlem gören toplam dokunun kapladığı alanı kapsayacak hacimden daha fazla fiksatif kullanılmaması halinde dokular içerisinde kısmen dahi olsa azalarak düşen fiksatif hacmi fiksayon etkinliğine tesir edebilmektedir. Bu yüzden, fiksatif hacmi (20) oranı / doku hacmi (1) olacak şekilde tasarlanmalıdır. Toplam doku hacmine göre fazladan fiksatif hacmi olması önemlidir fiksasyon ilerledikçe fiksatif tükenir ve küçük bir toplam hacimde bu fiksasyon kalitesini etkileyebilir.

Formaldehit (%10) fiksatifinin dokular içerisine penetrasyonu ağır gerçekleşen bir ilerlemedir. Farklı fiksatiflerin dokuya penetrasyon hızı difüzyonu farklıdır ve fiksasyonda optimum şartların sağlanma için dokunun ideal büyüklükte, ideal derinlikte olması, fiksasyon süresinin dokuya göre ayarlanması esastır. Ancak farklı dokularda fiksatiflerin farklı hızda difuze olabilecekleri dikkate alınmalıdır.

Medawar PB. sabitleyicilerin penetrasyon oranı $d = K\sqrt{t}$ formulu ile tasarlanmıştır. Formülde; (d) penetrasyon derinliği, (K) difüzyon katsayısı, (t) zaman olarak ifade edilmektedir. Formaldehit (%10) fiksatifinin dokulara difüzyonunda, penetrasyon derinliğinin K = 0,78 olup bu değer etanol için 1,0'dir. Farklı bir söylemle; Fiksatifin 1 saat içerisindeki difüzyon katsayısı, fiksatifin dokuya difüzyonunun milimetre cinsinden mesafesidir. Böylelikle %10 formaldehit 1 saatte bir saatte 1 mm nüfuz etmesi eklenir. Bu durum, 10 mm' lik bir numunenin orjinine kadar fiksatifin ilerleyebilmesi için 25 saatlik bir süreye ihtiyaç duyulacaktır (5 mm = 5² saat). (1)

Formaldehit solüsyonunun doku fiksasyonunda etkinliğinin sağlanmasındaki temel unsurlardan bir başkası ise Ph değeridir. %10 nötalize edilmiş tamponlu formaldehit fiksatifi; 1 küp şeker büyüklüğünde olan yumuşak doku içerisine difuze olma hızı; bir saatlik sürede yaklaşık 1,0 mm, 2 saatlik fiksasyon süresinde 1,4 mm, dört saatlik sürede 1,6mm, sekiz saatlik sürede, 2,2mm, sonrasında ise doku merkezine ulaşması beklenmelidir. Formaldehit %10 kullanıma hazır hale getirildikten hemen sonra şayet tamponlanmamış ise ortam ve solüsyon sıcaklığının da etkisi ile 1-2 sat içersinde asitlik derecesinde düşme başlar. Formaldehit fiksatifinin tamponlanarak taze kullanılması halinde benzer dokularda eş sürelerde; yakın özellikteki dokularda gerçekleşecek fiksasyon hızı öngörülebilmektedir. Formaldehit tamponlama işlemi Tablo 5 ve Tablo 6'da görüldüğü şekilde yapılmıştır.

Tablo.5


Sodyum Fosfat Tamponu Hazırlama Tablosu

A- di-Sodium Hydrogen Phosphate anhydrous, (6,5 gr)
B- Sodium di-Hydrogen Phosphate 2-Hydrate, (4 gr)
C- 1 lt için gerekli miktarlar böyle ama elinizdeki sodyum tuzlarının su oranı aynı değilse formül değişir.
D- 5 lt konsantre formol için 5 lt tampon çözeltisi ve 40 lt su olacak şekilde dozajlama cihazında yaptırıyoruz karışımı da.

Sodyum Fosfat tamponu Hazırlamak İçin Gereken Yolum	A- di-Sodium Hydrogen Phosphate anhydrous(Gram)	B- Sodium di-Hydrogen Phosphate 2-Hydrate(Gram)
1 Litre İçin	6,5	4
5 Litre İçin	32,5	20
10 Litre İçin	65	40
15 Litre İçin	97,5	60
20 Litre İçin	130	80

Tablo.6


Tamponlu Formaldehit Hazırlama Yönergesi

1-Menüden Hazırlama Tuşuna Bas.
2-Standart Tamponlu Hazırlamaya Bas.
3-Tamponlu Formol Volumunü Belirle(Örn 30Litre).
4-Başlat.
5-30 dak ile 60 dak arasında HAZIRLAMA SÜRESİNİ BEKLE.
6-Kullanıma Hazır.

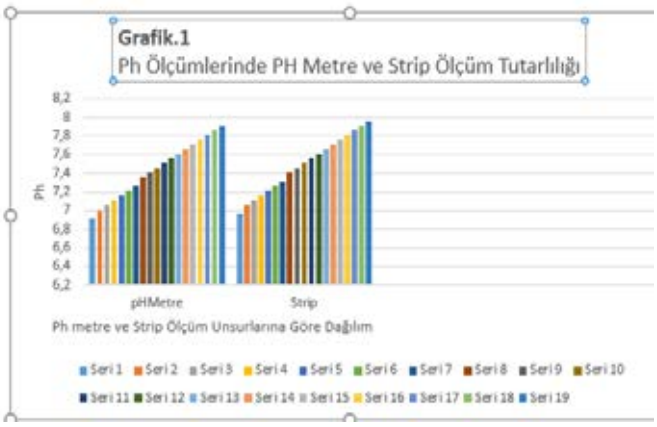
Böylelikle Ph kuvvetli asit değerlerine yaklaşır. Ph 5 ve altındaki değerlede formaldehit hızlıca formik asit formuna dönüşerek dokularda yer alan hemoglobin ile tepkimeye girmesi beklenir. Bu etkileşimden formaldehit hematin meydana gelir dokularda siyah veya kahverengi pigment şeklinde granuler oluşumlara neden olur, doku takip ve tanı kalitesine olumsuz etki etme etkinliğinin önlenmesi hedeflenmiştir.

Formaldehit fiksatifinin fiksasyon hemen öncesinde tamponlanmış olması, ideal şartlarda hazırlanması, karıştırılmış olması ve dozajlanması fiksasyon etkinliği açısından önemlidir. Çalışmamızda yer alan formaldehit fiksatifinin taze olması, üretim tarihinden itibaren tavsiye edilen son kullanım tarihini aşmamış olması, teknik şartnamemizde aranılan şartlara haiz istenilen evsafta olması sağlanmıştır.

Formaldehit konsantre formdan (%38-40) halden kullanıma hazır hale getirilmesi sonrasında doku fiksasyonunda etkinliğinin incelenmesi, takip işlemi yapılarak incelenen preparatlarda sıklıkla artefakların ekarte edilebilmesi için, 2022-2023 yılları arasında 434 gün boyunca Tıbbi Patoloji A.D. laboratuvarında formaldehit sodyum hidrojen fosfat ve di-sodyum hikroksi fosfat tuzlarının karışımı kullanılarak rutin protokol ile çözelti hazırlanmıştır.

III- BULGULAR

Solusyonun hazırlanmasında, Aquaderm revers osmosis su saflaştırma sisteminden elde edilen ve ortalama 9µS/saat iletkenlik değerinde ve en fazla 6,5 Ph değerinde olan su kullanılmıştır. Formaldehit otomatik formaldehit dozajlama makinesinde rutin tamponlu formaldehit programı kullanılarak doku fiksasyon ve takip işlemleri tam otomatik vakum-press kapalı sistem doku takip cihazı kullanılarak tamamlanmıştır. Tamponlama öncesinde kullanılmakta olan şebeke suyu her kullanımdan önce manuel şerit Ph kağıdı ve otomatik Ph metre ile 434 farklı ölçüm yapılmış, şerit Ph ölçer kağıtta asit bantında görüntü elde edilmiş, otomatik Ph metre de ise 5,5±0,7 aralığında ölçüm tespit edilmiştir.(Grafik.1)



Rutin tamponlama protokolü ve şebeke suyu kullanılarak otomatik formaldehit dozajlama makinesinde tamponlu formaldehit protokolunda üretilen tamponlu formaldehit hazırlanmıştır. Kullanıma hazır formaldehit proses öncesi manuel şerit Ph kağıdı ve otomatik Ph metre ile 434 farklı ölçüm yapılmış (Gfarik 2), şerit Ph ölçer kağıtta alkali bantında görüntü elde edilmiş, otomatik Ph metre de ise 7,6±0,7 aralığında ölçüm tespit edilmiş, elde edilen değerler Tablo 7 ve Tablo 8 'de gösterilmiştir.



Tablo.7	Şerit PhÖlçer	
Ölçüm Sayısı	434	434
Art Ort	Alkali Bant	7,6
St Sapma	***	0,7
İstatistiki Değer	Etkinlik%99,9	P< q 0,05
Sonuç	Ph metre Değerlerinin Doğrulayıcı Sonuç Elde edilmiştir.	Elde edilen formaldehit Ph doğruluk değerinin anlamlı olduğu anlaşılmıştır.

Tablo. 8 Ph metre değerleri

-416 örnekte Ph7,61-7,71

-8 Çalışmada 7,51-7,61 Arasında

-9 Çalışmada 7,71-7,81 arasında değer elde edildi.

6,91	7,01	7,06	7,11	7,16	7,21	7,26	7,31	7,36	7,41
6,96	7,06	7,11	7,16	7,21	7,26	7,31	7,36	7,41	7,45

7,45	7,51	7,56	7,61	7,66	7,71	7,76	7,81	7,86	7,91
7,51	7,56	7,61	7,66	7,71	7,76	7,81	7,86	7,91	7,96

IV- TARTIŞMA

Rutin doku takip makinesinde üretilerek kapalı otomatik takip cihazlarında kullanılarak sonlandırılan takip işlemlerinde, formaldehitin tamponsuz takiplerde beklendiği üzere parçalanarak formik asit oluşturmaması, asidik bir çözeltiden ziyade allkali ortamda takip işleminin tamamlanmış olması, formaldetten parçalanan pikrik ve formik asitin hemoglobinle reaksiyona girememesi nedeniyle eskisi gibi pigment oluşturmadığı görülmüştür. Böylece formaldehit fiksatifinin doğru, yerinde, taze hazırlanması ve uygun şekilde tamponlanarak kullanılması halinde dokularda analitik süreçte oluşması istenmeyen özellikler olan; doku bütünlüğünün bozulması, dokunun ıslak kalması, nükleer materyalin soluk kalması ve küçülmesi, sitoplazmanın silik kalması, küçülmesi gibi dokunun istenilen ideal şartlardan uzaklaşması, kesit alma

zorluğu yaşanmasının önüne geçilebileceği gösterilmiştir. Birim zamanda üretilen preparat sayısının artması, kesit ve görüntü kalitesinin üst düzeyde çıkartılması, harcanan sarf malzeme miktarının sınırlandırılması, tanı süresinin kısaltılarak ileri düzey (İHC, FİSH, nukleik Asit bazlı moleküler çalışmalar) analizler için ideal zemin(3) oluşturulmasına katkı sağladığı görülmüştür (Tablo 9).

Formaldehit fiksatifi kullanılarak yapılan doku fiksasyon etkinliği ve süresinin artırılması için formaldehit fiksatifi tam otomatik kapalı sistem takip cihazı üreticileri; rutin protokolunde 45-60 0C'ye kadar ısıtılması önerilebilmektedir. Böylelikle fiksasyon etkinliğinin arttığı bilinmekle birlikte dokularda sıcaklık ve fiksatif etkinliğinden kaynaklanabilecek reaktif değişimlere yol açabileceği göz ardı edilmemelidir.

Deney hayvanlarında yapılan deneysel çalışmalarda %10 konsantrasyonda formaldehit Ph derecesinin 5 ve aşağı inmesi halinde formik asit oluşması ile birlikte; formaldehite maruz kalan deney hayvanlarında mutajen etki yaptığı (Kawanishi, 1995)(4) belirlenmiştir. Bu durum ile karşılaşmamak için çalışmamızda kullanılan formaldehit aşağıda verilen oranlarda otomatik dozajlama ve karıştırma cihazında günlük tamponlanarak kullanılmıştır. Böylelikle formik asite dönüş ve dokularda pigmet oluşumu engellenmiştir. Formaldehit, memeli ve bakteri hücrelerinde mutasyonlarına oluşturabilmektedir.

Fransa hükümetinin formaldehit kullanımının belirlenmesiile alakalı çalışmaları esnasında %10 formaldehit fiksasyon etkinliği FineFIX ve RCL2 gibi farklı fiksatifler ile immunhistokimyasal çalışmlarda immun reaktivite etkinliği, moleküler çalışmalarda kullanılacak DNA ve RNA izolasyon değerleri ile karşılaştırılmıştır. Yapilan değerlendirmede; anılan fiksatiflerin kullanımı halinde hematoksilen-eosine boyama, immune reaktivite doku morfolojisi üzerine menfi etki görülmediği ancak moleküler patolojik çalışmalrda en iyi sonuçların FineFIX ve RCL2 fiksatifleri ile alındığı gösterilmiştir.(7)

Her ne kadar çalışmamızda control grubu kullanılmamış ise de; daha literatürde yer alan çalışmalara dayanılarak rutin protokol ile doku katip sisteminde formaldehitin taze halde kullanılması, verilen yol ve yöntem ile tamponlanması, fiksatifin arıtılmış nötral su ile dilue edilmesi halinde benzer sonuçlara ulaşılabileceğini göstermektedir.

Tablo.9 İdeal Formaldehit Fiksasyonu

	Çap Küçülmüş	Normal
Nükleus	-	+
Nükleer Materyal	**Açık Soluk Boyanmış**	**Koyu Parlak Boyanmış**
	-	+
Sitoplazma	**Küçülmüş**	**Normal**
	-	+
Nemli Doku	**Var**	**Yok**
	-	+
Dokuda Sertlik	**Kırılgan**	**İdeal Esnek**
	-	+
Kesitte Parçalanma	**Var**	**Yok**
	-	+
Pigment	**Var**	**Yok**
	-	+

V- SONUÇ

Her ne kadar formaldehit fiksatifi WHO (Dünya Sağlık Örgütü)tarafından sağlığa zararlı bir kimyasal olarak tanımlanmış ise de patoloji laboratuvarlarında doku fiksasyonu, histokimyasal boyama , immunhistokimyasal faaliyetlerde kullanım immunreaktivite için kullanılabilecek efektif, ekonomik bir fiksatif halen yaygın kullanıma girmemiştir. Bu nedenler göz önüne alınarak %10 formaldehit fiksatifinin yakın bir gelecekte de kullanımının devam edeceğini düşündürmektedir.

Tıbbi Patoloji rutin histopatoloji laboratuvarında fiksayon iş ve işlemlerinde vakum-press özelliğine sahip tam otomatik, kapalı sistem doku takip makinalarında yapılan takip faaliyetinde %10 konsantrasyonda kullanıma hazır formaldehit fiksatifinin; tavsiye edilen son kullanma tarihini aşmadan, otomatik dozajlama sisteminde protokole uygun halde tamponlanıp, Ph seviyesi rutin aralıklarla kontrol edilerek homojen, ideal, nötral %10 konsantrasyonda formaldehit fiksatif üretmek, kullanmak ve böylelikle dokularda ideal, sürdürülebilir, ekonomik takip işlemi uygulamak mümkün olduğu gösterilmiştir.

İdeal fiksasyonun uygulanması halinde tanı süresinin kısaltılması, birim maliyetin azaltılması, histopatolojik işlemler sonrasında yapılabilecek ileri tetkik ve analiz işlemlerine uygun zemin hazırlandığı ve bu çalışmanın uzun sürelere sari yapılabileceği anlaşılmıştır.

VI- KAYNAKLAR

[1] Medawar PB. Sabitleyicilerin penetrasyon oranı. J Royal Micros Soc 1941; 61 ;46-57

[2] Ö. Doğan. © 2005 Aegean Pathology Society, APJ, 2, 8–28.

[3] Kawanishi, M., Matsuda, T., Sasaki, G., Yagi, T., Matsui, S. ve Takebe, H. (1998a). İnsan hücrelerinde çoğaltılan mekik vektör plazmitlerinde krotonaldehit tarafından indüklenen mutasyonların bir spektrumu. Karsinogenez 19, 69–72. doi: 10.1093/carcin/19.1.690

[4] Auerbach, C, Genetıc And Cytogenetıcal Effects Of Formaldehyde And Related Compounds, *Mutatıon Research* ***39***: 317 (1977)

[5] Akasaka S, Yamamoto K (1991) Construction of Escherichia coli K12 phr deletion and insertion mutants by gene replacement. Mutat Res 254:27–35

[6] Auroma OI, Halliwell B, Gajewski E, Dizadaroglu M (1989) Damage to the bases in DNA induced by hydrogen peroxide and ferric ion chelates. J Biol Chem 264:20509–20512

[7] Sandra Lassalle 1, Véronique Hofman, Ilie Marius, Virginie Gavric-Tanga, Patrick Brest, Katia Havet, Catherine Butori, Eric Selva, José Santini, Baharia Mograbi, Paul Hofman (2009) Assessment of morphology, antigenicity, and nucleic acid integrity for diagnostic thyroid pathology using formalin substitute fixatives, Nov;19(11):1239-48. doi: 10.1089/thy.2009.0095.

Tıbbi Patoloji Laboratuvarında Otomatik Takip Sistemlerinde Etanol ve Ksilen Geri Kazanım Uygulamasında Verim Değerlendirilmesi

Efficiency Evaluation in Ethanol and Xylene Recovery Application in Automatic Tracking Systems in Pathology Laboratory

Habib Serdar Aday *https://orcid.org/0009-0004-1069-199X*
İstanbul Tıp Fakültesi
İstanbul Üniversitesi
Çapa(İstanbul, Türkiye)
saday@istanbul.edu.tr

Özetçe— Tıbbi Patoloji laboratuvarlarında rutin doku takip iş ve işlemlerinde günümüzde çoğunlukla kapalı, vakum pres özelliklerine sahip doku takip sistemleri ile yapılmaktadır. Doku takip cihazları, takip işlemini istenilen süre ve şekillerde tamamlamakla birlikte yoğun kimyasal solüsyon tüketen sistemlerdir. Takip işlemi esnasında kullanılmakta olan %10 kullanıma hazır formaldehit konsantre (%38-40) halden seyreltilerek kullanıma hazır (%10) hale getirilmesi, maliyetinin diğer kimyasal ajanlara göre düşük olması, birden fazla istasyonda kullanılması halinde münavebe usulune göre kullanılması, standardize ve pratik geri kazanım mekanizması olmaması tekrar kazanımı ekonomik olmaktan çıkarmaktadır. Ancak ksilen izomerleri (para-meta-orto) ve farklı konsantrasyonlar ve istasyon sayılarında takip sistemlerinde alan etil alkol ile birlikte değerlendirilmesi halinde, geri kazanım mekanizmaları ticari formatta temin edilebilen, geri kazanım halinde ekonomik olabilecek kimyasallardır. Çalışmamızda; 2012-2015 yılları arasında tıbbi patoloji laboratuvarında kurulu 3 adet tam otomatik doku takip cihazlarından elde edilen etil alkol ile karışık halde olan ksilen izomerlerinin tam otomatize alkol, ksilen ayrıştırma ve geri kazanım cihazı ile geri kazanılmasının efektif ve ekonomik bir yöntem olduğu kadar ekosistem için de kirliliğin azaltılması, olası tehlikeli atık miktarının minimum hale, katma değer ve elde edilen kamu menfaatinin maksimum düzeye uygulanabilir bir protocol ile ulaşılabileceği gösterilmiştir.

Ksilen izomerleri, envantere alınması ve kontrollü kullanımı sonrasında yakın çevreye dağılmaları veya kanalizasyon sistemine verilmeleri halinde; buharlaşarak atmosfere, yakın çevre yolu ile toprağa ve yeraltı kaynaklarına veya kanalizasyon yolu ile denizlere ve besin zincirine rahatlıkla karışabilir bu nedenle kullanımın kontrollü ve olabilecek en alt seviyede olması sağlanmalıdır.(1)

Ksilen yüksek oranda uçucu olduğundan, atmosferde rahatlıkla konsantrasyonu artar ve ortamda güneş ışığı tarafından daha az zararlı hale getirilse bile birkaç gün veya mikroorganizmalar tarafından parçalanmadan önce birkaç ay bulunduğu alanda kalabilir sonrasında daha zararlı kimyasallara parçalanmaktadır..(2)

Ksilen ve etanol gibi geri kazanılabilir kimyasal atıkların kontrolü ve yönetim sistemleri üzerinde Manfredini tarafından 2008 (3) yılında yapılan bir çalışmada ksilol gibi organik çözücüler, bir temizleme ve ardından bir saflık analizi ile öncelenen damıtma teknikleriyle geri kazanılabileceğini göstermiştir. Bu araştırma kapsamında kşşilen ve etanol gibi kimyasallar dilatasyon(damıtma) tekniği ile geri kazanılmış, kırılma indisi ile ultraviyole spektrofometresi kontrolü sonrasında kullanıma uygun görülerek laboratuvar faaliyetlerinde kullanılmıştır.(3)

Anahtar Kelimeler — Ksilen, etanol, dilatasyon, geri kazanım.

Abstract— In medical pathology laboratories, routine tissue tracking operations and procedures are mostly carried out with closed, vacuum press tissue tracking systems today. Tissue tracking devices complete the tracking process in the

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

desired time and manner, but they are systems that consume intensive chemical solutions. The fact that the 10% ready-to-use formaldehyde used during the tracking process is diluted from a concentrated (38-40%) state to a ready-to-use (10%) state, its cost is low compared to other chemical agents, it is used in more than one station according to the rotation method, and there is no standardized and practical recovery mechanism makes recovery uneconomical. However, xylene isomers (para-meta-ortho) and ethyl alcohol in different concentrations and station sa-tu-ons are chemicals that can be obtained in commercial format and can be economical in recovery. In our study; it has been shown that the recovery of xylene isomers mixed with ethyl alcohol obtained from 3 fully automatic tissue tracking devices installed in the medical pathology laboratory between 2012-2015 with a fully automated alcohol, xylene separation and recovery device is an effective and economical method as well as reducing pollution for the ecosystem, minimizing the amount of possible hazardous waste, and achieving maximum added value and public benefit with an applicable protocol.

Keywords — Xylene, ethanol, dilatation, recovery.

I. GİRİŞ

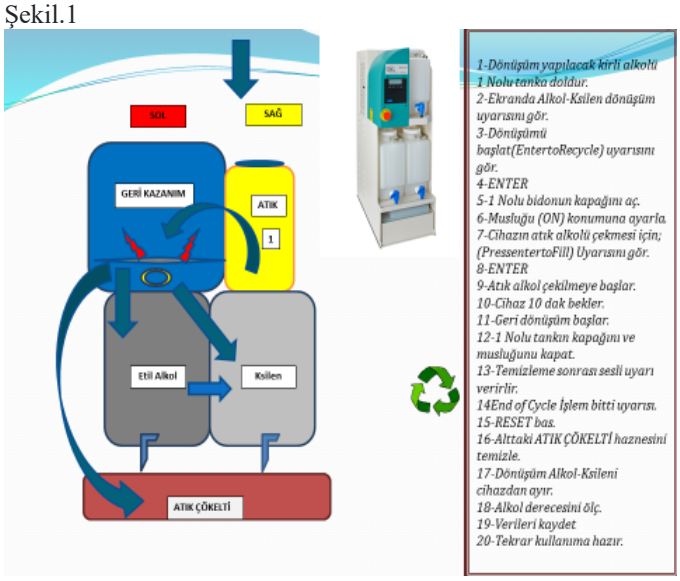
Tıbbi Patoloji laboratuvarlarında rutin doku takip iş ve işlemlerinde günümüzde çoğunlukla kapalı, vakum pres özelliklerine sahip doku takip sistemleri ile yapılmaktadır. Doku takip cihazları, takip işlemini istenilen süre ve şekillerde tamamlamakla birlikte yoğun kimyasal solüsyon tüketen sistemlerdir. Takip işlemi esnasında kullanılmakta olan %10 kullanıma hazır formaldehit konsantre (%38-40) halden seyreltilerek kullanıma hazır (%10) hale getirilmesi, maliyetinin diğer kimyasal ajanlara göre düşük olması, birden fazla istasyonda kullanılması halinde münavebe usulune göre kullanılması, standardize ve pratik geri kazanım mekanizması olmaması tekrar kazanımı ekonomik olmaktan çıkarmaktadır. Ancak ksilen izomerleri (para-meta-orto) ve farklı konsantrasyonlar ve istasyon sau-yılarında takip sistemlerinde alan etil alkol ile birlikte değerlendirilmesi halinde, geri kazanım mekanizmaları ticari formatta temin edilebilen, geri kazanım halinde ekonomik olabilecek kimyasallardır.

II. MATERYAL METOD

Çalışmamızda; 2012-2015 yılları arasında tıbbi patoloji laboratuvarında kurulu 3 adet tam otomatik doku takip cihazlarından elde edilen etil alkol ile karışık halde olan ksilen izomerlerinin tam otomatize alkol, ksilen ayrıştırma ve geri kazanım cihazı ile geri kazanılması amaçlanmıştır.

Takip sistemlerinden elde edilen kirli etil alkol ve ksilen atık olarak değerlendirilmeden geri kazanılması için takip cihazlarından sırası ile alınarak kapasitesi ölçüsünde geri dönüşüm sistemi tankına yerleştirilir. Ksilen ve etil alkol kaynama derecelerinin birbirinden farklı olması nedeni ile sistem dolum yapılması sonrasında ilk olarak kirlenmiş sıvı halde **atık solventi** 78,5 C^0 kadar ısıtır. Kaynama noktasına gelen etil alkol karışımdan ayrılır ve yoğunlaşır, rejenere olan ve filtreden geçirilerek temizlenen etil alkol kendi kısmına damıtma (distilasyon)" yöntemi ile temizleyerek ayrılarak depolanır. Böylelikle alkol ayrıştırması tamamlanır, partiküller ise atık ünitesine deport edilir ve tehlikeli atik bertaraf yüklenicisne teslim edilinceye kadar geçici tehlikeli atik depolama alaninda, uygun muhafaza kaplarina alinir.elde edilen veriler kaydedilir. Sistemin çalışma esası ve komponentleri aşağıda şekil.1'de gösterilmiştir.

Şekil.1



III. BULGULAR

Sistem geri kalan solüsyonu 138.5 C^0 kadar ısıtır, kaynama noktasına gelen ksilen ayrıştırma borusundan "damıtma (distilasyon)" ile ayrılarak ksilen tankına aktarılır. Ayrıştırma sonunda geri kazanılan alkol ve ksilen oda sıcaklığına soğutulur, alkol derecesi ölçülen; partiküller, formalin, çökmüş aldehit gruplar ve diğer kalıntılar ile kontamine olmuş etanol ve temizleme kabiliyet değerlendirmesi sonucu kulanılabilir hale gelen ksilen dilatasyon sonucu geri kazanılmış olur. Ürün, geri dönüşüm tankından alınır, atık tankı temizlenir ve system tekrar kulanılabilir hale getirilirek uygulama miktarları, tarih ve saat bilgisi kaydedilir. Bahse konu tarihler arasında geri kazanım ile elde edilen veriler Tablo.1'de verilmiştir.

Tablo.1

Takip A Takip B Takip C	**Kaynama Derecesi (C^0)**	**Haftalık Tüketim (Litre)**	**Aylık Tüketim (Litre)**	**Yıllık Tüketim (Litre)**
A-Etil Alkol(96)	**78,5**	**90**	**360**	**4.320**
B-Ksilen İzomerleri (Para Meta Orto)	**138.5**	**26**	**104**	**1.248**

2012 ile 2015 yılları arasında İstanbul Tıp Fakültesi Tıbbi Patoloji Anabilim Dalında doku takip laborauvarında yer alan 3 adet kapalı system doku takip cihazında kullanılmış ksilen ve

etil alkol toplam miktar, aylık geri kazanım, yıllık geri kazanım,aylık geri kazanım yüzdesi (%), yıllık geri kazanım yüzdesi (%) kaydedilmiş olup tablo 2'de verilmiştir.

Tablo.2

2012-2015 Toplam Tüketim (Litre)		Aylık	Yıllık	Aylık	Yıllık	Toplam
		Geri Kazanım (Litre)		Geri Kazanım Başarı Oranı(%)		
Etanol (%96)	17.280	313,2	3.844,80	87	89	88
Ksilen İzomer	4.992	97,76	1.148	94	92	93

IV. TARTIŞMA

Her ne kadar %96 alkol ve ksilen izomerlerinin atık olarak tanımlanabilmesi için kullanım amacı ve vasfını tekrar kullanımı mümkün olamayacak nitelikte kaybetmesi gerekmektedir. Ancak bu aşamada atık ünvanı kazanmaktadırlar. Tıbbi Patoloji laboratuvarları çalışma şekil ve yöntemleri dikkate alındığında özellikle doku takip süreçlerinde kapalı sistem doku takip cihazlarında yüksek miktarlarda alkol ve ksilen kullanmaktadırlar. Çalışmamızda kullanılmış olan alkol ve ksilen izomerlerini ayırabilen cihazlar veya farklı sistemler kullanılarak atık ünvanı kazanılmadan solusyonların tekrar kullanımı Dünyada yaygın olduğu görülmektedir.

Atık statüsü kaznmamış alkol ve ksilen izomeri miktarlarının azaltılması, çevreye yönelik etki yükünün düşürülmesi, kimyasal bertaraf maliyetlerinin azaltılması için evrensel ölçekte çalışmalar devam etmektedir. Bu konuda en fazla başvurulan yöntem alkol ve ksilen in farklı sıcaklık değerlerine maruz bırakılmaları halinde birbiriden ayrılmalarına ragmen farklı farklı kendi kimyasal özelliklerini sürdürebilir olmalarından kaynaklanmaktır. Ayrıca hükümetler düzeyinde bu çalışmaların çevre koruma, yaşam kalitesine olumlu etkide bulunmaları nedeniyle desteklendiği bilinmektedir. Ksilen kimyasalına maruziyet yanlızca tıbbi patoloji laboratuvarında kalmayıp endüstriyel üretim aşamalarında, üretim sonrası parselizasyon ve dağıtım sonrasında son kullanıcı ve uygunsuz kullanım ile kullanıcılarına ve yakın çevreye menfi etkisini göstermektedir. Ksilen kullanımı ile büyük miktarı yakın çevreye dağılarak etkin olmaktadır. Ardından atmosfer ve yakın çevreye etki etmektedir(4).

Geri kazanım sistemi kullanılmadan önce kullanım esnasında ortamın havalandırılması ve damıtmadan kaynaklanacak gazların ortamdan tahliyesine imkan sağlayabilecek ideal havalandırma sisteminin kurulmuş olması esastır. Sistemin çalışma alanlarının doğrudan gün ışığına maruz kalmaması, ortamda kıvılcım, elektrik açık akımına imkan vermeyecek düzenlemelerin alınmasına, elektrostatik kaynakların engellenmesine dikkat edilmelidir. Kimyasal geri kazanım ile etanol saflığı her bir geri kazanımda manuel alokolmere ile takip edilmiş ilk kazanım sonrasında yalnızca %1 oranında kayıp görülmüştür. Elde edilen %95 saflıktaki geri kazanılmış etanol; temizleme alkol istasyonunda güvenle kullanılmıştır.

V. SONUÇ

Günümüzde patoloji çalışanı olmasalar dahi ksilen veya farklı bir çözücüye maruz kalabilmektedir. Ancak düzenli olarak ksilen veya farklı bir kimyasal çözücüye profesyonel çalışanların maruz kalma oranı %12 kadardır. Costa ve arkadaşları (2007)'ye, tıbbi patoloji çalışanları ve sitoteknologlar ksilen ve diğer çözücülerin tehlikelerinden tam olarak haberdar değillerdir. Anılan unsurlar görevlilerin %6,6 kadarının idrarda metilhipürik asit tespit dozunda rastlanmış olup ksilen maruziyeti gösterilmiştir. Bunun yanı sıra Costa ve arkadaşlarının çalıştığı çoğu laboratuvarda ksilen izomerleri ve alkolun bertaraf yuklenicisine teslim edilmeden kanalizasyona deşarj edildiği anlaşılmıştır(5).

Her ne kadar %96 alkol ve ksilen izomerlerinin atık olarak tanımlanabilmesi için kullanım amacı ve vasfını tekrar kullanımı mümkün olamayacak nitelikte kaybetmesi gerekmektedir. Ancak bu aşamada atık ünvanı kazanmaktadırlar. Tıbbi Patoloji laboratuvarları çalışma şekil ve yöntemleri dikkate alındığında özellikle doku takip süreçlerinde kapalı sistem doku takip cihazlarında yüksek miktarlarda alkol ve ksilen kullanmaktadırlar. Çalışmamızda kullanılmış olan alkol ve ksilen izomerlerini ayırabilen cihazlar veya farklı sistemler kullanılarak atık ünvanı kazanılmadan solusyonların tekrar kullanımı Dünyada yaygın olduğu görülmektedir.

Atık statüsü kazanmamış alkol ve ksilen izomeri miktarlarının azaltılması, çevreye yönelik etki yükünün düşürülmesi, kimyasal bertaraf maliyetlerinin azaltılması için evrensel ölçekte çalışmalar devam etmektedir. Bu konuda en fazla başvurulan yöntem alkol ve ksilen in farklı sıcaklık değerlerine maruz bırakılmaları halinde birbinden ayrılmalarına rağmen farklı farklı kendi kimyasal özelliklerini sürdürebilir olmalarından kaynaklanmaktır. Ayrıca hükümetler düzeyinde bu çalışmaların çevre koruma, yaşam kalitesine olumlu etkide bulunmaları nedeniyle desteklendiği bilinmektedir.

Cazari ve arkadaşları tarafından yapılmış olan (2013) tarihli çalışmada, ksilen izomerlerinin histokimyasal boyamalarda kullanımının en aza indirilmiştir. Sonuçta; HE-EO (hematoksilen-eosine)boyama yapılan dokularda histokimyasal boyama etkinliğinde bir farka rastlanmadığı belirtilmiştir(6).

Adeniyi ve arkadaşları (2016) ise, ksilen izomerlerinin parlatma kabiliyetleri; Hindistan cevizi yağı, karanfil yağı, fıstık yağı ile karşılaştırılmıştır. Parlatma yeteneklerinde sayılan unsurların bir diğerine belirgin üstünlük sağlamadığı görülmüştür. Bu durum yalnızca parlatma yeteneği irdelenmiş olsa da farklı kimyasalların ksilen izomerleri yerine ikame edilebileceğini göstermektedir(7).

Otomatik doku takip sistemlerinde kullanılmakta olan etil alkolün alkol derecesi manuel alkol-metre ile kontrol edilmek üzere en az 500 cc etanol cam mezür içerisine alınır, manuel alkol-metre cam mezür içerisine alınır ve mezürde etanol bitiş noktasına kadar takip edilerek ölçüm alınır ve kaydedilir.

Elde edilen verilerin değerlendirilmesinde; 2012 ile 2015 yılları arasında doku takip ünitesinde kurulu ve faal 3 adet kapalı sistem doku takip sistemi çalıştırılmıştır. Bu sistemlerde doku takip işlemi için yıllık toplam 17.280 litre %96 saflıkta etanol, 4.992 iltre para, meta, orto ksilen izomeri kullanılmıştır. Takip işlemi sonuçlanması ile etanol ve ksilen ile kontamine hald ve tehlikeli atık vasfı kazanımış etanol ve ksilen geri dönüşüm cihazı ile tekrar kazanılmış ölçümleri ve incelemeleri yapılarak doku takip sisteminde temizleme ksileni ve etanolü olarak tekrar kullanılmıştır.

Etanol geri kazanım başarı yüzdesi aylık bazda %87 olurken yıllık bazda %88 olarak tespit edilmiştir. Ksilen geri kazanım başarısı aylık bazda %88 iken yıllık bazda; yıllık %89 olarak kaydedilmiştir. Ksilen geri kazanım başarısı aylık bazda %94 iken yıllık bazda %92 olarak kaydedilmiştir.

Her iki parametre ayrı ayrı değerlendirildiğinde; etanol geri kazanım halinde alkol saflığı ilk geri kazanımdan sonra; %96 dan %95 seviyesinde olduğu ancak bu değerin temizleme alkol kullanımında aksaklığa yol açmadığı anlaşılmıştır. Ksilen geri kazanılan ürünün temizleme aktivitesinde aksaklık kaydedilmemiştir.

Tüm bu veriler ışığında; etanol ve ksilen geri kazanım işlemlerinin efektif şekilde yapılabildiği, düşük maliyet ile yüksek kapasiteli ürün elde edildiği, kamu menfaatine dayalı katma değer sağlandığı, olası tehlikeli atık potansiyelinin bu yolla azaltılmasının mümkün olduğu görülmüştür(8).

VI- KAYNAKLAR

1- Agency for Toxic Substances and Disease Registry (ATSDR). Openly accessible at http://www.mastereditora.com.br/bjscrHernandes et al. /Braz. J. Surg. Clin. Res. V.17,n.2,pp.68-73 (Dec 2016- Feb 2017) Toxicological Profile for Xylenes. Atlanta, GA: Public Health Service, U.S. Department of Health and Human Services; 2007

2- *Albertini LB, Honorio DH. Tratamento de Resíduos de Xilol utilizado em Laboratórios Histológicos e Citopatológicos. In: 17˚ Simposio internacional de iniciação cientifica; 2009; nov 6; São Paulo: SIICUSP; 2009.*

3- *Manfredini KL. Estado atual e proposta de melhorias no gerenciamento de residuos de glutaraldeido, xilenos e formaldeido em um hospital escola e em um laboratorio universitario de anatomia. [Dissertação] Caxias do sul: Universidade de Caxias do Sul; 2014.*

4- *Brasil. Conselho Nacional do Meio Ambiente. Resolução n˚358, de 29 de abril de 2005. Dispõe sobre o tratamento e a disposição final dos resíduos dos serviços de saúde e dá outras providências. Brasília, DF. Ministério do Meio Ambiente, abril de 2005. Disponível em: . Acesso em: 01 ago. 2016.*

5- *Costa KNS, Pinheiro IO, Calazans GT, Nascimento MS. Avaliação dos riscos associados ao uso do xilol em laboratórios de anatomia patológica e citologia. Rev. bras. saúde ocup. [online] 2007; 32(116):50-56.*

6- *Cazari VR, Pereira TR, Romera AM, Brandão MC, Filho CZ, Favareto APA. Redução do uso do xilol na técnica de coloracão hematoxilina e eosina. Colloquium Vitae jul/dez 2013; 5(2):135-148.*

7- *Adeniyi IM, Adejoba RO, Akinlabi FM, Alao OJ. Vegetable Oils as Clearing Agents. Achievements In The Life Sciences Jun 2016. 10(1):1-4.*

8- *Levada JC. Gestão e gerenciamento de residuos quimicos e aplicação da tecnologia de destilação na recuperação de solventes orgânicos: estudo de caso da reciclagem do xileno. [Dissertação] São Carlos: Universidade de São Paulo; 2008.*

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Secure and Interpretable Dyslexia Detection Using Homomorphic Encryption and SHAP-Based Explanations

Mhd Raja Abou Harb
Computer Engineering Department
Işık University
Istanbul, Türkiye
21comp9001@isik.edu.tr

Baris Celiktas
Computer Engineering Department
Işık University
Istanbul, Türkiye
baris.celiktas@isikun.edu.tr

Gunet Eroglu
Computer Engineering Department
Bahçeşehir University
Istanbul, Türkiye
gunet.eroglu@eng.bau.edu.tr

***Abstract*—Protecting sensitive healthcare data during machine learning inference is critical, particularly in cloud-based environments. This study addresses the privacy and interpretability challenges in dyslexia detection using Quantitative EEG (QEEG) data. We propose a privacy-preserving framework utilizing Homomorphic Encryption (HE) to securely perform inference with an Artificial Neural Network (ANN). Due to the incompatibility of non-linear activation functions with encrypted arithmetic, we employ a dedicated approximation strategy. To ensure model interpretability without compromising privacy, SHapley Additive exPlanations (SHAP) are computed homomorphically and decrypted client-side. Experimental evaluations demonstrate that the encrypted inference achieves an accuracy of 90.03% and an AUC of 0.8218, reflecting only minor performance degradation compared to plaintext inference. SHAP value comparisons (Spearman correlation = 0.59) validate the reliability of the encrypted explanations. These results confirm that integrating privacy-preserving and explainable AI approaches is feasible for secure, ethical, and compliant healthcare deployments.**

***Index Terms*—Homomorphic Encryption, Encrypted Inference, Explainable Artificial Intelligence (XAI), SHAP, Quantitative EEG (QEEG), Dyslexia Detection**

I. Introduction

The adoption of Machine Learning (ML) in healthcare supports early diagnosis and clinical decision-making, yet raises privacy and security concerns—particularly when sensitive data are processed in cloud environments. EEG-based inference, for example, often requires outsourcing computations, exposing patient data to leakage and unauthorized access.

Privacy-Preserving Machine Learning (PPML) methods, notably Homomorphic Encryption (HE), enable computation on encrypted data without decryption. However, deploying HE is computationally intensive and incompatible with standard non-linear neural network operations.

Scheibner et al. [1] discussed the ethical promise of HE and blockchain for health data, while Munjal and Bhatia [2] analyzed trade-offs across PHE, SHE, and FHE—emphasizing FHE's strong privacy but high cost. In dyslexia detection, Usman and Muniyandi [3] introduced CryptoDL for encrypted MRI data, and later work [4] applied polynomial-approximated CNNs to encrypted fMRI, improving accuracy at the cost of scalability.

979-8-3315-5565-8/25/$31.00 © 2025 IEEE

Other privacy approaches include Federated Learning (FL) and Differential Privacy (DP). Mercier et al. [5] found FL+DP effective for ECG classification, though HE introduced latency. In parallel, explainable AI (XAI) methods have advanced dyslexia research. Gallego-Molina et al. [6] applied LIME to CNN-LSTM EEG models, while Robaa et al. [7] used Grad-CAM with transformer-based handwriting analysis—but both lacked privacy guarantees.

Despite EEG's promise, few studies integrate secure, interpretable ML frameworks for dyslexia detection. Behavioral and biometric methods are vulnerable to spoofing, while EEG captures subconscious neural patterns that are harder to manipulate. Still, challenges in encrypted computation and model transparency hinder EEG adoption. This study addresses this gap by introducing a homomorphic inference pipeline with SHAP-based explanations, extending our previous privacy-preserving QEEG classification framework [8], [9].

The rest of this paper is organized as follows: Section II details our methodology, Section III presents the results, Section IV discusses the findings, and Section V concludes the paper.

II. Methodology

A. Dataset and Preprocessing

The dataset used in this study consists of QEEG recordings from 200 children, equally divided into two groups: 100 diagnosed with dyslexia and 100 typically developing controls. Participants were selected based on the Test of Integrated Language and Literacy Skills (TILLS), with an average age of approximately 8.8 years in both groups. All subjects were medication-free and recruited from middle socioeconomic backgrounds.

EEG data were collected using the EMOTIV EPOC-X headset, initially sampled at 2048 Hz and downsampled to 128 Hz after artifact removal. Signals were acquired from 14 channels (AF3, F3, F7, FC5, T7, P7, O1, O2, P8, T8, FC6, F8, F4, AF4), and power spectral features were extracted across five frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta-1 (12–16 Hz), beta-2 (16–25 Hz), and gamma (25–45 Hz),

resulting in 70 features per participant.

Only resting-state EEG data were used to ensure consistency. Preprocessing included z-score filtering to remove outliers, followed by temporal segmentation and averaging. Feature values were standardized using the `StandardScaler` for the main classifier, and normalized using `MinMaxScaler` for the activation estimator.

The dataset was randomly split into 80% for training and 20% for testing. In addition, 5-fold Cross-Validation (CV) was applied during training to assess model robustness and reduce variance due to random sampling.

B. Encrypted Inference Framework

To preserve data privacy during inference, we implemented a secure framework based on the CKKS HE scheme. The system involves three main actors: the patient (data owner), the cloud service provider (CSP), and the key management service (KMS). After generating a public-private key pair via the KMS, the patient encrypts the preprocessed QEEG features using the public key and sends them to the CSP for processing.

The CSP performs inference entirely over encrypted data, using a pre-trained neural network model. The architecture comprises two components: a *Main ANN*, responsible for dyslexia classification, and an *ANN-based estimator*, described in detail in Section II-D. This design enables privacy-preserving inference without revealing intermediate or final values to the server.

The encrypted prediction results are returned to the patient, who decrypts them locally using the private key. This framework ensures end-to-end data confidentiality and supports encrypted arithmetic over real-valued data. Further details of the Main ANN and the activation estimator are presented in the following subsections.

C. ANN Classifier Architecture

The classification task is performed by a shallow ANN, referred to as the *Main ANN*, designed for compatibility with HE. The model consists of an input layer with 70 neurons (QEEG features), a hidden layer with 128 neurons using the Sigmoid activation, and an output layer for binary classification.

Training is conducted in plaintext using the Adam optimizer (learning rate 0.001) with L2 regularization and a learning rate scheduler to improve convergence. The use of a shallow structure helps mitigate noise accumulation during encrypted inference. Since the Sigmoid activation is not natively supported under HE, its output is approximated using a dedicated ANN-based estimator described in the next subsection.

D. ANN-Based Activation Function Estimation

To enable non-linear operations in the encrypted domain, we employ a lightweight ANN-based estimator to approximate the Sigmoid activation function. This estimator consists of a single hidden layer with 128 neurons and is trained on pre-activation outputs collected from the Main ANN's hidden layer.

During training, pre-activation values are normalized using `MinMaxScaler`, and the corresponding Sigmoid outputs serve as ground truth labels. The estimator is trained using the Adam optimizer (learning rate 0.0001) with weight clipping to maintain stability. This approach allows accurate Sigmoid approximation during encrypted inference without relying on polynomial or piecewise approximations.

E. SHAP-Based Explainability Under Encryption

To provide interpretability alongside privacy, we integrate SHAP [10] into the encrypted inference pipeline. Since the CKKS scheme supports approximate arithmetic over encrypted vectors, we adapted SHAP computation to operate homomorphically without decrypting feature values or predictions.

The SHAP estimation follows these steps: (1) the client encrypts the preprocessed input feature vector and sends it to the cloud; (2) the cloud generates perturbed versions of the encrypted input by selectively replacing features with encrypted baseline values; (3) inference is performed on each perturbed ciphertext using the encrypted ANN pipeline; (4) the CSP computes the average marginal contributions of each feature; and (5) the resulting SHAP vector is returned to the client for decryption and visualization.

This process enables secure, feature-level interpretability of encrypted predictions. We assess the fidelity of the homomorphic SHAP approximation by comparing it to plaintext SHAP results using Mean Absolute Error (MAE) and Spearman rank correlation. To mitigate the computational burden of SHAP estimation, we employed sampling-based Kernel SHAP with 50 perturbed samples per input, balancing interpretability fidelity and runtime [10].

F. CKKS Encryption Parameters

To support efficient and accurate computation on real-valued data under HE, we adopted the CKKS scheme [11]. The encryption parameters were carefully configured to balance security, noise tolerance, and computational efficiency, following best practices established in previous works such as [9]. The configuration is summarized as follows:

- **Polynomial Modulus Degree:** Set to 32,768 to ensure a high level of security (equivalent to 128-bit) and sufficient ciphertext capacity for deep computation.
- **Coefficient Modulus Sizes:** The modulus chain was defined using bit sizes [60, 50, 50, 50, 50, 50, 60], providing a total bit length suitable for precise fixed-point operations and ensuring ciphertext integrity during inference.
- **Global Scale:** Set to 2^{50} to preserve high precision in floating-point arithmetic.
- **Galois Keys:** Enabled for supporting vector rotation operations, which are essential for matrix multiplications during ANN and CNN inference on encrypted data.

This configuration enables seamless integration with the neural network architecture while supporting multi-step encrypted inference without bootstrapping, in line with the methodology described in [9].

III. Results

To evaluate the performance of the proposed ANN-based activation function estimators under HE, all experiments were conducted on Google Colab, leveraging the computational

capabilities of a TPU v2-8 instance. This environment provided sufficient processing power and memory to efficiently run encrypted neural network inference tasks and activation approximation models. Unless otherwise stated, all numerical results presented in this section were rounded to four digits after the decimal point to ensure consistency and clarity across reported metrics.

The results demonstrate the performance and interpretability capabilities of the proposed encrypted inference framework. Table I reports classification metrics for the Main ANN model in both plaintext and encrypted modes, including accuracy, F1-score, sensitivity, specificity, and Area Under Curve (AUC). Despite the computational overhead of privacy-preserving inference, the encrypted mode achieves competitive performance with an HE inference time of approximately 21.8740 seconds.

Figures 1 and 3 present the mean absolute SHAP values for the encrypted and plaintext inference, respectively, enabling a visual comparison of feature importance consistency. Figure 2 further highlights the EEG channel-wise importance during encrypted inference, offering spatial insight into the model's attention across the scalp.

Additionally, Table II quantifies the fidelity of SHAP value estimation under encryption by comparing the encrypted and plaintext SHAP vectors using MAE and Spearman rank correlation. Together, these results confirm that our framework maintains high classification performance while providing interpretable, privacy-preserving predictions.

TABLE I
EVALUATION METRICS RESULTS

Type	Acc.	Sens.	Spec.	F1	AUC	CV Acc.
Plaintext	0.9269	0.6991	0.6991	0.7631	0.8604	0.9246
HE	0.9003	0.6134	0.6134	0.6536	0.8218	0.8953

Note: Acc.=Accuracy, Sens.=Sensitivity, Spec.=Specificity, F1=F1-Score, AUC=Area Under Curve, CV=Cross Validation

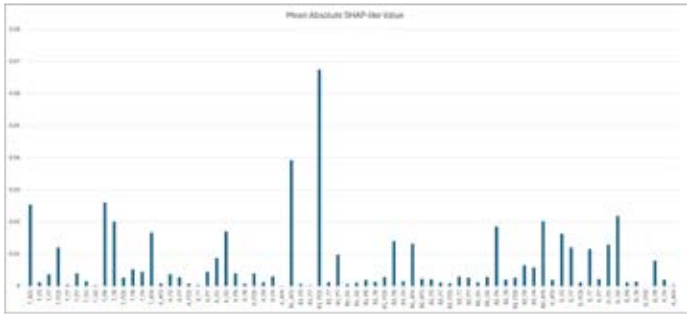


Fig. 1. **Interpretability Analysis: Mean absolute SHAP values per feature during homomorphic inference**

IV. DISCUSSION

As summarized in Table I, the encrypted inference pipeline maintained high classification performance, with minor reductions in accuracy and AUC relative to plaintext. Notably, the 5-fold CV average accuracy of 89.53% for the encrypted model confirms its robustness and generalizability across different training subsets, reinforcing the framework's reliability for real-world deployment.

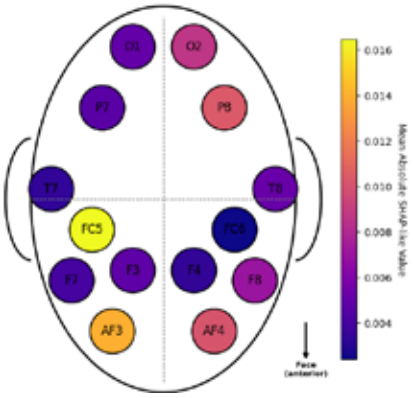


Fig. 2. **EEG channel-wise importance during HE inference.**

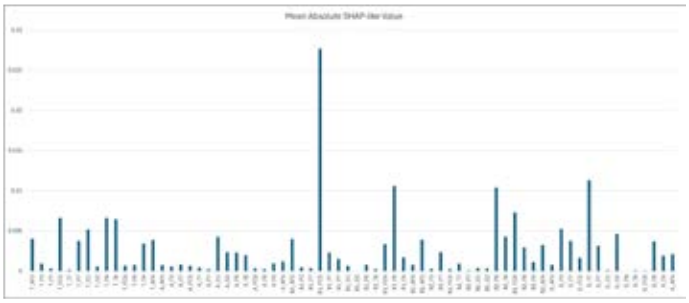


Fig. 3. **Interpretability Analysis: Mean absolute SHAP values per feature during plaintext inference.**

Figures 1 and 3 visualize the mean absolute SHAP values derived from encrypted and plaintext inference, respectively. These charts reveal a high degree of consistency in feature importance rankings across both modes, suggesting that interpretability is largely preserved under encryption. In addition, Figure 2 presents EEG channel-wise SHAP aggregation, highlighting key brain regions—such as the parietal and frontal lobes—that are commonly implicated in dyslexia. Notably, increased SHAP activity in frontal electrodes aligns with known associations between the prefrontal cortex and working memory or phonological processing. Likewise, heightened activity in parietal areas reflects the involvement of visuo-spatial processing, consistent with dyslexia-related neural patterns. This spatial representation provides clinically meaningful insight into which brain regions contribute most to model decisions, even when encrypted.

Further quantitative evidence of interpretability fidelity is

TABLE II
COMPARISON BETWEEN ENCRYPTED AND PLAINTEXT SHAP-LIKE VALUES

Metric	Value	Interpretation
MAE	4.95×10^{-3}	Moderate deviation in magnitude
Spearman Rank Correlation (ρ)	0.5885	Moderate positive correlation
p-value (for ρ)	8.41×10^{-8}	Highly statistically significant ($p \ll 0.01$)
Max Absolute Plaintext SHAP	≈ 0.0114	Context for interpreting MAE scale

provided in Table II, where the encrypted SHAP values are compared against plaintext values. The low Mean Absolute Error (4.95×10^{-3}) and moderate Spearman correlation ($\rho = 0.5885$) indicate a reliable approximation, while the highly significant p-value confirms the statistical strength of this correlation. These findings underscore that encrypted interpretability methods can preserve neurophysiological meaning and offer trustworthy clinical explanations.

Beyond performance and explainability, the security of the inference process is paramount. The CKKS scheme ensures that all computations on sensitive QEEG data are performed without decryption, mitigating risks such as data leakage, model inversion, or inference attacks. Since the cloud service provider (CSP) only interacts with ciphertexts, no raw signals or intermediate computations are exposed. The separation of encryption and decryption responsibilities—handled solely by the client through asymmetric keys—adds another layer of protection. Galois keys are tightly scoped to prevent misuse, ensuring the framework complies with data minimization and threat-resilient HE architectures.

However, this work also presents several limitations. EEG signals are inherently noisy and subject to inter-subject variability due to factors such as electrode placement, scalp impedance, or cognitive state, which may affect generalizability. Ethical concerns also arise when using EEG for cognitive monitoring, as such technology could be misused for intrusive surveillance or discriminatory profiling. In addition, the dataset size (200 subjects) is relatively small, which may restrict the strength of the conclusions. The current inference latency ($\sim$22s per sample) further limits real-time applicability. While a shallow ANN was intentionally employed to mitigate noise accumulation under homomorphic encryption, it may not capture more complex EEG patterns. Finally, the interpretability analysis was restricted to SHAP, leaving out other complementary explanation techniques.

The current latency ($\sim$21.9s per sample) remains a key limitation for real-time use. Although the use of a shallow ANN reduces homomorphic complexity, it may limit expressiveness for more nuanced EEG patterns. Lastly, the dataset used is limited in size and diversity; broader validation across multilingual and pediatric cohorts is necessary to assess the framework's clinical reliability.

V. Conclusion and Future Works

This paper presents a privacy-preserving and interpretable machine learning framework for dyslexia detection using QEEG data. By combining HE and neural network classification, the proposed system enables encrypted inference while maintaining data confidentiality.

The framework achieved 90.03% accuracy and an AUC of 0.8218 in encrypted mode, with a strong cross-validation performance (89.53%). Encrypted SHAP values demonstrated high consistency with plaintext explanations (MAE of 4.95×10^{-3}, Spearman correlation of 0.5885), validating interpretability under encryption. Additionally, SHAP aggregation highlighted frontal and parietal EEG regions aligned with known dyslexia markers.

Future research will address these limitations by extending evaluation to larger, more diverse datasets, optimizing homomorphic computations to reduce inference time, and benchmarking against deeper or alternative neural architectures. Moreover, incorporating other explainability approaches—such as LIME, Grad-CAM, or counterfactual methods—will enrich clinical interpretability. In parallel, optimization strategies such as ciphertext packing, batching, and hardware acceleration will be explored to reduce latency and enable more practical deployments.

In summary, this study shows that combining HE with XAI is feasible for secure and interpretable EEG-based dyslexia detection. With further optimization and validation on larger populations, the framework has the potential to support trustworthy AI-driven tools in healthcare.

Ethics Approval and Consent to Participation

The study protocol was approved by the Yeditepe University ethics committee, and the clinical trial was registered with the Turkey Pharmaceuticals and Medical Devices Agency (Nbr: 71146310-511.06, 2.11.2018). Prior to participating in the study, all participants were fully informed about the experimental procedure by the research ethics committee and provided their informed consent.

References

[1] J. Scheibner, M. Ienca, and E. Vayena, "Health data privacy through homomorphic encryption and distributed ledger computing: an ethical-legal qualitative expert assessment study," *BMC Medical Ethics*, vol. 23, no. 1, p. 121, 2022.

[2] K. Munjal and R. Bhatia, "A systematic review of homomorphic encryption and its contributions in healthcare industry," *Complex & Intelligent Systems*, vol. 9, no. 4, pp. 3759–3786, 2023.

[3] O. L. Usman and R. C. Muniyandi, "Cryptodl: Predicting dyslexia biomarkers from encrypted neuroimaging dataset using energy-efficient residue number system and deep convolutional neural network," *Symmetry*, vol. 12, no. 5, p. 836, 2020.

[4] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, "Privacy-preserving classification method for neural-biomarkers using homomorphic residue number system CNN: HORNS-CNN," in *Proc. 2022 Int. Conf. Business Analytics for Technology and Security (ICBATS)*, pp. 1–8, 2022.

[5] D. Mercier, A. Lucieri, M. Munir, A. Dengel, and S. Ahmed, "Evaluating privacy-preserving machine learning in critical infrastructures: A case study on time-series classification," *IEEE Trans. Ind. Informat.*, vol. 18, no. 11, pp. 7834–7842, 2021.

[6] N. J. Gallego-Molina, A. Ortiz, J. E. Arco, F. J. Martinez-Murcia, and W. L. Woo, "Unraveling brain synchronisation dynamics by explainable neural networks using EEG signals: Application to dyslexia diagnosis," *Interdisciplinary Sciences: Computational Life Sciences*, pp. 1–14, 2024.

[7] M. Robaa, M. Balat, R. Awaad, E. Omar, and S. A. Aly, "Explainable AI in handwriting detection for dyslexia using transfer learning," *arXiv preprint* arXiv:2410.19821, 2024.

[8] M. R. Abou Harb and B. Celiktas, "ANN activation function estimators for homomorphic encrypted inference," *IEEE Access*, vol. 13, pp. 103512–103528, June 2025, doi: 10.1109/ACCESS.2025.3579667.

[9] M. R. Abou Harb, B. Celiktas, and G. Eroglu, "Privacy-preserving dyslexia detection using homomorphic encryption for QEEG data with explainable AI insights via SHAP," unpublished manuscript, under review, *The Journal of Supercomputing*, 2025.

[10] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[11] J. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Proc. Advances in Cryptology – ASIACRYPT 2017*, vol. 10624, pp. 409–437, Springer, 2017.

Simplifying Hydration Monitoring: Wearable Sensor for Accessible Dehydration Detection in Vulnerable Populations

Dilay Memiş[1], Zeynep İlke Akbay[1], Gülizar Çalışkan Bilgin[2*]

[1] *İzmir University of Economics, Faculty of Engineering, Department of Biomedical Engineering, Türkiye*
[2] *İzmir University of Economics, Faculty of Engineering, Department of Genetics and Bioengineering, Türkiye*

Abstract— Dehydration is a common but often underestimated health concern, especially among the elderly, who often lose their perception of thirst. Current hydration monitoring technologies are typically complex, intrusive, or dependent on external devices such as smartphones, limiting their suitability for older individuals. In this study, we developed a wearable, noninvasive, standalone device to detect hydration status. Utilizing a Galvanic Skin Response (GSR) sensor, the device measures change in skin conductivity correlated with hydration levels. The Arduino Nano processes sensor data and provides immediate feedback via one of three methods, LED, buzzer, or vibration, according to user preference. Its simplicity, low cost, and independence from external devices make this solution especially accessible and practical for elderly users.

Keywords— dehydration; GSR sensor; wearable device; Arduino Nano

I. INTRODUCTION

Dehydration is a widespread health issue that profoundly impacts both mental and physical well-being, impacting organ function, nutrient transport, and body temperature regulation [1]. Even mild dehydration (1–2% fluid loss) affects memory, focus, and reaction time, while severe cases may lead to heatstroke or kidney failure. Vulnerable populations, such as athletes, outdoor workers, and the elderly, are more susceptible, and often fail to recognize symptoms until they have progressed [2]. Consequently, early detection and real-time hydration monitoring are more than beneficial, they are essential for effective health management [3].

Current hydration tracking relies on self-assessment or reminders, both of which are unreliable [4]. Although advanced systems utilizing Bluetooth, blood tests or urine analysis offer greater precision, they are generally complex, invasive and dependent on external devices such as smartphones, which severely limit their practical applicability, especially for older individuals [4]. There is therefore an urgent need for a simpler, more accessible and non-invasive solution for real-time hydration monitoring.

To address this gap, we propose a wearable hydration sensor integrating skin moisture detection through GSR. In this system, an Arduino Nano processes data and provides immediate feedback in the form of customizable LED, buzzer or vibration alerts. The device is standalone, battery-powered, and requires no apps or wireless connectivity. This study focuses on efficient hardware integration, sensor calibration, and firmware optimization to deliver a reliable and user-friendly product. It aims to reduce dehydration risks and promote hydration habits, especially among elderly users or those in high-heat environments.

As well as reducing the risks of dehydration and fostering better hydration habits, particularly among the elderly and those in high heat environments, this innovative project also aligns with five UN Sustainable Development Goals (SDGs). The five goals are: Goal 3: Good Health and Well-being by enabling real-time health monitoring; Goal 7: Affordable and Clean Energy through its low-power, standalone design; Goal 8: Decent Work and Economic Growth by enhancing safety and productivity for outdoor workers; Goal 9: Industry, Innovation, and Infrastructure by presenting a scalable and innovative health technology solution; and Goal 13: Climate Action by contributing to resilience against heat-related physiological stress.

II. PROBLEM STATEMENT

Dehydration is a serious health concern that can cause fatigue, headaches, dizziness, confusion, muscle cramps, and decreased cognitive and physical performance. In severe circumstances, it may result in heat exhaustion, kidney damage, or even life-threatening conditions such as heatstroke. Athletes, outdoor laborers, and elderly individuals are at higher risk, as they may not recognize the symptoms until they become serious [5]. Despite these risks, current hydration monitoring methods remain inadequate, often relying on subjective self-assessment or irregular reminders, both of which are prone to errors.

The lack of a simple, real-time, and non-invasive hydration monitoring solution leaves individuals vulnerable to dehydration-related health complications. This highlights the pressing need for a more effective and accessible means to manage hydration levels.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

III. LITERATURE REVIEW

Monitoring hydration status has become an increasingly relevant research area due to its importance in maintaining health and preventing dehydration-related complications. The literature reports various wearable systems, , each aiming to track hydration through various physiological signals or behaviors. However, most of these are either too invasive, dependent on smartphones, or unsuitable for elderly users who require simplicity and autonomy.

AutoHydrate, a wearable system combining a smartwatch and throat microphone to recognize drinking behavior, was introduced in [6]. The system achieved high accuracy (91.5%) but had the drawbacks of relying on Bluetooth communication with a smartphone, and requiring user interaction, reducing its practicality for elderly use.

Another system consisted of smartwatch-compatible device that used galvanic skin response (GSR) and photoplethysmography (PPG) sensors to estimate hydration based on sweat analysis, developed in [7]. Although accurate, the device required smartphone connectivity and complex signal processing.
An RF-based wristband that detected hydration levels by measuring variations in signal strength (RSSI) was explored in [8]. However, its requirement for environmental calibration made it impractical for casual or daily use.
A thermal-sensing device that evaluated urine temperature using infrared sensors was proposed in [9]. Its accuracy was offset by its dependence on urine sample collection limits its use in real-time, wearable applications.

A multi-sensor system involving a smart mouthguard and chest patch for comprehensive hydration and vital sign monitoring was presented in [10]. This technologically advanced solution, however, is highly complex, expensive, and relies on smartphone-based data processing.

Table 1. Comparison of existing hydration monitoring systems and the proposed device

Study	Technology Used	Limitations	Proposed Device Advantage
Auto Hydrate [6]	Throat microphone, smartwatch	Needs phone, not standalone	Fully autonomous; no connectivity required
SpectroPhon-DBM [7]	GSR + PPG	Needs phone, signal processing	Simple GSR only; low-cost system
BioMindR RF Band [8]	RF signal strength (RSSI)	Sensitive to environment	No calibration needed
Thermal Urinalysis Device [9]	IR sensor urine temperature measurement	Requires urine sample	Continuous,wearable without samples
Mouthguard + Chest Patch System [10]	Mouth + chest multi-sensor	Expensive; requires multiple wearables	Single wrist device with intuitive alerts
Our Wearable Dehydration Sensor Device	GSR sensor + Arduino Nano + LED, buzzer, vibration	GSR affected by external factors like humidity and stress	Standalone, low-cost, non-invasive, customizable alerts

Compared to previous systems, the proposed device offers a unique combination of simplicity, autonomy, and accessibility. It operates fully independently, without the need for smartphones, Bluetooth, or external data platforms, which enhances usability, particularly for elderly users. Relying solely on a GSR sensor and a straightforward microcontroller, the low-cost and energy-efficient system provides real-time hydration alerts through multiple customizable feedback options. This design approach addresses the key limitations of existing solutions and makes the device suitable for daily use in real-world scenarios.

IV. METHODOLOGY

This section details the components and steps needed for constructing our wearable dehydration alert system, aiming for real-time, non-invasive hydration detection using low-cost electronics.

4.1 Sensor Selection and Justification

The device uses a GSR sensor to monitor skin conductance, which reflects sweat gland activity and moisture levels. When skin is dehydrated, conductivity drops, enabling GSR to serve as a non-invasive proxy for hydration. The sensor was chosen for its simplicity, affordability, low power needs, and Arduino compatibility.

However, accuracy can be affected by factors such as temperature, humidity, stress, and skin type [11]. Studies such as [6] show that pairing GSR with deep learning improves classification of hydration states with up to 97% accuracy. GSR therefore offers effective early warnings, although it is not medically diagnostic. Future versions may integrate additional biosensors (e.g., temperature or PPG) and machine learning for better accuracy and personalization.

4.2 System Components

The key components are a GSR sensor, Arduino Nano, LED, buzzer, vibration motor, and three buttons for alert modes.

The device is breadboard-based and powered by a 9V battery.

4.3 Circuit Integration

The GSR sensor connects to Arduino's analog pin (e.g., A0) and is sampled periodically [12]. However, it is important to note that certain factors, such as temperature, humidity, stress, and skin type can affect accuracy. If readings fall below threshold, the chosen alert (LED, buzzer, or vibration) is activated. Debounce delays prevent repeated triggers.

4.4 Power Management

A 9V battery supplies power via Arduino's voltage regulator. All parts are optimized for low energy consumption.

4.5 Future Extension Possibilities

Potential improvements to the system include a more compact PCB design, added SMT components, and a Li-Po battery for better portability. A flexible, waterproof wristband and 3D-printed casing can enhance wearability.

Software-wise, AI could be used to personalize hydration alerts and reduce false positives. users with disabilities could be supported with accessibility features like haptic/voice output or app sync. Additionally, data logging via SD or BLE could enable remote monitoring and long-term trend analysis.

V. Results and Discussion

This section outlines the results obtained from the implementation and testing of the wearable dehydration monitoring prototype. Findings are presented with respect to hardware performance, the responsiveness of the alert mechanisms, and a comparative assessment against previously proposed hydration monitoring systems.

5.1. Functional Prototype

The wearable dehydration alert system was successfully assembled using low-cost, commercially available components. The prototype includes an Arduino Nano connected to a GSR sensor and three output devices: a red LED, a buzzer, and a vibration motor [13]. The circuit was mounted on a breadboard and powered by a 9V battery.

Users can select their preferred alert mode by selecting one of three pushbuttons. The selected output module is triggered once the GSR sensor detects a conductivity value below a predefined threshold. This modular setup ensures clear, immediate feedback in various real-world scenarios. The design facilitated switching between visual, auditory, and tactile alerts.

5.2. Observations and Performance

During testing, the device continuously measured skin conductance and compared values against the static threshold. When dehydration was simulated (e.g., using dry fingers or by reducing skin moisture), the sensor readings dropped, activating the corresponding feedback mechanism (LED, buzzer, or vibration motor). The observed GSR values typically ranged between 580–650 under normal skin conditions, falling to below 400 with a decrease in skin moisture. The system performed reliably in indoor environments and was easy to operate. A 5-second interval was implemented to reset the alert cycle and minimize false or repeated warnings.

5.3. Practical and Scientific Implications

The results of this study demonstrate that a simple GSR-based sensor, when integrated with a low-cost microcontroller, can provide a functional and accessible method for estimating hydration status in real time. Although not able to provide clinically diagnostic precision, GSR is effective in detecting relative changes in skin moisture that correspond to early signs of dehydration [14].

From a practical standpoint, the device offers a significant advantage for users, particularly the elderly, by eliminating the need for smartphones, wireless connectivity, or complex interfaces. The selectable alert modes (LED, buzzer, vibration) enhance accessibility and allow users to choose their preferred feedback method based on personal needs or sensory limitations.

Scientifically, the system shows that basic physiological signals, such as skin conductance, can be utilized for health monitoring, avoiding complexity and cost associated with more advanced wearable technologies. While more sophisticated systems using multiple sensors or cloud-based analytics may offer higher accuracy, the trade-off in simplicity and independence often renders them unsuitable for elderly or vulnerable users. Balancing technical feasibility with usability, our system is an effective tool for promoting hydration awareness and for reducing dehydration-related health risks in daily life.

5.4. Factors That Influence Results

Five factors can affect the accuracy of GSR-based hydration detection. First, environmental conditions such as ambient temperature and humidity may influence skin moisture levels, leading to false alerts in extremely dry or cold settings. Secondly, baseline variability can also result from user-specific physiological differences, including skin type and sweat gland activity. Thirdly, poor sensor contacts or motion during use may introduce noise into the measurements. fourth, the system uses a fixed threshold and a 5-second reset interval, which helps reduce repeated alerts; however, this may limit sensitivity to short-term hydration changes. Finally, the prototype does not perform personalized

calibration, therefore, it may not fully adapt to individual differences without further development.

5.5. Institutional Support and Applicability

In conclusion, the developed prototype demonstrates that a low-cost, non-invasive, and standalone dehydration monitoring device is both feasible and functional. Designed with elderly users in mind, the system successfully detects skin conductance changes associated with dehydration and provides immediate feedback through customizable alert modes. It is suitable for daily use due to its simplicity, portability, and independence from smartphones.

The current prototype is based on basic components and breadboard assembly, and its usability and reliability could be improved in areas such as sensor integration, ergonomic design, and data logging. Given the absence of government-supported dehydration monitoring tools in Turkey, this study also holds potential for institutional support and public health funding. With proper certification and clinical validation, the device could contribute to improved hydration awareness and reduced dehydration-related health risks in vulnerable populations.

VI. CONCLUSION

This study successfully demonstrated the feasibility of a standalone, non-invasive, and user-friendly wearable device for dehydration monitoring, particularly suited for elderly individuals. Using a GSR sensor and an Arduino Nano, the system detects changes in skin conductance and allows users to select alerts through LED, buzzer, or vibration modes.

Despite limitations in hardware resources and development time, the prototype performed reliably in indoor testing scenarios. The system can evolve into a viable product for real-world use with improvements such as miniaturization, multi-sensor integration, and data storage. Its simplicity, low cost, and accessibility make it a strong candidate for public health applications and potential institutional support.

REFERENCES

[1] B. M. Popkin, K. E. D'Anci, and I. H. Rosenberg, "Water, hydration, and health," *Nutrition Reviews*, vol. 68, no. 8, pp. 439–458, 2010.

[2] S. N. Thornton, "Thirst and hydration: Physiology and consequences of dysfunction," *Physiology & Behavior*, vol. 100, no. 1, pp. 15–21, 2010.

[3] O. R. Barley, D. W. Chapman, and C. R. Abbiss, "Reviewing the current methods of assessing hydration in athletes," *Journal of the International Society of Sports Nutrition*, vol. 17, no. 1, p. 52, 2020.

[4] J. C. Chiao, S. Bing, K. Chawang, and B. Crowe, "Thirsty for a noninvasive wearable to detect dehydration: A review," *IEEE Antennas and Propagation Magazine*, vol. 66, no. 1, pp. 54–68, 2024.

[5] D. J. Casa, L. E. Armstrong, S. K. Hillman, S. J. Montain, R. V. Reiff, and B. S. Rich, "National Athletic Trainers' Association position statement: Fluid replacement for athletes," *Journal of Athletic Training*, vol. 35, no. 2, pp. 212–224, 2000.

[6] Y. Mengistu, M. Pham, H. M. Do, and W. Sheng, "AutoHydrate: A wearable hydration monitoring system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, Korea, Oct. 2016, pp. 1857–1862. doi: 10.1109/IROS.2016.7759286.

[7] D. Rodin, Y. Shapiro, A. Pinhasov, A. Kreinin, and M. Kirby, "An accurate wearable hydration sensor: Real-world evaluation of practical use," *PLoS ONE*, vol. 17, no. 8, p. e0272646, 2022. doi: 10.1371/journal.pone.0272646.

[8] J. Wang, Z. Zilic, and Y. Shu, "Evaluation of an RF wearable device for non-invasive real-time hydration monitoring," in *Proc. IEEE Int. Conf. on Electrical and Computer Engineering (ICECE)*, 2017, pp. 91–94. doi: 10.1109/ICECE.2017.8262402.

[9] A. Yudhana, S. A. Akbar, F. Nuraisyah, and M. R. Sagita, "Self-health monitoring based on dehydration level and glucose content in urine using thermal sensing," in *Proc. 7th Int. Conf. on Engineering and Emerging Technologies (ICEET)*, Istanbul, Turkey, Oct. 2021, pp. 1–6. doi: 10.1109/ICEET53442.2021.9659731.

[10] K. R. Kim et al., "All-in-one, wireless, multi-sensor integrated athlete health monitor for real-time continuous detection of dehydration and physiological stress," *Advanced Science*, vol. 11, no. 33, p. e2403238, 2024. doi: 10.1002/advs.202403238.

[11] H. L. Tey, R. Wanitphakdeedecha, and F. Al-Niaimi, "Skin hydration and measurement methods," *Clinical, Cosmetic and Investigational Dermatology*, vol. 14, pp. 63–75, 2021.

[12] Raspberry Pi Foundation and Arduino LLC, *Arduino Sensor Kit Documentation*, 2022.

[13] M. Banzi and M. Shiloh, *Getting Started with Arduino*, 3rd ed., O'Reilly Media, 2014.

[14] U. Liaqat, M. Aamir, S. Kwon, and H. Kim, "Non-invasive estimation of human hydration levels using galvanic skin response and a hybrid deep learning model," *Scientific Reports*, vol. 12, p. 4534, 2022.

Evaluating Respiratory Effects of 3D-Printed Scoliosis Bracing in Different Physiological States

Büşra Sağlam
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
busrasgllm@gmail.com

Samet Çıklaçandır
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
samet.cikla@gmail.com

***Abstract*—Respiratory function is a critical physiological parameter often affected in individuals using scoliosis braces due to thoracic compression. Assessing respiratory parameters, especially Vital Capacity (VC), is essential to understand how orthotic interventions influence lung function. This study investigates the impact of scoliosis brace compression on VC under varying tightness levels and physiological states. A personalized 3D-printed orthosis was designed for a subject with lumbar scoliosis and fabricated using Fused Deposition Modeling (FDM) with polypropylene material. Respiratory measurements were conducted using the BIOPAC SS11LB airflow sensor while the subject was seated in a relaxed position. VC was recorded in both fasted and postprandial states across three brace tightness levels: loose, medium, and tight. Results showed a significant reduction in VC as brace tightness increased, with the highest reduction observed under tight brace conditions. Additionally, postprandial VC values were consistently lower than fasted measurements, indicating that physiological fullness further exacerbates thoracic restriction. These findings highlight the importance of considering both orthotic compression and nutritional status in respiratory assessments of scoliosis patients. The study provides valuable insights for developing patient-specific bracing strategies that balance spinal correction with minimal respiratory compromise.**

Keywords—Scoliosis brace, Vital Capacity, 3D printing, respiratory function, postprandial state, orthosis compression, personalized orthotics.

I. INTRODUCTION

Adolescent Idiopathic Scoliosis (AIS) is a complex three-dimensional deformity of the spine that not only alters spinal alignment but also significantly affects thoracic structure and pulmonary function. Particularly in thoracic scoliosis, the distortion of the rib cage, vertebral rotation, and reduced chest wall compliance contribute to restrictive ventilatory patterns [1,2]. Such respiratory limitations are especially critical during adolescence, a developmental stage essential for attaining full pulmonary potential [3].

In this context, the use of patient-specific scoliosis braces becomes increasingly important. Standard braces may not adequately accommodate the unique anatomical and physiological characteristics of each individual, potentially leading to excessive thoracic compression and further respiratory compromise. In contrast, personalized orthoses—designed based on the patient's spinal curvature, torso geometry, and clinical needs—offer a more balanced approach. These custom-made solutions aim to optimize spinal correction while minimizing negative impacts on respiratory function, thereby improving both therapeutic efficacy and patient comfort.

Conservative treatment approaches, notably rigid bracing, are widely used to prevent curve progression in AIS. However, the external compression applied by braces can further restrict thoracic expansion, potentially exacerbating existing pulmonary limitations [4]. Prior studies have demonstrated that braces can lead to reductions in lung volumes such as Forced Vital Capacity (FVC) and Forced Expiratory Volume in 1 second (FEV1), indicating a restrictive impact on breathing mechanics [5]. However, the specific effects of scoliosis braces on Vital Capacity (VC)—a direct measure of the maximal volume of air exhaled after a full inhalation—under relaxed, seated conditions remain insufficiently studied.

VC is a critical parameter in evaluating the maximal ventilatory capacity and thoracic mobility of individuals with scoliosis, especially under orthotic constraint. Unlike Tidal Volume, which measures normal, passive breathing, VC provides a clearer picture of the full expansion potential of the lungs, offering essential insights into the restrictive impact of brace application [1,5].

Monitoring respiratory function during brace use is particularly important because adolescents wearing scoliosis braces often engage in routine daily activities—such as sitting, walking, studying, and eating—while under orthotic constraint. During these activities, impaired thoracic mobility caused by the brace can lead to increased respiratory effort and reduced ventilatory efficiency. If unmonitored, such limitations may contribute to chronic discomfort, fatigue, and suboptimal pulmonary development. Therefore, assessing the impact of brace tightness on breathing mechanics under realistic conditions is crucial for ensuring both the clinical effectiveness and tolerability of scoliosis treatment.

To better illustrate the relationship between VC and other lung volumes, a graphical representation of the respiratory cycle is provided in Figure 1. This schematic clearly demonstrates how VC integrates the various lung capacities within the respiratory physiology framework.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

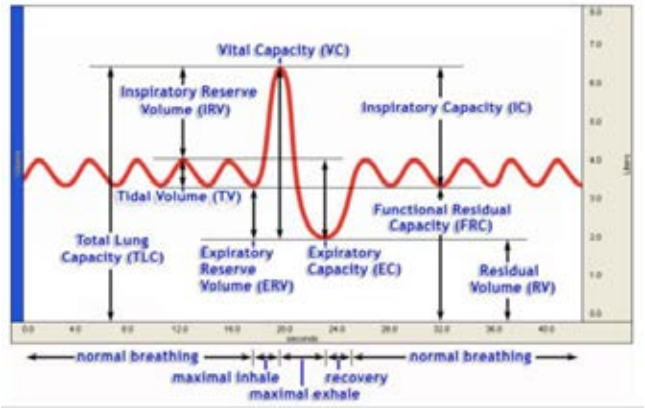


Figure 1. Graphical representation of lung volumes within the respiratory cycle [1]

By focusing solely on VC assessments under static conditions, this study aims to provide empirical evidence on how orthotic compression affects respiratory mechanics in AIS. The findings are anticipated to guide clinicians and orthotists in optimizing brace designs and adjustment protocols that safeguard respiratory health while ensuring effective spinal correction.

II. Materials and Methods

This study was conducted on a female subject presenting a 10° lumbar scoliosis curvature, verified through radiographic analysis based on the Cobb method, which is the gold standard for quantifying spinal curvature by measuring the angle between the most tilted vertebrae above and below the curve apex.. The experimental objective was to investigate the effects of varying scoliosis brace tightness on pulmonary function, specifically Vital Capacity, in a relaxed, seated posture.

Prior to respiratory data collection, the subject wore the custom-designed scoliosis orthosis for a continuous period of 42 hours to ensure physiological adaptation and minimize acute stress responses, surpassing the generally recommended 24-hour familiarization period. The orthosis was fabricated through a workflow beginning with 3D scanning, utilizing both photogrammetry and structured-light laser scanning, ensuring high anatomical fidelity. The obtained mesh data were processed in a Computer-Aided Design (CAD) environment, where the brace was parametrically modeled with ventilation zones, corrective force regions, and adjustable straps.

Manufacturing was carried out using Fused Deposition Modeling (FDM) with polypropylene (PP) filament. The print parameters included a nozzle temperature of 200–250°C, bed temperature of 80°C, 0.2 mm layer height, and 2% infill density. Post-processing involved sanding with P100 abrasive paper, further abrasion, and 600-grit wet polishing for skin contact safety.

For respiratory data collection, a BIOPAC SS11LB Airflow Transducer was connected to the BIOPAC MP36 data acquisition system. The airflow sensor, with a flow range of ±115 L/min and sensitivity of 0.5 mV/L/min, was externally positioned via a mouthpiece. A nose clip was applied to prevent nasal airflow, ensuring complete measurement of oral ventilation. During each trial, the subject performed a maximal inhalation followed by a maximal exhalation to determine VC. Measurements were conducted in a seated, relaxed posture to control for postural influence. The experimental configuration is depicted in Figure 2.



Figure 2. Experimental setup showing the BIOPAC SS11LB airflow transducer connected to the MP36 system for VC measurement during brace wear.

Three brace tightness conditions—Loose, Medium, and Tight—were applied by adjusting the Velcro straps, and VC was recorded for each condition. For every tightness level, the subject completed three measurement trials, each lasting approximately one minute, with the highest VC value retained for analysis. All procedures were conducted under standardized laboratory conditions to ensure data consistency and reliability.

The collected airflow data were sampled at a frequency of 2000 Hz and exported as .txt files for signal processing in MATLAB. The processing pipeline involved:

- Application of a low-pass filter to remove high-frequency noise.
- Detection of inspiratory and expiratory peaks to delineate respiratory cycles.
- Calculation of VC by measuring the maximal amplitude difference between full inhalation and exhalation.
- Visualization through time-domain plots to assess waveform integrity and comparative analysis across tightness conditions.

This data processing was performed using custom MATLAB scripts, ensuring reproducibility and standardized analysis across all experimental sessions. The processed data were subsequently analyzed statistically to identify the impact of orthotic compression on VC. Since the respiratory data were collected from a single subject across repeated trials under varying conditions, formal parametric tests such as t-tests were not applicable due to insufficient degrees of freedom. Instead, descriptive statistics—including mean, range, and percentage change—were calculated to interpret trends across conditions. Additionally, one-way ANOVA was applied illustratively to assess the variability in mean VC values under different brace tightness and nutritional states, providing preliminary insight into potential condition-dependent effects. Statistical methodologies are detailed in the subsequent section.

III. RESULTS

In this study, the effects of varying scoliosis brace tightness on Vital Capacity (VC) were examined under two physiological conditions: hungry and full states. Additionally, a baseline measurement was recorded in the without-brace condition to establish the subject's maximum pulmonary capacity.

In the hungry, without-brace condition, the subject achieved a VC of 3.884 liters, representing the highest capacity recorded throughout the study. This measurement served as a critical reference point to evaluate the restrictive effects of the brace. Although the VC measurement for the postprandial, without-brace condition was not available at the time of reporting, it will be included in future analyses to complete the dataset.

Upon application of the 3D-printed scoliosis brace, a clear and progressive reduction in VC was observed as the tightness of the brace increased. In the fasted state, the VC was measured as 2.362 L in the loose condition, 1.405 L in the medium condition, and 1.162 L in the tight condition. This reflects a 70% reduction in VC from the unbraced state to the tight brace condition, underscoring the significant mechanical restriction imposed by maximal brace compression.

Similarly, in the postprandial state, VC measurements exhibited a downward trend with increasing brace tightness. The VC was recorded as 1.382 L with the brace loosely fitted, 1.019 L at medium tightness, and 0.743 L under tight conditions. Notably, the postprandial VC values were consistently lower than the corresponding fasted values across all levels of brace tightness. This suggests that gastrointestinal distension following food intake may further restrict thoracic expansion, exacerbating the brace's restrictive effect on lung capacity. All recorded VC values across the experimental conditions are summarized in Table 1, enabling a direct comparison between states and brace tightness levels.

TABLE I. VITAL CAPACITY MEASUREMENTS BY BRACE TIGHTNESS AND NUTRITIONAL STATE.

Condition	VC (L)
Without Brace - Hungry	3.884
Without Brace - Full	2.319
Brace Loose - Hungry	2.362
Brace Medium - Hungry	1.405
Brace Tight - Hungry	1.162
Brace Loose - Full	1.382
Brace Medium - Full	1.019
Brace Tight - Full	0.743

To visually represent these trends, the data are plotted in Figure 3, illustrating the sharp decline in VC with increasing brace tightness and the additional restrictive influence of the postprandial state.

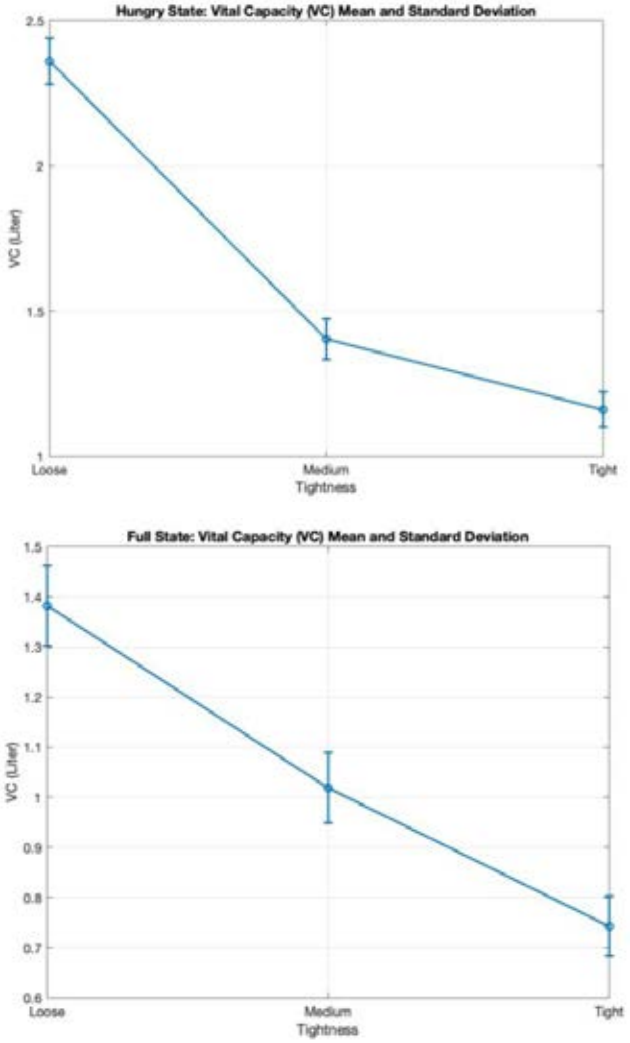


Figure 3. Comparison of Vital Capacity (VC) across brace tightness levels in both hungry and full states

These findings clearly demonstrate the restrictive impact of scoliosis bracing on pulmonary capacity, which is further compounded by physiological states such as postprandial fullness. The data highlight the importance of considering both brace adjustment and patient physiological state when evaluating respiratory function in brace-treated scoliosis patients.

IV. DISCUSSION AND CONCLUSION

This study presents an empirical investigation into the effects of scoliosis brace compression on Vital Capacity, uniquely contextualized within both fasted and postprandial physiological states. The findings corroborate and extend existing literature that underscores the restrictive impact of orthotic interventions on pulmonary function in individuals with scoliosis.

In line with previous studies, a progressive reduction in VC was observed with increasing brace tightness. For instance, Kennedy et al. (1989) demonstrated that thoraco-lumbar braces significantly reduce respiratory mechanics in adolescents with idiopathic scoliosis, with notable decrements in VC under rigid

bracing conditions [5]. Similarly, Noble-Jamieson et al. (1986) reported that the application of spinal braces in neuromuscular patients led to impaired respiratory capacity, attributed to diminished thoracic expansion [6]. The present findings, revealing up to a 70% reduction in VC from the unbraced to the tightest brace condition, are consistent with these earlier observations, albeit with a more personalized and systematically calibrated orthotic approach.

A distinct contribution of this study lies in the integration of the postprandial state as an experimental variable. The data demonstrated that VC values in the postprandial condition were consistently lower than their fasted counterparts across all brace tightness levels. This aligns with Malbrain (2007), who detailed the adverse respiratory effects of elevated intra-abdominal pressure, suggesting that gastric distension post-meal can further compromise diaphragmatic excursion and thoracic compliance [7]. This dual influence—orthotic compression combined with physiological fullness—highlights the importance of considering nutritional timing and patient state during orthotic fitting and respiratory assessments.

Another pivotal insight from this study is the use of a customized 3D-printed brace, offering a tailored fit that likely amplifies the restrictive effect compared to conventional, mass-produced orthoses. The precision of the 3D printing process and the inclusion of adjustable straps allowed for controlled modulation of tightness, which is often lacking in standard bracing approaches. This personalized design strategy may explain the relatively higher percentage of VC reduction compared to prior reports, such as the 15-20% VC reduction observed with Thoraco-Lumbo-Sacral Orthoses (TLSO) in traditional designs.

Moreover, by employing a BIOPAC SS11LB airflow sensor with a high sampling frequency and advanced MATLAB-based signal processing, this study ensured high-resolution quantification of VC changes. This methodological rigor enhances the validity of the findings and paves the way for real-time respiratory monitoring systems that could be embedded in future smart orthoses. Despite these strengths, the study acknowledges certain limitations. The analysis was conducted on a single subject, limiting the generalizability of the findings. However, the directional trends observed are robust and align with established biomechanical principles, suggesting that the interaction between brace compression and physiological states warrants further investigation.

In comparison to Rozek-Piechura et al. (2024), who studied the effects of the Cheneau brace combined with Schroth exercises on pulmonary function, this study isolates the mechanical impact of the brace itself without confounding rehabilitation exercises, offering a pure assessment of orthotic restriction on VC [8]. Clinically, these findings emphasize the necessity for orthotists and clinicians to consider patient-specific factors such as nutritional status, daily activity patterns, and respiratory function when prescribing and adjusting scoliosis braces. The observed exacerbation of restrictive effects postprandially suggests that brace wear schedules might benefit from synchronization with patients' feeding times to minimize respiratory burden. Future research should expand this inquiry with a larger cohort, incorporate continuous respiratory monitoring across different activities and postures, and explore adaptive orthotic designs capable of dynamically modulating pressure based on physiological feedback.

In summary, this study advances the understanding of how scoliosis braces interact with respiratory mechanics, particularly vital capacity (VC), under varying compression levels and physiological states. According to the results, VC dropped from 3.884 L (Without Brace – Hungry) to 1.162 L (Brace Tight – Hungry), which corresponds to a 70.1% reduction, clearly illustrating the substantial restrictive effect of high brace tightness. Likewise, the postprandial (full) state led to consistently lower VC values across all brace conditions—for example, in the tight brace condition, VC declined from 1.162 L (Hungry) to 0.743 L (Full), a 36% additional reduction due to physiological fullness. These findings strongly indicate that both brace tightness and nutritional state significantly influence pulmonary function, and should therefore be key considerations during orthosis design, adjustment, and wear-time scheduling. The integration of 3D printing, embedded sensors, and physiological context in this study represents a step toward more adaptive, patient-specific scoliosis management strategies. Future studies should expand this framework using larger sample sizes and continuous monitoring approaches.

ACKNOWLEDGMENT

This study was funded by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) through the 2209-B Industry-Oriented Undergraduate Research Projects Program. The authors also express their gratitude to FITED Orthotics and Prosthetics Company for their technical assistance and partnership.

REFERENCES

[1] Biopac Student Lab. (n.d.). Retrieved July 10, 2025, from https://bme.eng.ankara.edu.tr/wp-content/uploads/sites/47/2024/10/L13-Pulmonary-Function-II.pdf

[2] Tsiligiannis, T., & Grivas, T. (2012). Pulmonary function in children with idiopathic scoliosis. *Scoliosis*, *7*(1), 7. https://doi.org/10.1186/1748-7161-7-7

[3] Karaali, E., Çiloğlu, O., Görgülü, F. F., & Ekiz, T. (2021). Ultrasonographic measurement of diaphragm thickness in patients with severe thoracic scoliosis. *Journal of ultrasound*, *24*(1), 75–79. https://doi.org/10.1007/s40477-020-00536-w

[4] Weinstein, S. L., Dolan, L. A., Cheng, J. C., Danielsson, A., & Morcuende, J. A. (2008). Adolescent idiopathic scoliosis. *Lancet (London, England)*, *371*(9623), 1527–1537. https://doi.org/10.1016/S0140-6736(08)60658-3

[5] Yagci, G., Demirkiran, G., & Yakut, Y. (2019). In-brace alterations of pulmonary functions in adolescents wearing a brace for idiopathic scoliosis. *Prosthetics and orthotics international*, *43*(4), 434–439. https://doi.org/10.1177/0309364619839856

[6] Yurt, Y., Yatar, İ., Malkoç, M., Yakut, Y., Mıhçıoğlu, S., & Koltak, C. (2021). The effect of brace treatment on pulmonary functions in adolescent idiopathic scoliosis: An 8-month follow-up study. *Journal of back and musculoskeletal rehabilitation*, *34*(5), 887–893. https://doi.org/10.3233/BMR-200195

[7] Noble-Jamieson, C. M., Heckmatt, J. Z., Dubowitz, V., & Silverman, M. (1986). Effects of posture and spinal bracing on respiratory function in neuromuscular disease. *Archives of disease in childhood*, *61*(2), 178–181. https://doi.org/10.1136/adc.61.2.178

[8] Malbrain M. L. (2004). Different techniques to measure intra-abdominal pressure (IAP): time for a critical re-appraisal. *Intensive care medicine*, *30*(3), 357–371. https://doi.org/10.1007/s00134-003-2107-2

[9] Badowska, A., Okrzymowska, P., Piatek-Krzywicka, E., Ostrowska, B., & Rozek-Piechura, K. (2024). The Effect of the Cheneau Brace on Respiratory Function in Girls with Adolescent Idiopathic Scoliosis Participating in a Schroth Exercise Program. *Journal of clinical medicine*, *13*(23), 7143. https://doi.org/10.3390/jcm13237143

Force Mapping of 3D-Printed Scoliosis Braces at Varying Tightness Levels Using Calibrated FSR Sensors

Büşra Sağlam
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
busrasgllm@gmail.com

Samet Çıklaçandır
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
samet.cikla@gmail.com

***Abstract*—This study proposes a personalized, 3D-printed scoliosis brace integrated with Force-Sensitive Resistor (FSR) sensors to quantify biomechanical stress under varying clinical conditions. Using 3D scanning, CAD modeling, and Fused Deposition Modeling (FDM) with polypropylene, a patient-specific orthosis was fabricated for a subject with lumbar scoliosis. Two calibrated FSR sensors were embedded bilaterally at the lumbar-costal junction, where maximal compressive forces are typically concentrated. Data collection was performed under nine combined conditions of brace tightness (loose, medium, tight), posture (sitting, standing), and nutritional state (fasting, full), with each scenario repeated thrice over five-minute sessions. A three-way ANOVA revealed that tightness level, posture, and nutritional state each had statistically significant effects on applied force ($p < 0.001$), and Cohen's d values confirmed strong effect sizes across comparisons. The highest pressures occurred during tight brace use in a full, seated state. This study demonstrates the feasibility of real-time stress mapping in scoliosis braces, informing both orthotic customization and clinical guidelines for optimal brace adjustment. Future studies will extend this approach to larger patient cohorts and explore adaptive braces capable of dynamically modulating pressure based on physiological changes.**

Keywords—scoliosis brace, force-sensitive resistor, orthotic force mapping, 3D printing, polypropylene, biomechanical stress, sensor calibration, personalized orthosis.

I. INTRODUCTION

Scoliosis is a multifactorial, three-dimensional spinal deformity defined by a lateral curvature of the spine greater than 10°, typically accompanied by vertebral rotation and sagittal misalignment [1]. The condition predominantly arises during adolescence, coinciding with rapid growth periods, and is most commonly diagnosed as Adolescent Idiopathic Scoliosis (AIS), affecting approximately 2–3% of the adolescent population, with a higher prevalence among females [2,3]. If left untreated, scoliosis can lead not only to postural and aesthetic concerns but also to respiratory dysfunction, reduced physical performance, and psychosocial issues [2].

Bracing constitutes the primary conservative treatment for moderate scoliosis cases, particularly in skeletally immature individuals [4]. Traditional orthotic systems, such as the Milwaukee, Boston, Lyon, and Chêneau braces, have been widely utilized for decades, each employing different biomechanical correction strategies [5,6]. Despite their clinical efficacy in halting curve progression, these braces often suffer from limitations in patient compliance due to discomfort, suboptimal pressure distribution, restricted mobility, and social stigma related to their bulky appearance [7,8,9].

Recent advances in digital technologies, including computer-aided design (CAD), 3D body scanning, and additive manufacturing (3D printing), have enabled the development of personalized scoliosis braces that promise enhanced anatomical conformity and biomechanical effectiveness [10,11,12]. However, a significant gap remains in the objective quantification of the mechanical forces that these orthoses exert on the body. Traditional evaluations primarily rely on radiographic imaging and clinical observations, which do not provide continuous or dynamic insights into the pressure applied by braces during daily activities.

FSR sensors offer a practical and efficient solution to this limitation. FSRs are thin, flexible sensors that detect and quantify force through changes in electrical resistance when subjected to pressure [13]. Their compactness (<0.5 mm thickness), mechanical flexibility, and responsiveness to varying loads make them ideal for integration into wearable orthotic systems [13,14,15]. When embedded within a scoliosis brace, FSR sensors can provide real-time, continuous monitoring of the pressure distribution across the brace–body interface, thereby enabling a data-driven approach to orthosis optimization.

This study aims to fill the current gap in scoliosis treatment by quantitatively measuring the biomechanical stresses generated by scoliosis braces using calibrated FSR sensors integrated into a personalized 3D-printed orthosis produced via Fused Deposition Modeling (FDM). Existing clinical assessments rely heavily on subjective feedback and static imaging, lacking real-time data on the pressures exerted by braces during daily activities. By systematically evaluating force distribution across different brace tightness levels, postures, and nutritional states, this research provides objective insights into how orthotic pressure dynamically interacts with the body. The findings are expected to guide the development of more effective, comfortable, and adaptive brace designs tailored to individual patient needs.

II. MATERIALS AND METHODS

This study utilized a customized scoliosis orthosis designed for an individual with a 10° lumbar curvature. The subject's torso

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

geometry was captured using a combination of photogrammetry and structured-light laser scanning, ensuring a high-fidelity 3D model of the anatomy [17]. The scan data were processed in a CAD environment, where the orthosis was digitally modeled with attention to corrective force zones, ventilation areas, and sensor integration sites [10,11]. The final design was exported in STL format for fabrication.

The orthosis was produced via Fused Deposition Modeling (FDM) using polypropylene (PP) filament due to its biocompatibility and mechanical resilience [18]. Printing was performed with a nozzle temperature of 200–250°C, a bed temperature of 80°C, and a 0.2 mm layer height. Post-processing involved mechanical sanding and wet polishing to enhance surface comfort.

To measure the forces applied by the brace, two FSR 402 series FSR sensors were embedded symmetrically at the lumbar-costal transition region, identified as the point of highest compressive stress during brace tightening [13,14,15]. The sensors were strategically placed at the lumbar-costal transition region—an anatomical site identified as being subjected to the highest compressive stress during brace tightening [14,15]. Figure 1 illustrates the precise zones within the orthosis where the sensors were positioned to capture pressure dynamics during use.

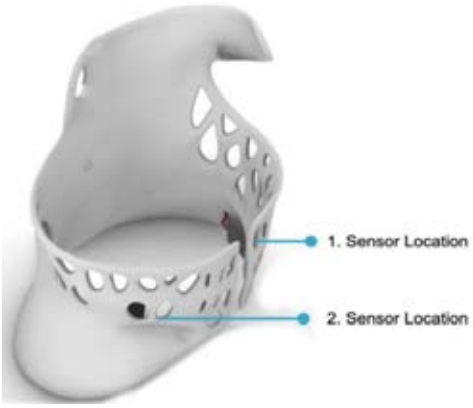


Figure 1. Visual representation of the sensor placement zones within the scoliosis orthosis

Additionally, Figure 2 shows the actual appearance of the personalized 3D-printed brace while worn by the subject, demonstrating the fit and anatomical conformity of the orthosis.

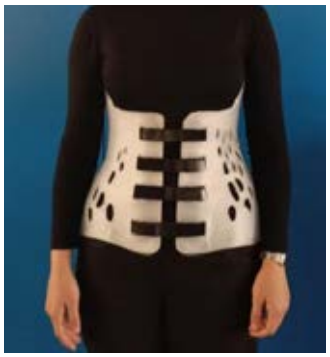


Figure 2. The personalized 3D-printed scoliosis brace worn by the subject during data acquisition.

The physical characteristics and specifications of the FSR sensors employed in this study are shown in Figure 3, highlighting their suitability for integration into orthotic designs without compromising structural integrity or patient comfort.

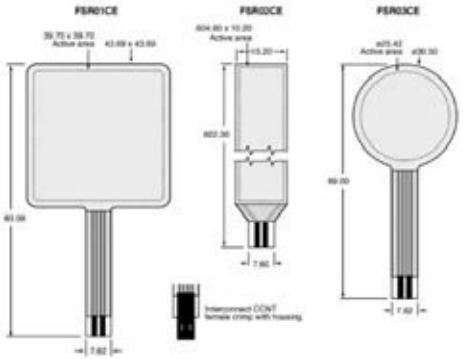


Figure 3. Dimensions of Different FSR Types [16]

Sensor calibration was carried out using a digital force gauge by applying known weights incrementally on each FSR sensor. Corresponding voltage outputs were recorded via Arduino's analog input pins, which were connected using a standard voltage divider circuit. The sensors were powered through the 5V pin of the Arduino, with signal output linked to analog pins A0 and A1 for FSR1 and FSR2, respectively. Calibration data were then processed in MATLAB, where a second-degree polynomial regression model was fitted to convert voltage readings into Newtons with high accuracy.

Force data were collected across different brace tightness levels (loose, medium, tight), under sitting and standing postures, and during fasting and full nutritional states. Each condition was recorded for five minutes and repeated three times to ensure consistency. The Arduino system continuously logged analog voltage outputs, which were then converted to force values via the calibration equation in MATLAB. Visualization and initial preprocessing of the data were also conducted in MATLAB.

For the statistical analysis, a three-way ANOVA was employed to examine the effects of brace tightness, posture, and nutritional state, and their possible interactions on the measured forces. Three-way ANOVA was chosen as it allows evaluation of not only the main effects of each independent variable but also their interaction effects on the dependent variable (force) [1]. Additionally, Cohen's d effect size was calculated for pairwise comparisons to quantify the magnitude of differences between tightness levels, providing a standardized measure of effect strength [2]. Pearson correlation coefficients were computed to assess the consistency between the two FSR sensors, ensuring the reliability of measurements across all conditions.

III. RESULTS

The force measurements obtained from the embedded FSR sensors revealed a consistent and significant increase in applied forces as the brace tightness progressed from loose to tight configurations. The highest force values were recorded under the tight-full-sitting condition, where the combined effects of

orthotic compression, posture, and postprandial abdominal expansion contributed to maximal stress on the torso.

Descriptive analysis showed that the average force measured by FSR1 increased from 1.99 N (loose) to 5.26 N (tight), while FSR2 increased from 1.30 N to 2.89 N under the same conditions. The minimum and maximum force values for each sensor across all conditions are presented in Figure 4, indicating the range of forces applied.

Analysis of stage (tightness) and condition (posture and nutritional state) interactions is depicted in Figure 5, illustrating how force values escalate, particularly under tight and full conditions.

Condition-wise comparisons of the measured forces are shown in Figure 6, highlighting the differential impacts of posture and feeding state on pressure levels. The full-sitting condition consistently produced the highest readings across both sensors.

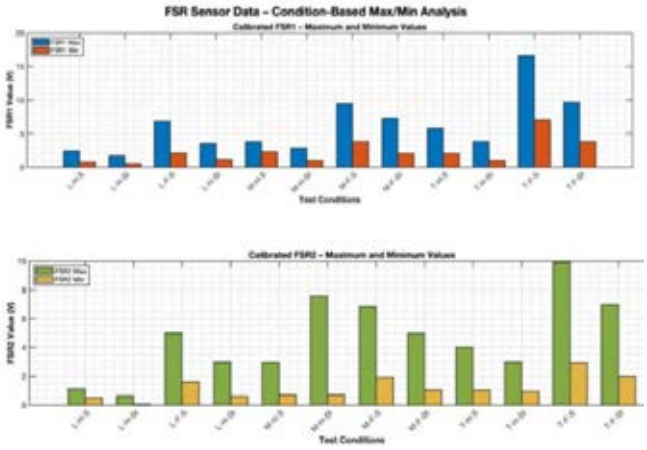


Figure 4. Minimum and Maximum Values of Force Sensors

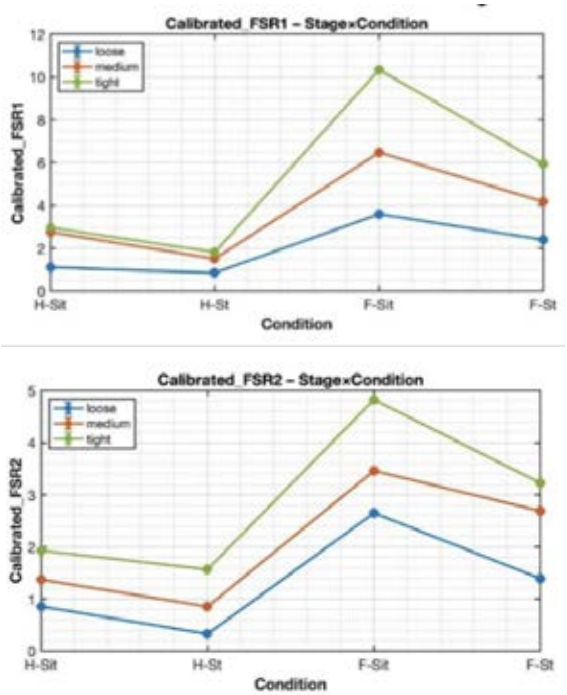


Figure 5. Stage x Condition Interaction of Force Sensors

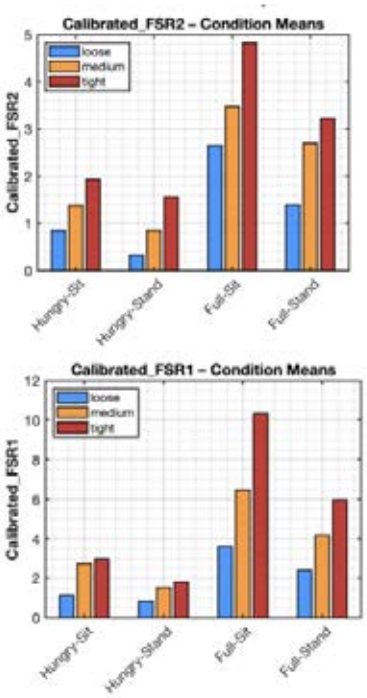


Figure 6. Condition x Wise Comparisons of Force Sensors

Statistical evaluation via three-way ANOVA confirmed that brace tightness, posture, and nutritional state each had significant main effects on applied force ($p < 0.001$), with large effect sizes ($\eta^2 > 0.20$). Moreover, strong positive correlations were observed between FSR1 and FSR2 outputs across all conditions ($r > 0.90$, $p < 0.001$), indicating measurement reliability, as shown in Table 1.

Collectively, these findings demonstrate that the magnitude of mechanical stress induced by the scoliosis orthosis is modulated not only by brace tightness but also by the user's physiological state and posture. The FSR sensor system effectively captured these dynamics, providing a comprehensive and quantifiable assessment of orthotic pressure behavior under varying conditions.

TABLE I. STATISTICAL SUMMARY OF MAIN EFFECTS AND PAIRWISE COMPARISONS

Comparison	Mean Force (N)	Cohen's d	p-value
Loose vs Medium (FSR1)	1.99 → 3.73	1.05	< 0.001
Loose vs Tight (FSR1)	1.99 → 5.26	1.26	< 0.001
Medium vs Tight (FSR1)	3.73 → 5.26	0.54	< 0.001
Loose vs Medium (FSR2)	1.30 → 2.10	0.71	< 0.001
Loose vs Tight (FSR2)	1.30 → 2.89	1.25	< 0.001
Medium vs Tight (FSR2)	2.10 → 2.89	0.57	< 0.001

IV. Discussion

This study offers a quantitative approach to evaluating scoliosis braces by integrating calibrated FSR sensors within a custom-designed, 3D-printed orthosis. While previous research has predominantly focused on radiographic assessments or qualitative clinical observations, the biomechanical stresses generated by braces during use remain insufficiently quantified in dynamic conditions. Although some prior studies have proposed pressure sensor systems for orthotic evaluation [14,15], the precise impact of variables such as brace tightness, posture, and nutritional state on localized pressure dynamics is not yet fully elucidated.

Our results demonstrate that increasing brace tightness leads to significant elevations in the applied forces, aligning with fundamental biomechanical principles that tighter adjustments increase corrective pressure. These findings are consistent with Fuss et al. (2021), who highlighted the potential of integrating pressure sensors into customized scoliosis braces for real-time monitoring, though their study lacked systematic evaluations across physiological states [15]. Furthermore, the observed modulation of applied forces by posture and nutritional state suggests that the internal body environment and positioning critically influence orthotic stress distributions—an aspect not extensively covered in prior literature.

The strong correlation between the two symmetrically placed FSR sensors ($r > 0.90$) underscores the reliability and consistency of the measurement system developed in this study. The statistical significance of all three independent variables—brace stage, posture, and nutritional status—supports the robustness of the sensor-based methodology for orthotic evaluation. These insights align with the broader call in the literature for more dynamic and patient-responsive assessment methods in conservative scoliosis treatment [14, 15].

By enabling real-time biomechanical data acquisition, this approach addresses a critical gap in clinical practice: the lack of continuous and dynamic monitoring of orthotic pressures during daily activities. Such data-driven insights could inform the development of adaptive or intelligent brace systems capable of modulating force application based on the patient's physiological and postural variations, as advocated by Giggins et al. (2014) in the context of rehabilitation technologies [16].

Compared to traditional assessments that often rely on subjective patient feedback or static imaging, this study contributes an empirical foundation for optimizing orthotic designs and personalization strategies. Nevertheless, further research with larger sample sizes and diverse patient profiles is warranted to generalize these findings and to explore the feasibility of integrating real-time feedback systems into routine clinical practice.

In conclusion, the present findings contribute to the evolving understanding of biomechanical interactions in scoliosis bracing and provide a stepping stone for future developments in sensor-enhanced orthotic solutions. This knowledge is expected to aid in refining brace designs, improving patient-specific adjustments, and advancing the broader field of intelligent, sensor-integrated medical devices.

Acknowledgment

This research was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under the 2209-B Industry-Oriented Undergraduate Research Projects Program. The authors also thank FITED Orthotics and Prosthetics Company for their technical support and collaboration.

References

[1] Negrini, S., et al. (2018). 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. *Scoliosis and spinal disorders*, *13*, 3. https://doi.org/10.1186/s13013-017-0145-8

[2] Weinstein, S. L., Dolan, L. A., Cheng, J. C., Danielsson, A., & Morcuende, J. A. (2008). Adolescent idiopathic scoliosis. *Lancet (London, England)*, *371*(9623), 1527–1537. https://doi.org/10.1016/S0140-6736(08)60658-3

[3] Lonstein J. E. (2006). Scoliosis: surgical versus nonsurgical treatment. *Clinical orthopaedics and related research*, *443*, 248–259. https://doi.org/10.1097/01.blo.0000198725.54891.73

[4] Negrini, S., Donzelli, S., Lusini, M., & Zaina, F. (2012). Bracing can reduce high-degree curves and improve aesthetics immediately after the end of growth. Final results of a retrospective case series. *Studies in health technology and informatics*, *176*, 393–396.

[5] Blount, W. P., Schmidt, A. C., Keever, E. D., & Leonard, E. T. (1958). The Milwaukee Brace In The Operative Treatment Of Scoliosis. *The Journal Of Bone And Joint Surgery. American Volume*, *40-A*(3), 511–525.

[6] Nathan, P., Chou, S. M., & Liu, G. (2023). A review on different methods of scoliosis brace fabrication. *Prosthetics and orthotics international*, *47*(4), 424–433. https://doi.org/10.1097/PXR.0000000000000195

[7] Moe, J. H., & Kettleson, D. N. (1970). Idiopathic scoliosis. Analysis of curve patterns and the preliminary results of Milwaukee-brace treatment in one hundred sixty-nine patients. *The Journal of bone and joint surgery. American volume*, *52*(8), 1509–1533.

[8] Stagnara, P. (1985). The Lyon brace: principles and results. *Scoliosis Research Society*.

[9] Reichel, D., & Schanz, J. (2003). Developmental psychological aspects of scoliosis treatment. *Pediatric rehabilitation*, *6*(3-4), 221–225. https://doi.org/10.1080/13638490310001644593

[10] Rigo, M. D., Villagrasa, M., & Gallo, D. (2010). A specific scoliosis classification correlating with brace treatment: description and reliability. *Scoliosis*, *5*(1), 1. https://doi.org/10.1186/1748-7161-5-1

[11] Rigo, M., & Weiss, H. R. (2008). The Chêneau concept of bracing--biomechanical aspects. *Studies in health technology and informatics*, *135*, 303–319.

[12] Hacioğullari, S. Ö., & Babaarslan, O. (2022). Güç Tutuşurluk Katki Maddesi İle Üretilen Polipropilen Filament İpliklerin Mekanik, Yapisal Ve Renk Özelliklerinin İncelenmesi. *Journal Of Textiles And Engineer*, *2*.

[13] Interlink Electronics. (2020). *FSR Force Sensing Resistor Integration Guide*. Retrieved from https://www.interlinkelectronics.com/fsr

[14] Fuss, F. K., Ahmad, A., Tan, A. M., Razman, R., & Weizman, Y. (2021). Pressure Sensor System for Customized Scoliosis Braces. *Sensors*, *21*(4), 1153. https://doi.org/10.3390/s21041153

[15] Giggins, O. M., Sweeney, K. T., & Caulfield, B. (2014). Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. *Journal of neuroengineering and rehabilitation*, *11*, 158. https://doi.org/10.1186/1743-0003-11-158

[16] Ohmite FSR Force-Sensing Resistors. (n.d.). Retrieved July 9, 2025, from https://www.mouser.com.tr/new/ohmite/ohmite-fsr-force-resistors/?srsltid=AfmBOooBHpbpxAAWXH9TmKUUFpDDRRlw8cioZ8hS30GkLwsnLocYgsBw

[17] Lauria, G., Sineo, L., & Ficarra, S. (2022). A detailed method for creating digital 3D models of human crania: an example of close-range photogrammetry based on the use of structure-from-motion (SfM) in virtual anthropology. *Archaeological and Anthropological Sciences*, *14*(3), 42.

[18] Shahrubudin, N., Te Chuan, L., & Ramlan, R. (2019). An overview on 3D printing technology: Technological, materials, and applications. *Procedia Manufacturing*, *35*, 1286–1296. https://doi.org/10.1016/j.promfg.2019.06.089

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Optimized Radix Sort for Efficient Health Data Sorting in Biomedical Backend Systems

Tugay Direk*
*Department of Software Engineering
Izmir University of Economics, Izmir, Turkiye
Email: tugay.direk@ieu.edu.tr

***Abstract*—Efficient data processing is essential in modern biomedical systems, especially as hospital information systems, genomic pipelines, and centralized clinical data platforms generate increasingly large volumes of health related information. This paper presents an optimized variant of the radix sort algorithm tailored for use in backend biomedical data handling systems. The algorithm maintains linear time complexity O(n) and offers improved performance over comparison based algorithms in specific scenarios involving high volume, mixed digit size integer inputs typical of centralized health data aggregation. Our method outperforms quicksort in runtime efficiency under large scale, heterogeneous numeric inputs, such as those found in genomic data sorting or patient record integration. However, due to its use of additional memory buffers, including an extra bucket for digit length categorization, our approach incurs moderately higher memory usage compared to in place quicksort implementations. We evaluate this trade off and demonstrate the performance advantage in use cases where memory availability is not a primary constraint. This contribution supports the development of faster biomedical informatics infrastructures, enabling more responsive analytics and scalable integration of heterogeneous medical data sources.**

***Keywords*—Radix Sort, Biomedical Informatics, Health Data Processing, Sorting Algorithms, Backend Systems, Genomic Data Analytics**

I. Introduction

Healthcare systems increasingly rely on large scale data collected from biomedical devices, hospital information systems, and research platforms. These data must be processed rapidly and accurately by centralized systems to support clinical decision making, research, and health record integration [1]. Efficient sorting of large scale genomic data is essential for numerous bioinformatics tasks including variant analysis, interval overlap detection, and feature merging. Sorting algorithms play a big role in this task. For instance, AIList data structure introduced by Feng et al. optimized interval search in genomics. Part of its design includes sorting intervals by start coordinate and then structuring them with precomputed end values to speed up overlap queries [2]. Another study by Mao et al. evaluates augmented two dimensional range trees for genomic interval queries, including refined operations based on Allen's interval algebra. The implementation builds range trees using interval start positions as primary keys, implicitly relying on sorted start coordinate data to optimize search operations [3]. Therefore, efficient sorting algorithms is a need for these types of work pipelines.

The study of sorting algorithms has a rich history, which has been extensively included in resources such as from Knuth [4] to Cormen et al. [5]. Sorting remains a cornerstone of data processing and analysis, with direct applications across biomedical engineering, particularly in genomic data processing, electronic health records (EHR) management, clinical signal analysis, and medical imaging [6], [7], [8]. Efficient sorting is also critical in feature alignment and biological sequence comparison where the order of elements carries biological or clinical meaning [9]. Traditional sorting algorithms, such as quick sort and merge sort, exhibit varying degrees of efficiency depending on the nature of the comparison approach. These algorithms rely on comparisons for all input types, which limits their potential for optimizations that could achieve linear time complexity. Radix sort, a non comparative integer sorting algorithm, has demonstrated superior performance in specific contexts. Firstly conceptualized by Harold H. Seward in 1954 [10], sorts integers by processing individual digits. It basically uses digit values from least significant digit through the most significant one and places the numbers in correct order after comparison of every digit. However, its effectiveness diminishes when handling arrays with integers of widely varying digit lengths. The algorithm proceeds as follows: it first identifies the number with the maximum digit length in the input array. This value determines the total number of digit wise passes to perform. Regardless of how many elements share this maximum digit length, even if only one does, all numbers are treated as having the same length by padding shorter values with leading zeros during the sorting process. This feature of the method makes it very inefficient on varying types of input. Because of this reason, radix sort faces performance degradation when applied to data arrays containing heterogeneous digit lengths, a common case in real world biomedical applications. For instance, the human genome exhibits considerable variability in chromosome sizes, ranging from large chromosomes such as chr1 (248 million base pairs) to relatively shorter ones like chr21 (46 million base pairs) [11]. Consequently, genomic coordinates span a wide range of digit lengths, from six digit to nine digit integers, creating challenges for radix based sorting algorithms when processing heterogeneous data. Similarly, hospital logs

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

and clinical telemetry data often include event timestamps with inconsistent digit representations due to varying time resolutions [12]. In these cases, radix sort treats all elements as having the maximum digit length in the array, padding shorter values with leading zeros. This introduces unnecessary computation and leads to performance inefficiencies.

Previous research has explored various enhancements to radix sort. Several optimizations were proposed for parallel radix sort algorithms, improving efficiency on multicore processors. Advancements by Lee et al. [13], Cho et al. [14] and Satish et al. [15] have focused on optimizing radix sort for modern hardware architectures. However, the challenge of handling integers with varying digit lengths remains relatively unexplored. To the extent of our research, digit length based grouping has not been formally studied or applied in the context of radix sorting, including in biomedical data applications. Our work introduces and evaluates this approach specifically for genomic intervals and health record identifiers, demonstrating clear performance gains.

This paper proposes a digit length aware radix sorting algorithm designed to better handle digit diverse integer arrays found in biomedical datasets. By introducing an initial categorization phase that segments data by digit length, the method executes radix sorting more efficiently on each subgroup. This strategy reduces redundant passes over shorter integers and improves runtime under practical data conditions.

We demonstrate the real world applicability of our method in genomic interval analysis, a routine task in bioinformatics and biomedical research, where start-end intervals from variant files (such as BED file type) must be sorted for intersection, merging, and feature mapping. These coordinates often span chromosomes with digit length imbalance, making the proposed method well suited for this domain [16].

II. Method

Genomic data analysis often involves operations like detecting overlapping features and computing feature intersections [16]. These tasks require coordinate based interval sorting where chromosome positions range from a few digits (10-500 base pairs (bp)) to 15 digit representations for long chromosomes or genome assemblies [11]. Tools like BEDTools [17] need sorted BED files for operations like intersect, merge, and subtract for rapid execution, making efficient sorting vital [18]. Another tool BEDOPS also requires sorted input for high performance genomic feature operations [19]. Sorting genomic intervals accurately and efficiently has real world implications in biomedical research. For example, when comparing Single Nucleotide Variants (SNVs) across samples, sorted inputs enable fast binary-search–style lookups and efficient overlap detection, as implemented by tools like Tabix which require position-sorted inputs for interval retrieval [20].

Since sorting the numbers is a significant task in mentioned real world problems, optimizing radix sort algorithm which has promising outcomes is a contribution to the field. The main problem in traditional radix sort arises from the initial assumption which requires every number must be assumed to have the same digit length of the number with maximum digit length. For this reason, zero is assigned for the missing digits for each number. The run time of radix sort is linear theoretically, but because of this assumption, it never reaches to its best case when even one number with high digit length is present in the input array. This assumption forces the algorithm for unnecessary iterations although we do not need. Proposed method solves the problem of having numbers with highly varying numbers of digits by separating each number to a different categorization based on their number of digits. The optimized radix sort algorithm consists of three main steps:

Firstly, the digit length categorization phase initializes a separate empty list to hold categories of integers based on their digit lengths and the input array is traversed, categorizing each integer by its digit length. Next, radix sort is applied to each category independently. Lastly, the sorted categories are merged to produce the final sorted array. This procedure actually mimics the human recognition for the numbers. To compare the numbers with different digit lengths, we just look at the digit length and know the one with higher value is greater. We do not need to compare the numbers with different digit lengths. This method also does not consider the numbers with different digit lengths and only focuses on numbers with the same digit lengths. By this key idea, the best case running time of radix sort can always be achieved independent from the varying input types. Pseudocode of the method is given in Algortihm 1:

Algorithm 1 OptimizedRadixSort(arr)

```
Initialize separate lists for each digit length, where digit_classes[i] will hold numbers with (i + 1) digits
For each number in arr:
        Determine the number of digits d in the number
        Append the number to digit_classes[d-1]
For each digit_class in digit_classes:
        If the digit_class is not empty:
                Apply RadixSort(digit_class)
Initialize sorted_arr as an empty list
For each digit_class in digit_classes:
        Append all numbers in the digit_classes to sorted_arr
```

III. Discussion

The running time of the optimized radix sort algorithm can be analyzed as follows;

Digit Length Categorization: Categorizing numbers by digit length has a time complexity of O(n), where n is the number of integers. For each number, determining its digit length requires at most 15 simple integer comparisons (assuming a maximum of 15 digits), which is much faster than costly string conversions.

Radix Sort on Categories: For each category, radix sort runs in O(d * (n + b)), where d is the number of digit, n is the input size and b is the base of the number. Since the b and d values

are small values for high input sizes, this asymptotic run time is considered to be linear. However, as mentioned in previous sections, for small number of integers with high digit lengths, this slows down the algorithm in considerable manner.

Merging Sorted Categories: This step involves concatenating the sorted arrays. Since each category would be sorted after radix sort. It will require only linear time to concatinate them.

The key idea for the best case running time in radix sort is achieved through the categorization of each number according to its digit length. By this procedure, unnecessary iterations for smaller numbers with less digit lengths are avoided and best running time is always guaranteed. Thus, the overall time complexity of the optimized radix sort is O(n). To evaluate the performance of the algorithm, experiments with varying input sizes were conducted. The results in Table I, II and III demonstrate the robust performance of the proposed method across different dataset sizes, validating the theoretical analysis. Merge sort, quick sort and traditional radix sort methods are considered for the evaluation of the method. For the comparison of these algorithms, an everyday personal computer that any person can have access was used and it has Intel core i5-6600K CPU at 3.30 GHz and 8 GB RAM.

We designed test arrays of integers for evaluation purposes. The majority are five digit or less, while a minority have up to 15 digits, mimicking genomic coordinate distributions. Analyses using the GenomicDistributions package on BED formatted data demonstrate that interval lengths in real datasets exhibit long tailed distributions, spanning from small (hundreds of base pairs) to very large genomic regions (hundreds of billions) [21]. This variability naturally translates into digital length heterogeneity when represented as integer coordinates for sorting, a key challenge addressed by our proposed digit-length-aware radix sort algorithm. Our test arrays replicate BED data characteristics from organisms with large genomes or long read sequencing technologies. These data are also available in the BEDTools sample intervals repository[1]. We synthetically generated BED files data exhibiting such features to realistically model digit length heterogeneity.

Genomic intervals in BED files often consist of a chromosome label and two integers representing start and end positions. For example:

```
chr1 123 150
chr2 45678 46000
chr3 123456789012345 123456789012360
```

Sorting is critical to run tools like BEDTools merge, which merges overlapping or adjacent intervals. These tools require lexicographical order on chromosome names and numerical order on start positions. Since intervals on different chromosomes can vary greatly in length (from hundreds to billions), sorting performance may degrade due to variable digit lengths. Our method improves this by focusing on the numeric digit length component.

[1]https://github.com/arq5x/bedtools2/blob/master/data/knownGene.hg18.chr21.bed

TABLE I: Running time comparison of sorting methods by varying input and digit sizes. Input array consists of equal number of integers for each digit length (s: seconds)

Input Size	Max Digit Length	Merge Sort	Quick Sort	Radix Sort	Optimized Radix Sort
1,000,000	15	8.7092 s	4.4872 s	16.4054 s	8.0623 s
1,000,000	12	9.3178 s	4.6163 s	11.7712 s	6.3523 s
1,000,000	8	9.3198 s	2.9651 s	7.2773 s	4.3360 s
1,000,000	5	9.1830 s	2.2092 s	4.7842 s	3.1801 s
10,000,000	15	106.8788 s	48.7653 s	165.4651 s	97.3188 s
10,000,000	12	126.5181 s	48.3542 s	164.3529 s	92.7429 s
10,000,000	8	117.3067 s	45.6204 s	84.8983 s	54.5895 s
10,000,000	5	111.7021 s	28.8122 s	46.8787 s	35.0197 s
10,000,000	3	108.2604 s	15.9188 s	26.1440 s	22.0906 s

Table I shows the results of different methods with varying input and digit sizes. When input size is one million, four different maximum digit lengths were considered. Input array consists of equal number of integers for each digit length.

As can be detected from the table, proposed method outperforms merge sort and radix sort in all maximum digit lengths. Quick sort performs better in this comparison. However, while the length of maximum digit length is reduced, our method closely approaches quick sort. The same phenomena can be observed for the input size of ten million.

TABLE II: Running time comparison of sorting methods by varying input and digit sizes. Input array consists of only 100 integers with at most 15 digits length and the others are at most 5 digits length.

Input Size	Merge Sort	Quick Sort	Radix Sort	Optimized Radix Sort
10,000,000	128.3958 s	79.6171 s	164.2375 s	66.8402 s
1,000,000	10.2567 s	5.4478 s	15.3449 s	6.1318 s
100,000	0.8134 s	0.3913 s	1.2569 s	0.5317 s

TABLE III: Running time comparison of sorting methods by varying input and digit sizes. Input array consists of only 10,000 integers with at most 15 digits length and the others are at most 5 digits length.

Input Size	Merge Sort	Quick Sort	Radix Sort	Optimized Radix Sort
10,000,000	127.2835 s	78.7963 s	167.2343 s	67.8085 s
1,000,000	9.9886 s	5.0389 s	14.9283 s	6.3450 s

Table II and III demonstrates the clear superiority of our method. Table II consists of inputs with only 100 integers with at most 15 digits length and the others are at most 5 digits length. Input sizes are one thousand, one million and ten million. When the input size is ten million, our method outperforms the quick sort by more than 13 seconds. This corresponds to a performance gain exceeding 15% in execution time. On the input sets with one thousand and one million, the proposed method and quick sort are very close to each other. The same pattern can also be seen in Table III. These results highlight the clear performance advantage of the proposed method.

The biomedical engineering domain, especially in genomics, provides a compelling use case for this optimization. Genomic interval processing tasks, such as merging overlapping intervals or identifying intersections between genomic features and known variants, frequently involve large datasets encoded in BED or VCF formats. These datasets inherently contain coordinate values with varying digit lengths (positions ranging from a few hundred to over a billion base pairs). Accurate and efficient sorting of these genomic intervals is critical for tools like BEDTools, BEDOPS, and others that rely on linear ordering of features to perform interval comparisons and overlaps. By employing the proposed sorting algorithm, such preprocessing steps can become significantly more time and memory efficient, particularly when scaled to whole genome analyses.

While the primary application context discussed has been genomic data, the proposed method may also be applicable in clinical data systems where sorting variable length numeric keys (such as national IDs or hospital specific patient codes) is required. Although these identifiers might not exhibit the same degree of digit variability or sorting load as genomic coordinates, systems with large scale patient data could still benefit from the performance enhancements offered by digit aware sorting methods.

Future works can include extending this method to parallel or distributed computing environments, which are increasingly used in genomics and biomedical data processing. It can also be integrated into established genomic libraries or clinical informatics platforms to evaluate real world system level performance.

IV. Conclusion

This paper presents an optimized radix sort algorithm for arrays containing integers with varying digit lengths. By categorizing integers based on digit length, applying radix sort to each category, and merging the results, the algorithm achieves a time complexity of O(n) theoretically. Experimental results confirm the algorithm's efficiency across different input sizes, making it a robust choice for sorting large datasets with diverse digit lengths. Close and sometimes better results of quick sort algorithm is based on its implementation details which does not require high overhead. However, based on the types of inputs, the proposed method outperforms quick sort although it requires more overhead such as the inclusion of extra arrays to hold each category and merging at the last step. Parallel computing with threading may be a future work to increase its time efficiency even more. A different thread can be initialized for each digit class and better results can be achieved. This work opens up a promising direction for bridging the efficiency gap in sorting operations across biomedical domains where numerical inputs with variable digit lengths are common.

References

[1] B. Kornelia and S. Andrzej, "The use of big data analytics in healthcare," *Journal of Big Data*, 2022. [Online]. Available: https://doi.org/10.1186/s40537-021-00553-4

[2] J. Feng, A. Ratan, and N. C. Sheffield, "Augmented interval list: a novel data structure for efficient genomic interval search," *Bioinformatics*, vol. 35, no. 23, pp. 4907–4911, 05 2019. [Online]. Available: https://doi.org/10.1093/bioinformatics/btz407

[3] M. Chengsheng, E. Alal, and L. Yuan, "Efficient genomic interval queries using augmented range trees," *Scientific Reports*, 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-41451-3

[4] D. E. Knuth, *The art of computer programming, volume 3: (2nd ed.) sorting and searching*. USA: Addison Wesley Longman Publishing Co., Inc., 1998.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 2009.

[6] B. A. G. et al, "Genomic data in the all of us research program," *Nature*, 2019. [Online]. Available: https://doi.org/10.1038/s41586-023-06957-x

[7] R. MK, W. W, and O.-M. L., ""big data" and the electronic health record," *Yearb Med Inform.*, 2014. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4287068/

[8] M. Santos and N. P. Rocha, "A big data approach to explore medical imaging repositories based on dicom," *Procedia Computer Science*, vol. 219, pp. 1224–1231, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923004143

[9] S. H. et al, "Next-generation sequencing technology: Current trends and advancements." *PubMed-not-MEDLINE*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10376292/

[10] A. Cassar and Emmanuel, "In search of the fastest sorting algorithm," *Symposia Melitensia*, 2018. [Online]. Available: https://www.um.edu.mt/library/oar//handle/123456789/30001

[11] W. L. G, G. J. G. R, H. K, T. S, H. T, and H. J. L, "The vertebrate genome annotation (vega) database," *Nucleic Acids Research*, 11 2007. [Online]. Available: https://doi.org/10.1093/nar/gkm987

[12] G. A. et al, "Timing errors and temporal uncertainty in clinical databases-a narrative review." *PubMed-not-MEDLINE*, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9433547/

[13] S.-J. Lee, M. Jeon, D. Kim, and A. Sohn, "Partitioned parallel radix sort," *Journal of Parallel and Distributed Computing*, vol. 62, no. 4, pp. 656–668, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731501918088

[14] M. Cho, D. Brand, R. Bordawekar, U. Finkler, V. Kulandaisamy, and R. Puri, "Paradis: an efficient parallel algorithm for in-place radix sort," *Proc. VLDB Endow.*, vol. 8, no. 12, p. 1518–1529, aug 2015. [Online]. Available: https://doi.org/10.14778/2824032.2824050

[15] N. Satish, M. J. Harris, and M. Garland, "Designing efficient sorting algorithms for manycore gpus," *2009 IEEE International Symposium on Parallel & Distributed Processing*, pp. 1–10, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:7697594

[16] O. Burçak and C. Tolga, "Joa: Joint overlap analysis of multiple genomic interval sets," *BMC Bioinformatics*, 2019. [Online]. Available: https://doi.org/10.1186/s12859-019-2698-4

[17] Q. A. R. and H. I. M., "Bedtools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, 01 2010. [Online]. Available: https://doi.org/10.1093/bioinformatics/btq033

[18] Q. A. R., "Bedtools: The swiss-army tool for genome feature analysis," *Current Protocols in Bioinformatics*, vol. 47, no. 1, pp. 11.12.1–11.12.34, 2014. [Online]. Available: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471250953.bi1112s47

[19] N. Shane, K. M. Scott, R. A. P., H. Eric, T. R. E., J. A. K., R. Eric, M. M. T., V. Jeff, T. Sean, S. Richard, H. Richard, and S. J. A., "Bedops: high-performance genomic feature operations," *Bioinformatics*, vol. 28, no. 14, pp. 1919–1920, 05 2012. [Online]. Available: https://doi.org/10.1093/bioinformatics/bts277

[20] L. Heng, "Tabix: fast retrieval of sequence features from generic tab-delimited files," *Bioinformatics*, vol. 27, no. 5, pp. 718–719, 01 2011. [Online]. Available: https://doi.org/10.1093/bioinformatics/btq671

[21] K. K. et al, "Genomicdistributions: fast analysis of genomic intervals with bioconductor," *BMC Genomics*, 2022. [Online]. Available: https://doi.org/10.1186/s12864-022-08467-y

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

The Effects of Different Knee Braces on Gait-Related Muscle Activity

Serra Gedik
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
serragedik0@gmail.com

Fatma Günseli Yaşar Çıklaçandır
Dept. of Computer Engineering
Izmir Katip Celebi University
Izmir, TURKEY
fatmagunseli.yasar@ikcu.edu.tr

Özlem Karabiber Cura
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
ozlem.karabiber@ikcu.edu.tr

Ahmet Emir Öğüt
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
emirogut@hotmail.com

Samet Çıklaçandır
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
samet.cikla@gmail.com

Abstract*—Knee braces are prescribed to healthy athletes for proprioceptive support and to patients with knee disorders for pain reduction/joint realignment, yet their muscle-level effects lack direct comparison. Using surface EMG on peripatellar muscles in four healthy women (age: 22.8 ± 3 years; height: 166.0 ± 3.6 cm; mass: 63.5 ± 7.7 kg) walking at 4 km·h^{-1}, 5% incline for 30 second sessions, a basic sleeve (Brace-1) and patella-stabilizing brace (Brace-2) were compared. Using nine time-domain features (RMS, Energy, Activity, Mobility, Complexity, Mean, MAD, Skewness, and Peak-to-Peak), a quadratic-kernel SVM distinguished between brace-on and brace-off epochs, obtaining mean accuracies of 87.8% for Brace-1 and 83.9% for Brace-2, and 85.6% for the Brace-1 and Brace-2 comparison. Divergent activation strategies were revealed by violin plots. Whereas Brace-2 displayed upward-shifted, tighter bands indicating steadier activation, Brace-1 produced taller, broader RMS/energy curves indicating variable bursts, meaning Brace-1 amplifies the EMG signals but diversifies, while Brace-2 has a more stabilizing effect on peripatellar muscle activities. Differential neuromuscular signatures were confirmed by machine-learning metrics (low FDR, high sensitivity/specificity). Evidence-based brace selection is informed by the fact that the patella brace allows mechanical realignment with moderate bursts, while the neoprene sleeve enhances proprioceptive drive through diversified recruitment.

Keywords—knee braces; electromyography; gait analysis; muscle activity; machine learning, classification.

I. Introduction

Knee braces are a great way to prevent injuries, have better proprioception and suppress pain [1]. Recent studies show that the type of knee brace may have different demands on quadriceps and hamstring coactivation during simple activities like walking [2, 3]. Knee braces are therefore applied well beyond everyday fitness: they are fitted after ligament injuries or reconstructive surgery to protect healing tissue and restore proprioceptive acuity [4]; they are prescribed in degenerative conditions such as medial knee osteoarthritis to off-load painful compartments and improve functional stability [5]; and they are often recommended during return-to-sport phases to limit excessive joint excursions while users rebuild strength and confidence [2]. In each case the brace acts as an external constraint that moderates mechanical load and provides cutaneous feedback and gives the wearer a tangible sense of security, thereby complementing muscle retraining and physiotherapy.

A basic knee sleeve is a lightweight, circumferential compression cuff that supplies constant low-grade pressure and warmth; this continuous cutaneous input heightens joint-position sense and offers mild global support during early rehabilitation or low-impact sport [1, 2]. The patella-stabilizing brace integrates a semi-rigid frame and medial pull-straps that exert a directed force on the patella, guiding it centrically through the trochlear groove; in doing so, it redistributes patellofemoral contact stress and attenuates anterior knee pain [4, 6]. Thus, whereas the basic sleeve principally amplifies sensory feedback and global comfort, the patella-tracking brace delivers targeted biomechanical realignment for patellofemoral control.

Studies on basic sleeves demonstrate both sensory and functional benefits. *Bottoni et al.* reported that wearing a sleeve significantly reduced joint-position error and enhanced proprioception in young, non-professional athletes [1]. In people with knee osteoarthritis, *Cudejko et al.* found that a soft sleeve lowered pain, improved dynamic stability, and thereby lessened activity limitations [2]. During high-volume resistance exercise, *Hatfield et al.* observed higher RMS and energy values—as well as better overall performance—in participants who wore a neoprene sleeve, attributing these gains to improved kinesthetic awareness [7].

Research on patella-tracking braces focuses mainly on patellofemoral pain. *Salarie Sker et al.* showed that an elastic patellar brace reduced pain and significantly modified leg-muscle EMG patterns during stance [4]. Similarly, *Kölle et al.* reported an immediate pain decrease, improved patellar kinematics, and adjusted quadriceps–hamstring co-contraction

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

after brace application [6]. Collectively, the evidence suggests that sleeves enhance proprioceptive feedback and overall muscle engagement, whereas patella-tracking braces achieve pain relief and mechanical realignment through targeted patellar guidance.

This study aims to determine how two common knee-brace designs differ in the way they modulate activity of the peripatellar muscles. Literature identifies vastus medialis (VM, patellar alignment) and semimembranosus (SM, tibial stabilization) as optimal EMG targets for detecting brace-induced neuromuscular changes. [8, 3].

II. Materials and Methods

The study utilized a BIOPAC MP36 electromyography system (BIOPAC Systems Inc., USA) for surface EMG acquisition, featuring four channels with 24-bit resolution and a sampling rate of 1,000 Hz. Disposable Ag/AgCl electrodes (20 mm diameter) were employed for signal capture.

Two knee brace models were included in the experiments. Brace-1, the basic sleeve (Orthocare 6120 Genucare Basic, Turkiye) provided circumferential compression without rigid components, designed for mild proprioceptive enhancement and thermal retention. Brace-2, the patellofemoral knee brace (Orthocare 6158 Genucare Luxa, Turkiye) brace incorporated a medical-grade polyamide frame with medial straps to correct lateral patellar displacement through controlled medial force application.

Four healthy female volunteers (aged 22.8 ± 3 years; height 166.0 ± 3.6 cm; body mass 63.5 ± 7.7 kg) without any knee disorders participated in this study. The sample size was constrained to four subjects due to scheduling limitations and data quality considerations.

Disposable Ag/AgCl electrodes were placed on both right and left legs' VM and SM, with skin preparation, ensuring electrode-skin impedance <5 kΩ. Two knee braces were tested, respectively, for experimental tasks. Knee braces were changed after the walking trial of a brace type was finished. Participants performed treadmill walking trials under conditions of a speed of 4 km/h paired with incline of 5% Each 30-second trial was repeated for both brace types 9 times, yielding 18 trials per participant. Between trials, 45-second rest intervals prevented fatigue while maintaining thermal consistency of the braces.

A zero-phase bandpass filter (0.5–500 Hz, 2nd-order Butterworth) was used to remove motion artifacts and high-frequency noise from raw EMG signals. To ensure stationarity, the filtered data were divided into non-overlapping 5-second epochs (5,000 samples per window). MATLAB was then used to extract time-domain features from each epoch. These were the following: Energy (sum of squared samples, reflecting mechanical work) [11], Skewness (asymmetry in amplitude distribution, highlighting burst-like or silent-gap behavior) [10], Mean (average rectified value, indicating overall muscle activation) [9], and Hjorth parameters (Activity as signal variance, Mobility approximating mean frequency, and Complexity indicating abrupt frequency changes) [10]. Furthermore, calculations were made for Mean-Absolute Deviation (which tracks variability related to fatigue) [12], Root-Mean-Square (RMS), which correlates power-equivalent amplitude with force output [13], and Peak-to-Peak Amplitude, which indicates instantaneous burst strength.

For classification, a quadratic support vector machine (SVM) was selected. Unlike more complicated kernels (such as Gaussian), this kernel offers a moderate amount of flexibility and successfully models the curvature frequently found in EMG datasets without overfitting. The results of trials with other classifiers were comparable to those of *Toledo-Pérez et al.*, who found that quadratic SVM produced better validated accuracy in lower-limb EMG studies with limited time-domain features [14].

Sokolova & Lapalme's confusion-matrix-derived metrics were used to assess the model's performance. Precision (positive predictive value), sensitivity (true positive rate), specificity (true negative rate), accuracy (overall correctness), and False Discovery Rate (FDR) (1-precision). Together, these evaluate robustness against false alarms and missed detections [15].

III. Results

Brace-specific neuromuscular signatures were quantified using SVM classification. Significant brace-induced modulations were found by the quadratic SVM classifier in every condition (Table 1). The most noticeable EMG changes were caused by Brace-1, demonstrated by a lower FDR (12.3 ± 6.2%) and a higher brace-detection accuracy (87.8 ± 2.6%) in comparison to Brace-2 (83.9 ± 3.9% accuracy; 16.0 ± 4.0% FDR). Importantly, direct brace-to-brace discrimination confirmed essentially distinct activation strategies with an accuracy of 85.6 ± 6.4%. The feature distributions in Figs. 1-3 are consistent with these machine-learning metrics.

Table 1 Quadratic SVM classifications: Performance metrics (mean ± SD) of 3 conditions: Brace-1, Brace-2, and Brace-1 vs. Brace-2.

Metrics	SVM (Quadratic SVM)		
	Brace-1	**Brace-2**	**Brace-1 vs. Brace-2**
Accuracy	87.8 ± 2.6 %	83.9 ± 3.9 %	85.6 ± 6.4 %
Sensitivity	88.1 ± 3.4 %	84.0 ± 6.5 %	81.5 ± 3.5 %
Specificity	87.6 ± 4.1 %	83.8 ± 6.1 %	81.6 ± 5.7 %
Precision	87.7 ± 3.6 %	84.0 ± 4.0 %	82.1 ± 3.3%
FDR	12.3 ± 6.2 %	16.0 ± 4.0 %	17.9 ± 3.3 %

Data from all trials revealed Brace-1 consistently increased EMG amplitude and energy in the braced leg. Median RMS rose by 28% and energy by 34% versus the unbraced leg (Fig. 1), with wider interquartile ranges confirming higher stride variability. Complexity features (Hjorth Activity/Mobility/Complexity) showed similar increases. A quadratic SVM achieved 87.8% accuracy classifying braced/unbraced states. This demonstrates the sleeve induces

machine-detectable EMG modifications without requiring muscle- or task-specific analysis.

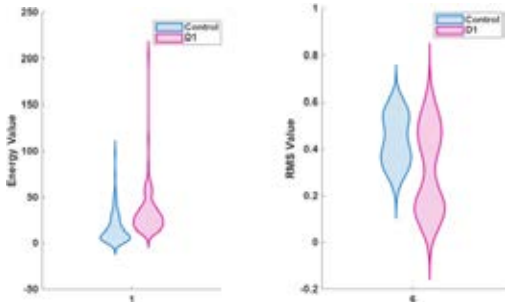


Figure 1: Violin graphs of energy and RMS features, comparing the left (braceless) leg and the right (with brace-1) leg.

Aggregated data showed Brace-2 induced a moderate EMG amplitude increase in the braced leg. Median RMS rose by 22% and energy by 30% versus control (Fig. 2), with wider interquartile ranges confirming heightened variability. Complexity features (skewness, kurtosis, and Hjorth activity) exhibited extended distribution tails. A quadratic SVM achieved 83.9% accuracy classifying braced/unbraced states, indicating the patella brace induces a milder but detectable EMG modification.

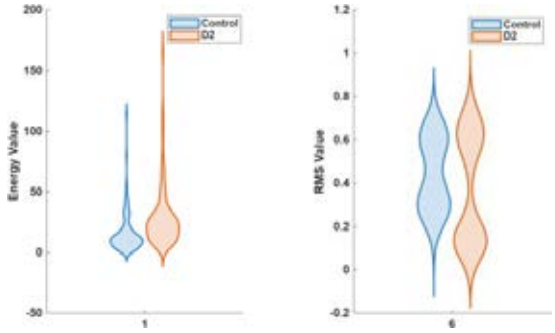


Figure 2: Violin graphs of energy and RMS features, comparing the left (braceless) leg and the right (with brace-2) leg.

A direct brace-to-brace comparison (Fig. 3) clarifies these relative effects. Brace-1 retained the largest median energy and the broadest 90th-percentile tail, signaling occasional high-power bursts; it also yielded higher, more dispersed PTP and Activity values, suggesting a permissive strategy that accentuates peak contractions. By contrast, Brace-2 showed only a modest right-shift in RMS (+3 %) but a conspicuously tighter inter-quartile band, and its Mobility/Complexity metrics clustered around mid-range values, consistent with a levelling effect that raises baseline activation while curbing extremes. These converging patterns—together with the aggregate brace-to-brace classifier performance—explain the hierarchy: Brace-1 leaves the most distinctive spectral–temporal fingerprint, whereas Brace-2 produces a subtler, more homogeneous modulation of neuromuscular output.

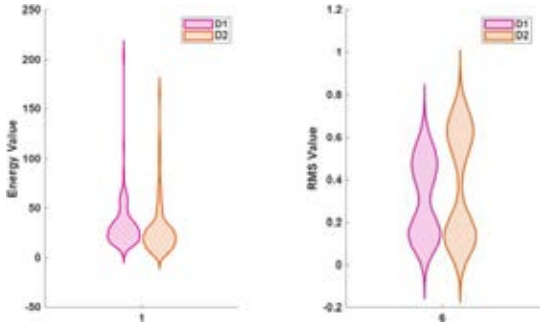


Figure 3: Violin graphs of energy and RMS features, comparing Brace-1 and Brace-2.

Across four healthy females and nine walking conditions, both braces measurably altered the surface-EMG of the braced limb. The basic sleeve elicited the largest rises in RMS and energy and the most distinctive classifier signature; the patella-stabilizing brace produced a more moderate, uniformly distributed activation shift.

IV. Discussion

Both braces altered muscle activity differently. Brace-1 increased EMG variability with broader RMS/energy distributions, indicating amplified motor-unit bursts. This matches *Heß et al. (2025)*, where sleeves enhanced proprioceptive drive [16]. Conversely, Brace-2 narrowed distributions by suppressing extremes and shifting activation uniformly—consistent with *Salarie Sker et al.*'s "levelling" effect [3] and *Gülling et al.*'s reduced EMG during extensions [17]. Mechanistically, Brace-1 boosts afferent input for joint sensing, while Brace-2 promotes uniform engagement to offload quadriceps and limit lateral glide.

Unlike single-brace studies, this multi-feature analysis reveals distinct brace "fingerprints" in healthy knees. Clinically, sleeves (e.g., Brace-1) suit mild support/sensory augmentation needs, while patellar-tracking braces (e.g., Brace-2) optimize precise patellar control and energy smoothing.

V. Conclusion

This study confirms that braces significantly alter peripatellar muscle activity in healthy knees, with effects dictated by design. For general proprioceptive support during low-impact activities, elastic sleeves optimize benefits by diversifying activation. When precise patellar alignment or lateral-motion restraint is critical (e.g., post-patellofemoral pain rehabilitation or mal-tracking-prone knees), patella-stabilizing braces are preferred: they elevate baseline activation within a controlled range while suppressing extremes. Clinically, begin with sleeves for mild instability; if malalignment persists under load, switch to patella-tracking braces for targeted control without excessive muscular strain.

ACKNOWLEDGEMENT

This study was supported by İzmir Katip Çelebi University Scientific Research Council Agency as project number 2024-KDP-MÜMF-0013. The authors gratefully acknowledge the financial support provided, which facilitated the acquisition of research equipment, data collection, and analysis necessary for the successful completion of this work.

REFERENCES

[1] G. Bottoni, A. Herten, P. Kofler, M. Hasler, and W. Nachbauer, "The effect of knee brace and knee sleeve on the proprioception of the knee in young non-professional healthy sportsmen," The Knee, vol. 20, no. 6, pp. 490–492, Dec. 2013, doi: https://doi.org/10.1016/j.knee.2013.05.001.

[2] T. Cudejko et al., "Decreased Pain and Improved Dynamic Knee Instability Mediate the Beneficial Effect of Wearing a Soft Knee Brace on Activity Limitations in Patients With Knee Osteoarthritis," Arthritis Care & Research, vol. 71, no. 8, pp. 1036–1043, Jul. 2019, doi: https://doi.org/10.1002/acr.23722.

[3] F. Salarie Sker, M. Anbarian, A. H. Yazdani, P. Hesari, and A. Babaei-Ghazani, "Patellar bracing affects sEMG activity of leg and thigh muscles during stance phase in patellofemoral pain syndrome," Gait & Posture, vol. 58, pp. 7–12, Oct. 2017, doi: https://doi.org/10.1016/j.gaitpost.2017.06.271.

[4] B. D. Beynnon, L. Good, and M. A. Risberg, "The Effect of Bracing on Proprioception of Knees with Anterior Cruciate Ligament Injury," Journal of Orthopaedic & Sports Physical Therapy, vol. 32, no. 1, pp. 11–15, Jan. 2002, doi: https://doi.org/10.2519/jospt.2002.32.1.11.

[5] C. H. Fantini Pagani, S. Willwacher, B. Kleis, and G.-P. Brüggemann, "Influence of a valgus knee brace on muscle activation and co-contraction in patients with medial knee osteoarthritis," Journal of Electromyography and Kinesiology, vol. 23, no. 2, pp. 490–500, Apr. 2013, doi: https://doi.org/10.1016/j.jelekin.2012.10.007.

[6] T. Kölle, W. Alt, and D. Wagner, "Immediate effects of an elastic patellar brace on pain, neuromuscular activity and knee kinematics in subjects with patellofemoral pain," Archives of Orthopaedic and Trauma Surgery, vol. 140, no. 7, pp. 905–912, Feb. 2020, doi: https://doi.org/10.1007/s00402-020-03378-7.

[7] D. L. Hatfield, A. M. Stranieri, L. M. Vincent, and J. E. Earp, "Effect of a Neoprene Knee Sleeve on Performance and Muscle Activity in Men and Women During High-Intensity, High-Volume Resistance Training," Journal of Strength and Conditioning Research, vol. 35, no. 12, pp. 3300–3307, Sep. 2021, doi: https://doi.org/10.1519/jsc.0000000000004072.

[8] S. Pal et al., "Patellar Maltracking Correlates With Vastus Medialis Activation Delay in Patellofemoral Pain Patients," The American Journal of Sports Medicine, vol. 39, no. 3, pp. 590–598, Nov. 2010, doi: https://doi.org/10.1177/0363546510384233.

[9] A. D. Vigotsky, I. Halperin, G. J. Lehman, G. S. Trajano, and T. M. Vieira, "Interpreting Signal Amplitudes in Surface Electromyography Studies in Sport and Rehabilitation Sciences," Frontiers in Physiology, vol. 8, no. 1, Jan. 2018, doi: https://doi.org/10.3389/fphys.2017.00985.

[10] S. Abbaspour, M. Lindén, H. Gholamhosseini, A. Naber, and M. Ortiz-Catalan, "Evaluation of surface EMG-based recognition algorithms for decoding hand movements," Medical & Biological Engineering & Computing, vol. 58, no. 1, pp. 83–100, Nov. 2019, doi: https://doi.org/10.1007/s11517-019-02073-z.

[11] J. Jie, K. Liu, H. Zheng, B. Wang, and R. Dai, "High dimensional feature data reduction of multichannel sEMG for gesture recognition based on double phases PSO," Complex & Intelligent Systems, vol. 7, no. 4, pp. 1877–1893, Mar. 2021, doi: https://doi.org/10.1007/s40747-020-00232-6.

[12] M. Aviles, Juvenal Rodríguez-Reséndiz, and D. Ibrahimi, "Optimizing EMG Classification through Metaheuristic Algorithms," Technologies, vol. 11, no. 4, pp. 87–87, Jul. 2023, doi: https://doi.org/10.3390/technologies11040087.

[13] S. Karlsson and B. Gerdle, "Mean frequency and signal amplitude of the surface EMG of the quadriceps muscles increase with increasing torque—a study using the continuous wavelet transform," Journal of Electromyography and Kinesiology, vol. 11, no. 2, pp. 131–140, Apr. 2001, doi: https://doi.org/10.1016/s1050-6411(00)00046-8.

[14] D. C. Toledo-Pérez, J. Rodríguez-Reséndiz, R. A. Gómez-Loenzo, and J. C. Jauregui-Correa, "Support Vector Machine-Based EMG Signal Classification Techniques: A Review," Applied Sciences, vol. 9, no. 20, p. 4402, Oct. 2019, doi: https://doi.org/10.3390/app9204402.

[15] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: https://doi.org/10.1016/j.ipm.2009.03.002.

[16] T. Heß, T. L. Milani, J. Stoll, and C. Mitschke, "Running and Jumping After Muscle Fatigue in Subjects with a History of Knee Injury: What Are the Acute Effects of Wearing a Knee Brace on Biomechanics?," Bioengineering, vol. 12, no. 6, p. 661, Jun. 2025, doi: https://doi.org/10.3390/bioengineering12060661.

[17] L. K. Gulling, S. M. Lephart, D. A. Stone, J. J. Irrgang, and D. M. Pincivero, "The effects of patellar bracing on quadriceps EMG activity during isokinetic exercise," Isokinetics and Exercise Science, vol. 6, no. 2, pp. 133–138, Oct. 1996, doi: https://doi.org/10.3233/ies-1996-6209.

Designing a Novel Implant for Preventing the Risk of Postoperative Pancreatic Fistula in Whipple Surgery

Ahmet Emir Öğüt
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
emirogut@hotmail.com

Samet Çıklaçandır
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
samet.cikla@gmail.com

***Abstract*—The Whipple operation (pancreaticoduodenectomy) is a complex surgical procedure that involves the removal of the head of the pancreas, duodenum, gallbladder, and sometimes part of the stomach and bile duct. This procedure is performed to remove the tumorous structures surrounding the pancreas, leaving healthy tissue intact. To ensure the continuity of the digestive system, the remaining parts of the pancreas, bile duct, and stomach are reconnected to the intestines after the structures are removed. Due to the complexity of this operation, there are significant potential risks, one of which is pancreatic fistula, defined as the leakage of pancreatic fluid. Pancreatic fistula is the uncontrolled leakage of pancreatic fluid into the body cavities and surrounding tissues. Since pancreatic fluid contains potent digestive enzymes such as amylase, lipase, and trypsin, this can lead to severe tissue damage and infections. This situation typically occurs when the stitches placed by the physician in the anastomosis area, connecting the pancreatic duct to the duodenum, do not provide sufficient stability. The main reason for the formation of a fistula is that the suture line does not heal completely. To prevent serious complications in the patient, drug therapy, endoscopic stent applications, and revision surgeries may be required. This project aims to develop a new implant to prevent pancreatic fistulas. This new solution aims to connect the pancreatic duct directly to the duodenum without the use of sutures.**

***Keywords*—Pancreatic leakage, implant, pancreatic surgery, pancreaticoduodenectomy, pancreatic fistula.**

I. INTRODUCTION

A complicated surgical procedure called the Whipple procedure (pancreaticoduodenectomy) is frequently used to treat cancers of the duodenum, bile duct, or pancreatic head. The pancreatic head, duodenum, gallbladder, and a portion of the bile duct are removed, and several anastomoses are used to reassemble the organ. One of the most dangerous side effects after pancreatic surgery is pancreatic leakage [1]. The health of the patient may be seriously threatened by leakage from the pancreatojejunostomy anastomosis, especially following intricate procedures like the Whipple operation [2]. Severe inflammatory processes are triggered by the unchecked release of pancreatic enzymes into surrounding tissues, and this significantly raises the rates of postoperative morbidity and mortality [3]. Up to 33% of patients experience postoperative pancreatic fistula (POPF), which is still a widespread and difficult complication, even in specialized surgical centers [4,5,6].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

The surgical procedure on the afflicted tissues is depicted in Figure 1. The anastomosis of healthy tissues occurs after the removal of the purple-colored areas, which represent the tumorous tissues.

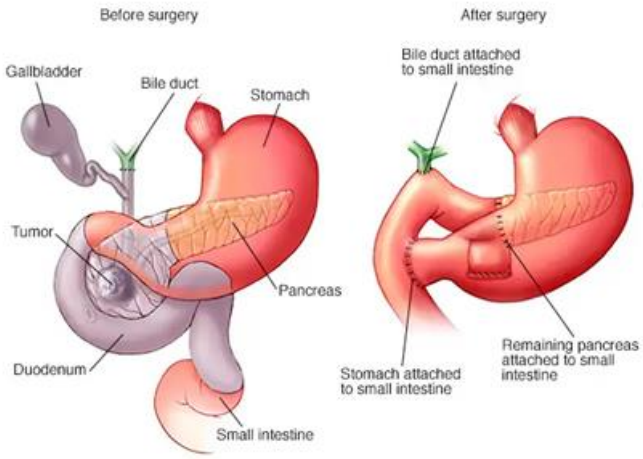


Figure 1: In Whipple surgery, the tumorous structure is removed and healthy tissues are anastomosed to ensure secretion.

An important anatomical feature that joins the common bile duct (ductus choledochus) with the Wirsung duct (pancreatic duct) before entering into the duodenum is the Ampulla of Vater. The constant passage of pancreatic and bile secretions through this region makes it extremely prone to problems. As a result, creating a biocompatible implant that stops pancreatic leakage without interfering with the Ampulla of Vater's ability to function should greatly enhance patient outcomes and lower the possibility of problems following surgery [6]. While avoiding leaking at the anastomosis site, such an implant must maintain the normal flow of pancreatic secretions.

When choosing materials for the implant, biocompatible and biomechanically stable materials must be evaluated. The materials chosen should be appropriate for long-term implantation and compatible with the surrounding tissues. Additionally, the implant needs to maintain the natural flow balance between the bile and pancreatic fluids without obstructing it and be physically compatible with the Ampulla of Vater [7].

There are currently very few implant-based methods for treating pancreatic fistula in the literature [8,9,10]. 3D-printed stents were used in one investigation; they demonstrated better recovery quality in animal models [9]. These stents' flow and leakage capabilities were not assessed, though. Despite the leak of the amount of design, the first thought design about this project has shown in Figure 2.

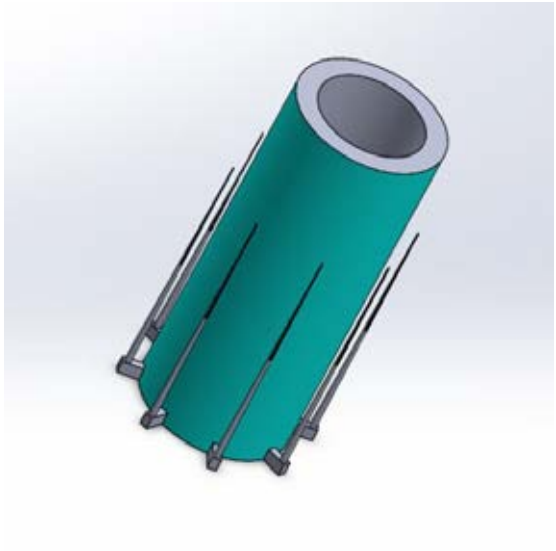


Figure 2: The CAD design of implant

It is anticipated that this effort will improve the quality of recovery for patients following Whipple surgery and lower the risk of postoperative pancreatic fistula.

II. MATERIALS AND METHODS

A. Mechanical Properties of Pancreas

The pancreas is a complex organ that has both exocrine and endocrine functions. It is placed retroperitoneal in the abdominal space and has three main anatomical parts: head, body, and tail [11]. The exocrine part includes the enzymes that are released from the pancreatic acini. These enzymes (amylase, lipase, and trypsin) play an essential role in digestion, and the transmission occurs in the Wirsung channel [12]. The endocrine part secretes insulin, glucagon, and somatostatin from the Langerhans cells. Mechanically, the pancreas has soft and easily damaged tissue.

There are no exact measurements available for the density of pancreatic tissue; however, in research, the density of all soft tissue has been standardized at 1000 kg/m³ [13]. Due to its low elastic modulus of around 1.4 ± 0.47 kPa, it exhibits high deformation ability [14]. It has a 0.495 Poisson's ratio, indicating that pancreas tissue is almost incompressible mechanically [15]. These mechanical properties make it challenging to maintain tissue stability after surgical procedures and increase the risk of leakage at the anastomosis site [3].

At the Whipple surgery, the head of the pancreas is removed, and a representation of the remaining part is designed using CAD software, as shown in Figure 3. The length of the remaining pancreas can be altered due to the operation and the patient's condition. In this design, the total length of the body and tail has been determined to be 11 cm, and the thickness of the wall is 10.75 mm at the body and 6 mm at the end of the tail. Furthermore, the channel diameter is 3.5 mm at the body and 1.5 mm at the tail.



Figure 3: Pancreas design after Whipple surgery.

B. Mechanical Properties of Duodenum/Jejunum

The small intestine has three parts: the duodenum, jejunum, and ileum. The jejunum is the middle segment, which begins with the ligament of Treitz and continues into the ileum [11]. The duodenum is the first section of the small intestine, located between the stomach and the jejunum. It is approximately 25-30 cm long [16]. In this region, the bile duct and pancreatic duct open into the duodenum through the ampulla of Vater [7].

The wall of the duodenum consists of the mucosal layer, submucosa, muscle layer, and serosa. The average wall thickness is 2-3 mm [17].

The jejunum has an approximate length of 2.5 meters and plays a crucial role in the digestion and absorption of nutrients [18]. The jejunum has an elastic, flexible, and porous structure. In research, its density is reported as approximately 1120 kg/m³ [19]. Its elasticity module ranges around 4 MPa, demonstrating its high elasticity and ability to change shape under tension [19]. The passion ratio is 0.45, indicating that the jejunum doesn't have significant volumetric changes when it is stretched [19].

During the surgery, most of the duodenum is removed and connected again to the pancreas. A representation of the duodenum is shown in Figure 4. The length of the duodenum is determined as 8 cm, the outer diameter is 5 cm, and the thickness of the wall is 3 mm. The duct diameter is 2.5 cm, which is determined by the width of the pancreatic duct.



Figure 4: CAD representation of duodenum.

C. Design of Implant and Organ Geometries

For this study, three-dimensional (3D) models of the implant and the surrounding anatomical elements were produced using SolidWorks 2024 CAD software. The implant was specifically designed to fit the pancreaticojejunostomy site smoothly following the Whipple procedure. For simple insertion during surgery, the implant was made with a snap-fit mechanism. The goal of this design was to establish a secure and leak-proof bond without harming the surrounding tissues. Figure 5 provides a graphic representation and placement of the finished implant.

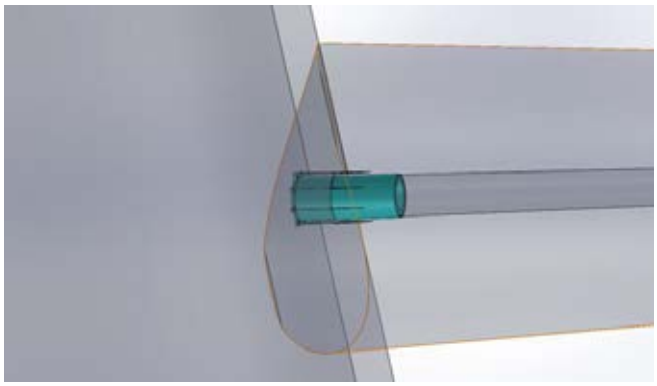


Figure 5: Placement of Implant: From the inside of the duodenum to the pancreatic duct.

D. Determination of Material Properties

The physical characteristics of the implant, pancreatic fluid, and surrounding tissues were carefully chosen during the modeling procedure. The biocompatibility and resistance to pancreatic enzymes of the materials used in the implant's manufacture were assessed. The goal was to prevent the structure from being damaged by digestive enzymes, such as trypsin, lipase, and amylase. The intestinal tissue's porous and flexible structure was noted; its flexibility is crucial for modeling dynamic motions in the regions that will come into contact with the implant following surgery. In contrast, a softer tissue with a low elastic modulus was used to simulate the pancreas. This delicate structure highlights the need for a sutureless solution. Despite having a low viscosity, pancreatic fluid is quite biochemically active. For the sake of simplicity, it was treated as a Newtonian fluid in this study, even though it is typically described as a non-Newtonian fluid in the literature. The flow environment was also used to replicate a second fluid flow that came from the duodenum, in addition to pancreatic fluid. It was determined that this second fluid shared the same physical characteristics as water. Table 1 provides a summary of the mechanical and physical parameters.

Table 1. Mechanical properties of biological tissues

Mechanical Properties	**Biological Part**		
	Pancreatic Liquid	***Pancreas***	***Duodenum***
Density (ρ)	1010	1050 kg/m3 [13]	1087 kg/m3 [19]
Elasticity Module (E)	-	1.40 ± 0.47 KPa [14]	4 MPa [19]
Passion Ratio (v)	-	0.495 [15]	0.495 [19]
Viscosity (u)	1.17-1.72 mPa.s [17, 18]	-	-
Modelling	Not Newtonian	-	-
Substance	Amylase, Lipase, and Trypsin	-	-

III. RESULTS

The final implant model was designed to facilitate sutureless placement at the anastomosis site following the Whipple procedure. The implant, created using three-dimensional modeling, was optimized and placed based on the anatomical dimensions of the pancreatic duct and duodenum. Material properties and parameters were researched for future use and research. The implant's needle-based insertion mechanism was designed to allow for rapid placement and secure fixation during surgery. The goal was to minimize tissue damage caused by this structure. The model's surface geometry was adjusted for direct placement within the Wirsung duct. The implant was designed to ensure no gaps were left. Because duct widths can vary from patient to patient, future studies may aim to offer individual designs for each patient. The design prioritized both ease of placement and retention of the postoperative position. Due to the gap in the literature, reference sources were consulted, and the implant, designed to address the leakage problem, was mechanically enhanced to prevent removal from its original location, thereby prioritizing surgical time and ease of use.

IV. DISCUSSION

POPF, one of the most serious complications after the Whipple procedure, prolongs patient recovery time and increases the risk of mortality. Because traditional anastomosis techniques cannot always provide adequate protection against leakage due to the soft tissue of the pancreas and the small diameter of the pancreatic duct, an implant design that facilitates this situation is of significant importance in both the

literature and the clinic. The implant designed in this study aims to eliminate the need for surgical sutures and provide a seamless connection. This aims to reduce the risk of leakage from enzymatic fluids passing between the pancreatic duct and duodenum, as well as the risk of fistula problems resulting from the destructive effects of leakage. The snap-fit mechanism ensures implant retention and postoperative stability while potentially causing minimal tissue damage. While 3D-printed stent designs developed for similar purposes exist in the literature, the geometry of this implant has been optimized to adapt to the anastomotic anatomy specifically. While the design demonstrates the correct approach, analyses must be conducted to support its reliability. Further studies are particularly focused on leak tests and CFD analyses. Thanks to analyses and leakage tests, the clinical suitability of the implant will be checked, and the door can be opened for future studies.

ACKNOWLEDGMENT

This study was supported by the İzmir Katip Çelebi University University Scientific Research Council Agency, project number 2025-TYL-FEBE-0009. The authors gratefully acknowledge the financial support provided, which facilitated the acquisition of research equipment, data collection, and analysis necessary for the successful completion of this work.

REFERENCES

[1] Mayo Clinic. (n.d.). *Whipple procedure (pancreaticoduodenectomy)*. Mayo Clinic. Retrieved February 13, 2025

[2] Unegbu, F. C., & Anjum, F. (2023). Pancreatic Fistula. In *StatPearls [Internet]*. StatPearls Publishing.

[3] Bassi, C., Marchegiani, G., Dervenis, C., Sarr, M., Hilal, M. A., Adham, M., ... & Buchler, M. (2017). The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 years after. *surgery*, *161*(3), 584-591.

[4] Tewari, M., Hazrah, P., Kumar, V., & Shukla, H. S. (2010). Options of restorative pancreaticoenteric anastomosis following pancreaticoduodenectomy: a review. *Surgical oncology*, *19*(1), 17-26.

[5] Ramacciato, G., Mercantini, P., Petrucciani, N., Nigri, G. R., Kazemi, A., Muroni, M., ... & Ravaioli, M. (2011). Article commentary: risk factors of pancreatic fistula after pancreaticoduodenectomy: a collective review. *The American Surgeon*, *77*(3), 257-269.

[6] Yeo, C. J., Cameron, J. L., Lillemoe, K. D., Sohn, T. A., Campbell, K. A., Sauter, P. K., ... & Hruban, R. H. (2019). Does the type of pancreaticojejunostomy after the Whipple procedure make a difference? *Journal of Gastrointestinal Surgery, 23*(5), 852-860.

[7] Gaujoux, S., Cortes, A., Couvelard, A., Noullet, S., Clavel, L., Rebours, V., ... & Sauvanet, A. (2011). Postoperative pancreatic fistula: ductal anatomy and pancreatic consistency matter. *Surgery, 149*(2), 175-182.

[8] Hackert, T., Werner, J., Weitz, J., Schmidt, J., & Büchler, M. W. (2018). Postoperative pancreatic fistula. *Surgeon, 16*(1), 5-12.

[9] Xiang, F., Yao, C., Guan, G., & Luo, F. (2024). Application of 3D Printing to Design and Manufacture Pancreatic Duct Stent and Animal Experiments. *Bioengineering*, *11*(10), 1004.

[10] Pan, M., Zhao, C., Xu, Z., Yang, Y., Teng, T., Lin, J., & Huang, H. (2021). radiopaque chitosan ducts fabricated by extrusion-based 3D printing to promote healing after pancreaticoenterostomy. *Frontiers in Bioengineering and Biotechnology*, *9*, 686207.

[11] David, J., & Susan, S. (2008). Gray's anatomy: the anatomical basis of clinical practice.

[12] Beger, H. G., Buchler, M. W., Kozarek, R., Lerch, M. M., Neoptolemos, J. P., Warshaw, A. L., ... & Shiratori, K. (Eds.). (2009). *The pancreas: an integrated textbook of basic science, medicine, and surgery*. John Wiley & Sons.

[13] Nabavizadeh, A., Payen, T., Iuga, A. C., Sagalovskiy, I. R., Desrouilleres, D., Saharkhiz, N., ... & Konofagou, E. E. (2020). Noninvasive Young's modulus visualization of fibrosis progression and delineation of pancreatic ductal adenocarcinoma (PDAC) tumors using Harmonic Motion Elastography (HME) in vivo. *Theranostics*, *10*(10), 4614.

[14] Pagnanelli, M., De Gaetano, F., Callera, A., Nappo, G., Capretti, G., Carrara, S., ... & Zerbi, A. (2024). Analysis of the Mechanical Characteristics of Human Pancreas through Indentation: Preliminary In Vitro Results on Surgical Samples. *Biomedicines*, *12*(1), 91.

[15] Rubiano, A., Delitto, D., Han, S., Gerber, M., Galitz, C., Trevino, J., ... & Simmons, C. S. (2018). Viscoelastic properties of human pancreatic tumors and in vitro constructs to mimic mechanical properties. *Acta biomaterialia*, *67*, 331-340.

[16] Standring, S., Ellis, H., Healy, J., Johnson, D., Williams, A., Collins, P., & Wigley, C. (2005). Gray's anatomy: the anatomical basis of clinical practice. *American journal of neuroradiology*, *26*(10), 2703.

[17] Hounnou, G., Destrieux, C., Desmé, J., Bertrand, P., & Velut, S. (2002). Anatomical study of the length of the human intestine. *Surgical and radiologic anatomy*, *24*(5), 290-294.

[18] Hall, J. E., & Hall, M. E. (2020). *Guyton and Hall Textbook of Medical Physiology E-Book: Guyton and Hall Textbook of Medical Physiology E-Book*. Elsevier Health Sciences.

[19] Liu, C., & Ma, J. (2023). Simulation Analysis of Small Intestinal Peristalsis Based on Finite Element. *Advances in Engineering Technology Research*, *4*(1), 480-480.

[20] Tajikawa, T., Aoki, K., & Fukuzawa, H. (2023). Investigation of pancreatic juice reflux mechanism in high confluence of pancreaticobiliary ducts and pancreaticobiliary maljunction (Development and validation of a mathematical model for pancreatic and bile juice flow based on fluid mechanics). Journal of Biorheology, 37(2), 44-55.

[21] Tympner, F., & Rösch, W. (1978). Viscosity of pure pancreatic juice in chronic pancreatitis. Klinische Wochenschrift, 56, 421-422.

Attention Augmented U-Net for Robust Thyroid Ultrasound Segmentation

Özgür Ömer Gülsoy[1]
ozguromergulsoy@gmail.com
Istanbul Okan University, Computer Engineering, Istanbul, Turkey[1]

Mohammed Qusay Mawlood[2]
mohammawlood@stu.okan.edu.tr
Istanbul Okan University, Artificial Engineering, Istanbul, Turkey[2]

Mahsa Mikaeili[3]
mahsa.mikaeili@okan.edu.tr
Istanbul Okan University, Mechatronic Engineering, Istanbul, Turkey[3]

Sina Alp[4]
Sina.alp@okan.edu.tr
Istanbul Okan University, Electric and Electronics Engineering, Istanbul, Turkey[4]

***Abstract*— Thyroid nodule segmentation in ultrasound imaging faces challenges from speckle noise, low contrast, and morphological variability. We propose an EfficientNet-B5-backboned UNet++ architecture enhanced with Dual-Path Attention Modules (DPAM) integrating Global Context, Efficient Channel, and Spatial Attention mechanisms. Our model employs a composite loss function combining Focal Tversky, Dice, and Boundary losses with temporal ramp-up strategy to address class imbalance. Training features a three-phase progressive schedule with automatic threshold optimization and progressive resizing to 512×512, complemented by ultrasound-specific data augmentations. Evaluated on the TN3K dataset using 5-fold cross-validation, our model achieves state-of-the-art performance: 0.8875 Dice, 0.7969 IoU, 0.9688 accuracy, representing improvement over recent CNN and Transformer baselines while maintaining computational efficiency with less than 50M parameters. The tight training-validation gaps demonstrate robust generalization, indicating reliable boundary delineation and clinical applicability for computer-aided diagnosis pipelines.**

Keywords— ***Thyroid Segmentation, Ultrasound Imaging, UNet++, Dual-Path Attention Module, Medical Image Analysis***

I. INTRODUCTION (*HEADING 1*)

Thyroid nodules rank among the most common endocrine disorders, with B-mode ultrasound (US) serving as the primary modality for detection, characterization, and follow-up due to its real-time capability, low cost, and safety. However, US images suffer from speckle noise, low contrast, heterogeneous texture, and ill-defined lesion borders, making manual delineation time-consuming and operator-dependent. Accurate automated segmentation is therefore critical for clinical workflows and computer-aided diagnosis (CAD) systems [1,2,3].

UNet++ with redesigned skip connections for multiscale feature fusion [5], ResUNet with residual encoders [1], and attention-augmented variants that dynamically recalibrate feature maps [6,7]. EfficientNet backbones [8] have proven effective in data-scarce scenarios, while Transformer-based hybrids like Enhanced-TransUNet [7] and BPAT- feature extraction. Addressing thyroid ultrasound's unique challenges—blurred boundaries and severe class imbalance—recent methods employ boundary-aware supervision, dual-path attention, shape priors, and composite loss functions. Focal Tversky [13], boundary loss [11], and integrated multi-component losses like those in DPAM-UNet++ [10] have demonstrated superior edge precision compared to standard Dice and cross-entropy losses.

This work advances the field by combining an EfficientNet-B5-backed UNet++ architecture with Dual-Path Attention Mechanism (DPAM), trained using a composite loss integrating Focal Tversky, Dice, and boundary terms. Our pipeline

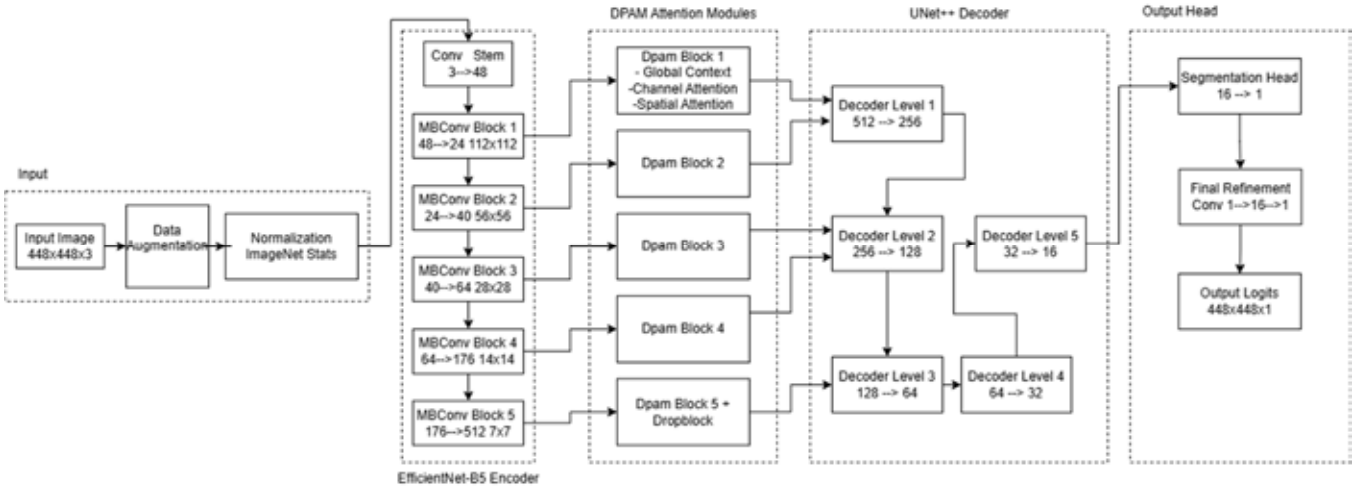


Fig. 1. Proposed models architecture diagram.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

incorporates domain-specific augmentations and a three-stage training schedule: frozen encoder initialization, full fine-tuning, and progressive resizing. The paper is organized as follows: Section II presents the proposed methodology, Section III discusses experimental results, and Section IV provides conclusions.

II. METHODS AND MATERIALS

A. Model Architecture

Our model employs EfficientNet-B5 as an encoder within a UNet++ architecture, chosen for its strong feature extraction capabilities and excellent segmentation performance in medical imaging tasks. The key enhancement of our architecture as shown in Fig. 1, is the strategic integration of a Dual-Path Attention Module (DPAM) after each encoder stage except the initial stem stage resulting in five DPAM blocks at critical feature extraction points. Each DPAM combines three complementary attention mechanisms: Global Context Attention [15] for capturing long-range dependencies, Efficient Channel Attention [16] for modeling inter-channel relationships with minimal computational overhead, and Spatial Attention [17] for emphasizing diagnostically relevant regions while suppressing background noise. These attention pathways operate in parallel, allowing simultaneous focus on different ultrasound feature aspects. The outputs from the three attention mechanisms are concatenated and merged using 1×1 convolutional layers with batch normalization, effectively learning optimal feature combinations. The decoder leverages UNet++'s densely connected skip connections with nested pathways, which help maintain fine-grained spatial details and ensure efficient gradient flow—both essential for accurate boundary detection in ultrasound images. Finally, a refinement module with additional convolutional layers further processes the segmentation logits, producing precise binary segmentation masks that match the original 448×448 pixel input size.

B. Loss Function

In this study, we employ an Enhanced Composite Loss that integrates multiple specialized losses: Focal Tversky Loss [13], Dice Loss, and Boundary Loss [11]. The composite loss function is defined as:

$$L = \alpha_c L_{FT} + \beta_c L_{Dice} + \gamma_c(t) L_{Boundary} \quad (1)$$

Where L_{FT} is the Focal Tversky Loss, L_{Dice} is the Dice Loss, and $L_{Boundary}$ is the Boundary Loss. The coefficients α_c, β_c, and $\gamma_c(t)$ balance these components. Focal Tversky Loss (L_{FT}): This loss emphasizes challenging pixels, formulated as:

$$L_{FT} = \left[1 - \frac{TP}{TP + \alpha FN + \beta FP + \epsilon}\right]^{\gamma} \quad (2)$$

Where: TP is the true positive, FN is the false negative, FP is the false positive parameter, α and β control sensitivity to false negatives and false positives, respectively. We set α=0.7, β=0.3, and γ=1.5 to enhance attention on difficult-to-segment regions. ϵ is a small constant ($1e^{-6}$) ensuring numerical stability. Dice Loss (L_{Dice}): Dice Loss ensures good overlap between predicted masks and ground truth masks, defined as:

$$L_{Dice} = 1 - \frac{2 \times (P \cap T)}{|P| + |T| + \epsilon} \quad (3)$$

where P is the predicted segmentation and T is the ground truth. Boundary Loss ($L_{Boundary}$): This loss focuses explicitly on the boundaries between segmented regions, calculated using gradient edge maps derived using Sobel filters:

$$L_{Boundary} = \frac{1}{N} \sum_{i=1}^{N} |\nabla P_i - \nabla T_i| \quad (4)$$

where ∇P and ∇T are gradient maps of predictions and targets, respectively, and N denotes the total number of pixels. To balance region segmentation and boundary precision dynamically during training, we employ a ramp-up strategy for the boundary loss coefficient $\gamma_c(t)$:

$$\gamma_c(t) = \gamma_{start} + (\gamma_{end} - \gamma_{start}) \times \min\left(1, \frac{t}{t_{ramp}}\right) \quad (5)$$

Where: γ_{start} and γ_{end} define initial and maximum weights of boundary loss. In this study, these are set as 0.15 and 0.35 respectively. T is the current training epoch, and t_{ramp} is the epoch threshold to reach the maximum boundary loss contribution set as 20 epochs.

C. Datasets

In our experiments, we utilize the open-access Thyroid Nodule TN3K dataset [14], which is specifically curated for thyroid nodule segmentation tasks. The TN3K dataset consists of 3,493 B-mode ultrasound images with corresponding pixel-wise nodule annotations, collected from 2,421 patients at Zhujiang Hospital of Southern Medical University. The images were acquired between January 2016 and August 2020 using three different ultrasound systems—GE Logiq E9, Hitachi ARIETTA 850, and Mindray RESONA 70B—to ensure diversity in imaging equipment and acquisition settings.

III. EXPERIMENTS

A. Training and Evaluation

To evaluate model generalization, we conduct five-fold cross-validation. The dataset is divided into five folds, each containing approximately 698–699 images. In each iteration, four folds (2,794 images) are used for training, while the remaining fold (699 images) is used for validation. We use metrics such as Dice, IoU, Accuracy, Precision and Recall for our performance metrics. Across all five folds, our model achieves the following mean on the held-out validation sets Fig.2 and Table.1. These results demonstrate consistency across folds, with sub-1% standard deviations on all primary metrics.

Fig. 2 shows the evolution of Dice, IoU and accuracy on both the training and validation sets over 48 epochs. We employ a three-phase training strategy with 5-fold cross-validation. Phase 1: 5 epochs with frozen encoder, AdamW (lr=1e-4, wd=1e-4), 5-epoch warmup followed by cosine annealing. Phase 2: 35 epochs with all layers unfrozen, automatic threshold search every 5 epochs starting at epoch 15, early stopping (patience=10). Phase 3: 8 epochs at 512×512 resolution (lr=1e-5, batch size 8), threshold search every 2 epochs. All phases use morphological post-processing (opening kernel 3, closing kernel 5) and retain the largest

Table.1. Performance on validation data.

Metric	Mean
Dice	0.8875
IoU	0.7969
Accuracy	0.9688
Precision	0.8827
Recall	0.8955
Optimal Threshold	0.544

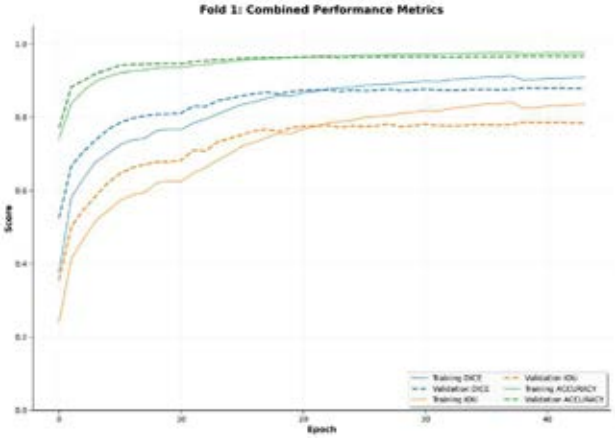


Fig. 2. Fold 1 the combined learning curves (Dice, IoU, Accuracy). Solid lines: training; dashed: validation.

connected component. The model with highest validation Dice is saved.

Fig. 3 illustrates the impact of our morphological post-processing. The bottom row particularly demonstrates how this approach handles fragmented predictions: the raw output shows two disconnected regions, but the morphological operations eliminate the spurious smaller fragment while preserving the main nodule region However, as illustrated in Fig. 4, morphology can also slightly shrink to thin or highly under-segmented regions, especially in very small nodules, and may occasionally eliminate true detections if the nodule is fragmented into multiple tiny islands.

B. Ablation study

To systematically evaluate the contribution of each proposed component, we conducted a comprehensive ablation study by progressively adding modules to a baseline UNet architecture. Table II shows the combination of DPAM and adaptive threshold search achieves the best performance with a Dice coefficient of 0.8875, representing a 2.5% relative improvement over the baseline. The negative impact of morphological post-processing underscores the importance of careful component selection, as additional processing steps do not always guarantee performance improvements in medical imaging applications. DPAM provides the largest single contribution highlighting the importance of attention mechanisms for medical image segmentation tasks involving complex anatomical structures.

C. Comparitive Experiment

To validate the effectiveness of our proposed architecture we conducted a comparative analysis against several thyroid segmentation methods. Our model achieved superior

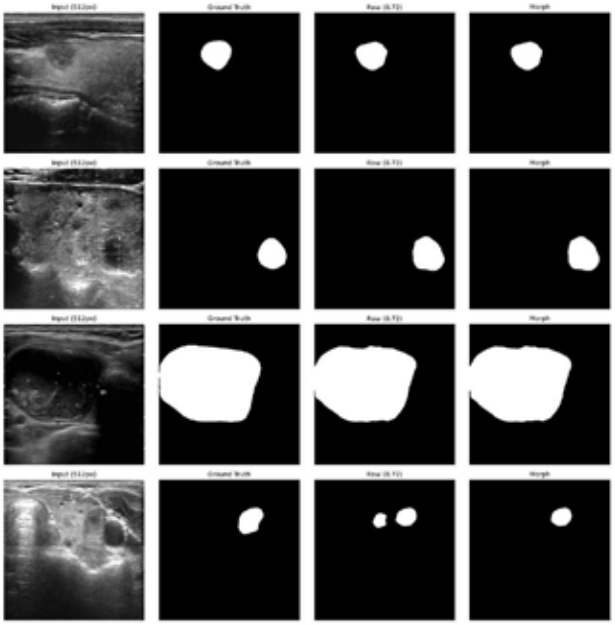


Fig. 3. Validation examples from Fold 1 comparing raw predictions with morphological post-processing. Columns show (left to right): input ultrasound image, ground truth mask, raw model output, and result after morphological post-processing.

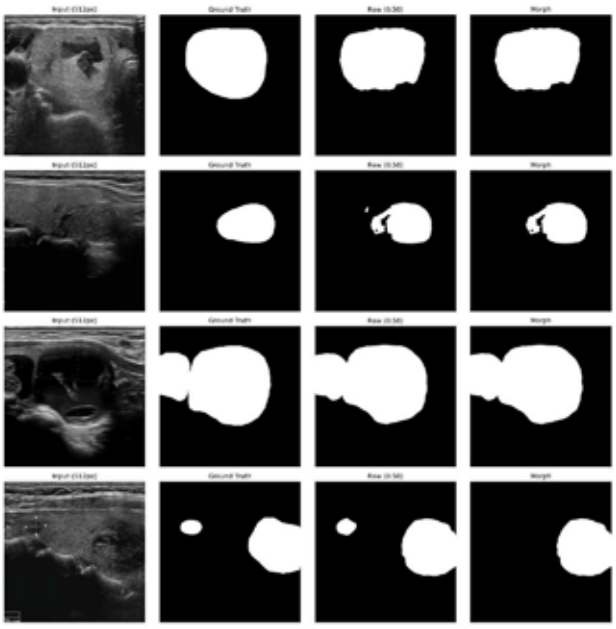


Fig. 4. Validation examples from Fold 2 comparing raw predictions with morphological post-processing. Columns show (left to right): input ultrasound image, ground truth mask, raw model output, and result after morphological post-processing

performance, demonstrating the highest Dice (0.8875), IoU (0.7969), Recall (0.8827), and Precision (0.8955) compared to all other tested models. Notably, the improvement over standard approaches like UNet and more recent enhancements such as BPAT-UNet, DPAM-UNet++ and Improved V-Net highlights the robustness and effectiveness of integrating

Table. 2. Ablation study results

	Dice	IoU	Acc	Recall	Precision
Baseline	0.8658	0.7691	0.9620	0.8700	0.8750
DPAM	0.8853	0.7931	0.9688	0.8775	0.8956
DPAM + Threshold	**0.8875**	**0.7969**	**0.9688**	**0.8827**	0.8955
DPAM + Threshold + Morph	0.8836	0.7914	0.9681	0.8770	**0.8966**

Table . 3. Performance comparison of different models

	DICE	IoU	Recall	Precision
Proposed Model	**0.8875**	**0.7969**	**0.8827**	**0.8955**
BPAT-UNet [9]	0.8423	0.7187	0.8557	0.8294
DPAM-UNet++ [10]	0.8310	0.7451	0.8702	0.8443
Improved V-Net [12]	0.8338	0.7550	-	-
UNet [4]	0.7829	0.6395		
SGUNet [2]	0.7323	0.6298	0.7120	0.8522

EfficientNet-B5, DPAM, and composite loss functions specifically tailored for thyroid ultrasound image segmentation.

IV. Conclusion

In this study, we introduced an EfficientNet-B5-based UNet++ segmentation architecture augmented with Dual-Path Attention Modules (DPAM), optimized for automated thyroid nodule segmentation in ultrasound images. The proposed model demonstrates strong performance, effectively addressing challenges such as low contrast, speckle noise, and blurred boundaries through dual-path attention mechanisms and composite loss optimization. Experimental results validate significant improvements over existing segmentation methods, indicating the robustness and clinical applicability of our approach.

However, our study is limited by the size and diversity of the dataset, suggesting potential improvements through training on larger and multi-center data. Future research directions include exploring advanced domain adaptation techniques and integrating explainability modules, further bridging the gap between automated segmentation and clinical practice.

References

[1] A. Prochazka and J. Zeman, "Thyroid Nodule Segmentation in Ultrasound Images Using U-Net With ResNet Encoder: Achieving State-of-the-Art Performance on All Public Datasets," AIMS Med. Sci., vol. 12, no. 2, pp. 124–144, Apr. 2025.

[2] H. Pan, Q. Zhou, and L. J. Latecki, "SGUNet: Semantic Guided UNet for Thyroid Nodule Segmentation," in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), Venice, Italy, Apr. 2020, pp. 72–75.

[3] J. Sun et al., "TNSNet: Thyroid Nodule Segmentation in Ultrasound Imaging Using Soft Shape Supervision," Comput. Methods Programs Biomed., vol. 215, Art. no. 106600, Aug. 2022.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI), Munich, Germany, Oct. 2015, pp. 234–241.

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "U-Net++: Redesigning Skip Connections to Exploit Multi-Scale Features in Image Segmentation," IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, 2019.

[6] H. Gong et al., "Thyroid Region Prior-Guided Attention for Ultrasound Segmentation of Thyroid Nodules," Comput. Biol. Med., vol. 155, Art. no. 106389, Mar. 2023.

[7] A. Ozcan, Ö. Tosun, E. Donmez, and M. Sanwal, "Enhanced-TransUNet for Ultrasound Segmentation of Thyroid Nodules," Biomed. Signal Process. Control, vol. 95, Art. no. 106472, 2024.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, May 2019, pp. 6105–6114.

[9] H. Bi et al., "BPAT-UNet: Boundary Preserving Assembled Transformer UNet for Ultrasound Thyroid Nodule Segmentation," Comput. Methods Programs Biomed., vol. 238, Art. no. 107614, Jan. 2023.

[10] P. Dong et al., "An Ultrasound Image Segmentation Method for Thyroid Nodules Based on Dual-Path Attention Mechanism-Enhanced U-Net++," BMC Med. Imaging, vol. 24, Art. no. 341, 2024.

[11] H. Kervadec et al., "Boundary Loss for Highly Unbalanced Segmentation," in Proc. Med. Imaging with Deep Learn. (MIDL), Innsbruck, Austria, July 2019, pp. 285–296.

[12] B. Yetginler and İ. Atacak, "An Improved V-Net Model for Thyroid Nodule Segmentation," Appl. Sci., vol. 15, no. 7, Art. no. 3873, 2025.

[13] N. Abraham and N. M. Khan, "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation," in Proc. IEEE 16th Int. Symp. Biomed. Imaging (ISBI), Venice, Italy, Apr. 2019, pp. 672–676.

[14] H. Gong et al., "TN3K: Thyroid Nodule 3493 A Public Dataset for Thyroid Nodule Segmentation," Comput. Methods Programs Biomed., vol. 215, Art. no. 106613, 2022.

[15] Y. Cao et al., "GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Seoul, Korea, Oct. 2019, pp. 1971–1980.

[16] Q. Wang et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, June 2020, pp. 11534–11542.

[17] S. Woo et al., "CBAM: Convolutional Block Attention Module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, Sept. 2018, pp. 3–19.

TIPTEKNO'25
MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Scolifix: Integrated Wearable Solution for Remote Scoliosis Rehabilitation

Fırat Matur
Biomedical Engineering
Bahcesehir University
Istanbul, Türkiye
firat.matur@bau.edu.tr

Gözde Hilal Öçalan
Biomedical Engineering
Bahcesehir University
Istanbul, Türkiye
gozde.ocalan@bahcesehir.edu.tr

Bora Büyüksaraç
Biomedical Engineering
Bahcesehir University
Istanbul, Türkiye
bora.buyuksarac@bau.edu.tr

***Abstract*—Scoliosis is a complex three-dimensional spinal deformity characterized by lateral curvature and vertebral rotation, often requiring individualized therapeutic interventions to prevent progression or to improve posture. Personalized Schroth exercises allow prescription of patient specific Schroth exercises, and they are often used in scoliosis treatment. In traditional methods, exercise instructions are typically conveyed verbally or through printed materials, which often lead to misunderstandings or incomplete execution by the patient. Digitally prescribed patient-specific exercises allow exercise critical parameters such as direction, velocity, target angles, number of repetitions, and duration to keep at the desired position, specified in detail through an electronic platform. With its client-server architecture, an integrated wearable technology allows telemedicine in scoliosis rehabilitation. The exercise program is digitally transmitted to the patient's device in a standardized format via wearable technology, enabling clear and consistent follow-through. This technical approach ensures accurate delivery of the exercises and allows for a precisely personalized and monitored treatment protocol. Using telemedicine and wearable technology in Schroth exercise allows physicians to monitor patients remotely and provide remote prescriptions to patients if needed. This article defines such a wearable technology named Scolifix (the patent application: 2025/007293-2025-GE-356269) and explains the standards that it complies with.**

***Keywords*—Wearable device, scoliosis, telemedicine.**

I. Introduction

Scoliosis is a musculoskeletal disorder characterized by a sideways curvature of the spine, typically beginning during childhood or adolescence, and, if not noticed early, can gradually worsen over time. Normally, the spine should appear as a straight line when viewed from the front; however, in scoliosis, it curves to the right or left, taking on an "S" or "C" shape [1].

The symptoms of scoliosis often involve changes in body symmetry: shoulder height asymmetry, one-sided waist or back prominence, or more noticeable curvature of the spine when the patient bends forward. Occasionally, patients experience back or lower back pain related from their scoliosis. In more severe cases, the curved spine can compress the rib cage, leading to shortness of breath. Those symptoms cause not only cosmetic concerns but also physical discomfort [2].

979-8-3315-5565-8/25/$31.00 © 2025 IEEE

The impact of scoliosis on quality of life can vary. Mild cases often do not cause significant health problems, but moderate to severe curvatures can lead to both physical and emotional challenges. Early diagnosis and treatment of scoliosis are critical during childhood to help prevent the condition from worsening [3].

There is no one-size-fits-all treatment: treatment for scoliosis is patient-specific, designed based on the degree of curvature, the patient's age, and the likelihood of progression. Regular checkups may be sufficient in mild scoliosis, whereas, for moderate cases, wearing a brace might be an option. The purpose of braces is to prevent the spine from curving further, especially when children are still growing. Physical therapy and targeted Schroth exercises also play a key role in managing scoliosis. These exercises aim to improve posture, strengthen the muscles around the spine, and promote muscular balance [4]. When conservative treatments are ineffective or with more severe cases, even a spinal fusion surgery may be necessary [5].

Physical therapy in scoliosis treatment focuses on preventing the curvature from worsening, improving posture, restoring muscular balance, and enhancing the patient's quality of life. These methods are personalized based on the type and degree of scoliosis that the patient is experiencing, and the patient's age. One of the most widely used rehabilitation techniques is the Schroth exercises. This 3D therapy includes targeted postures, breathing exercises, and muscle strengthening techniques specifically tailored to the individual's spinal curve.

The main goal of the Schroth therapy is to realign the spine in the opposite direction of the curve, reduce body asymmetry, increase lung capacity, and improve posture. Patients are trained on specific movements and positions based on the direction of their spinal curve. The exercises prescribed aim to counteract the rotation and lateral bending caused by their scoliosis. Training may also include rotational breathing in which the patient takes deep breaths into the compressed side of the rib cage, helping to expand the limited side and improve lung function. The therapy also helps patients become more aware of their bodies and learn how to maintain proper posture in daily life. Because Schroth therapy is customized for each patient, a unique program should be developed for each patient based on their scoliosis type and curvature. When done

consistently, Schroth exercises can slow the progression of the curve and, in some cases, even reduce it. Hence, for patients seeking non-surgical treatment options, Schroth exercises may offer a promising alternative [6].

However, implementing patient-specific Schroth exercises has some challenges. First of all, the exercises must be tailored to each individual. Every scoliosis case involves a different curvature pattern, rotation angle, and postural imbalance, which makes it difficult for patients to identify and perform the necessary positions independently. Moreover, since the exercises involve not just muscular work but also complex 3D breathing techniques, learning the correct breathing rhythm can be time-consuming and error-prone without professional guidance [7]. Maintaining consistency is another challenge: the exercises require discipline and routine that can be hard to achieve, especially for children and teens who may lack motivation or patience. Due to the technical complexity of some movements, doing Schroth exercises correctly at home may be difficult, and improper execution could reduce effectiveness or even worsen posture. For these reasons, traditionally, it is recommended that Schroth therapy initially be conducted under the supervision of an experienced physical therapist and then gradually transitioned to a home-based program as the patient gains confidence [8].

Because of these challenges, there is a growing need for a device that can assist patients in performing Schroth exercises correctly and independently when they are away from the clinic. Such a device could offer positioning support and guided instruction, helping patients follow their personalized therapy plans more effectively at home, which will bring in necessary motivation. Furthermore, having this device connected to physicians' systems will help physicians follow progress remotely and intervene on time if needed.

II. Device Design

Wearable telemedicine technology for Schroth exercise, i.e., Scolifix, has a client-server architecture and consists of three main components: two software components (with client and server architecture), and one wearable sensor module (Fig. 1). Scolifix is currently under patent review (Application - 2025/007293-2025-GE-356269).

A. Server module

The server module is a software designed as a multi-platform and multi-language application that also keeps an encrypted centralized patient database. The server module handles the following tasks:

- Provides the specialist doctor with CRUD services (create, read, update, and delete) on patient and exercise records.
- Respond to requests coming from the client module.
- CRUD services to patient-specific exercises, and push them to the client module.
- Maintain a history of exercise records and generate related reports.
- Allows data mirroring, backup, or restore if needed.

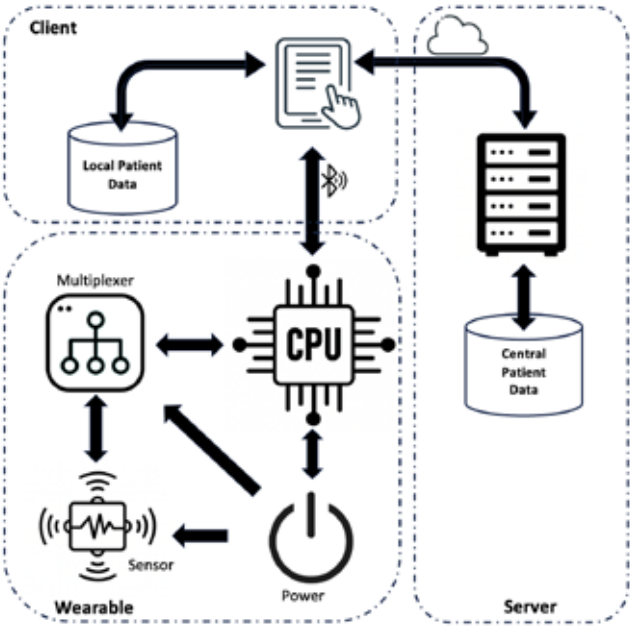


Fig. 1: Scolifix is a wearable telemedicine technology for Schroth exercise (the system architecture is under patent review - Application number: 2025/007293-2025-GE-356269).

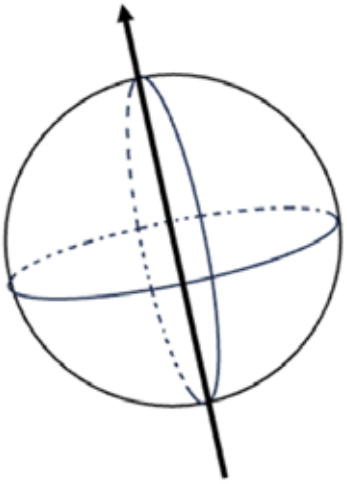


Fig. 2: Gimbal interface enables the client application to give visual directions on exercise. For each sensor, there is one gimbal, and with its colored visual alerts, the patient can see how the sensor and the attached anatomic point are moving. This visual real-time feedback allows the patient to complete exercises correctly.

- Push new firmware updates to client and wearable modules.
- Decode and encode industry-standard encrypted two way communication data to ensure secure communication with the client module.

B. Client module

The client module is a multi-platform application running on portable devices, such as iOS or Android. Also, it has an internal encrypted database to keep patient data assigned to devices. The client module handles the following tasks:

- Check for new firmware for the client and wearable modules, and push the updated firmware to the wearable sensor module.
- Receive patient-specific exercises created by the specialist from the server and update the locally stored patient exercise patterns.
- Keep a local copy of exercise definitions for the patient for offline operation, which is linked with the unique wearable identification number.
- Remind and initiate personalized exercises.
- Use the data received from the wearable module to monitor whether the patient is performing the prescribed exercises correctly, and notify the patient visually and audibly in real time. The target and real-time status of the IMUs are displayed using gimbal spheres (Fig. 2). The belts of the gimbal spheres turn red, yellow, or green depending on how close the real-time orientation matches the targets.
- Securely save the generated exercise records and push them to the server module encrypted. In case of a communication failure, retain data until communication is reestablished.
- Decode and encode industry-standard encrypted data to ensure secure communication with the server module.

C. Wearable module

The wearable module is a hardware platform with six subcomponents: the central processing unit (CPU), channel multiplexer (MUX), the power unit, inertial measurement sensor units (IMUs), data cables, and a special adjustable garment to keep all these components securely in their prescribed anatomic locations. Wearable module has at least 8 IMUs, which can be increased to a maximum of 64 if needed, and incresed number of IMUs allow high-resolution tracking for complex 3D movements. One of those IMUs is used as the geometric reference point for the others, and another is placed on the chest for breath monitoring. The properties and tasks of wearable unit are listed as follows:

- Equipped with a garment, made of flexible and washable fabric with pockets to keep all components (IMUs, power, CPU, and MUX) safely in their prescribed anatomic locations.
- Garment has flexible cable channels with hook-and-loop straps to secure connecting cables without restricting patient movements.
- Shielded data cables carry data between MUX and IMUs, and energy from the power unit to MUX and IMUs.
- IMUs continuously measure the direction, velocity, and acceleration of anatomical points.
- MUX allows communication between the CPU and IMUs.
- CPU selects the IMU channel, reads data, calculates relative acceleration, velocity, and direction with respect to the reference sensor, including breath volume and frequency.
- CPU keeps a unified device identifier (UID) which is also linked to the patient.
- CPU packs IMU data, includes time stamp and labels with UID, and wirelessly transfers to client module using Bluetooth technology.
- Power unit is equipped with a rechargeable battery, and its battery management circuitry allows safe charge, short circuit, and thermal protection of the battery.
- Power unit displays the battery charge level and energizes all electronic units of the wearable module.
- All components in contact with the patient are biocompatible, and all electronic components have IP65 dust and water spill resistance.

III. Device Compliance

The sensors, the client and server are expected to comply with Class IIb or relevant standards based on the intended market and categorized under four major headings:

- Medical device classification
- Quality and risk standards
- Wireless or telecommunication standards
- Data privacy and security standards

The standards are summarized in Tab. I for convenience.

A. Medical device classification

The wearable module will be actively influencing the treatment with real-time collected data, and later provide feedback to the patient as audio or visual feedback for scoliosis exercises, through the client module. Incorrect feedback could worsen scoliosis by improper posture reinforcement; therefore, the standards applicable to the device are Class IIb for EU (MDR2017/745) and Class II per 21 CFR 807.92 (510(k)) for US (FDA) with a predicate device category of 21 CFR 890.5050 or CFR 890.2000.

B. Quality and risk standards

For the EU market, to obtain the CE mark, ISO 13485 is mandatory. Similarly, ISO 14971 is needed for risk analysis and possible risk filing. Furthermore, the EU mandates a unified device identifier (UDI) per Article 27, Annex VI. FDA also recognizes ISO 14971 for risk assessment and requires UDI per 21 CFR 830. Possible software failures may result in medium injury; therefore, for the EU market, the IEC 62304 Class B standard will be applicable. Similarly, the audiovisual feedback through the user interface may affect patient safety, forcing IEC62366 compliance. The US also expects these standards to be met; furthermore, FDA Cybersecurity Guidance, including patchability and vulnerability monitoring, should be met.

TABLE I: Required medical device compliance summary, before marketing Scolifix in different regions

Region	Device Class	Software Class	UDI Requirement	Medical Standards	Data Privacy/Security	Wireless Standards
EU (MDR)	Class IIb	IEC 62304 Class B	Mandatory (MDR Art 27, EUDAMED)	MDR 2017/745 (Rule 13) ISO 13485 ISO 14971 EN 62366-1 EN 1041 (Labeling)	GDPR (Art. 32) ISO 27001 (cloud)	RED Directive ETSI EN 300 328 (BLE) EN 301 489 (EMC)
USA (FDA)	Class II	No formal class IEC 62304 used voluntarily	Mandatory (21 CFR 830, GUDID)	21 CFR 820 (QSR) FDA Cybersecurity Guidance (2023) ISO 13485 (MDSAP) ANSI/AAMI SW91	HIPAA (BAAs required) NIST SP 800-53 (cloud) UL 2900-1	FCC Part 15B (EMC) FCC Part 15C (Intentional radiators) Bluetooth SIG
UK (MHRA)	Class IIb	Class B	Planned (Mirrors EU, not enforced)	UK MDR 2002 BS EN 62304 BS EN 60601-1 (Safety)	UK GDPR NHS DSP Toolkit	UKCA (via RED compliance) BS EN 300 328
Canada	Class II	Class B	Voluntary (MDSAP alignment)	SOR/98-282 ISO 13485 (MDSAP) ISO 14971	PIPEDA PHIPA (Ontario)	ISED RSS-247 (BLE) CAN/CSA-C22.2 No. 60601-1
Australia	Class IIb	Class B	In development (Consultancy phase)	TGA Act 2002 ISO 13485 ISO 14971	Privacy Act 1988 (APP 8) OAIC Guidelines	ACMA RCM (BLE) AS/NZS 60601.1
Japan	Class II	IEC 62304 applied; class determined by risk (Class B)	Mandatory (PMDA, IMDRF-aligned)	PMD Act JIS Q 13485 JIS T 62304 JIS T 14971	APPI MHLW Cloud Guidelines	MIC Ordinance (BLE) JIS C 60601-1
China	Class II	Class B	Mandatory (NMPA database)	RSAMD (revised 2021) YY/T 0287 GB 9706.1 YY 0316	PIPL MLPS 2.0 (cybersecurity framework)	SRRC (BLE) GB 9706.1 (Safety)

C. Data privacy and security standards

Since the device processes and transmits real-time or offline health data, stricter safeguards apply. In the EU, GDPR and MDR compliance are expected. Furthermore, per Article 32 GDPR, pseudonymization of patient data is required. Since the data or server will most likely be hosted as a cloud/server structure, ISO 27001 will also be required. For the US, HIPAA security implementation (NIST SP 800-53; NIST SP 800-66) applies, and a Business Associate Agreement (BAA) is required if cloud providers are utilized for hosting and storing protected health information (PHI).

D. Wireless and Telecommunication standards

Due to the use of Bluetooth data communication in wearable devices and client applications, BLE Certification is needed (ETSI EN 300 328 / FCC Part 15). Furthermore, for the EU, per the RED Directive, a CE marking is required for radio equipment. For the US, FCC Part 15B (EMC) and 15C intentional radiators compliance are required.

IV. Conclusion

The device design has reached a mature stage, with its operational mechanism clearly defined and aligned with relevant international compliance standards. A physical prototype has not yet been entirely constructed; on the other hand, its detailed design documentation supports its technical feasibility and regulatory alignment. A patent application has been filed (2025/007293-2025-GE-356269) to secure the intellectual property of the Scolifix.

A comprehensive usability and clinical testing stage will be the critical future steps in evaluating the effectiveness and safety of the Scolifix. Clinical tests of Scolifix with patient populations or in real-world therapeutic contexts have not been started; so, the clinical relevance has not been assessed yet. However, despite this pending stage, with its comprehensive design, including non-invasive operation, user-specific therapeutic feedback, integration with digital health platforms, and remote monitoring and patient treatment by the physician, Scolifix has promising features in telemedicine. Further studies will focus on prototype construction, followed by usability testing and clinical validation to support its translational readiness.

References

[1] S. de Reuver, J. F. Homans, T. P. Schlösser, M. L. Houben, V. F. Deeney, T. B. Crowley, R. Stücker, S. Pasha, M. C. Kruyt, D. M. McDonald-McGinn *et al.*, "22q11. 2 deletion syndrome as a human model for idiopathic scoliosis," *Journal of Clinical Medicine*, vol. 10, no. 21, p. 4823, 2021.

[2] T. Cheng, E. Diarbakerli, A. Simony, M. Ø. Andersen, A. Danielsson, J. Kere, E. Einarsdottir, and P. Gerdhem, "Pain in idiopathic scoliosis not associated with known genetic variants for pain," *Pain reports*, vol. 10, no. 1, p. e1227, 2025.

[3] N. Lalechou, K. Chandolias, A. Foti, N. Karavidas, T. Apostolou, I. Kallistratos, T. Besios, and A. Hristara-Papadopoulou, "The correlation between clinical and epidemiological characteristics of adolescents with idiopathic scoliosis and their quality of life: A pilot cross-sectional study," *International journal of health sciences*, vol. 8, no. S1, pp. 789–802, 2024.

[4] Y. Fan, Q. Ren, M. K. T. To, and J. P. Y. Cheung, "Effectiveness of scoliosis-specific exercises for alleviating adolescent idiopathic scoliosis: a systematic review," *BMC musculoskeletal disorders*, vol. 21, no. 1, p. 495, 2020.

[5] I. Saleh, W. D. Hutami, D. Librianto, M. Prasetyo, A. J. Rahyussalim, A. Hendriarto, E. A. Noor, and A. Rifki, "The development of new scoring system to define the presence of instability and the need of fusion in degenerative lumbar spinal stenosis-jakarta instability score," *Global Spine Journal*, vol. 15, no. 1, pp. 241–250, 2025.

[6] S. Schreiber, E. C. Parent, D. L. Hill, D. M. Hedden, M. J. Moreau, and S. C. Southon, "Patients with adolescent idiopathic scoliosis perceive positive improvements regardless of change in the cobb angle–results from a randomized controlled trial comparing a 6-month schroth intervention added to standard care and standard care alone. sosort 2018 award winner," *BMC musculoskeletal disorders*, vol. 20, no. 1, p. 319, 2019.

[7] S. Schreiber, E. C. Parent, E. K. Moez, D. M. Hedden, D. Hill, M. J. Moreau, E. Lou, E. M. Watkins, and S. C. Southon, "The effect of schroth exercises added to the standard of care on the quality of life and muscle endurance in adolescents with idiopathic scoliosis—an assessor and statistician blinded randomized controlled trial:"sosort 2015 award winner"," *Scoliosis*, vol. 10, no. 1, p. 24, 2015.

[8] A. A. Abdel-Aziem, O. R. Abdelraouf, S. A. Ghally, H. A. Dahlawi, and R. E. Radwan, "A 10-week program of combined hippotherapy and schroth's exercises improves balance and postural asymmetries in adolescence idiopathic scoliosis: a randomized controlled study," *Children*, vol. 9, no. 1, p. 23, 2021.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Functional Brain Network Analysis In Dyslexia Using Graph Theory Metrics

Ozlem Karabiber Cura
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
ozlem.karabiber@ikcu.edu.tr

Sevde Dervisoglu
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
dervisoglusevde@gmail.com

Günet Eroglu
Dept. of Computer Engineering
Bahçeşehir University
Istanbul, Turkey
gunet.eroglu@eng.bau.edu.tr

***Abstract*—This study investigated brain network changes in 14 children diagnosed with developmental dyslexia over a 60-day cognitive training period using EEG-based functional connectivity and graph theory analysis. Features such as CPSD, MSC, correlation coefficient, and entropy-based measures are extracted from EEG recordings. Subsequently, seven graph metrics (assortativity, transitivity, clustering coefficient, global efficiency, local efficiency, modularity, and characteristic path length) are calculated. The extracted features are classified using machine learning models. The analyses showed increased brain network integration and strengthened long-range connectivity by Day 60. These findings suggest that cognitive intervention may promote neuroplasticity in children with dyslexia and that EEG-based graph metrics could serve as potential biomarkers for monitoring treatment progress.**

***Index Terms*—Developmental dyslexia, EEG, functional connectivity, graph theory, neuroplasticity, cognitive training, machine learning, CPSD, coherence, entropy, brain network metrics.**

I. INTRODUCTION

Developmental dyslexia is a prevalent and heterogeneous neurodevelopmental disorder characterized by persistent difficulties with reading, decoding, and spelling, despite normal intelligence and adequate educational conditions [1]. The prevalence of dyslexia in society is not well understood by families or educators. According to the Turkish Dyslexia Association, 41,600 of Turkey's 85 million citizens have been diagnosed with dyslexia, which represents just 0.05% of the country's entire population, despite the fact that the illness is thought to affect at least 10% of people in every society [2]. Although traditionally viewed as a phonological processing disorder, growing evidence suggests that dyslexia stems from a more complex interplay of deficits in phonological awareness, working memory, attentional control, and executive functioning. These deficits indicate the involvement of multiple, distributed brain regions and network-level dysfunctions, rather than a single localized abnormality [3].

Teacher reports, observations, and standardized tests are used to diagnose dyslexia in youngsters. Frequent reading lag may be a sign of dyslexia; however, timing and criteria can differ, requiring further neuropsychological or medical testing [4]. Electroencephalography (EEG) has emerged as a powerful, non-invasive, and cost-effective tool for investigating the temporal dynamics of brain function in neurodevelopmental disorders. Owing to its excellent temporal resolution, EEG is particularly suited for capturing the fast-paced neural processes involved in language and reading. Potential implications for early dyslexic identification are suggested by Castillo-Barnes et al.'s work, which examines dyslexia in Spanish readers using cross-frequency coupling analysis and EEG signals [4]. Significant differences in fixation-related potential amplitudes between chronological age controls and dyslexics during serial Rapid-Automatized Naming (RAN) tasks are shown by an electroencephalography and eye tracking investigation, indicating a link between dyslexia and neurocognitive processes [5]. According to a study that uses EEG data to examine brain mechanisms in low-level auditory processing, dyslexic patients differ from control patients in their network features, including a loss of small-world topology [6]. Using a k-nearest neighbor classifier, a study of dyslexic children revealed a 96.7% classification accuracy, along with poor functional segregation and disrupted information transfer in their brain networks [7].

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE

A. Objectives of the proposed study

The primary objective of this study is to investigate the neuroplastic effects of a 60-day cognitive training protocol on children diagnosed with developmental dyslexia using EEG-based functional connectivity and graph theoretical analysis. Specifically, the study aims to:

1) Quantify changes in functional brain connectivity over time by extracting both linear and nonlinear EEG-based connectivity measures (CPSD, MSC, Coef, Covv, RE_Coef, CE_Coef).
2) Evaluate alterations in large-scale brain network topology by computing seven graph theory metrics (assortativity, transitivity, clustering coefficient, global efficiency, local efficiency, modularity, and characteristic path length).
3) Classify pre- and post-training brain network states using supervised machine learning algorithms (Fine Tree, Logistic Regression, Fine KNN, Linear SVM, Bagged Trees) to identify patterns indicative of training-induced neuroplasticity.

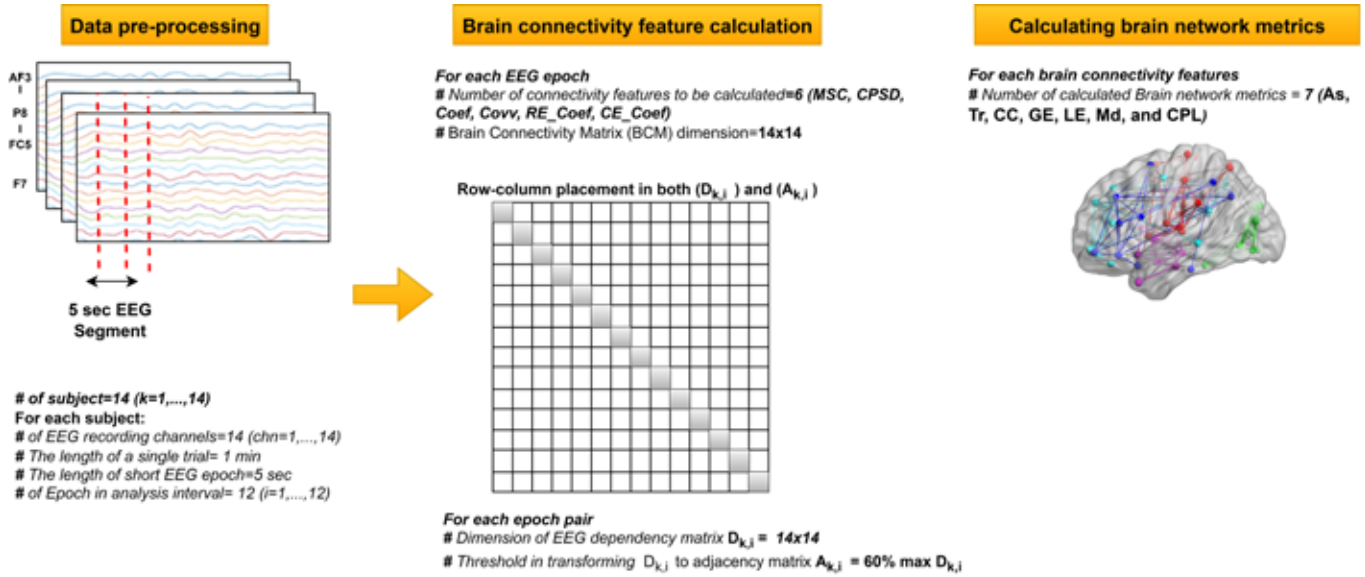


Fig. 1: The main steps of the EEG analysis procedure.

4) Assess the potential of EEG-derived graph metrics as biomarkers for monitoring the effectiveness of cognitive interventions in dyslexia.

Through this multi-modal approach, the study seeks to deepen the understanding of the neural basis of dyslexia and contribute to the development of objective, EEG-based tools for tracking treatment progress in educational and clinical settings.

II. MATERIALS AND METHODS

Figure 1 illustrates that EEG signals are recorded from 14 subjects across 14 channels, with each recording lasting 1 minute. These 1-minute recordings are analyzed by dividing them into 12 short periods (epochs) of 5 seconds. A 14x14 dependency matrix $(D_{k,i})$ is calculated for each epoch pair, and this matrix is converted to an adjacency matrix $(A_{k,i})$ by taking the threshold as 60% of the maximum value. This process is performed to compare brain connectivity before and after treatment using six different brain connectivity features.

A. Subjects and Experimental Data

EEG data are acquired from 14 children diagnosed with developmental dyslexia across two sessions: Day 1 (pre-training) and Day 60 (post-training). Each participant underwent a 1-minute EEG recording in a relaxed, eyes-open state, conducted in a quiet room with minimal external distractions. Data collection is performed using a 14-channel EEG system with electrode placement conforming to the international 10–20 system [8]. The electrode positions included AF3, AF4, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6,F4, and F8. Signals are recorded at a sampling rate of 500 Hz with an online band-pass filter of 0.5–45 Hz. The ground electrode is located at FPz, and the reference was placed on the left mastoid [2].

Preprocessing of EEG data is conducted using MATLAB R2024b and the EEGLAB toolbox. Raw EEG signals are band-pass filtered between 0.5 Hz and 45 Hz to eliminate slow drifts and high-frequency noise, preserving activity across the delta, theta, alpha, beta, and low gamma bands relevant to cognitive processing. Artifacts such as eye blinks and muscle activity were removed using Independent Component Analysis (ICA). The 1-minute continuous EEG data are then segmented into twelve non-overlapping epochs of 5 seconds each. Epochs with voltage fluctuations exceeding $\mp 100\mu V$ are excluded. Baseline correction is applied to all valid epochs before further feature extraction and analysis.

B. Functional Connectivity Calculation

To capture inter-regional communication, six different pair-wise connectivity measures are calculated between all 91 electrode pairs (computed from 14 electrodes) for each 5-second epoch. These measures included cross-power spectral, magnitude-squared coherence, Pearson correlation coefficient, covariance, correntropy coefficient, and cohentropy coefficient. These features are selected to capture both linear (Coef, MSC, Covv) and nonlinear (RE_Coef, CE_Coef) dependencies among brain regions. Feature values are computed for each epoch

C. Network Analysis Based on Graph Theory

In this study, graph theoretical analysis is applied to EEG-derived functional connectivity networks across Day 1 and Day 60. Functional connectivity matrices are constructed per subject and EEG epoch using multiple estimators. To ensure connected and meaningful network structures, proportional thresholding retaining 60% of the strongest connections (i.e., threshold range from 0.6 to 1.0) is applied. The thresholded adjacency matrix is represented as an undirected weighted

graph with 105 segments as nodes. Seven graph metrics are computed for each network: assortativity (As), transitivity (Tr), clustering coefficient (CC), global efficiency (GE), local efficiency (LE), modularity (Md), and characteristic path length (CPL). These metrics are calculated for each connectivity method, producing a feature set of 7 metrics × 6 connectivity features = 42 features. Lastly, these features are inputs to supervised classification models to evaluate temporal changes in network topology [8], [9].

D. Machine Learning Classification

In this study, supervised machine learning algorithms are applied to distinguish between Day 1 and Day 60 EEG data using a dataset consisting of 42 features. These included seven graph metrics calculated across six connectivity features. Fine KNN classification is used, with performance evaluated using 5-fold cross-validation. The performance of the classifier is assessed in this study using the following performance criteria: Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), Positive Predictive Value (PPV), False Discovery Rate (FDR), and F1 Score (F1-S).

In addition to classification, circular connectivity graphs are created to visualize functional connectivity patterns. The brain is divided into 105 anatomical regions, and connections between these regions are mapped to illustrate changes in brain network organization after training. This combined approach allowed for both quantitative and visual interpretation of the neural changes over time.

III. RESULTS AND DISCUSSION

This study examined EEG-based functional connectivity and brain network changes in 14 children with developmental dyslexia before and after a 60-day cognitive training program. Linear and nonlinear connectivity measures are extracted from 14-channel EEG data, and graph theoretical metrics are computed to evaluate network topology. Using 42 features derived from six connectivity types and seven graph metrics, machine learning models classified Day 1 and Day 60 recordings with high performance. Circular connectivity graphs further visualized enhanced network integration and long-range connections post-training, suggesting cognitive intervention-induced neuroplasticity.

Table I summarizes the classification performance of the Fine KNN algorithm across six EEG-based connectivity measures (CE_Coef, Coef, Covv, CPSD, MSC, RE_Coef), averaged across 14 subjects in distinguishing between pre- and post-training sessions. Among the features, RE_Coef achieved the highest performance with 99.16% accuracy and an F1-S of 99.3%, followed by MSC and Coef with similarly strong results. In contrast, CPSD showed the lowest performance, with reduced specificity and the highest false discovery rate. Overall, the results highlight that nonlinear connectivity metrics, particularly RE_Coef, are more effective in capturing training-related neuroplastic changes in children with dyslexia when used with Fine KNN classification.

TABLE I: Classification results of the fine KNN model using six EEG connectivity features, evaluated with performance metrics.

Connectivity Features	ACC	SEN	SPE	PPV	FDR	F1-S
CPSD	88.41	89.79	84.99	91.19	8.81	90.36
MSC	98.43	99.10	97.22	98.39	1.61	98.72
Coef	98.13	99.59	95.89	97.51	2.49	98.52
Covv	91.86	92.54	90.26	94.25	5.75	93.34
CE_Coef	96.29	96.67	95.47	97.19	2.81	96.88
RE_Coef	99.16	99.08	99.34	99.53	0.47	99.30

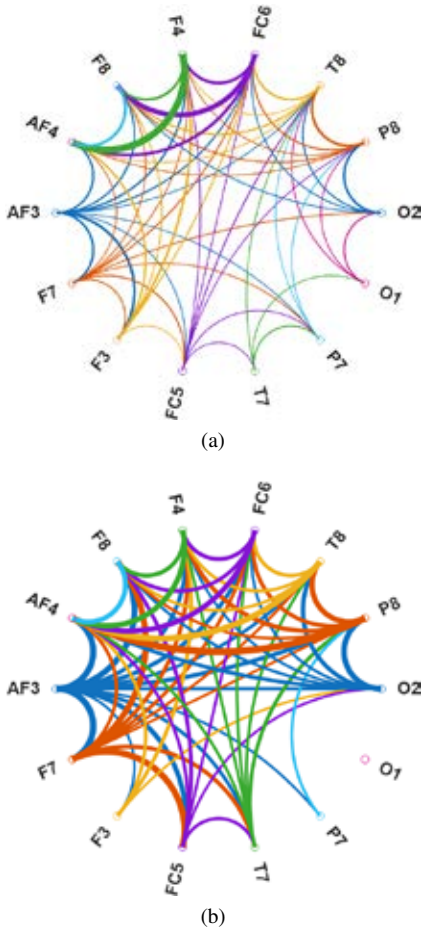


Fig. 2: Circular Connectivity Graph of RE_Coef for (a) Day 1 and (b) Day 60

Circular connectivity graphs of one epoch of one subject based on the RE_Coef metric, comparing functional brain networks at Day 1 and Day 60 of the cognitive training protocol, are given in Figure 2 to visualize functional connectivity patterns. Each node represents one of the 14 EEG electrode positions (AF3, AF4, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8), and the colored lines (edges) indicate functional connections derived from nonlinear RE_Coef values. The Day 1 graph shows sparser and more localized connectivity patterns, with fewer long-range connections between distant brain regions. In contrast, the Day 60 graph demonstrates a clear increase in both the number and distribution of inter-regional connections, particularly between frontal, temporal,

and occipital lobes. This denser and more integrated network structure observed after training suggests improved global communication efficiency and reflects training-induced neuroplastic changes

The graph in Figure 3 examines the subject-based ACC and SEN distributions of different connectivity metrics (MSC, CPSD, Coef, Covv, RE_Coef, and CE_Coef). The MSC, RE_Coef, and CE_Coef metrics have high median values and narrow distributions across both performance measures, providing the most successful and consistent results. In contrast, CPSD and Covv have the widest distribution and low outliers, demonstrating the lowest consistency. Coef, on the other hand, exhibited a more limited success rate and a wider variance, demonstrating a moderate performance. These results demonstrate that the MSC, RE_Coef, and CE_Coef metrics are particularly reliable and stable.

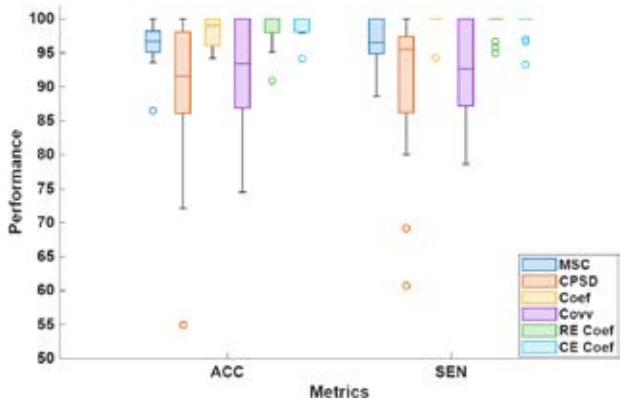


Fig. 3: Box plot of subject-based performance distribution for ACC and SEN values.

IV. CONCLUSION

This study demonstrates that a 60-day cognitive training program led to measurable changes in the brain network organization of children with developmental dyslexia, as assessed through EEG-based functional connectivity and graph theoretical analysis. The observed increases in global and local efficiency, clustering, and long-range connections—particularly evident in nonlinear metrics like RE_Coef—indicate enhanced network integration and suggest that targeted interventions can promote neuroplasticity. Machine learning models successfully classified pre- and post-training brain states, supporting the potential of EEG-derived graph metrics as objective biomarkers for monitoring cognitive treatment outcomes. Overall, the findings highlight the value of combining EEG, graph theory, and machine learning to understand and track functional brain changes in dyslexia.

It is particularly noteworthy that there is a significant increase in both the number and distribution of intraregional connections between the frontal, temporal, and occipital lobes. This denser and more integrated network structure observed after training suggests improved global communication efficiency and reflects training-induced neuroplastic changes. The RE_Coef metric, which captures higher-order statistical dependencies, appears particularly sensitive to these functional reorganizations in the brains of dyslexic children.

ACKNOWLEDGMENT

Following an explanation of the experimental procedure by the research ethics committee, approval of the study protocol by the Yeditepe University Ethics Committee, and registration of the clinical trial with the Turkey Pharmaceuticals and Medical Devices Agency (TİTÇK) (Nbr: 71146310-511.06,2.11.2018), all participants gave their informed consent.

REFERENCES

[1] R. L. Peterson and B. F. Pennington, "Developmental dyslexia," *The lancet*, vol. 379, no. 9830, pp. 1997–2007, 2012.

[2] G. Eroglu, M. Köprülü, and B. Karabacak, "Dyslexia biomarker detection with qeeg data in children: Feasibility, acceptability, economic impact, pilot study and survey results," *Qeios*, pp. 1–19, 2022.

[3] M. J. Snowling and C. Hulme, "Annual research review: The nature and classification of reading disorders–a commentary on proposals for dsm-5," *Journal of child psychology and psychiatry*, vol. 53, no. 5, pp. 593–607, 2012.

[4] D. Castillo-Barnes, N. J. Gallego-Molina, M. A. Formoso, A. Ortiz, P. Figueiredo, and J. L. Luque, "Probabilistic and explainable modeling of phase–phase cross-frequency coupling patterns in eeg. application to dyslexia diagnosis," *Biocybernetics and Biomedical Engineering*, vol. 44, no. 4, pp. 814–823, 2024.

[5] C. Christoforou, A. Fella, P. H. Leppänen, G. K. Georgiou, and T. C. Papadopoulos, "Fixation-related potentials in naming speed: A combined eeg and eye-tracking study on children with dyslexia," *Clinical Neurophysiology*, vol. 132, no. 11, pp. 2798–2807, 2021.

[6] N. J. Gallego-Molina, A. Ortiz, F. J. Martínez-Murcia, M. A. Formoso, and A. Giménez, "Complex network modeling of eeg band coupling in dyslexia: An exploratory analysis of auditory processing and diagnosis," *Knowledge-based systems*, vol. 240, p. 108098, 2022.

[7] N. G. Seshadri, B. K. Singh, and R. B. Pachori, "Eeg based functional brain network analysis and classification of dyslexic children during sustained attention task," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4672–4682, 2023.

[8] H. Chen, Y. Song, and X. Li, "A deep learning framework for identifying children with adhd using an eeg-based brain network," *Neurocomputing*, vol. 356, pp. 83–96, 2019.

[9] S. Aydın, F. H. Cetin, M. Ç. Uytun, Z. Babadagi, A. S. Gueven, and Y. Işık, "Comparison of domain specific connectivity metrics for estimation brain network indices in boys with adhd-c," *Biomedical Signal Processing and Control*, vol. 76, p. 103626, 2022.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Determination of the Effect of Anticancer Laser Therapies on the Cancer Cell Migration and Invasion via Artificial Intelligence Methods

1st Sümeyra Arslan
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Turkey
sum.rsln@gmail.com

2nd Özlem Karabiber Cura
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Turkey
ozlem.karabiber@ikcu.edu.tr

3rd Büşra Sirek
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Turkey
busra.sirek@ikcu.edu.tr

4th Nermin Topaloglu Avsar
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Turkey
nermin.topaloglu@ikcu.edu.tr

***Abstract*—Photodynamic treatment (PDT) and photobiomodulation (PBM) demonstrate potential in targeting cancer cells and reducing their motility. This research examines the impact of photodynamic therapy (PDT) using Chlorin e6 (Ce6) as a photosensitiser, both with and without low-level laser therapy, on the migration of human prostate cancer cells (PC3), assessed through in vitro scratch wound healing assays. This integrative methodology facilitated objective and consistent measurement of cellular motility, offering insights into the effects of photodynamic therapy and PBM on prostate cancer cell migration, and illustrating their potential synergy as combinatory anti-metastatic therapies. The results demonstrate that the proposed method effectively facilitates the monitoring of cell migration dynamics and the evaluation of treatment efficacy.**

***Index Terms*—Cancer Research, Photodynamic Therapy, Cell Migration, Image Processing, Prostates Cancer, Scratch Assay**

I. INTRODUCTION

Cancer metastasis is the leading cause of cancer-related deaths worldwide, highlighting the urgent need for effective therapeutic options targeting the pathways that facilitate tumour cell migration and invasion [1]. Photodynamic therapy (PDT) is a promising minimally invasive treatment that uses photosensitizers activated by certain light wavelengths to make reactive oxygen species (ROS), which then kill cancer cells [4, 9]. PDT not only induces direct cellular apoptosis but also modulates various cellular processes, including migration, invasion, and adhesion [7, 8]. Recent research has examined the combination of PDT with photobiomodulation (PBM), a low-level laser treatment known for its impact on cellular metabolism and signalling, to improve therapeutic efficacy and diminish migratory consequences [7, 10]. It is important to look at how medicines affect the movement of cancer cells in order to improve treatment plans that will lower the risk of metastasis and improve patient outcomes. Accurate quantification of cell migration dynamics is essential for evaluating the efficacy of antimetastatic treatments. The in vitro scratch wound healing experiment is an established technique for examining collective cell migration by creating a uniform wound area in a confluent cell monolayer and monitoring its closure over time [5, 11]. Manual examination of wound healing images is time-consuming, subjective and prone to variability between observers, limiting its reliability and efficiency [6. To overcome these constraints, contemporary digital image processing methodologies integrated with machine learning algorithms have been formulated to automate and standardise the evaluation of cell migration tests [1, 3, 12]. Automated pipelines like these cut down on human error, make it easier to repeat experiments, and make it easier to screen a lot of therapeutic interventions at once. To quantify in vitro cell migration tests, the initial crucial step is the segmentation of input pictures to distinguish cellular regions from the wound area (non-cellular parts). After achieving this binary classification, the cell migration process can be quantitatively evaluated by determining the rate of scratch wound closure over time [2, 5].

This study investigated the impact of Chlorin e6-mediated PDT, both alone and in combination with PBM, on the migratory behaviour of PC3 cells. Cells were treated with diode lasers at 655 nm and 808 nm wavelengths at specified energy levels. A wound scratch assay was carried out, with imaging using light microscopy. A custom automated image analysis pipeline, which included grayscale conversion, noise reduction, segmentation, wound area masking, skeletonization, and feature extraction, allowed for precise quantification of cell migration dynamics, highlighting PDT and PBM's modulatory influence as potential anti-metastatic combinations.

II. MATERIALS AND METHODS

A. Cell Culture Experiments

PC3 cells were grown in RPMI-1640 medium with 10% FBS, 1% L-glutamine and 1% penicillin-streptomycin under normal conditions $(37C, 5\%CO_2)$. Cells from passages 5–15 were sown at 110 cells per well in 96-well plates and adhered for 24 hours before treatment. Diode lasers at 655 and 808 nm (400 mW output) were used to achieve energy densities of 50 J/cm² and 5 J/cm², respectively.The laser parameters were chosen in accordance with previously reported ranges for PDT

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE

and PBM in prostate cancer models, providing reliable photobiological responses while minimizing the risk of thermal artifacts [1,15]. Chlorin e6 (Ce6) was used as a photosensitiser in PDT methods. Five experimental groups were formed: control (no treatment), laser (655 nm irradiation), PBM (808 nm irradiation), PDT (Ce6 incubation followed by 655 nm irradiation), and PBM + PDT (808 nm pre-irradiation, Ce6 incubation, then 655 nm irradiation).

The scratch wound experiment was carried out by making consistent scratches in confluent monolayers with a pipette tip. Following irradiation, cells were cultured under usual conditions, and scratch wound closure was examined at predetermined intervals using inverted light microscopy. A unique MATLAB image processing pipeline was used to provide quantitative analysis of cell migration, which was validated with manual measurements in ImageJ.

B. Image Preprocessing and Feature Extraction

To make the intensity distribution easier to understand and eliminate color information, all of the microscopic images were first changed to grayscale. This made the calculations less complicated. Figure 1. shows the whole pre-processing and feature extraction method. A Gaussian filter was used to get rid of high-frequency noise while keeping the structural boundaries. The Gaussian filtering technique is mathematically described as a convolution:

$$I_{\text{filtered}}(x, y) = (I * G_\sigma)(x, y) \tag{1}$$

where $I(x, y)$ is the intensity of the image at point (x, y) and G_σ represents the Gaussian kernel with standard deviation σ.

To enhance the visibility of scratch wound boundaries and improve segmentation accuracy, histogram equalization was used. This process redistributes pixel intensities by transforming the image histogram into a near-uniform distribution, defined by:

$$s_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n} \tag{2}$$

where r_k is the input intensity level, n_j is the number of pixels with intensity r_j, and n is the total number of pixels. Adaptive thresholding, guided by Otsu's method, was employed for binarization. Otsu's method determines an optimal threshold t^* that maximizes inter-class variance:

$$t^* = \arg\max_t \left[\omega_0(t)\omega_1(t)\left(\mu_0(t) - \mu_1(t)\right)^2\right] \tag{3}$$

where ω_0 and ω_1 are class probabilities and μ_0, μ_1 are class means. We then used a zero-crossing method to find the edges of the consolidated cellular areas and the scratch wound. This method finds boundaries by looking for places where the intensity of a filtered image varies. The length of the major axis was calculated for each edge that was found. The cellular region and the scratch wound were separated by the two longest edges.

Finally, a distance map was generated by computing the minimum Euclidean distances between opposing boundaries, and the average value was used as a quantitative metric for scratch wound width.

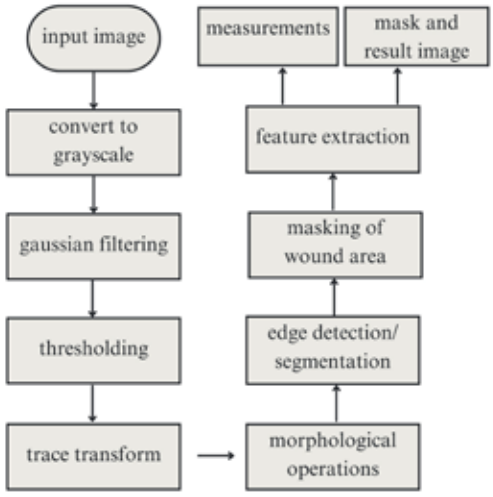


Fig. 1: Flowchart of the proposed image preprocessing and feature extraction algorithm

A local energy map was made to look at the shape of cells and the properties of the scratch wound area. This method looks at the local structural complexity in the greyscale image by looking at how the intensity and edge information changes in different areas. This gives us an idea of how the texture of the cells is spread out. The Trace Transform was also utilised to extract features. This is a more general version of the Radon transform that combines functions along straight lines at different angles and offsets. It lets you get invariant shape features that work even when the shape is rotated, scaled, or moved. Mathematically, it can be expressed as:

$$T(\phi, \rho) = \mathcal{T}[f(x, y)] = \int_{L_{\phi,\rho}} \tau(f(x, y))ds \tag{4}$$

where $L_{\phi,\rho}$ is a line parameterized by angle ϕ and distance ρ from the origin, and τ is a trace functional such as mean, max, or entropy.

From the segmented scratch wound and cell regions, several quantitative features were extracted, including area (A), convex area (CA), eccentricity (Ecc), equal diameter (EqD), equal Ellipse area (EqElpA), minimum Feret diameter (MinFD), factor R (Rf), modified ratio (ModR), boundary box area (BBA) and shape factor 1 (SF1) [14]. The scratch wound area at each time point was used to calculate closure percentage as:

$$\text{Scratch wound Closure } (\%) = \left(\frac{A_0 - A_t}{A_0}\right) \times 100 \tag{5}$$

where A_0 is the initial calculated scratch gap feature and A_t is the calculated scratch gap feature at time t [22]. Cell migration distance and width changes are also tracked longitudinally to assess therapeutic impacts.

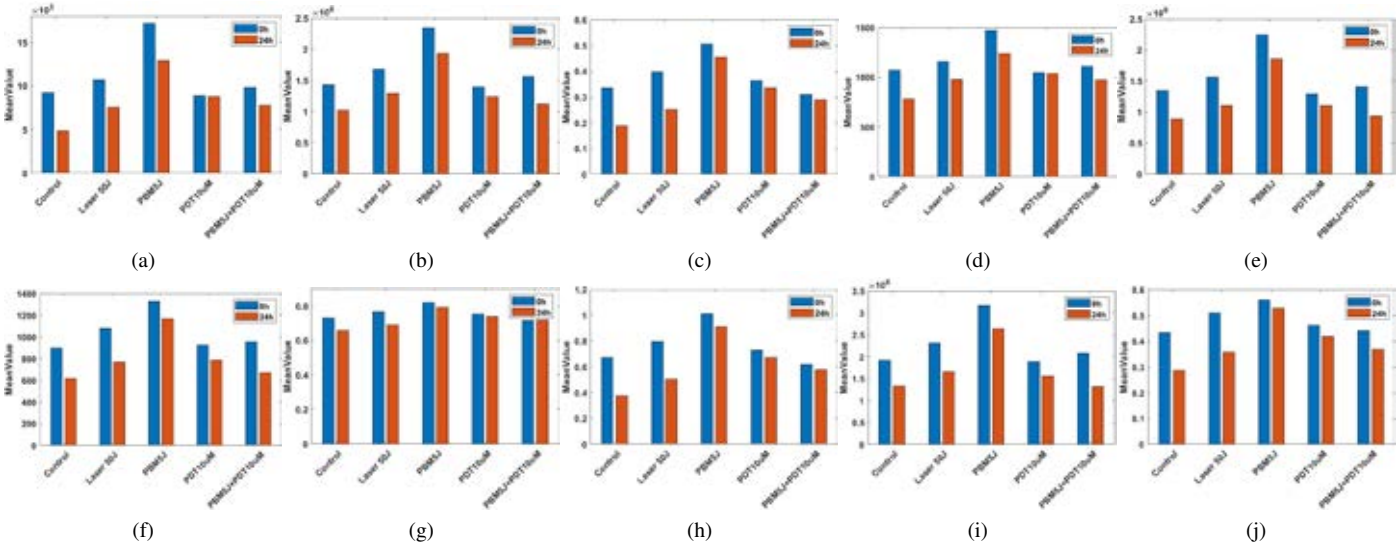


Fig. 2: Extracted quantitative features from segmented scratch gap and cell regions: (a) A, (b) CA, (c) Ecc, (d) EqD, (e) EqElpA, (f) MinFD, (g) Rf, (h) ModR, (i) BBA, and (j) SF1.

III. RESULTS

The image processing pipeline (Figure 1) encompassed a sequence of preprocessing operations, contrast enhancement, Gaussian smoothing, and histogram equalization followed by segmentation procedures incorporating adaptive thresholding and zero crossing edge detection, and culminating in quantitative feature extraction. This integrated methodology enabled precise delineation of scratch gap boundaries under variable imaging conditions and facilitated the derivation of morphological and geometric descriptors for subsequent quantitative analysis[3].

Quantitative features extracted from scratch assays at 0 h and 24 h were analysed across the five experimental groups (Control, Laser 50J, PBM5J, PDT10 µM, and PBM5J+PDT10 µM). As illustrated in Figure 2, the control group exhibited the most pronounced variations between 0 h and 24 h, whereas PDT-treated groups demonstrated minimal changes, indicating a significant inhibition of cell migration. Size-related features, including A, CA, EqD, EqElpA, MinFD and BBA [14], showed the largest reductions in the control group, reflecting rapid cell migration and scratch wound closure. In contrast, PDT10 µM and PBM5J+PDT10 µM groups maintained higher values at 24 h, consistent with delayed closure. Shape-related features (Ecc, Rf, ModR, SF1) decreased markedly in the control group, suggesting isotropic and uniform gap filling, while PDT-treated groups displayed only minor reductions, indicating irregular and incomplete migration.

To allow direct comparison of migration dynamics across treatments and features, percentage changes from 0 h to 24 h were calculated using Equation (6). In brief, for each treatment group and each quantitative feature, the mean value at 24 h was compared to the mean value at 0 h, and the relative change was expressed as a percentage. This approach, based on the formula originally applied for wound area closure, was extended here to all extracted features, enabling normalization of differences in absolute scale between parameters [13]. The resulting values, summarised in Figure 3, show that the control group exhibited the largest relative changes in most parameters, while PDT-based treatments produced minimal changes, indicating substantial inhibition of migration. Laser 50 J/cm^2 and PBM 5 J/cm^2 groups displayed intermediate effects.

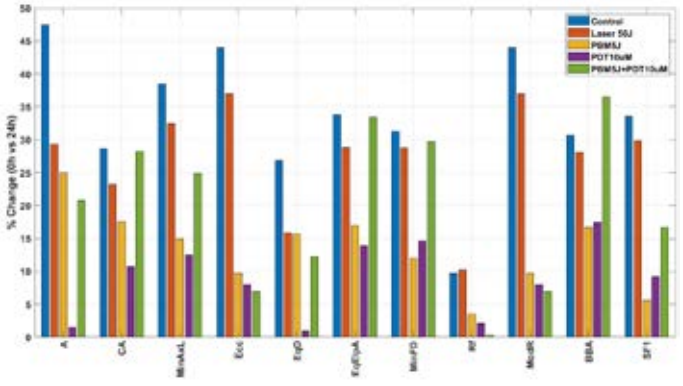


Fig. 3: Percentage change of measured features from 0 h to 24 h across the five experimental groups (Control, Laser 50 J/cm^2, PBM 5 J, PDT 10 µM, and PBM + PDT). Changes in size- and shape-related metrics provide a quantitative assessment of cell migration dynamics and treatment effects.

Visual inspection of segmented scratch areas (Figure 4) supports the quantitative results. Control cells showed almost complete closure at 24 h, while Laser 50J and PBM5J groups had partial closure. In contrast, PDT10 µM and PBM5J+PDT10 µM retained wide gaps, confirming suppressed migration under PDT-based treatments.

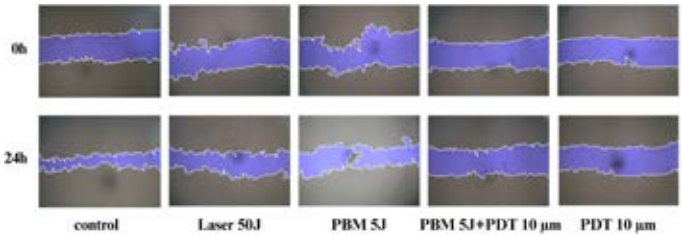


Fig. 4: Segmentation of cancer cell migration in five experimental groups at 0 h (top) and 24 h (bottom). Blue regions represent the segmented scratch wound areas.

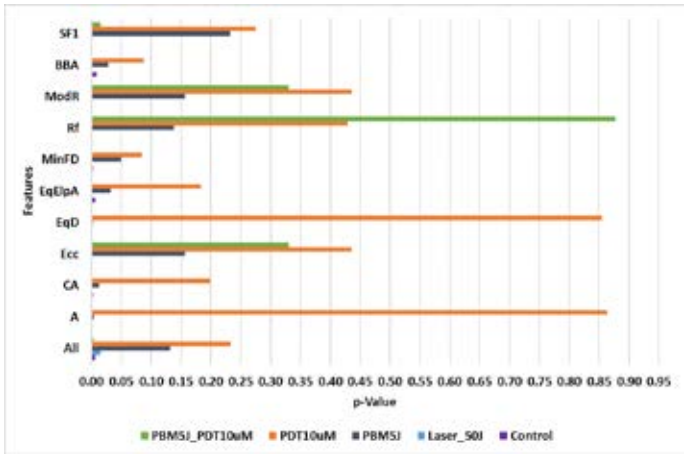


Fig. 5: ANOVA test results showing the p-values for different extracted features. "All" represents the p-values obtained by considering all concatenated size- and shape-related features, while the other labels correspond to individual feature sets. Significant differences ($p \leq 0.05$) indicate variation in migration dynamics among the experimental groups.

Quantitative features from scratch assays at 0 h and 24 h were analysed using one-way ANOVA ($\alpha = 0.05$). As shown in Figure 5, most features differed significantly between groups ($p \leq 0.05$), with the control group showing the largest changes, while PDT-treated groups maintained higher values, confirming significant inhibition of cell migration.

IV. CONCLUSION

This study applied a multi-step image processing pipeline to quantify migration in PC3 cancer cells under different treatments. PDT 10 µM, alone or combined with PBM 5 J, strongly inhibited migration, while Laser 50 J/cm² and PBM 5 J alone had moderate effects. The pipeline provides a reproducible, quantitative assessment of migration and lays the groundwork for future AI-based studies, where feature-based classification and machine learning models can enhance automated analysis of cell migration.

ACKNOWLEDGMENT

This study was conducted using infrastructure provided by the Department of Biomedical Engineering, Faculty of Science, Izmir Katip Celebi University. No ethical approval was required.

REFERENCES

[1] "A Deep Learning Pipeline for the Segmentation of In Vitro Wound Healing Microscopy Images following Laser Therapy," *2022 Medical Technologies Congress (TIPTEKNO)*, Antalya, Turkey, Nov. 2022, doi: 10.1109/TIPTEKNO56568.2022.9960169.

[2] Y. S. Erdem, Ö. Yalçın Özuysal, D. Pesen Okvur, B. Töreyin, et al., "An Image Segmentation Method for Wound Healing Assay Images," *Natural and Applied Sciences Journal*, vol. 4, no. 1, pp. 30–37, 2021, doi: 10.38061/idunas.853356.

[3] Z. Zhang et al., "Morphology-based prediction of cancer cell migration using an artificial neural network and a random decision forest," *Integr. Biol. (Camb)*, vol. 10, no. 12, pp. 758–767, Dec. 2018, doi: 10.1039/c8ib00106e.

[4] L. C. de Melo Gomes et al., "Photodynamic therapy reduces cell viability, migration and triggers necroptosis in prostate tumor cells," *Photochem. Photobiol. Sci.*, vol. 22, no. 6, pp. 1341–1356, Jun. 2023, doi: 10.1007/s43630-023-00382-9.

[5] A. Grada et al., "Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay," *J. Invest. Dermatol.*, vol. 137, no. 2, pp. e11–e16, Feb. 2017, doi: 10.1016/j.jid.2016.11.020.

[6] J. Oldenburg et al., "Methodology for comprehensive cell-level analysis of wound healing experiments using deep learning in MATLAB," *BMC Mol. Cell Biol.*, vol. 22, no. 1, p. 32, Jun. 2021, doi: 10.1186/s12860-021-00369-3.

[7] D. Mokoena, B. P. George, and H. Abrahamse, "Cannabidiol Combination Enhances Photodynamic Therapy Effects on MCF-7 Breast Cancer Cells," *Cells*, vol. 13, no. 2, p. 187, Jan. 2024, doi: 10.3390/cells13020187.

[8] L. Huang et al., "MPPa-PDT suppresses breast tumor migration/invasion by inhibiting Akt-NF-B-dependent MMP-9 expression via ROS," *BMC Cancer*, vol. 19, p. 1159, Nov. 2019, doi: 10.1186/s12885-019-6374-x.

[9] V. Vithanage, C. d. J, M. d. p. De. C, and S. Rajendram, "Photodynamic Therapy: An Overview and Insights into a Prospective Mainstream Anticancer Therapy," *JOTCSA*, vol. 9, no. 3, pp. 821–848, 2022.

[10] H. Umar et al., "Prediction of cell migration potential on human breast cancer cells treated with Albizia lebbeck ethanolic extract using extreme machine learning," *Sci. Rep.*, vol. 13, p. 22242, Dec. 2023, doi: 10.1038/s41598-023-49363-z.

[11] J. E. Jonkman et al., "An introduction to the wound healing assay using live-cell microscopy," *Cell Adh. Migr.*, vol. 8, no. 5, pp. 440–451, 2014, doi: 10.4161/cam.36224.

[12] E. Jalilian, M. Linortner, and A. Uhl, "Impact of Image Compression on In Vitro Cell Migration Analysis," *Computers*, vol. 12, no. 5, p. 98, May 2023, doi: 10.3390/computers12050098.

[13] M. Glaß et al., "Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines," Zentrum für Angewandte Medizinische und Humanbiologische Forschung (ZAMED), Martin Luther University, Germany.

[14] Petrović N, Moyà-Alcover G, Jaume-I-Capó A, González-Hidalgo M. Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images. Comput Biol Med. 2020 Nov;126:104027. doi: 10.1016/j.compbiomed.2020.104027. Epub 2020 Oct 10. PMID: 33075715.

[15] Sirek B, Topaloğlu N. Red wavelength-induced photobiomodulation enhances indocyanine green-based anticancer photodynamic therapy. Med Oncol. 2024 Nov 19;42(1):8. doi: 10.1007/s12032-024-02558-4. PMID: 39560842.

Antibacterial Activity of Green Synthesized Silver Nanoparticles from Aloe Vera Plant

Imene Belhout, Canan Bağcı, Hakan Solmaz
Department of Biomedical Engineering
Bahçeşehir University
Istanbul, Türkiye
hakan.solmaz@bau.edu.tr

***Abstract*— This study offers an environmentally friendly and cost-effective method for synthesizing silver nanoparticles (AgNPs) using *Aloe vera* leaf extract, without the use of harmful chemicals. The phytochemicals, such as polysaccharides and flavonoids, acted as reducing and stabilizing agents to convert silver ions (Ag^{+}) to metallic silver (Ag^{0}). Best synthesis conditions were found by mixing 90 mL of 1 mM $AgNO_3$ with 10 mL of 10% *Aloe vera* extract at pH 8, followed by a 48-hour incubation. The mixture color changed from light yellow to dark brown, indicating AgNPs formation, confirmed by UV-Vis spectroscopy with a peak at 411 nm. Characterization via Dynamic Light Scattering (DLS), Zeta potential, and Fourier Transform-Infrared Spectroscopy (FTIR) showed that AgNPs had an average size of 147.6 nm, a polydispersity index of 0.387, and a zeta potential of -40.3mV, indicating high stability. Antibacterial efficacy was evaluated using the disc diffusion assay against *Staphylococcus aureus (S. aureus)* and *Escherichia coli (E. coli)* at 2, 5, and 10 µg/µL concentrations. The test showed dose-dependent inhibitory zone diameters of 7.4-10.2 mm against *S. aureus* and 8-9.1 mm against *E. coli*. These results highlight the potential of green synthesized AgNPs in antibacterial activity against resistant tested strains. Future studies are recommended to evaluate long-term stability and biocompatibility.**

Keywords—green synthesis; silver nanoparticles; Aloe vera; antibacterial activity.

I. INTRODUCTION

Antibiotic resistance is considered one of the major public health challenges. As bacteria are becoming increasingly resistant to traditional antibiotics, the World Health Organization (WHO) listed multiple priority strains in 2017 for which new, safe, and effective therapeutic agents are urgently needed. *Staphylococcus aureus (S. aureus)* is one of the high-priority antibiotic resistant bacteria [1]. *Escherichia coli (E. coli)* is not in the list; however, recent studies reported its growing resistance [2].

Nanotechnology is an emerging approach that has proven its efficacy in many biomedical applications, including antibacterial resistance. Nanomaterials are one of the products of nanotechnology research, where nano-sized particles ranging between 10 to 100 nm are used increasingly due to their characteristic physicochemical features [3]. Nanoparticles are known for their large surface area, ability to interact with bacteria membranes, and subsequently cytotoxic and antibacterial potential [4]. Metal nanoparticles are a promising subtype of nanomaterials where metals such as silver, gold, and copper are used to synthesize nanoparticles with characteristic optical, catalytic, and antimicrobial properties [5].

Silver nanoparticles (AgNPs) are a type of metal NPs that have been extensively employed as an antibiotic agent in many health conditions such oral infections and wound healing. While AgNPs are well-documented to inhibit bacterial growth, the underlying mechanisms are not yet fully understood. Some studies report that AgNPs can penetrate bacterial wall and cause DNA damage and increase reactive oxygen species (ROS) inside the bacteria, leading to cell death [6-7].

The physicochemical properties, especially the size and shape of AgNPs plays a crucial role in dictating their bactericidal activity. These features mainly rely on the synthesis approach. There are two synthesis methods, chemical synthesis and biosynthesis. The latter involves producing NPs from various living organisms such as bacteria, fungi and plant extracts [8].

Green synthesis is an eco-friendly, sustainable, cost-effective, and non-toxic approach of producing nanoparticles. It refers to the use of plant extracts as reducing and stabilizing agents for converting silver ions (Ag^{+}) to metallic silver (Ag^{0}). These agents are the phytochemical components of plants such as flavonoids, phenolics, alkaloids, and proteins. Green synthesis is reported to produce NPs with higher surface area to volume ratio compared to other methods, leading to better bioactivity [8-9].

Aloe vera plant (AV) belongs to the *Liliacea* family. It is a xerophyte, arborescent, and succulent plant known for its medicinal properties, traditionally used for wound healing and burns. It contains various bioactive compounds, including polysaccharides, glycoproteins, and anthraquinones, which contribute to its therapeutic effects. These compounds possess antimicrobial properties, making *Aloe vera* an ideal candidate for the green synthesis of silver nanoparticles in the context of antibiotic resistance [9-10].

The present study aims to develop a green and sustainable method for synthesizing silver nanoparticles (AgNPs) using *Aloe vera* plant extract. This biosynthesis approach harnesses the natural reducing and stabilizing agents present in *Aloe vera* to produce environmentally friendly nanoparticles. The synthesized AgNPs are then thoroughly characterized to assess their physicochemical properties. Additionally, the paper aims to evaluate the antibacterial efficacy of the AgNPs against both

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

gram-positive (*S. aureus*) and gram-negative (*E. coli*) bacterial strains using the disc diffusion method. This research explores the potential application of green-synthesized AgNPs as an effective antibacterial agent, particularly in addressing the growing challenge of antibiotic resistance.

II. METHODS

A. *Synthesis and Characterization of AgNPs from Aloe vera plant extract.*

AgNPs were synthesized based on previously published protocol [11], with some modifications. First, 10 g of washed *Aloe vera* (AV) leaves were chopped into small pieces in a mortar and pestle, then transferred into a glass beaker containing 100 mL of distilled water. The solution was heated for 30 min at 60°C with continuous stirring until it turned light yellow. After cooling, the extract was purified using filter paper. The resulting plant extract solution was stored at 4°C for future use. Silver nitrates ($AgNO_3$) solution was prepared at 1mM concentration by mixing 0.017g of $AgNO_3$ powder with 100 mL of distilled water. For AgNPs synthesis, 90 mL of $AgNO_3$ solution was mixed with 10 mL of 10% AV extract and few drops of 1M NaOH to reach pH 8. The resulting 100 mL of NPs were left in the magnetic stirrer at 300 rpm for 1 hour and then incubated at 40°C for 48h in dark. After incubation period, the AgNPs were centrifuged at 4100 rpm for 30 min. The supernatant was discarded and residual pellet was resuspended with distilled water and put in a petri dish and dried in the oven at 40°C. The collected AgNPs powder was stored at 4°C for further applications. The experimental framework for AgNPs synthesis is schematically presented in Fig. 1.

The biosynthesized silver nanoparticles from *Aloe vera* plant were characterized using the following methods: UV–Vis spectroscopy to understand the optical properties of AgNPs within the wavelength range of 200-900 nm, Dynamic Light Scattering (DLS) to measure the hydrodynamic diameter of the particles, Zeta potential to assess the surface charge and stability of nanoparticles, and Fourier transmission-infrared spectroscopy (FTIR) to analyze the functional groups resulting from the reduction and stabilization of AgNPs. DLS and Zeta measurements were conducted three times. Data are presented as mean ± standard error.

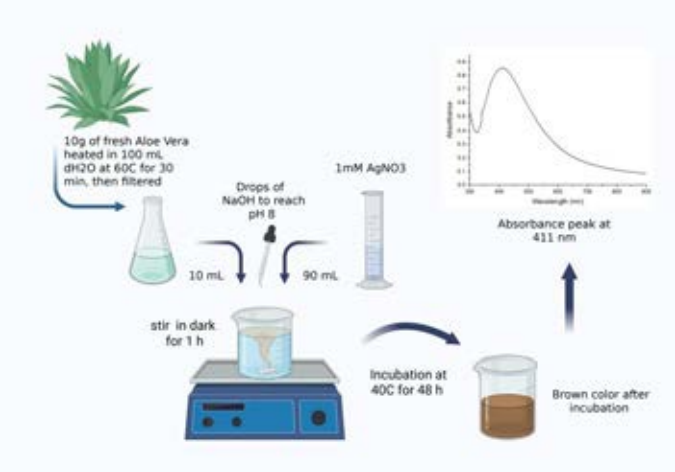


Fig. 1. The experimental framework of AgNPs synthesis

B. *Antibacterial Activity of silver nanoparticles*

The antibacterial activity of green synthesized AgNPs was assessed using agar disc diffusion assay [12]. One gram-negative (*E. coli* ATCC-8739) and one gram-positive (*S. aureus* ATCC-6538) bacterial strains were used.

38 grams of Mueller-Hinton agar were weighed, dissolved in 1 liter of distilled water, sterilized, then cooled and poured into petri dishes. Three AgNPs concentrations were prepared (2, 5, 10 µg/ µL), and 6 mm discs were added into the NPs dilutions (discs immersed together into a NPs solution of 1 ml and left to absorb it) and allowed at 4°C for absorption. Bacteria were incubated overnight in the broth liquid medium. After incubation, the young bacteria were diluted and 100 ul of the diluted sample was taken and spread throughout the petri dish. The AgNPs-absorbed discs and Tetracycline discs (30 µg) for a positive control were placed on top of the bacteria and incubated overnight at 37°C. The inhibition zones were determined by measuring the diameters of the zones around the discs. Each AgNPs concentration was tested three times. Data are expressed as mean ± standard error and a paired two-tailed Student s t- test was employs for P-value. Each concentration was compared to the antibiotic control group (Tetracycline).

III. RESULTS AND DISCUSSION

Silver nanoparticles were successfully synthesized using *Aloe vera* (AV) leaves extract. The solution color changed from light yellow to dark brown (Fig. 2a, 2b), indicating the conversion of silver ions (Ag +) to silver metal. This change is justified by the presence of chemical reducing agents in the AV extract [13].

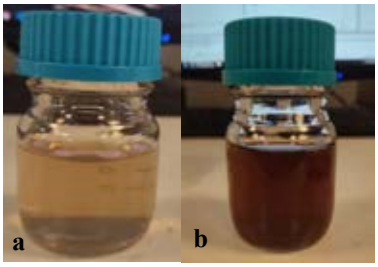


Fig. 2. Color change of AgNPs. a. Initial mixture. b. after 48h incubation.

In addition to the color testing, UV–Vis spectroscopy was used to confirm the formation of silver nanoparticles. The absorbance peak between 400 and 450 nm is an indicator of AgNPs formation [14]. In our context, a peak was observed at 411 nm (Fig. 3).

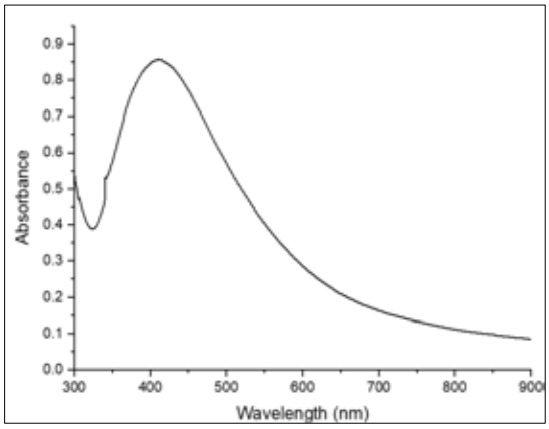


Fig. 3. Absorption spectra of AgNPs.

The FTIR results highlight the chemical composition of the dried AgNPs synthesized from *Aloe vera* (Fig. 4). From the observed peaks, the bio-compounds in the *Aloe vera* were essential for good nanoparticles as reducing, capping, and stabilizing agents. Table I gives more details about the present functional groups. Based on published literature, the wavenumbers were associated to their corresponding active groups [11,13].

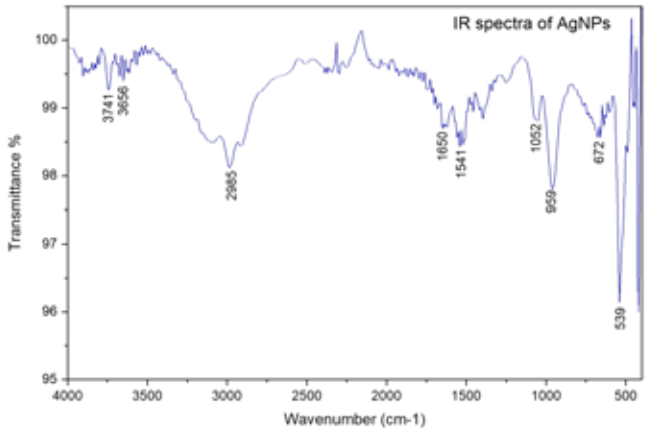


Fig. 4. FTIR spectra of AgNPs synthesized from *Aloe vera*.

Table I. Active compounds involved in AgNPs stabilization [11,13].

Wavenumber (cm^{-1})	Functional group	Comment
3656 and 3741	O–H stretching	Hydroxyl groups in polysaccharides or phenols
2985	C–H stretching	Aliphatic groups
1650	C=O stretching	carbonyl group aldehydes and aliphatic amides
1541	C = O and C–N stretching	nitro compounds
1052	C–O stretching	alcohols
959	C = C bending	Polysaccharides or aromatic compounds
672	C–H bending	Aromatic compounds
539	Ag–O or Ag–N bond stretching	Metal–ligand bond

Dynamic Light Scattering (DLS) and zeta potential measurements were employed to assess the size distribution and surface charge stability of AgNPs (Table II). The average diameter was found to be 147.6 nm, with a polydispersity index (PDI) of 0.387. This indicates the moderate size distribution homogeneity and the fairly uniform nanoparticles suitable for biological applications. Previous studies reported that AgNPs with moderate size and low-to-moderate PDI demonstrate acceptable antimicrobial/antibacterial activities [15,16].

Zeta potential showed a surface charge of –40.3mV, suggesting strong electrostatic repulsion between particles. This high negative value indicates good colloidal stability and low aggregation potential, which is essential for prolonged suspension and effective interaction with bacterial membranes [15].

Table II. Size, Polydispersity Index, and Zeta Potential of AgNPs (n=3)

Measurement	Hydrodynamic diameter (DLS) (Mean ± SD)	Polydispersity index (PDI)	Surface charge (Zeta potential) (Mean ± SD)
AgNPs	147.6 ± 3.08 nm	0.387	–40.3 ± 0.66 mV

The antibacterial activity of green-synthesized silver nanoparticles from *Aloe vera* plant was evaluated using the disc diffusion assay against *S. aureus* and *E. coli* at three concentrations (2 µg, 5 µg, and 10 µg). Results from images (Fig. 5) were quantified and illustrated in the bar graph (Fig. 6).

The inhibition zone diameters clearly indicate that AgNPs exhibited dose-dependent antibacterial effects against both tested strains. At 2 µg, AgNPs showed moderate inhibition zones of 7.4 ± 0.67 mm for *S. aureus* and 8 ± 0.25 mm for *E. coli*. When the concentration increased to 10 µg, the diameters extended to 10.2 ± 0.76 mm and 9.1 ± 0.17 mm, respectively.

This observed antibacterial effect of AgNPs is explained by their ability to attach to bacterial membrane proteins, triggering permeability and entrance of nanoparticles inside the cell. Subsequently, enzyme deactivation, protein denaturation, and DNA damage occur, leading to bacteria death. Another mechanism is the release of silver ions (Ag^{+}) from the nanoparticles. These ions can disrupt the uptake and release of essential physiological ions required for cellular function [13,17]. In Fig. 7, the possible antibacterial effect of AgNPs is schematically illustrated.

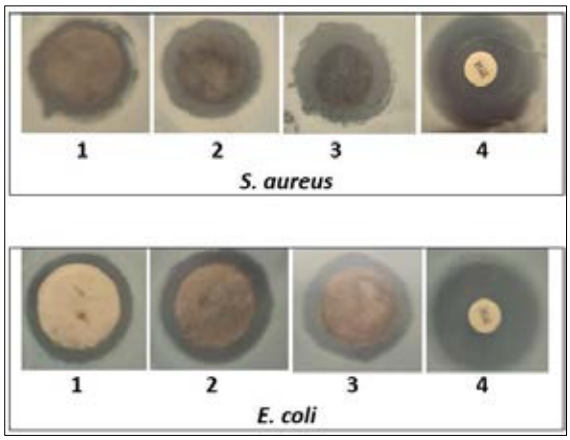


Fig. 5. Antibacterial activities of different concentrations of AgNPs. 1 (2 µg), 2 (5 µg), 3 (10 µg), 4 (Tetracycline 30 µg).

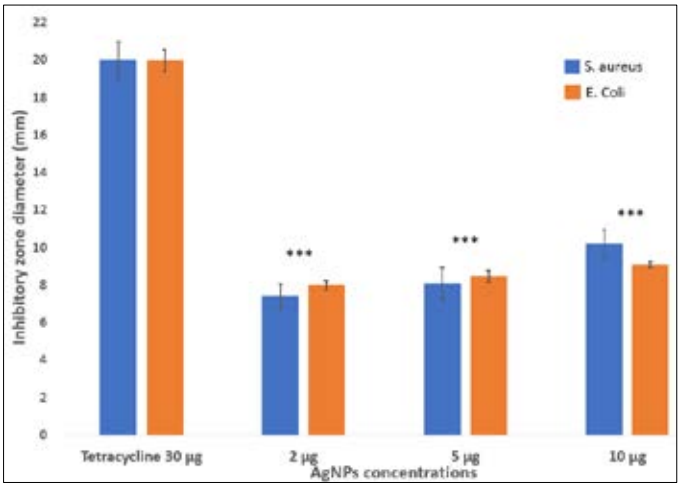


Fig. 6. Inhibitory zone diameter of AgNPs synthesized from *Aloe vera* against *S. aureus* and *E. coli.* (n=3) (***p < 0.001).

Table III. Inhibitory zone diameter of AgNPs (n=3) (***p < 0.001)

Micro-organism	Inhibition zone diameter in millimeters (Mean ± SD)			
	Positive control	AgNPs (µg/µL)		
	Tetracycline 30 µg	2	5	10
S. aureus	20 ± 1	7.4 ± 0.67***	8.1 ± 0.85***	10.2 ± 0.76***
E. coli	20 ± 0.58	8 ± 0.25***	8.5 ± 0.32***	9.1 ± 0.17***

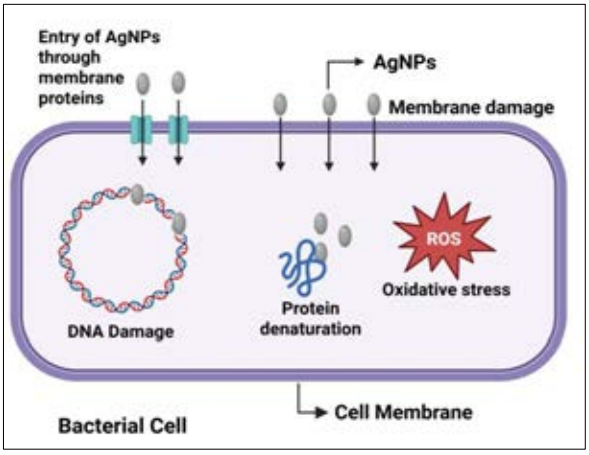


Fig. 7. Schematic illustration of possible antibacterial mechanisms of AgNPs, created by the author based on concepts described in the literature [13, 17]. Created by BioRender.

IV. CONCLUSION

In the present study, a simple, eco-friendly, and inexpensive green synthesis method was proven to be effective in producing biologically active silver nanoparticles using *Aloe vera* plant extract. The successful formation of AgNPs was confirmed by the brown color and the characteristic absorbance peak at 411 nm in UV-Vis spectroscopy. DLS, Zeta potential, and FTIR analyses verified the nanoparticles' size (147.6 nm), stability (-40.3 mV), and functional groups composition, respectively. The biosynthesized AgNPs exhibited dose-dependent antibacterial activity against both gram positive (*S. aureus*) and gram negative (*E. coli*) bacterial strains. This study highlights the biomedical potential of AgNPs against antibiotic-resistant microorganisms. Further studies are needed to check nanoparticles long-term stability and biocompatibility.

REFERENCES

[1] Z. Breijyeh, B. Jubeh, and R. Karaman, "Resistance of gram-negative bacteria to current antibacterial agents and approaches to resolve it," Molecules, vol. 25, no. 6, p. 1340, Mar. 2020.

[2] M. S. Aljohni, M. Harun-Ur-Rashid, and S. Selim, "Emerging threats: Antimicrobial resistance in extended-spectrum beta-lactamase and carbapenem-resistant Escherichia coli," Microb. Pathog., vol. 107275, Jan. 2025.

[3] G. Angeles Flores, G. Cusumano, R. Venanzoni, and P. Angelini, "Advancements in antibacterial therapy: Feature papers," Microorganisms, vol. 13, no. 3, p. 557, Mar. 2025.

[4] T. Gul, L. Tabassam, A. Basharat, A. Amir, Z. Baqar, and M. J. Khan, "In vitro and in vivo efficacy of zinc oxide green nanoparticles against multidrug-resistant Salmonella Typhi," Braz. J. Microbiol., vol. 55, no. 4, pp. 3839–3848, Dec. 2024.

[5] M. Hancharova, K. Halicka-Stępień, A. Dupla, A. Lesiak, J. Sołoducho, and J. Cabaj, "Antimicrobial activity of metal-based nanoparticles: A mini-review," Biometals, vol. 37, no. 4, pp. 773–801, Jan. 2024.

[6] N. Tripathi and M. K. Goshisht, "Recent advances and mechanistic insights into antibacterial activity, antibiofilm activity, and cytotoxicity of silver nanoparticles," ACS Appl. Bio Mater., vol. 5, no. 4, pp. 1391–1463, Mar. 2022.

[7] S. Zhou, Y. Qin, A. Lei, H. Liu, Y. Sun, J. Zhang, and Y. Chen, "The role of green synthesis metal and metal oxide nanoparticles in oral cancer therapy: a review," J. Drug Target., vol. 33, no. 6, pp. 853–876, Feb. 2025.

[8] S. Dong, X. Li, Q. Pan, K. Wang, N. Liu, W. Yutao, and Y. Zhang, "Nanotechnology-based approaches for antibacterial therapy," Eur. J. Med. Chem., vol. 116798, Aug. 2024.

[9] K. Khaldoune, N. Fdil, and M. Ait Ali, "Exploring Aloe vera: A comprehensive review on extraction, chemical composition, biological effects, and its utilization in the synthesis of metallic nanoparticles," Biocatal. Agric. Biotechnol., vol. 57, p. 103052, Apr. 2024.

[10] S. S. Martin, L. Zhang, Y. Wang, Z. Zhou, R. Ma, H. Mo, Z. Tan, H. Shi, and T. Wang, "Advances in preparation techniques and antimicrobial properties of Aloe vera extract," Fitoterapia, vol. 106696, Jun. 2025.

[11] J. M. Ashraf, M. A. Ansari, H. M. Khan, M. A. Alzohairy, and I. Choi, "Green synthesis of silver nanoparticles and characterization of their inhibitory effects on AGEs formation using biophysical techniques," Sci. Rep., vol. 6, no. 20414, Feb. 2016.

[12] M. R. Zaidan, A. Noor Rain, A. R. Badrul, A. Adlin, A. Norazah, and I. Zakiah, "In vitro screening of five local medicinal plants for antibacterial activity using disc diffusion method," Trop. Biomed., vol. 22, no. 2, pp. 165–170, Dec. 2005.

[13] T. Liknaw, Y. Belay, R. Ramesh, and R. P. D. M., "Aloe vera leaf extract as a sustainable route for silver nanoparticle synthesis with enhanced antimicrobial activity," Sci. Rep., vol. 15, no. 1, p. 22481, Jul. 2025.

[14] W. K. A. Wan Mat Khalir, K. Shameli, S. D. Jazayeri, N. A. Othman, N. W. C. Jusoh, and N. M. Hassan, "Biosynthesized silver nanoparticles by aqueous stem extract of Entada spiralis and screening of their biomedical activity," Front. Chem., vol. 8, p. 620, Aug. 2020.

[15] J. A. Badmus, S. A. Oyemomi, O. T. Adedosu, T. A. Yekeen, M. A. Azeez, E. A. Adebayo, A. Lateef, U. M. Badeggi, S. Botha, A. A. Hussein, and J. L. Marnewick, "Photo-assisted bio-fabrication of silver nanoparticles using Annona muricata leaf extract: exploring the antioxidant, anti-diabetic, antimicrobial, and cytotoxic activities," Heliyon, vol. 6, e05413, 2020.

[16] W. H. S. Alwis, V. Murthy, H. Wang, R. Khandanlou, and P. K. Mandal, "Green synthesis of Terminalia ferdinandiana Exell-mediated silver nanoparticles and evaluation of antibacterial performance," Biomolecules, vol. 14, no. 12, p. 1516, Nov. 2024.

[17] T. C. Dakal, A. Kumar, R. S. Majumdar, and V. Yadav, "Mechanistic basis of antimicrobial actions of silver nanoparticles," Front. Microbiol., vol. 7, p. 1831, Nov. 2016.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Assessment of Oxidative Effects of Cold Atmospheric Plasma via Image Processing Techniques

Elcin Erkal
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Türkiye
elcinerkal@hotmail.com

Mehmet Akif Ozdemir
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Türkiye
makif.ozdemir@ikcu.edu.tr

Gizem Dilara Ozdemir
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Türkiye
gizemdilara.ekimci@ikcu.edu.tr

Utku Kursat Ercan
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Türkiye
utkuk.ercan@ikcu.edu.tr

Onan Guren
Dept. of Biomedical Eng.
Izmir Katip Celebi University
Izmir, Türkiye
onan.guren@ikcu.edu.tr

***Abstract*—Cold atmospheric plasma (CAP) produces reactive oxygen and nitrogen species (ROS/RNS) that induce oxidative modifications on microbial surfaces. This study quantitatively assessed CAP-induced oxidative effects on KI-starch agar using image processing techniques. Plasma treatments were applied with varying discharge gaps, duty cycles, frequencies, and treatment times, resulting in 648 images. From these images, 106 features related to color, texture, and frequency were extracted and analyzed using Pearson and Spearman correlations, multiway ANOVA, and multicollinearity checks. Plasma treatment time emerged as the primary factor influencing visual features, while interactions among all plasma parameters were statistically significant** ($p < 0.001$)**. A final set of 30 non-redundant features captured the key surface modifications, offering a reliable framework for predictive modeling. These results provide a quantitative understanding of plasma-induced oxidative effects and highlight the potential of image-based markers for optimizing plasma-based antimicrobial applications.**

***Index Terms*—Cold Atmospheric Plasma, Image Processing, Antimicrobial Activity, Plasma Medicine, Statistical Analysis**

I. Introduction

Cold Atmospheric Plasma (CAP) is a type of plasma generated by ionizing gases at low temperatures [1]. Its ability to operate at low temperatures enables its application without causing damage to living cells, tissues, and microorganisms [2]. The generation of active components such as reactive oxygen species (ROS), reactive nitrogen species (RNS), free radicals, ultraviolet light, and electric fields during plasma ionization significantly influences biological and chemical processes [3]. These components influence biological and chemical processes, leading to structural changes in microorganisms, the development of antimicrobial effects, regulation of cell proliferation, and modifications of certain biological structures. Particularly, reactive species such as (Reactive Oxygen Nitrogen Species (RONS) can cause oxidative and nitrosative stress in cells, leading to cell membrane damage and disruption of metabolic functions [4].

One of the most significant benefits of CAP in biomedical applications is its strong antimicrobial properties. Studies have shown that CAP can effectively eliminate bacteria, viruses, and fungi [5], [6]. Furthermore, studies have shown that direct CAP applications subject many microorganisms, including biofilms, to powerful electrical fields that can damage their physicochemical structure [7]. Traditional sterilization and disinfection methods often become ineffective as microorganisms develop resistance over time. Moreover, many chemical disinfectants can lead to toxic side effects. CAP overcomes these issues with its broad-spectrum antimicrobial activity, providing sterilization without leaving residues on the treated surface [8].

Despite the widespread application of CAP, there is still a need to accurately determine the plasma parameters to optimize its antimicrobial efficacy. Therefore, this study aims to analyze the oxidative effects of CAP under different plasma conditions using image processing and statistical analysis methods to determine optimal treatment parameters. This study quantitatively reveals oxidative effects through image processing, thereby enabling future bacterial assays to establish a direct link between oxidative visualization and antimicrobial efficacy. To effectively assess the oxidative effects of CAP, a reliable visualization method is required. Among various techniques, the KI-starch agar method stands out for its ability to provide a rapid and direct assessment of plasma-generated

This work was supported by the Izmir Katip Celebi University Scientific Research Projects Coordination Unit [grant number: 2025-TYL-FEBE-0001].

979-8-3315-5565-8/25/31.00 © 2025 IEEE

reactive species [9], [10]. In this technique, reactive species produced by CAP oxidize the KI-starch complex, resulting in a visible color change in the agar medium. The intensity and spread of the color change may be directly related to parameters such as plasma treatment time, frequency, duty cycle value, and discharge gap. In this study, statistical analysis was performed to elucidate the correlation between color changes and plasma treatment parameters.

II. Materials and Methods

A. Data Collection

1) Preparation of Agar Medium and KI-Starch: For the preparation of the agar medium, 3 grams of agar powder were mixed with 200 mL of deionized water (DIW). The mixture was then heated on a hot plate at 95-100 ° C and homogenized using a magnetic stirrer. The solution was cooled to 50-55°C. For the preparation of the KI-starch solution, 1.2 grams of starch (Sigma Aldrich) were gradually added to 200 mL of preheated DIW and mixed using a magnetic stirrer. Once homogenized, the solution was allowed to reach room temperature, after which 2.4 grams of Potassium Iodide (KI, TEKKIM) were added and the solution was homogenized again. The prepared KI-starch solution was mixed with the agar solution, which had been cooled to approximately 50°C, and the resulting mixture was poured into sterile Petri dishes at 15 mL per dish. The agar was then allowed to solidify at room temperature for 30-60 minutes.

2) Plasma Device and Experimental Setup: A custom-made plasma device was used to apply plasma treatments to the KI-Starch agars. The plasma was generated using a specially manufactured copper electrode in the Dielectric Barrier Discharge (DBD) method to ensure controlled plasma generation and uniform application. Four fundamental plasma parameters were examined: discharge gap (0.5,1,1.5 mm), duty cycle(40%, 60%, 80%, 100%), frequency (1, 2, 3 kHz), and treatment time(5-10-15-30-45-60 sec). Each parameter plays a determining role in the antimicrobial efficacy by influencing the intensity and distribution of reactive species generated by plasma.

3) Imaging Setup: Plasma treatment was applied to the KI-starch agar medium using the predefined experimental parameters. The ROS/RNS generated by the plasma application caused visible color changes on the agar surface. Images were captured using a fixed camera angle under uniform lighting conditions. Example of images shown in Fig. 1. All experiments were carried out in triplicate to ensure the reliability and reproducibility of the results. 648 agar image were collected from 216 different experimental combinations, each with 3 repetitions.

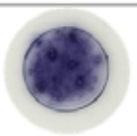
Images	Plasma Parameters
	Discharge gap: 1.5 mm Duty Cycle: 40% Frequency: 1000 Hz. Plasma Treatment Time: 5 sec
	Discharge gap: 1 mm Duty Cycle: 100% Frequency: 1000 Hz. Plasma Treatment Time: 5 sec
	Discharge gap: 1 mm Duty Cycle: 100% Frequency: 3000 Hz. Plasma Treatment Time: 15 sec
	Discharge gap: 1.5 mm Duty Cycle: 100% Frequency: 2000 Hz. Plasma Treatment Time: 60 sec

Fig. 1. Examples of obtained images.

B. Image Processing

1) Preprocessing: Raw images obtained from experimental studies were subjected to various preprocessing steps to make them suitable for analysis. First, to reduce the influence of irrelevant areas such as labels and background, and to standardize the image dimensions, all images were resized to 512×512 pixels. To focus on the central region where the plasma effect was most intense, a circular cropping was applied. In this process, a circular mask was created at the center of the image. To reduce color imbalances and lighting variations, the White Patch Retinex (WPR) method was employed. WPR is a color balancing and illumination correction technique that rescales the color channels based on the brightest (white) regions of the image [11]. This minimizes color distortions caused by lighting sources, enabling more consistent and accurate color analysis. To enhance image contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. CLAHE divides the image into small regions and enhances local contrast in each region, making plasma-induced color changes more distinguishable. It also prevents noise amplification caused by excessive contrast enhancement [12].

2) Feature Extraction: To numerically evaluate the effects of plasma treatment on the agar surface, multidimensional features based on color, texture, and frequency were extracted. To ensure that the analysis focused solely on plasma-affected regions, Region of Interest (ROI) masks were created for each image using alpha channel–based binary masking to exclude irrelevant areas. Regarding color features, statistical measures including mean, standard deviation, skewness, kurtosis, and entropy were calculated for the red, green, and blue channels of the RGB color space. To capture color variations more

accurately, RGB images were converted to HSV [13] and LAB color spaces. In the HSV space, mean, standard deviation, skewness, kurtosis, and entropy values were extracted from the hue, saturation, and value channels. Similarly, in the LAB space, these statistical features were computed for the lightness (L), green-red (A), and blue-yellow (B) channels. This comprehensive approach allowed a more detailed and reliable assessment of plasma-induced color changes. Additionally, a purple color ratio was calculated from the HSV space to quantify plasma-generated purple color variations. In order to quantify the relationships between different color channels, R/G and R/B ratios were calculated in RGB, S/V in HSV, and A/B in LAB. Intensity-based features were extracted from grayscale images, including fundamental statistics such as mean, standard deviation, skewness, kurtosis, and entropy, as well as histogram-based features including width, energy, entropy, and contrast. This enabled a detailed quantification of changes in intensity levels caused by plasma treatment. Texture features were derived using the Gray Level Co-occurrence Matrix (GLCM) method. GLCM models the spatial relationships between pixels to provide statistical texture parameters such as contrast, homogeneity, energy, correlation, and dissimilarity [14]. For frequency analysis, Fast Fourier Transform (FFT) was applied to examine structural regularities and periodicities induced by plasma on the surface. Mean and standard deviation values of the amplitude and phase spectra were calculated to numerically characterize frequency components.

3) Dataset Construction: All color, intensity, texture, and frequency features extracted from images were compiled into a dataset along with sample IDs. The target variable was defined as the purple area induced by plasma exposure ("Purple Area"), which was measured in mm² using ImageJ.

4) Data Normalization: Because the resulting features have different scales and distributions, they pose a risk of creating imbalances in the analyses. To address this issue and ensure comparability across features, all numerical data were normalized using the Z-score method. In this method, each feature was rescaled internally using its own mean and standard deviation. Identity information and output variables were excluded from this process.

C. Statistical Analysis

Statistical analyses were performed to investigate the relationships between plasma parameters and the numerical features extracted from KI-starch images. A feature-wise approach was applied, evaluating each feature individually against the plasma parameters (Discharge Gap, Duty Cycle, Frequency, and Plasma Treatment Time) to identify the most sensitive features while excluding irrelevant or noisy ones. Pearson and Spearman correlations were calculated to examine relationships between each parameter and feature, and those with significant correlations ($p < 0.05$) and strong associations ($|r| > 0.5, p < 0.01$) were selected for further analysis. Multi-way ANOVA was then used to assess both main and interaction effects, treating all parameters as categorical variables and applying a significance threshold of $p < 0.001$. Finally, highly correlated feature pairs ($|r| > 0.9$) were identified and one feature from each pair was removed, resulting in an independent, multicollinearity-free feature set suitable for further analysis.

III. Results and Discussion

A. Correlation Analysis

Pearson and Spearman correlation coefficients were calculated to evaluate the relationships between plasma parameters and 106 extracted image features. After filtering by statistical significance ($p < 0.05$), 318 significant correlations were identified with Pearson and 272 with Spearman.

Among the parameters, Plasma Treatment Time consistently showed the strongest relationships with visual features. Notably, high correlations were observed for Mean R ($r_{\text{Pearson}} \approx -0.57$, $r_{\text{Spearman}} \approx -0.72$), Entropy R ($r_{\text{Pearson}} \approx -0.58$, $r_{\text{Spearman}} \approx -0.72$), and Skewness R ($r_{\text{Pearson}} \approx 0.56$, $r_{\text{Spearman}} \approx 0.72$). These results suggest that longer treatment reduces brightness and entropy while increasing asymmetry in pixel distributions, indicating systematic surface modifications.

Other parameters (Discharge Gap, Duty Cycle, Frequency) exhibited weaker but still significant effects, particularly on features related to texture complexity and distribution (e.g., negative correlation between Discharge Gap and Entropy R). Overall, the correlation analysis highlights treatment time as the dominant experimental factor influencing surface morphology.

TABLE I
NUMBER OF SIGNIFICANT FEATURES AND HIGHEST PEARSON AND SPEARMAN CORRELATIONS PER PARAMETER

Parameter	Number of Significant Features	$r_{\max}$ (Pearson)	$r_{\max}$ (Spearman)
Discharge Gap	85	0.304	0.314
Duty Cycle	91	0.210	0.212
Frequency	51	0.136	0.104
Plasma Treatment Time	101	0.704	0.822

B. Multi-Way ANOVA

The multi-way ANOVA analysis provided a comprehensive understanding of how plasma parameters influenced the extracted features. Examination of the main effects confirmed that each parameter individually contributed to significant variations in color, texture, histogram, and frequency-based descriptors. Among these, Treatment Time consistently emerged as the most dominant factor, producing strong and systematic changes across nearly all feature categories, particularly in intensity- and histogram-related measures.

However, the analysis also revealed that the plasma process cannot be adequately explained by single-parameter effects alone. The four-way interaction among Discharge Gap, Duty Cycle, Frequency, and Treatment Time was found to be highly significant ($p < 0.001$) and often produced stronger influences than the individual parameters themselves. This result highlights that the observed modifications on the agar surface are not merely additive effects of independent variables, but rather

the outcome of a synergistic and nonlinear interplay among all four plasma parameters.

For instance, while Treatment Time increased the purple color area in general, its magnitude and distribution were strongly dependent on the simultaneous configuration of Discharge Gap, Duty Cycle, and Frequency. In other words, a long treatment time could either amplify or diminish the plasma impact depending on how the other three parameters were combined. Such findings demonstrate that the plasma–surface interaction is inherently multifactorial, where the collective balance of gap distance, energy delivery, oscillation frequency, and exposure duration determines the resulting feature patterns.

This strong four-way interaction emphasizes that optimizing plasma treatment requires a holistic consideration of all parameters simultaneously, rather than adjusting them in isolation.

C. Multicollinearity

To prepare the extracted image features for potential machine learning applications, multicollinearity was carefully assessed among the statistically significant features. Feature pairs exhibiting very high correlations ($|r| > 0.99$) were considered redundant, and only one representative from each pair was retained. This process resulted in a final set of 30 non-redundant features, encompassing a broad spectrum of visual characteristics:

Color-based metrics: Mean R, Kurtosis R, Skewness B, Std H, Kurtosis H, Entropy H, Std S, Skewness S, Kurtosis S, Entropy S, Kurtosis B_lab, Entropy B_lab
Histogram-derived features: Hist_Width R, Hist_Width G, Hist_Width B, Hist_Width V, Hist_Width L, Hist_Energy H, Hist_Energy S, Hist_Energy B_lab, Hist_Contrast H, Hist_Contrast S, Hist_Contrast A
Texture descriptors: GLCM_Contrast, GLCM_Dissimilarity
Frequency-domain measures: FFT_Amplitude_Mean
Color ratios: A/B, Gray_Hist_Width

The resulting correlation matrix confirmed that no feature pair exceeded the multicollinearity threshold, ensuring sufficient independence among features. This carefully selected and diverse set provides a robust foundation for subsequent modeling and analysis while capturing the key visual characteristics influenced by plasma treatment.

IV. Conclusion

This study demonstrated that plasma applications induce significant and systematic changes in agar surface characteristics, as evidenced by both colorimetric and texture-based image features. Correlation analyses identified strong relationships between plasma treatment parameters—especially plasma treatment time—and key visual descriptors, while multi-way ANOVA confirmed that both main effects and the four-way interaction of Discharge Gap, Duty Cycle, Frequency, and Treatment Time significantly influenced color, texture, and frequency-based features. These findings highlight the multidimensional impact of plasma on surface properties, providing quantitative insight into plasma-induced oxidative effects.

Although direct microbial inactivation was not measured, the results offer a robust foundation for future studies. Applying the same plasma parameters in bacterial assays could establish a link between oxidative visualization and antimicrobial efficacy, supporting the optimization of plasma treatment conditions.

Moreover, the curated set of 30 non-redundant image features provides a comprehensive quantitative framework for data-driven approaches. Machine learning models can utilize these descriptors to predict treatment outcomes, optimize parameters, or guide quality control in plasma-based applications.

References

[1] A. Kojtari, U. Ercan, J. Smith, G. Friedman, R. Sensenig, S. Tyagi, S. Joshi, H. Ji, and A. Brooks, "Chemistry for antimicrobial properties of water treated with non-equilibrium plasma," *J. Nanomed. Biotherapeutic Discovery*, vol. 4, no. 1, p. 120, 2013.

[2] D. Braný, D. Dvorská, E. Halašová, and H. Škovierová, "Cold atmospheric plasma: A powerful tool for modern medicine," *International Journal of Molecular Sciences*, vol. 21, no. 8, 2020.

[3] H. Zhang, C. Zhang, and Q. Han, "Mechanisms of bacterial inhibition and tolerance around cold atmospheric plasma," *Applied Microbiology and Biotechnology*, vol. 107, pp. 1–16, 07 2023.

[4] S. Kalghatgi, C. M. Kelly, E. Cerchar, B. Torabi, O. Alekseev, A. Fridman, G. Friedman, and J. Azizkhan-Clifford, "Effects of non-thermal plasma on mammalian cells," *PLOS ONE*, vol. 6, no. 1, pp. 1–11, 01 2011.

[5] S.-y. Zhai, M. G. Kong, and Y.-m. Xia, "Cold atmospheric plasma ameliorates skin diseases involving reactive oxygen/nitrogen species-mediated functions," *Frontiers in Immunology*, vol. 13, 2022.

[6] Y. Rao, W. Shang, Y. Yang, R. Zhou, and X. Rao, "Fighting mixed-species microbial biofilms with cold atmospheric plasma," *Frontiers in Microbiology*, vol. 11, 2020.

[7] Y. W. Kim, S. Subramanian, K. Gerasopoulos, H. Ben-Yoav, H.-C. Wu, D. Quan, K. Carter, M. Meyer, W. Bentley, and R. Ghodssi, "Effect of electrical energy on the efficacy of biofilm treatment using the bioelectric effect," *npj Biofilms and Microbiomes*, vol. 1, p. 15016, 09 2015.

[8] A. S. Katsigiannis, D. L. Bayliss, and J. L. Walsh, "Cold plasma for the disinfection of industrial food-contact surfaces: An overview of current status and opportunities," *Comprehensive Reviews in Food Science and Food Safety*, vol. 21, no. 2, pp. 1086–1124, 2022.

[9] K. Kosumsupamala, P. Thana, N. Palee, K. Lamasai, C. Kuensaen, A. Ngamjarurojana, P. Yangkhamman, and D. Boonyawan, "Air to h 2 -n 2 pulse plasma jet for in-vitro plant tissue culture process: Source characteristics," *Plasma Chemistry and Plasma Processing*, vol. 42, 05 2022.

[10] F. Mitsugi, M. Wago, R. Sakamoto, K. Nishida, and T. Kawasaki, "Influence of retrogradation on ki-starch visualization of reactive oxygen species emitted by plasma jets," *IEEE Transactions on Plasma Science*, vol. PP, pp. 1–7, 06 2021.

[11] M. Singh and D. S. Sharma, "Enhanced color correction using histogram stretching based on modified gray world and white patch algorithms," 2014.

[12] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.

[13] Z. Liu, W. Chen, Y. Zou, and C. Hu, "Regions of interest extraction based on hsv color space," in *IEEE 10th International Conference on Industrial Informatics*, 2012, pp. 481–485.

[14] M. A. Ozdemir, G. D. Ozdemir, and O. Guren, "Classification of covid-19 electrocardiograms by using hexaxial feature mapping and deep learning," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 170, 2021.

Smartphone Platform for Image-Based Cell Confluency Analysis

Meryem Beyza Avci*†‡, Arif E. Cetin‡§, Mehmet Turkan*
*Dept. of Electrical and Electronics Engineering, Izmir University of Economics, Izmir, Turkiye
†Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Izmir, Turkiye
‡Izmir Biomedicine and Genome Center, Izmir, Turkiye
§Dept. of Biophysics, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkiye

***Abstract*—Monitoring cell confluency is a fundamental step in cell culture workflows, offering insights into cell growth dynamics, health, and density. In this study, we present a smartphone-based image analysis framework for automated confluency estimation using brightfield images. Unlike conventional systems that rely on deep learning or high-end microscopy, our approach applies image processing techniques based on variance and entropy filtering to identify and segment cell-covered regions. These filters provide robust texture-based differentiation between cell-populated and background areas without requiring prior training or hardware calibration. Binary masks are generated to compute confluency as the ratio of segmented area to total field of view. The method is lightweight, cost-effective, and adaptable to varying image conditions, making it ideal for portable and resource-limited settings.**

***Keywords*—cell analysis, confluency, smartphone microscopy, image processing.**

I. Introduction

Confluency analysis gauges the surface area covered by cells, a valuable metric in studies where uniform cell coverage is crucial [1], [2], [3]. The maintenance of uniform confluency is important for the standardization of cell culture protocols, the reproducibility of the experimental results, and the minimization of the variability in cell behavior [4]. For example, the evaluation of potential therapeutics can be performed by measuring cell proliferation. By comparing confluency levels between treated and control groups, researchers can determine the impact of a drug on cell growth [5]. Furthermore, bio-manufacturing processes, for instance, the generation of monoclonal antibodies, vaccines, and recombinant proteins, necessitate strict control of confluency for enhancing productivity as well as yield [6], [7], [8], [9]. Monitoring confluency within stem cell-based treatments facilitates the optimal differentiation, expansion, and viability for transplantation purposes [10], [11]. Tissue engineering tasks include seeding cells onto scaffolds for the development of artificial functional tissues or organs [12]. In such cases, confluency assessment maintains a uniform cell distribution, preventing inconsistencies in tissue formation that can affect structural integrity and function [13], [14]. Optimal confluency levels are crucial in genetic engineering and transfection study to maximize both the reproducibility and the efficiency of the transfection process [15], [16]. Cancer biology and toxicology research are largely dependent upon confluency measurements to standardize cell growth-based proliferation assays and assess the effects of chemotherapeutic agents under uniform growth conditions [17], [18], [19], [20]. All these applications highlight the pivotal role of confluency monitoring as a quality parameter for experiment consistency, reproducibility, and maximizing the efficiency of cell-based research and production processes.

Manual confluency assessment, typically performed through visual inspection under a microscope, remains a widely used and label-free approach in cell culture studies. Although this method is simple and non-destructive, it is highly subjective, time-consuming, and prone to inter-operator variability. To overcome these limitations, various commercial systems have been developed that offer automated confluency analysis using either image processing algorithms or AI-based techniques. For example, platforms like CELENA® X and ZEISS Labscope utilize artificial intelligence for robust, morphology-independent analysis, while systems such as CKX-CCSW and Axion BioSystems apply classical image processing [21], [22], [23], [24]. However, these solutions often involve specialized infrastructure [25].

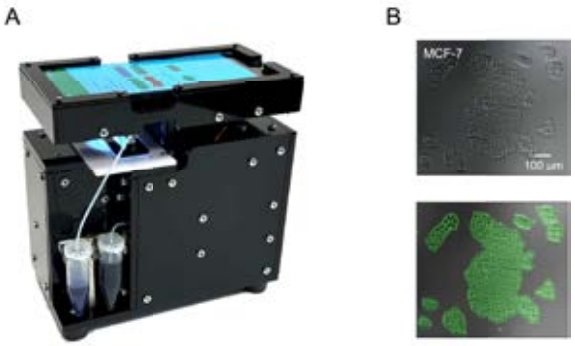


Fig. 1: (A) Image of the confluency analysis platform. (B) Segmentation result for breast cancer cells (MCF-7) cell culture image, showing green-highlighted regions indicating cell-covered areas.

In this study, a smartphone-based automated platform is proposed for accurate and reproducible cell confluency analysis (Figure 1). Unlike traditional manual methods that are time-consuming and prone to subjectivity, the system integrates a lightweight image processing algorithm that eliminates the

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

need for deep learning or complex hardware. By minimizing user intervention and operator-dependent parameters, the platform ensures reliable segmentation of cell-covered regions, making it a practical and accessible tool for both research and clinical environments.

II. Methodology

A. High-Contrast Imaging

The confluency calculations offered by the platform are performed by creating high-contrast cell images through an integration of a custom-designed grid slab. In Figure 2A, the working principle of this grid slab is illustrated (see photograph), which contains a white paper with printed black grids placed between five microscope slides; for instance, four glass slides above the paper and one underneath, all of them securely bonded together to ensure stability and alignment. The grid slab with a thickness of 4 mm and a flow cell with a 3 mm acrylic layer are combined, creating a total distance of 7 mm between each sample and the grid. With this design, generation of a high-contrast background without capturing the grid in the imaging process is achieved. The visibility of transparent cells is augmented by black grids, which highlight their edges and features. The detection of other transparent cells is enhanced by creating variations of light intensity in such areas, where light can pass through or is blocked by the grid. The working principle of this technique is similar to phase contrast or differential interference contrast microscopy [26]. A comparison of the same samples of MCF-7 cells imaged by phase contrast microscopy is shown in Figure 2B, which clearly demonstrates the platform's capability of high-contrast imaging (Olympus CKX41) with a much simpler and more cost-effective methodology.

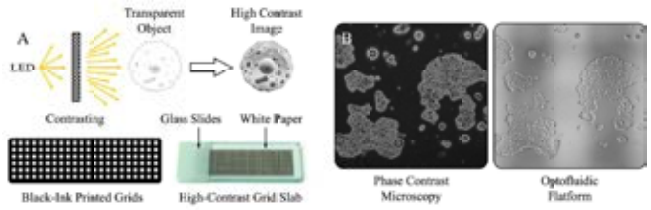


Fig. 2: (A) Schematic representation of the grid slab's working principle. (B) High-contrast confluency images produced by the optofluidic platform, offering visibility similar to phase-contrast microscopy.

B. Adherent Cell Culture

MCF-7 cells were cultured in RPMI 1640 medium (Sigma-Aldrich®), supplemented with 10% FBS (Sigma-Aldrich®), 100 units/mL penicillin (Sigma-Aldrich®), and 100 µg/mL streptomycin (Sigma-Aldrich®).

C. Segmentation Algorithm for Confluency Analysis

The flowchart in Figure 3 depicts the steps of the proposed image segmentation algorithm for confluency analysis. To attenuate noise and possible artifacts due to JPEG compression, the input image with dimensions of 4000×3000×3 pixels (8 bits per pixel) is filtered with a 2D Gaussian smoothing filter (SD=2) in the preprocessing step. Thereafter, variance and entropy filters, two distinct filtering techniques, are applied to augment image features for segmentation. The local variance information is extracted from the image [27], which highlights the regions with prominent intensity variations, i.e., textures, corners, and edges. Due to the sharp intensity variations at cell boundaries, a variance filter is employed to improve the visibility of these regions. This filter calculates the local variance in a small area of each pixel within the Gaussian-smoothed image. For each target pixel, a 3×3 pixel window is extracted in the developed segmentation algorithm, and the output value of the central pixel of this pixel window is computed from the variance of these nine pixels. As a result, the variance of each local pixel window is represented by its output. To correct illumination and improve the contrast of the variance-filtered output, a gamma transformation with γ = 0.6 is applied. To ensure consistent intensity scaling, the final values are normalized to a range of [0, 255] [28]. Then, Sobel operators are applied to compute the gradient magnitude of the resulting image, highlighting intensity variations along cell boundaries [29]. Two 3×3 kernels are used by Sobel operators to estimate the image gradients in both x (horizontal) and y (vertical) dimensions, emphasizing rapid intensity changes at cell boundaries.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{1}$$

Here, G_x and G_y are the gradient extraction kernels along the x- and y-directions, where their magnitude at each pixel location (i, j) is calculated via the Pythagorean sum as:

$$G_{\text{mag}}(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \tag{2}$$

To ensure consistent intensity scaling, the gradient image is rescaled to a range of [0,255] and then processed by hysteresis-based double thresholding technique. This approach is more effective for segmentation of cell boundaries and reveals finer details, since it increases two intensity thresholds to differentiate strong edges while maintaining weaker, connected edges. This method attenuates the noise and isolates the most prominent edges adopting a higher threshold as "marker" (set to 0.8), and a lower threshold as "mask" (set to 0.5), capturing finer details and weaker edges. By using morphological reconstruction, the two output images are merged afterwards for more precise edge detection. By merging strong edge information with augmented details, this method enables more precise segmentation of cell boundaries, which balances noise/artifacts attenuation and preserves details. An approach based on entropy filtering is also introduced in this study, reinforcing the outcomes from the variance filtering technique.

The randomness or variability in local pixel intensity is evaluated by the entropy filter, which is a commonly used technique. This technique sets higher values for regions with

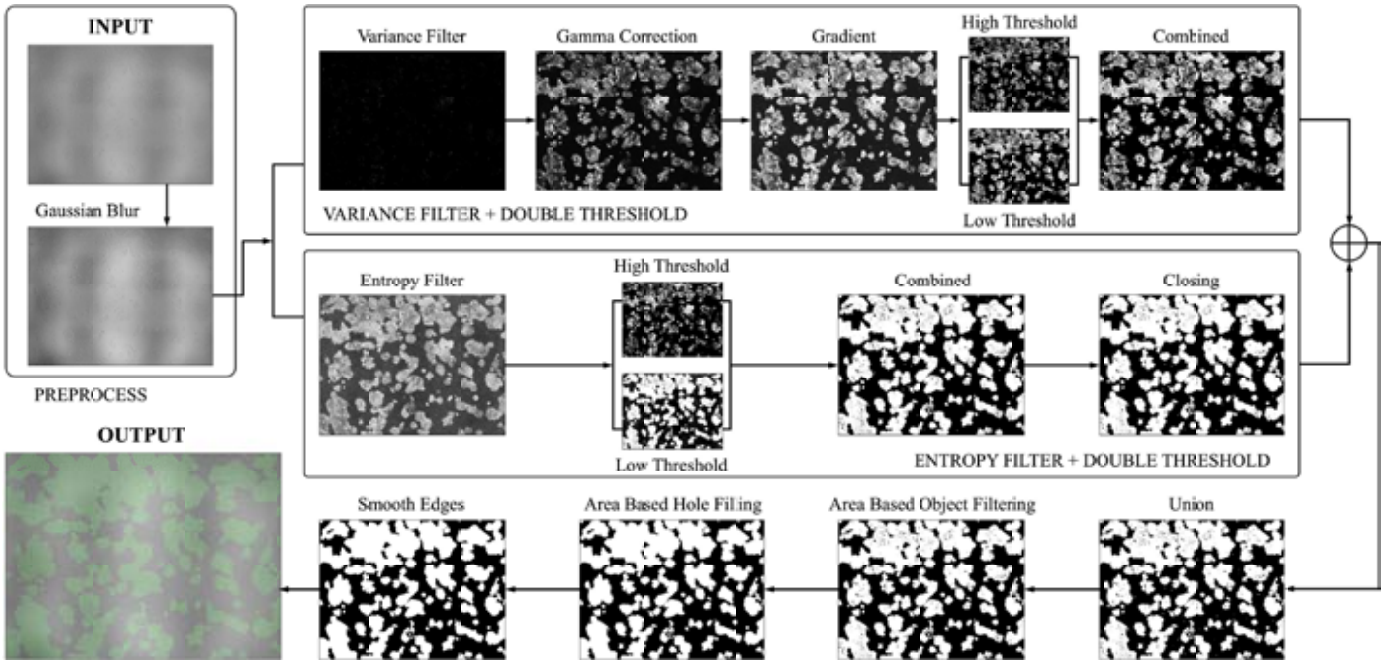


Fig. 3: Flowchart depicting the workflow of the image segmentation algorithm developed for confluency analysis.

intense variation. The entropy filter is a widely used technique for quantifying the randomness or variability in local pixel intensity [30]. With this technique, higher values are set to regions with marked intensity variation to accentuate areas rich in information, i.e., edges, corners and textures. From another perspective, complex textures and varying intensities of cells produce higher entropy values compared to the uniform background, which has lower entropy values. Owing to this difference, the selection of cell boundaries from the relatively uniform background is facilitated. In this study, this filter is applied through the computation of the local entropy within a 9x9 neighborhood around each target pixel. To partition regions of varying intensity, the filtering output is double-thresholded by using thresholds of 0.7 and 0.4. Then, the output images are merged using morphological reconstruction, followed by a morphological closing operation with a disk-shaped kernel of radius 3. Small internal holes across cell boundaries are filled by the closing operation, which augments the partition and preserves the integrity of the identified cell structures. A "union" operation is applied to merge binary outputs of variance an entropy filters for creation of a unified segmentation result. To enhance this output, small-connected objects are eliminated by implementation of an area-based object filter utilizing a connectivity value of 8. Thereafter, the internal sides of the identified cell boundaries are filled with area-based hole filling, providing a complete segmentation. For further augmentation of the output, a mean filter with a window size of 21 is applied to smooth the cell boundaries. Moreover, the resultant processed image is converted to the RGB color space, marking the detected cells in green. This green mask is then combined with the input image for a clearer visual representation of detected cells. Ultimately, confluency (%) is computed by division of the number of white pixels by the total number of pixels in the final binary output image, i.e., *White Pixels / Total Pixels*. Ultimately, confluency (%) is defined as the ratio of white pixels to total pixels, where the values are measured from the image of final binary output.

III. Results

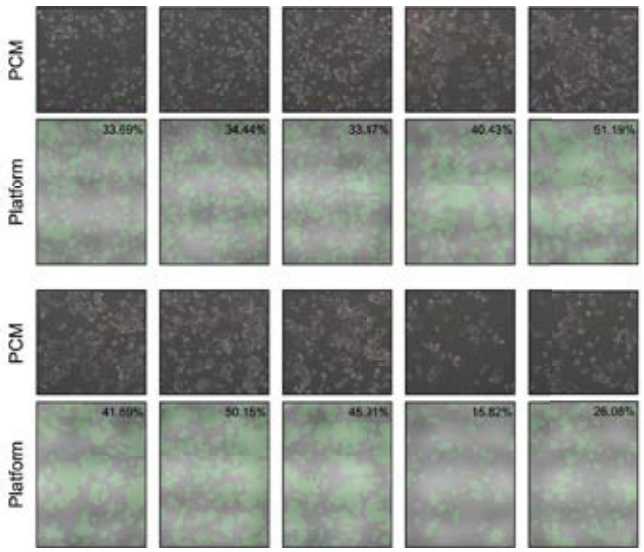


Fig. 4: Confluency analysis results of MCF-7 cells at different densities.

In Figure 4, the results of confluency analysis of MCF-7 cells at different confluency levels are illustrated. A phase contrast microscopy is utilized to capture the images in the first row (PCM), while the platform is used to acquire corresponding images shown in the second row, overlaid with

the segmentation results. The identified areas covered by cells are emphasized in green to represent confluency estimation visually.

IV. Conclusion

This study presented a smartphone-based automated platform for reliable and reproducible cell confluency analysis. By combining variance and entropy filtering with morphological operations, the system enables accurate segmentation without the need for complex hardware or deep learning. As the global demand for accessible and efficient cell analysis tools increases, the proposed platform offers a cost-effective alternative that bridges the gap between advanced laboratory systems and resource-limited environments. Its versatility and simplicity make it a promising solution for widespread adoption in both research and clinical settings.

Acknowledgment

M.B.A. acknowledges the Scientific and Technological Research Council of Turkiye (TUBITAK) 2210/D National Industrial MSc/MA Scholarship Program & TUBITAK 2224-A Grant Program for Participation in Scientific Meetings Abroad.

References

[1] O. Hayes, B. Ramos, L. L. Rodriguez, A. Aguilar, T. Badia, and F. O. Castro, "Cell confluency is as efficient as serum starvation for inducing arrest in the G0/G1 phase of the cell cycle in granulosa and fibroblast cells of cattle," *Animal Reproduction Science*, vol. 87, no. 3–4, pp. 181–192, 2005.

[2] R. T. Tavaluc, L. S. Hart, D. T. Dicker, and W. S. El-Deiry, "Effects of low confluency, serum starvation and hypoxia on the side population of cancer cell lines," *Cell Cycle*, vol. 6, no. 20, pp. 2554–2562, 2007.

[3] Z. Xue, J. Zeng, Y. Li, B. Meng, X. Gong, Y. Zhao, and X. Dai, "Proteomics reveals that cell density could affect the efficacy of drug treatment," *Biochemistry and Biophysics Reports*, vol. 33, p. 101403, 2023.

[4] C. P. Segeritz and L. Vallier, "Cell Culture: Growing Cells as Model Systems In Vitro," in *Basic Science Methods for Clinical Researchers*, Elsevier, 2017, pp. 151–172.

[5] B. C. Heng, X. Zhao, S. Xiong, K. W. Ng, F. Y. C. Boey, and J. S. C. Loo, "Cytotoxicity of zinc oxide (ZnO) nanoparticles is influenced by cell density and culture format," *Archives of Toxicology*, vol. 85, pp. 695–704, 2011.

[6] M. Ye, M. Wilhelm, I. Gentschev, and A. Szalay, "A modified limiting dilution method for monoclonal stable cell line selection using a real-time fluorescence imaging system: A practical workflow and advanced applications," *Methods and Protocols*, vol. 4, no. 1, p. 16, 2021.

[7] H. Zhang, X. Duan, G. Liu, Y. Li, S. Dong, J. Lin, R. Zhang, X. Cai, and H. Shan, "Comparative transcriptomic analysis of PK15 cells infected with a PRV variant and the Bartha-K/61 vaccine strain," *Frontiers in Microbiology*, vol. 14, p. 1164170, 2023.

[8] X. Y. Wang, Q. J. Du, W. L. Zhang, D. H. Xu, X. Zhang, Y. L. Jia, and T. Y. Wang, "Enhanced transgene expression by optimization of poly A in transfected CHO cells," *Frontiers in Bioengineering and Biotechnology*, vol. 10, p. 722722, 2022.

[9] C. Y. Lai, A. To, T. A. S. Wong, M. M. Lieberman, D. E. Clements, J. T. Senda, A. H. Ball, L. Pessaint, H. Andersen, W. Furuyama, and A. Marzi, "Recombinant protein subunit SARS-CoV-2 vaccines formulated with CoVaccine HT™ adjuvant induce broad, Th1 biased, humoral and cellular immune responses in mice," *Vaccine: X*, vol. 9, p. 100126, 2021.

[10] D. Lisini, S. Frigerio, S. Nava, and S. Pogliani, "Stem Cell Production: Processes, Practices, and Regulation: Stem Cell Culture and Expansion: Role of Culture, Types of Cells, Growth Conditions, Media Nutrients, Growth Factors, Growth Phase Cycle," in *Stem Cell Production: Processes, Practices and Regulations*, Springer Singapore, Singapore, 2022, pp. 125–158.

[11] A. Jamalpoor, S. Hartvelt, M. Dimopoulou, T. Zwetsloot, I. Brandsma, P. I. Racz, T. Osterlund, and G. Hendriks, "A novel human stem cell-based biomarker assay for in vitro assessment of developmental toxicity," *Birth Defects Research*, vol. 114, no. 19, pp. 1210–1228, 2022.

[12] A. Marchini and F. Gelain, "Synthetic scaffolds for 3D cell cultures and organoids: Applications in regenerative medicine," *Critical Reviews in Biotechnology*, vol. 42, no. 3, pp. 468–486, 2022.

[13] D. A. Matoz-Fernandez, K. Martens, R. Sknepnek, J. L. Barrat, and S. Henkes, "Cell division and death inhibit glassy behaviour of confluent tissues," *Soft Matter*, vol. 13, no. 17, pp. 3205–3212, 2017.

[14] O. M. Wroblewski, M. H. Nguyen, P. S. Cederna, and L. M. Larkin, "Impact of cell seeding density and cell confluence on human tissue engineered skeletal muscle," *Tissue Engineering Part A*, vol. 28, no. 9–10, pp. 420–432, 2022.

[15] F. Krötz, H. Y. Sohn, T. Gloe, C. Plank, and U. Pohl, "Magnetofection potentiates gene delivery to cultured endothelial cells," *Journal of Vascular Research*, vol. 40, no. 5, pp. 425–434, 2003.

[16] P. Agrawal, N. P. Ingle, W. S. Boyle, E. Ward, J. Tolar, K. D. Dorfman, and T. M. Reineke, "Fast, efficient, and gentle transfection of human adherent cells in suspension," *ACS Applied Materials & Interfaces*, vol. 8, no. 14, pp. 8870–8874, 2016.

[17] V. Laurent, A. Fraix, T. Montier, S. Cammas-Marion, C. Ribault, T. Benvegnu, P. A. Jaffres, and P. Loyer, "Highly efficient gene transfer into hepatocyte-like HepaRG cells: New means for drug metabolism and toxicity studies," in *Drug Testing in Vitro: Breakthroughs and Trends*, vol. 5, Weinheim: WILEY-VCH Verlag, 2010, pp. 314–320.

[18] J. Conde, M. Larguinho, A. Cordeiro, L. R. Raposo, P. M. Costa, S. Santos, M. S. Diniz, A. R. Fernandes, and P. V. Baptista, "Gold-nanobeacons for gene therapy: evaluation of genotoxicity, cell toxicity and proteome profiling analysis," *Nanotoxicology*, vol. 8, no. 5, pp. 521–532, 2014.

[19] M. Emadi Baygi, Z. S. Soheili, I. Schmitz, S. Sameie, and W. A. Schulz, "Snail regulates cell survival and inhibits cellular senescence in human metastatic prostate cancer cell lines," *Cell Biology and Toxicology*, vol. 26, pp. 553–567, 2010.

[20] T. Singh and S. K. Katiyar, "Green tea polyphenol, (-)-epigallocatechin-3-gallate, induces toxicity in human skin cancer cells by targeting β-catenin signaling," *Toxicology and Applied Pharmacology*, vol. 273, no. 2, pp. 418–424, 2013.

[21] Logos Biosystems, "CELENA® X High Content Imaging System," 2024. [Online]. Available: https://logosbio.com/celena-x/ [Accessed: 31 January 2025].

[22] ZEISS Microscopy, "ZEISS Labscope," 2024. [Online]. Available: https://www.zeiss.com/microscopy/en/products/software/zeiss-labscope/zeiss-labscope-lab.html [Accessed: 31 January 2025].

[23] Olympus Life Science, "CKX-CCSW Confluency Checker Software," 2024. [Online]. Available: https://www.olympus-europa.com/company/en/news/press-releases/new-olympus-confluency-checker-software.html [Accessed: 31 March 2025].

[24] Axion BioSystems, "Confluency Module," 2025. [Online]. Available: https://www.axionbiosystems.com/products/imaging/imaging-software/confluency-module [Accessed: 10 March 2025].

[25] M. B. Avci, S. D. Yasar, and A. E. Cetin, "An optofluidic platform for cell-counting applications," *Analytical Methods*, vol. 15, no. 18, pp. 2244–2252, 2023.

[26] V. C. Frohlich, "Phase Contrast and Differential Interference Contrast (DIC) Microscopy," *Journal of Visualized Experiments (JoVE)*, vol. 18, p. e844, 2008.

[27] J. Yang, G. Zhu, and Y. Q. Shi, "Analyzing the effect of JPEG compression on local variance of image intensity," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2647–2656, 2016.

[28] X. Li, M. Liu, and Q. Ling, "Pixel-wise gamma correction mapping for low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 2, pp. 681–694, 2023.

[29] G. Moreno, J. S. Ascaneo, J. O. Ricardo, L. T. De La Cruz, Y. Arias, J. R. Strickler, and R. M. Lopes, "A new focus detection criterion in holograms of planktonic organisms," *Pattern Recognition Letters*, vol. 138, pp. 497–506, Oct. 2020.

[30] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. Springer, 2013.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Self-assembling Peptide Hydrogel Design for Enhanced Osseointegration and Vascularization of Dental Implants

Elif Deniz Toprakçı[1], Buse Sezer[2]

[1]Biomedical Engineering, İzmir Katip Çelebi University, Balatçık, Çiğli,35620, Turkey

denizetoprakci@gmail.com

busezer9@gmail.com

Günnur Pulat[1]

[2]Biomedical Technologies Department, İzmir Katip Çelebi University, Balatçık, Çiğli,35620, Turkey

onak.gunnur@ikcu.edu.tr

***Abstract*— Successful dental implant integration relies on osseointegration, a process requiring both bone formation (osteogenesis) and vascularization (angiogenesis) at the implant site. While conventional surface modification strategies primarily aim to enhance bone cell attachment, they often neglect the critical role of vascular support in tissue regeneration. Peptide hydrogels have emerged as promising materials due to their ability to enhance cell adhesion, mimic the natural environment, and be injectable. In this project, we developed a multifunctional hydrogel incorporating the osteoinductive peptide KLD-12 with the endothelial cell-selective REDV sequence. This KLD-12-REDV hydrogel is designed to create a biomimetic microenvironment that simultaneously promotes osteogenic and vasculogenic processes. The hydrogel's biocompatibility was assessed in vitro using a co-culture system of mesenchymal stem cells and endothelial progenitor cells. Rheological analysis confirmed the hydrogel's structural stability, while viability assays demonstrated high cell survival rates. This strategy offers a novel and minimally invasive approach that extends beyond traditional implant surface modifications by integrating bioactive cues to support comprehensive tissue regeneration.**

Keywords—dental implants, self-assembling peptide hydrogel, osteogenic differentiation, vascularization

I. INTRODUCTION

Dental implants are a conventional solution for replacing missing or diseased teeth, which offers functional and aesthetic benefits. The clinical success of these implants largely depends on their ability to osseointegrate, which is a process of achieving stable integration with the surrounding bone tissue [1]. Osseointegration is the process in which an implant's surface becomes directly bonded to living bone, forming both a structural and functional link without the presence of a fibrous tissue layer in between. This stable interface allows for long-term fixation and functional load-bearing of the implant [2]. Despite advancements in implant designs and surgical techniques, the natural process of osseointegration can be prolonged, often taking several months to achieve sufficient stability for prosthetic loading. Prolonged healing periods may lead to patient discomfort, longer treatment durations, and increased risk of implant failure. Therefore, accelerating the osseointegration process remains a major focus in the field of dental implantology. While conventional surface modification methods such as plasma spraying and hydroxyapatite coatings improve the initial cellular adhesion of osteogenic cells, they often fail to address another critical factor: vascularization [3]. Adequate vascularization is essential for transporting oxygen, nutrients, and metabolic waste, thereby supporting healthy bone remodeling and long-term implant stability. The limited ability of current surface modification strategies to promote both osteogenesis and angiogenesis simultaneously creates a significant gap in achieving optimal and rapid osseointegration. Consequently, the development of multifunctional biomaterials that can support both bone regeneration and vascular network formation is of great importance [3].

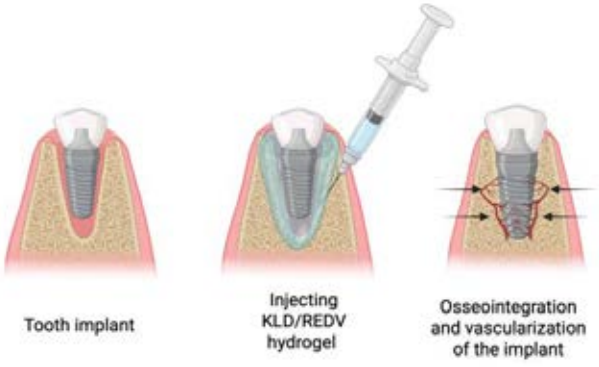


Fig 1. Schematic illustration of the designed project, showing the possible mechanism of action of the KLD-12-REDV hydrogel (Created in BioRender. Onak, G. (2025) https://BioRender.com/h94832z)

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Hydrogels have emerged as promising candidates in tissue engineering due to their high water content, biocompatibility, injectability, and capability of mimicking the extracellular matrix (ECM). Among them, self-assembling peptide hydrogels offer additional advantages such as tunable mechanical properties, controlled bioactivity, and ease of chemical modification [4]. KLD-12 peptide (Ac-(KLDL)$_3$-COOH), composed of alternating hydrophobic and hydrophilic amino acids, can spontaneously form nanofiber networks in aqueous environments, resulting in stable hydrogel structures. These hydrogels provide a supportive microenvironment for cell attachment, migration, and differentiation, which makes them attractive for bone tissue engineering applications [5]. Conversely, the REDV peptide (Arg-Glu-Asp-Val) has a well-recognized affinity for endothelial cells, enabling targeted binding and promoting the formation of new blood vessels. Incorporating REDV into self-assembling hydrogels may enhance angiogenic responses, contributing to improved tissue integration and accelerated healing [5,6]. In this study, we designed and developed a multifunctional self-assembling peptide hydrogel that can simultaneously promote osteogenesis and vascularization, thereby enhancing osseointegration of dental implants was designed and developed. For this purpose, the KLD-12 sequence was modified with the REDV motif to create a novel KLD-12-REDV peptide hydrogel. The developed hydrogel was synthesized using solid-phase peptide synthesis (SPPS) and characterized through rheological and in vitro cell viability studies by co-culturing human bone marrow-derived mesenchymal stem cells (hBMSC) and human umbilical vein endothelial cells (HUVECs). Developed hydrogels possessed a hydrogel structure suitable for the injectable applications, and it also enhanced cell viability, which confirmed the bioactivity of the hydrogel. Future studies will include the vasculogenic and osteogenic differentiation of these cells to confirm the regenerative potential aimed by the project.

II. Methodology

A. Solid-Phase Peptide Synthesis (SPPS)

The peptides Ac-KLDLKLDLKLDLKLDL-COOH and Ac-KLDLKLDLKLDLKLDLREDV-COOH (K: Lysine, L: Leucine, D: Aspartic acid, R: Arginine, E: Glutamic acid, D: Aspartic acid, V: Valine) were synthesized using SPPS based on Fmoc (9-fluorenylmethyloxycarbonyl) chemistry. The synthesis was conducted on Rink Amide MBHA resin (loading capacity: 0.53 mmol/g). Amino acid couplings were performed using HBTU/HOBt as activating agents and DIEA in DMF as the base. Deprotection of Fmoc protection groups was realized with 20% piperidine solution in DMF. When the peptide sequences were completed, the resins were left in contact with 20% acetic anhydride in DMF for acetylation to neutralize the terminal charge of the peptide. Following chain elongation and final deprotection, cleavage of the peptides from the resin using a cleavage cocktail consisting of TFA:TIPS:DIW (95:2.5:2.5:2.5 v/v) was realized. The peptides were precipitated in cold diethyl ether after being kept in contact with the cleavage solution for 2 hours and finally centrifuged at 4100 rpm for 10 minutes, and lyophilized [6].

B. Self-assembling Peptide Hydrogel Preperation

To prepare the hydrogels, lyophilized KLD-12 and KLD-12-REDV peptides were initially weighed and transferred into sterile eppendorf tubes at a final concentration of 6% (w/v). Prior to use, ultraviolet (UV) light was used for the sterilization of peptide powders under a laminar flow hood for 45 minutes. For reconstitution, sterile distilled water and Dulbecco's Modified Eagle Medium (DMEM) were used at a ratio of 1:1 volume. Firstly, 1 ml of DIW was added to 60 mg of peptide powder. The mixture was then sonicated until a homogeneous solution was obtained. To initiate gelation, an additional volume of DMEM was added immediately before use [6].

C. Rheology Analysis

The viscoelastic properties of KLD-12 and KLD-12-REDV hydrogels were analyzed using a rotational rheometer (e.g., Anton Paar MCR series) with 25 mm parallel plate geometry. Approximately 500 µL of each hydrogel was placed between the plates, and the upper plate was lowered to a set gap. A frequency sweep between 0.1 and 10 Hz was carried out at 1% strain, remaining within the linear viscoelastic range, with all tests conducted at 25 °C.The storage modulus (G′) and loss modulus (G″) were recorded to evaluate the mechanical stiffness and viscoelastic behavior. Measurements were repeated in triplicate for statistical validation.

D. Cell Encapsulation and Cell Proliferation Analysis

hBMSCs and HUVECs were used for evaluating the biocompatibility and dual-functionality (osteogenic and vasculogenic) of the hydrogels. hBMSCs and HUVECs were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Evaluation of the cell viability conducted using the MTT assay at days 1, 3, and 7. The assay measures mitochondrial metabolic activity of both cell types under co-culture conditions.

MTT analyses were performed at each time point, by removing medium on the hydrogels and adding 100 µL of 0.5 mg/mL MTT solution in DMEM. MTT solution was kept in contact with hydrogels for 2-4 hours at 37 °C. Following incubation, 100 µL of DMSO was added to dissolve the resulting formazan crystals. Absorbances was obtained by measuring at 570 nm using a microplate reader. Cell viability data were normalized to day 1 values, and statistical comparisons were performed to evaluate differences between hydrogel formulations and time points.

E. Statistical Analysis

All results are expressed as mean ± standard deviation (SD). Group comparisons were conducted using either an unpaired Student's t-test or a one-way analysis of variance (ANOVA) with Tukey's post hoc test when applicable. Statistical significance was defined as $p < 0.05$. Graphs and statistical evaluations were generated using GraphPad Prism software.

III. Results and Discussion

Rheological properties of hydrogels are significant for their applications in tissue engineering, 3D printing, and drug delivery. These hydrogels should exhibit tunable rheological properties, which are crucial for their injectability [7]. For an injectable hydrogel, the storage modulus (G') should possess a higher value than the loss modulus (G''), to ensure that the material maintains its structure and mechanical stability at the target site. The rheological behavior of KLD-12 and KLD-12-REDV hydrogels was assessed to determine their mechanical stability and viscoelastic properties. As shown in Figure 1, both hydrogels demonstrated dominant elastic behavior, with G′ values significantly higher than G″ across all tested frequencies. This result confirms that the self-assembled networks maintained solid-like characteristics, suitable for tissue scaffold applications. However, a comparative analysis between the two formulations revealed that KLD-12-REDV exhibited a lower G′ than KLD-12, indicating a reduction in stiffness following REDV peptide conjugation. This reduction in modulus may be attributed to two possible factors: the REDV motif may slightly disrupt β-sheet stacking or hydrogen bonding, weakening nanofiber crosslinking, and the hydrophilic and charged nature of the REDV extension may increase water content or spacing between fibers, thereby reducing overall rigidity [8]. Despite the reduction in the rigidity, the KLD-12-REDV hydrogel still displayed a G′ in the range typical for injectable tissue engineering materials, suggesting that the mechanical compromise remains within acceptable limits for in vivo use. The gel's ability to recover from shear stress and its solid-like behavior over physiological frequency ranges make it a promising candidate for use in peri-implant applications.

The biocompatibility of the developed hydrogels was investigated by encapsulating a co-culture of hBMSCs and HUVECs within the KLD-12 and KLD-12-REDV matrices. Cells were encapsulated at a density of 1×10^6 cells/mL and cultured in a 1:1 mixed medium of DMEM. Cell viability of cell-encapsulated hydrogels was investigated at days 1, 3, and 7 using the MTT assay, which quantifies mitochondrial enzymatic activity as an indirect measure of viable cell number. MTT absorbance readings at 570 nm were normalized to KLD-12 day 1 to enable comparison of metabolic trends over time.

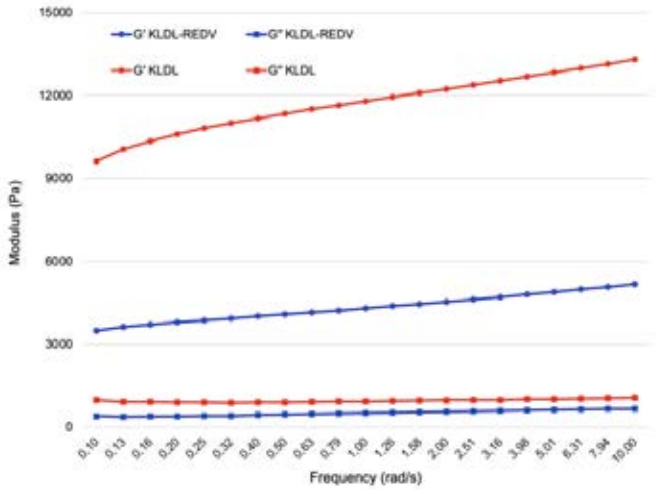


Fig 2. Rheology analysis results: Frequency sweep showing storage modulus (G′) and loss modulus (G″) of KLD-12 and KLD-12-REDV hydrogels (1% w/v) measured at 25 °C. Data shown are mean ± SD, n = 3.

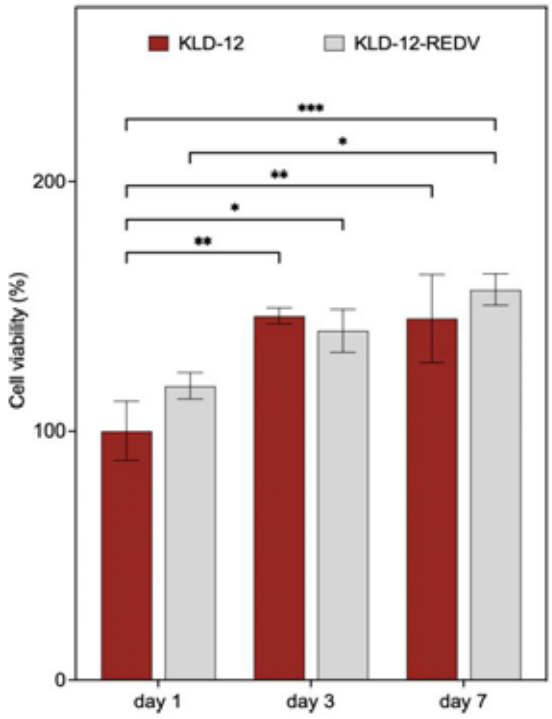


Fig 3. Cell proliferation assay results. Relative metabolic activity of hBMSC + HUVEC co-culture encapsulated in KLD-12 and KLD-12-REDV hydrogels, measured at days 1, 3, and 7 using the MTT assay. Error bars represent mean ± SD, n = 3.

As illustrated in Figure 2, in both KLD-12 and KLD-12-REDV groups, there was a progressive increase in absorbance from day 1 to day 7, indicating cellular proliferation and metabolic activity over time. On day 1, the KLD-12-REDV hydrogel exhibited an 18% higher cell viability compared to the KLD-12 hydrogel. By day 3, cell viability was comparable between the two groups. However, by day 7, the KLD-12-REDV group

again showed a notable improvement with a 10% increase in cell viability relative to the KLD-12 hydrogel. Notably, the KLD-12-REDV hydrogel showed a higher metabolic activity by day 7 compared to unmodified KLD-12. This observation confirms improved viability or proliferation of endothelial cells and mesenchymal stem cells in response to the REDV modification. Although these results are preliminary, we anticipate that functionalizing the hydrogel with REDV may improve the biological performance of the material in vascularized tissue models, considering the previously proven activity of REDV [9].

Hydrogels and peptides are promising agents used in enhancing the bioactive properties of orthopedic implants. Studies generally include surface modification of dental implants with peptides for antimicrobial effect and osseointegration [10,11]. Xu et al. developed a self-assembling peptide hydrogel for infected bone defects and used RADA-16 SAP and BMSC-derived exosomes for this purpose. BMSCs were used for the cell proliferation assay, and the developed hydrogel system enhanced the cell proliferation compared to the control group. However, osteogenic differentiation and vascularization capacity were not investigated [12]. Achieving osseointegration and vascularization in dental implants remains a significant challenge. As there is currently a lack of studies specifically focusing on this approach for dental implants, the development of injectable peptide-based hydrogel systems to support these biological processes holds great potential for enhancing implant biocompatibility and long-term success.

IV. Conclusion

This study successfully designed, synthesized, and evaluated a novel self-assembling peptide hydrogel aimed at enhancing osseointegration in dental implant applications through the combined stimulation of osteogenic and vasculogenic processes. The peptide hydrogel was constructed by integrating the KLD-12 sequence, known for its osteoinductive properties, with the REDV motif, a short peptide that promotes selective adhesion and survival of endothelial cells. Preliminary rheological analysis confirmed that both KLD-12 and KLD-12-REDV hydrogels exhibited dominant elastic behavior ($G' > G''$), consistent with well-formed viscoelastic scaffolds suitable for tissue engineering. The addition of the REDV motif caused a slight reduction in mechanical stiffness, yet the hydrogel remained within the optimal modulus range for injectable biomaterials. Furthermore, MTT assay results demonstrated high cell viability of co-cultured hBMSCs and HUVECs within both hydrogel formulations over a 7-day period, with a slightly enhanced metabolic activity observed in the REDV-modified group. These findings suggest that the dual-functional hydrogel provides a supportive environment for both stem cells and endothelial cells, which is critical for early osseointegration and neovascularization around dental implants. Future work will focus on comprehensive characterization of the KLD-12-REDV hydrogel, including osteogenic and vasculogenic assays, 3D co-culture studies, degradation and release profiling, and optimization of formulation parameters to enhance its therapeutic potential.

Acknowledgment

This work is funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under the 2209B Program for Industrial Undergraduate Research Projects.

This work is supported by the TÜBİTAK International Conference Participation Grant 2224-A.

References

[1] Nandal, S., Ghalaut, P., Shekhawat, H., & Nagar, P. (2014). Osseointegration in dental implants: a literature review. *Med. Sci*, *4*, 411-413.

[2] Albrektsson, T., & Johansson, C. (2001). Osteoinduction, osteoconduction and osseointegration. *European spine journal*, *10*(Suppl 2), S96-S101.

[3] Buser, D., Sennerby, L., & De Bruyn, H. (2017). Modern implant dentistry based on osseointegration: 50 years of progress, current trends and open questions. *Periodontology 2000*, *73*(1), 7-21.

[4] Cui, H., Webber, M. J., & Stupp, S. I. (2010). Self-assembly of peptide amphiphiles: From molecules to nanostructures to biomaterials. *Peptide Science: Original Research on Biomolecules*, *94*(1), 1-18.

[5] Hubbell, J. A., Massia, S. P., Desai, N. P., & Drumheller, P. D. (1991). Endothelial cell-selective materials for tissue engineering in the vascular graft via a new receptor. *Bio/technology*, *9*(6), 568-572.

[6] Onak, G., Gökmen, O., Yaralı, Z. B., & Karaman, O. (2020). Enhanced osteogenesis of human mesenchymal stem cells by self-assembled peptide hydrogel functionalized with glutamic acid templated peptides. *Journal of Tissue Engineering and Regenerative Medicine*, *14*(9), 1236-1249.

[7] Herrada-Manchón, H., Fernández, M. A., & Aguilar, E. (2023). Essential guide to hydrogel rheology in extrusion 3D printing: how to measure it and why it matters?. *Gels*, *9*(7), 517.

[8] Warren, J. P., Culbert, M. P., Miles, D. E., Maude, S., Wilcox, R. K., & Beales, P. A. (2023). Controlling the self-assembly and material properties of β-sheet peptide hydrogels by modulating intermolecular interactions. *Gels*, *9*(6), 441.

[9] Liu, Y., Tan, T. T. Y., Yuan, S., & Choong, C. (2013). Multifunctional P (PEGMA)–REDV conjugated titanium surfaces for improved endothelial cell selectivity and hemocompatibility. *Journal of Materials Chemistry B*, *1*(2), 157-167.

[10] Chen, X., Zhou, X. C., Liu, S., Wu, R. F., Aparicio, C., & Wu, J. Y. (2017). In vivo osseointegration of dental implants with an antimicrobial peptide coating. *Journal of Materials Science: Materials in Medicine*, *28*(5), 76.

[11] Chen, X., Zhou, X. C., Liu, S., Wu, R. F., Aparicio, C., & Wu, J. Y. (2017). In vivo osseointegration of dental implants with an antimicrobial peptide coating. *Journal of Materials Science: Materials in Medicine*, *28*(5), 76.

[12] Xu, H., Feng, J., Dai, N., Han, Q., Zhou, B., Yang, G., & Hu, R. (2024). Self-assembling peptide hydrogel scaffold integrating stem cell-derived exosomes for infected bone defects. *Journal of Biomaterials Science, Polymer Edition*, *35*(10), 1511-1522.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Classification of Emotions Under Multiple Olfactory Stimuli Using EEG Signals and Machine Learning

Ozge Ada Bozbas, Burak Akbugday and Aydin Akan*
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Izmir, TURKEY
ozge.ada@std.ieu.edu.tr, {burak.akbugday, akan.aydin}@ieu.edu.tr

Sude Pehlivan Akbugday
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Izmir, TURKEY
sude.pehlivan@ieu.edu.tr

***Abstract*—Accurately predicting emotional states is crucial in various fields, including cognitive sciences, artificial intelligence, human-computer interaction, and neuromarketing. In particular, the development of emotion-focused technologies requires the analysis of emotional responses with objective biological data. In this study, the emotional responses of individuals to different odor stimuli were examined using electroencephalogram (EEG) signals and classified using machine learning methods. A total of 46 participants were exposed to an odorless condition and four different odor stimuli (cinnamon, citrus, green tea, and lavender). After each exposure, participants filled out self-report questionnaires based on the valence-arousal model. Emotional states were predicted using time-domain features obtained from EEG signals, and various machine learning algorithms were applied for classification. The results show that EEG-based approaches can classify emotional responses with high accuracy, with lavender being the odor that created the most potent effect, achieving an accuracy rate of 80.14%. This study demonstrates that emotion analysis using EEG signals combined with subjective assessment has significant potential in areas such as neuromarketing and therapeutic applications.**

***Index Terms*—Emotion Estimation, Olfactory stimulus, Electroencephalogram (EEG), Machine learning, Valence-Arousal Model**

I. Introduction

Emotions are essential components of human cognition and physiology, playing a crucial role in decision-making, perception, social interaction, and intelligence [1]. Accurate recognition and analysis of emotions have become increasingly important in various domains, including artificial intelligence, neuroscience, human-computer interaction (HCI), education, mental health, entertainment, and neuromarketing [2]. Emotion recognition involves determining an individual's affective state by analyzing their physiological and neurological responses. Commonly used biosignals include electrocardiography (ECG), galvanic skin response (GSR), respiration, and electroencephalography (EEG) [3]. Among these, EEG has attracted significant attention due to its high temporal resolution and its ability to reflect cortical activity associated with affective processing directly. EEG enables the extraction of features related to emotional states in both time and frequency domains, making it particularly suitable for real-time and fine-grained emotion analysis [4].

Machine learning-based techniques have been relatively successful in identifying emotions using EEG. Using data taken from EEG signals, algorithms including Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and k-nearest Neighbors (k-NN) accurately characterize emotional states [5]. Dai et al. showed that integrating multichannel feature extraction and machine learning methods effectively improves accuracy rates when predicting emotion from EEG signals [6]. Advances in artificial intelligence and neuroscience have enabled the study of the relationship between olfactory stimuli and emotional responses using EEG. Studies such as [7] and [8] analyzed the non-linear nature of EEG signals and demonstrated the effectiveness of the Multivariate Empirical Mode Decomposition (MEMD) method in identifying emotional states.

These studies reveal significant findings in classifying emotional responses to olfactory stimuli using EEG signals. In particular, multi-modal machine learning approaches have the potential to improve the classification of EEG signals. Recent studies have further validated the effectiveness of integrating multiple EEG domains for emotion recognition, especially in wearable and real-time contexts [9]. By focusing on the cerebral processing of olfactory experiences, the studies provide new insights into how the human brain processes different olfactory experiences. However, these studies lack consideration of emotion in all dimensions, namely, arousal and valence, in tandem with feedback from the participants. Consideration of feedback received from participants, alongside the EEG signals, allows for gaining further and more robust insight into the emotions of individuals.

This study proposes a method to classify emotional responses to four different odor stimuli using machine learning techniques and features derived from EEG signals. Unlike many prior works, which often rely on conventional stimuli (e.g., visual or auditory) or limited sensor/channel sets, this approach leverages olfactory stimuli. It incorporates personalized labeling via subject-specific thresholds derived from valence–arousal questionnaires. Time-domain features extracted from multi-segment EEG recordings are used in conjunction with various machine learning classifiers to model emotional

*This study was supported by Izmir University of Economics, Scientific Research Projects Coordination Unit. Project number: 2022-07.

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE

responses. By combining EEG features with subjective reports within the valence–arousal framework, this study provides a richer and more individualized understanding of emotion, offering novel insights into EEG-based olfactory emotion recognition.

II. Material and Method

A. Experimental Setup and the Dataset

The dataset consists of EEG recordings from 46 right-handed participants. Each participant was exposed to four odor stimuli (cinnamon, citrus, lavender, and green tea) as well as an unscented control condition. The stimuli were presented in randomized order using opaque boxes containing odor-saturated absorbent papers. To avoid olfactory habituation, coffee beans were provided between trials as a neutralizer. During each trial, participants were instructed to close their eyes and remain still for 1 minute, during which EEG signals were recorded using a Bitbrain Versatile-32 system. All experimental procedures were approved by the Izmir Katip Çelebi University Non-Interventional Clinical Research Ethics Committee (Decision No: 14, Date: 08.02.2024).

Each participant completed a single trial for each of the four olfactory conditions and an additional trial for the odorless baseline, resulting in a total of five trials. After each trial, participants completed a self-report questionnaire assessing their emotional responses across two core affective dimensions: valence and arousal. The questionnaire consisted of two tables with six bipolar adjective pairs (e.g., Happy–Unhappy, Calm–Tense), each rated on a scale from -2 to +2. Individual valence and arousal scores were calculated by averaging the corresponding item ratings. Emotional labeling was performed by comparing these scores with each participant's baseline values obtained under the odorless condition. EEG segments recorded during odor exposure that showed higher valence and/or arousal than the baseline were labeled as 1, indicating a positive emotional shift in Fig. 1. Segments from the odorless condition were labeled as 0, representing a neutral reference. This participant-specific, data-driven labeling strategy enabled the robust modeling of EEG-based emotional changes induced by olfactory stimuli, as shown in Fig. 2.

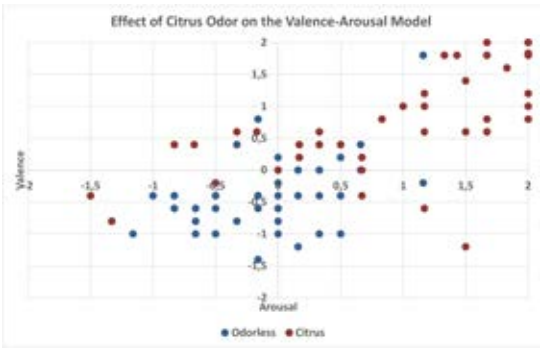


Fig. 1: Distribution of valence and arousal scores for citrus odor and odorless Conditions.

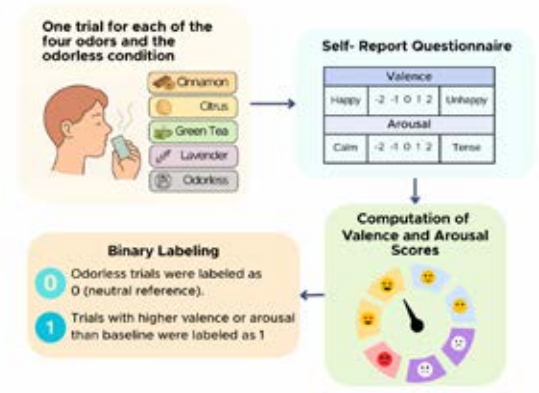


Fig. 2: Experimental setup and labeling before classification

B. EEG Preprocessing

Raw EEG signals were filtered using a 5th-order Butterworth band-pass filter with a passband of 0.5-45 Hz to remove low-frequency drifts, high-frequency muscle artifacts, and other physiological interferences unrelated to cortical activity. Segmentation was then performed by extracting five consecutive 1-second epochs immediately following the onset of each olfactory stimulus. Accordingly, five segments per odor condition were obtained from each participant, resulting in a balanced dataset structure suitable for feature extraction and classification.

C. Feature Extraction and Labeling

From each 1-second EEG segment, nine time-domain features were extracted to characterize both the statistical and dynamic properties of the signal. These included the mean, energy, skewness, and kurtosis to capture distributional aspects; root mean square (RMS) and mean absolute deviation (MAD) to reflect signal magnitude and variability; and the Hjorth parameters—activity, mobility, and complexity—which describe the amplitude variance, dominant frequency content, and frequency variation over time, respectively. These features are widely used in EEG-based emotion and neural signal analysis due to their computational efficiency and interpretability.

Two EEG channel configurations were used: the complete 23-channel set and a reduced 10-channel subset (F3–F4, F7–F8, FC1–FC2, FC5–FC6, FP1–FP2) based on prior affective EEG research [10].

Outliers were detected using z-score analysis, where a feature value x was considered an outlier if its absolute z-score exceeded 3; $z = \frac{x-\mu}{\sigma}$ where μ is the mean and σ is the standard deviation.

Detected outliers were replaced via mean imputation within the same feature vector. Each feature vector was labeled according to participants' questionnaire scores. Valence and arousal scores were binarized using subject-specific medians as thresholds to account for inter-individual variability. Separate classifiers were trained for each emotional dimension and odor type.

III. Emotion Estimation by Using EEG Features

To investigate the relationship between olfactory stimuli and emotional states, EEG-based classification models were developed within the valence-arousal framework. For model input, EEG features were extracted from two spatial configurations:

1) Complete 23-channel configuration: EEG features were extracted from all available channels to capture a comprehensive representation of brain activity across the scalp.
2) Reduced 10-channel configuration: A subset of frontal and central channels (F3–F4, F7–F8, FC1–FC2, FC5–FC6, FP1–FP2) was selected based on their known relevance to affective state processing, as supported by frontal asymmetry literature [11]. This approach aimed to reduce dimensionality while preserving emotion-relevant information.

Classification was performed separately for each odor condition and each emotional axis. A total of 17 machine learning classifiers, which offer competitive emotion classification performance with relatively low computation costs, were evaluated [12]. These are various types of SVM (Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, Coarse Gaussian SVM), k-NN (Fine k-NN, Medium k-NN, Coarse k-NN, Cosine k-NN, Cubic k-NN, Weighted k-NN), DT (Fine Tree, Medium Tree, Coarse Tree), and Kernel Approximation Methods (SVM Kernel and Logistic Regression Kernel). The performance of the trained model was evaluated using 5-fold cross-validation, as well as standard metrics such as accuracy, sensitivity, and specificity. An 80% to 20% train-test split was used on the whole dataset to ensure reliable performance estimation across different emotional states.

III. Results and Discussion

This study utilizes EEG time-domain features to evaluate the performance of machine learning models that analyze the effect of four different odors on individuals' valence and arousal levels. The accuracy rates of models trained using all EEG channels and those trained using selected EEG channels are compared, and the performances of different machine learning classifiers are analyzed.

The results in Tables I and IV show that SVM-based models provide the highest accuracy rates and are the most successful in classifying odor-induced emotional responses. According to the classification results, lavender odor achieved the highest accuracy rates in predicting both arousal and valence levels (see 3 and 4). In contrast, the cinnamon odor had a less pronounced effect on EEG signals than the other odors, resulting in relatively lower accuracy rates. This difference may be attributed to the more consistent emotional response lavender elicits across participants, which supports clearer EEG patterns for machine learning models to capture.

These findings align with earlier research demonstrating the effectiveness of traditional classifiers such as SVM, DT, and k-NN in EEG-based emotion recognition tasks [13], [14]. In contrast to prior studies that predominantly utilize visual or auditory stimuli, this work uniquely centers on olfactory triggers, which interact more directly with the limbic system and are thus well-suited for eliciting emotional responses. Additionally, subject-specific valence–arousal labels, derived from individualized self-assessment scores, were used to enhance model sensitivity and account for personal variation in emotional perception.

The comparison between 23-channel and 10-channel EEG configurations further revealed that the selected 10-channel setup targeting frontal and fronto-central regions (F3–F4, F7–F8, FC1–FC2, FC5–FC6, FP1–FP2) achieved superior classification performance. This selection was informed by prior findings indicating that these regions are critically involved in affective processing. Reducing the number of electrodes not only improved accuracy but also supports the development of more efficient and deployable emotion recognition systems.

Among the tested odors, lavender produced the highest classification accuracy in both emotional dimensions, reaching 80.14% for valence and 79.19% for arousal. These outcomes suggest that lavender elicits more consistent and distinguishable EEG patterns, thereby improving model reliability. Taken together, the results confirm the value of combining olfactory stimulation with personalized EEG analysis and contribute to expanding the scope of emotion recognition beyond conventional stimulus modalities.

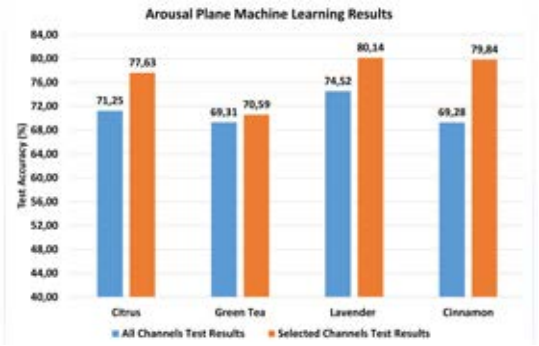


Fig. 3: Arousal estimation results using different classifiers

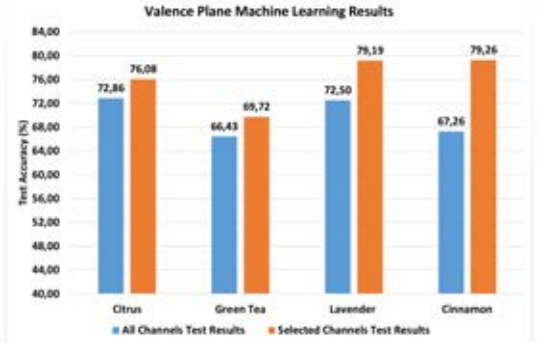


Fig. 4: Valence estimation results using different classifiers

TABLE I: Classification Results of Valence and Arousal for Citrus

Parameter	All 23 EEG Channels			Selected 10 EEG Channels		
	Model	**Accuracy % (Validation)**	**Accuracy % (Test)**	**Model**	**Accuracy % (Validation)**	**Accuracy % (Test)**
Arousal	Medium Gaussian SVM	**71,22**	**71,25**	Cubic SVM	**75,66**	**77,63**
	Fine Tree	66,17	66,36	Medium Tree	68,00	71,38
	Weighted k-NN	64,59	65,16	Coarse k-NN	64,66	67,25
	SVM Kernel	57,98	61,20	SVM Kernel	60,25	59,63
Valence	Medium Gaussian SVM	**71,12**	**72,86**	Cubic SVM	**76,82**	**76,08**
	Weighted k-NN	64,12	66,63	Fine Tree	71,52	67,3
	Fine Tree	66,07	66.45	Medium k-NN	66,26	66,22
	Logistic Regression Kernel	58.52	58.75	Logistic Regression Kernel	62,40	61,08

TABLE II: Classification Results of Valence and Arousal for Green Tea

Parameter	All 23 EEG Channels			Selected 10 EEG Channels		
	Model	**Accuracy % (Validation)**	**Accuracy % (Test)**	**Model**	**Accuracy % (Validation)**	**Accuracy % (Test)**
Arousal	Medium Gaussian SVM	**65,62**	**69,31**	Cubic SVM	**69,45**	**70,59**
	Cosine k-NN	61,13	65,03	Fine Tree	60,55	64,12
	Fine Tree	61,60	63,24	Weighted k-NN	62,65	63,09
	SVM Kernel	57,78	62,72	Logistic Regression Kernel	57,76	57,35
Valence	Fine Gaussian SVM	**64,79**	**66,43**	Cubic SVM	**69,86**	**69,72**
	Medium k-NN	60,49	63,95	Fine Tree	62,36	63,47
	Fine Tree	62,30	63,22	Coarse k-NN	60,24	62,92
	SVM Kernel	58,38	58,88	Logistic Regression Kernel	57,29	59,58

TABLE III: Classification Results of Valence and Arousal for Lavender

Parameter	All 23 EEG Channels			Selected 10 EEG Channels		
	Model	**Accuracy % (Validation)**	**Accuracy % (Test)**	**Model**	**Accuracy % (Validation)**	**Accuracy % (Test)**
Arousal	Medium Gaussian SVM	**71,83**	**74,52**	Cubic SVM	**79,31**	**80,14**
	Fine Tree	69,64	72,64	Fine Tree	71,88	72,08
	Coarse k-NN	65,37	68,36	Cosine k-NN	64,97	68,89
	Logistic Regression Kernel	58,95	62,32	Logistic Regression Kernel	62,36	64,31
Valence	Medium Gaussian SVM	**72,50**	**72,50**	Cubic SVM	**77,94**	**79,19**
	Weighted k-NN	67,63	68,63	Fine Tree	73,38	73,11
	Fine Tree	69,45	66,63	Weighted k-NN	66,89	67,03
	Logistic Regression Kernel	59,72	61,87	SVM Kernel	60,95	65,54

TABLE IV: Classification Results of Valence and Arousal for Cinnamon

Parameter	All 23 EEG Channels			Selected 10 EEG Channels		
	Model	**Accuracy % (Validation)**	**Accuracy % (Test)**	**Model**	**Accuracy % (Validation)**	**Accuracy % (Test)**
Arousal	Cubic SVM	**67,95**	**69,28**	Cubic SVM	**77,18**	**79,84**
	Weighted k-NN	63,38	66,90	Medium k-NN	62,14	65,32
	Fine Tree	61,48	62,90	Medium Tree	61,69	63,06
	SVM Kernel	57,31	60,80	Logistic Regression Kernel	57,26	60,00
Valence	Medium Gaussian SVM	**66,50**	**67,26**	Cubic SVM	**76,62**	**79,26**
	Weighted k-NN	63,99	65,66	Weighted k-NN	62,72	65,29
	Fine Tree	60,85	61,96	Medium Tree	62,54	62,35
	SVM Kernel	57,24	57,29	Logistic Regression Kernel	58,13	61,32

IV. Conclusions and Future Aspects

The experimental findings of this study demonstrate that both single and ensemble machine learning models can effectively classify emotional states elicited by olfactory stimuli, using time-domain EEG features. When benchmarked against recent research such as [15] and [16], the proposed methodology yields comparable accuracy levels, reaffirming the role of EEG as a viable modality in emotion recognition under sensory stimulation. In contrast to conventional approaches that rely on visual or auditory stimuli, this study solely focuses on olfaction. It enriches the existing literature by validating performance through valence–arousal–based classification.

Building upon earlier work [17], [18], [19], the integration of personalized labels and channel-specific modeling provided new insights into the emotional implications of odor stimuli on EEG dynamics. These results are also supported by recent efforts highlighting the efficacy of channel selection and feature domain hybridization for real-time and wearable emotion recognition systems [20], [21]. However, there is still room for improvement as the classification results are relatively low, especially for green tea. Moreover, due to time and scope constraints, this study only utilizes ML methods, whereas DL methods offer more accurate and robust classifiers, overall.

Looking forward, several enhancements are envisioned to improve model robustness and expand applicability. In particular, applying mode decomposition techniques and feature map representations [22] alongside DL methods will allow for generating augmented training data by decomposing the EEG into intrinsic mode functions, which could enhance classification accuracy and generalization. Furthermore, other feature sets, such as non-linear and entropy features, will enable a better representation of data, leading to more generalized classifiers. Clinically, the study contributes to a better representation for emotion modeling, the establishment of a baseline for the development of lightweight, high-accuracy models, as well as improved techniques for neuromarketing [17], affective computing [23], and cognitive interface [24] systems.

References

[1] A. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*. Random House, 2008.

[2] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE transactions on affective computing*, vol. 10, no. 3, pp. 374–393, 2017.

[3] V. M. Alvarez, R. Velázquez, S. Gutiérrez, and J. Enriquez-Zarate, "A method for facial emotion recognition based on interest points," in *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. IEEE, 2018, pp. 1–4.

[4] V. Kuperman, Z. Estes, M. Brysbaert, and A. B. Warriner, "Emotion and language: valence and arousal affect word recognition." *Journal of Experimental Psychology: General*, vol. 143, no. 3, p. 1065, 2014.

[5] R. Alhalaseh and S. Alasasfeh, "Machine-learning-based emotion recognition system using eeg signals," *Computers*, vol. 9, no. 4, p. 95, 2020.

[6] W. Dai, X. Li, Z. Xia, J. Zhou, L. Song, H. Mao, and Y. Kang, "Olfactory affective computation based on eeg signal data," in *2021 IEEE International Conference on Progress in Informatics and Computing (PIC)*. IEEE, 2021, pp. 432–436.

[7] S. P. Akbugday, B. Akbugday, F. Yeganli, A. Akan, and R. Sadikzade, "Decoding olfactory eeg signals using multi-domain features and machine learning," in *2024 Medical Technologies Congress (TIPTEKNO)*. IEEE, 2024, pp. 1–4.

[8] A. Mert and A. Akan, "Emotion recognition from eeg signals by using multivariate empirical mode decomposition," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 81–89, 2018.

[9] Y. Liu, W. Zhang, and C. Wang, "Multi-domain eeg feature fusion for emotion recognition in wearable applications," *Biomedical Signal Processing and Control*, vol. 85, p. 104224, 2024.

[10] R. Yuvaraj, P. Thagavel, J. Thomas, J. Fogarty, and F. Ali, "Comprehensive analysis of feature extraction methods for emotion recognition from multichannel eeg recordings," *Sensors*, vol. 23, no. 2, p. 915, 2023.

[11] O. Bazgir, Z. Mohammadi, and S. A. H. Habibi, "Emotion recognition with machine learning using eeg signals," in *2018 25th national and 3rd international iranian conference on biomedical engineering (ICBME)*. IEEE, 2018, pp. 1–5.

[12] M. Soroushnejad, M. Rezaei, and B. M. Asl, "A survey on eeg-based emotion recognition methods with a focus on feature extraction and classification techniques," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 319–339, 2021.

[13] O. K. Cura, A. Akan, G. C. Yilmaz, and H. S. Ture, "Detection of alzheimer's dementia by using signal decomposition and machine learning methods," *International journal of neural systems*, vol. 32, no. 09, p. 2250042, 2022.

[14] O. Fasil and R. Rajesh, "Time-domain exponential energy for epileptic eeg signal classification," *Neuroscience letters*, vol. 694, pp. 1–8, 2019.

[15] X. Xia, X. Liu, W. Zheng, X. Jia, B. Wang, and Y. Shi, "Recognition of odor and pleasantness based on olfactory eeg combined with functional brain network model," *International Journal of Machine Learning and Cybernetics*, vol. 14, pp. 2761–2776, 2023.

[16] A. Panteli, E. Kalaitzi, and C. A. Fidas, "A review on the use of eeg for the investigation of the factors that affect consumer's behavior," *Physiology & Behavior*, vol. 278, p. 114509, 2024.

[17] S. Pehlivan, B. Akbugday, A. Akan, and R. Sadighzadeh, "Detection of olfactory stimulus from eeg signals for neuromarketing applications," in *2022 30th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2022, pp. 1–4.

[18] B. Akbugday, A. Akan, S. Pehlivan, and R. Sadighzadeh, "An assessment of linear and nonlinear features for detecting olfactory stimulus in eeg," in *2022 Medical Technologies Congress (TIPTEKNO)*. IEEE, 2022, pp. 1–4.

[19] B. Akbugday, S. Pehlivan Akbugday, R. Sadikzade, A. Akan, and S. Unal, "Detection of olfactory stimulus in electroencephalogram signals using machine and deep learning methods," *ELECTRICA*, 2024.

[20] N. Sharma and R. Verma, "Lightweight emotion recognition model using selected eeg channels for real-time applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–10, 2025.

[21] P. Rajput and A. Singh, "Multiscale entropy and power-based hybrid features for robust eeg emotion recognition," *Biomedical Signal Processing and Control*, vol. 85, p. 105000, 2025.

[22] S. P. Akbugday, O. K. Cura, B. Akbugday, and A. Akan, "Detection of alzheimer's dementia by using eeg feature maps and deep learning," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024.

[23] E. Sipahioglu, B. Akbugday, S. P. Akbugday, and A. Akan, "Investigating the effect of noise levels on mental tasks using artificial intelligence," in *2024 Medical Technologies Congress (TIPTEKNO)*, 2024.

[24] E. Yurdakul, E. Ergun, C. Yildiz, B. Akbugday, A. Akan, T. Sokucu, and S. Pehlivan, "Realtime determination of mood in classroom environment," in *2022 Medical Technologies Congress (TIPTEKNO)*, 2022.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Olfactory Emotion Recognition Using EEG Spectral Topographic Heatmaps and CNNs

Faezeh Yeganli, Aydin Akan*
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Izmir, TURKEY
{faezeh.yeganli, akan.aydin}@ieu.edu.tr

Riza Sadikzade
School of Applied Management Sciences
Izmir University of Economics
Izmir, TURKEY
riza.sadikzade@ieu.edu.tr

***Abstract*—Recognizing emotions evoked by olfactory stimuli using electroencephalogram (EEG) signals is a growing area of interest in affective computing and neuromarketing. This study proposes a lightweight framework that classifies EEG responses to scented and unscented conditions along arousal and pleasure dimensions using single-band spectral topographic heatmaps and a convolutional neural network (CNN). EEG signals were recorded from 57 participants and processed to generate frequency-specific scalp maps for delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma ($>$ 30 Hz) bands. These heatmaps were used as CNN inputs, achieving up to 98.39% accuracy for arousal (Theta band) and 96.55% for pleasure (Beta and Theta band). The results demonstrate that spectral-domain features alone provide highly discriminative information for olfactory emotion recognition, while reducing computational complexity compared to multi-domain approaches. This approach shows strong potential for real-time neuromarketing, affective computing, and brain–computer interface applications.**

***Keywords*—EEG signal, olfactory emotion recognition, spectral features, topographic heatmaps, deep learning**

I. Introduction

The investigation of emotional states via brain activity has gained significant attention in affective computing and consumer neuroscience. Because emotions shape decision-making, attention, and memory, accurate recognition offers clinical, commercial, and technological benefits [1, 2]. Although surveys and facial expressions remain common, they suffer from subjectivity and cultural bias; thus neurophysiological methods—especially electroencephalography (EEG)—are preferred for reliable, real-time insight into emotional processes [3, 4]. Compared to vision and hearing, olfaction exerts a uniquely strong influence on emotion due to its direct pathway to limbic and memory-associated structures [5, 6]. Consequently, olfactory-induced EEG studies have grown, particularly in consumer behavior and neuromarketing, where identifying odor-evoked emotional responses can guide product design and marketing strategies. Evidence shows that ambient scent substantially affects customer perception and behavioral responses in real-world settings (e.g., airport terminals, retail) [7, 8], and EEG studies demonstrate that scented products and store odors modulate both arousal and pleasure, providing a reliable neurophysiological indicator of emotional response [9, 10, 11].

Tracking brain function using the non-invasive, inexpensive, high- temporal- resolution method known as electroencephalography or EEG. Since this technology can track quick neural activity, it is perfect for watching brief emotional reactions to sensory input. Many recent studies have classified emotional states by extracting multi-domain EEG characteristics, using time-domain features such as variance and Hjorth parameters, frequency-domain traits such as spectral band powers and nonlinear features such as entropy and fractal dimensions [12, 13]. While Lanata et al. [5] accurately recognized pleasant against unpleasant aromas around 75% of the time using EEG characteristics, Becerra et al. [6], for instance, claimed that SVM-based odor pleasantness categorization had as much as 99.9% accuracy. The shear amount of features and preprocessing steps in these methods, however, sometimes demands computationally intensive pipelines, which might limit their real-time application in embedded or portable systems [14, 15].

Frequency-domain (spectral) characteristics as well as spatial topography are increasingly highlighted in EEG-based emotion recognition research. By examining the power spectral density (PSD) of canonical frequency bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma ($>$ 30 Hz), it is possible to capture the dominant neural oscillations related to particular cognitive and emotional processes. Moreover, projecting these spectral powers onto scalp topographic heatmaps enhances the interpretability of brain activation patterns and provides an effective input for machine learning and deep learning algorithms. Studies such as ERENet [12], DSSA-Net [13], and TSGCN [14] have demonstrated that spectral topography-based methods can achieve 94–96% accuracy on benchmark emotion datasets like DEAP and SEED, confirming the potential of spectral-focused approaches for emotion recognition.

Building on these insights, the present study extends the previous multi-domain approaches [15, 16] and focuses exclusively on spectral features to create a lightweight yet accurate classification framework. EEG signals recorded from

*This study was supported by Izmir University of Economics, Scientific Research Projects Coordination Unit. Project number: 2022-07.

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE

participants under scented and unscented conditions were transformed into frequency-specific topographic heatmaps, which were then classified using deep learning model. This approach reduces the feature set and computational load while maintaining high recognition performance. Our results show that spectral-only features can achieve up to 98.3% accuracy for Theta-band arousal classification and 96.55% for Beta and Theta band pleasure classification, which is comparable to or better than multi-domain methods reported in the literature.

The key contributions of this study are as follows:1- Generation of spectral EEG topographic heatmaps for olfactory emotion recognition. 2- Implementation of deep learning classifiers to distinguish scented versus unscented responses in the arousal and pleasure dimensions. 3- Demonstration that spectral-domain features alone provide high classification accuracy while offering computational efficiency compared to multi-domain approaches.

II. MATERIALS AND METHODS

The methodology of the study can be broken down into four main steps which are data acquisition, preprocessing, feature extraction, and classification.

A. Dataset and Experimental Setup

This study employed the same dataset and experimental protocol as our previously published work [16]. In brief, 57 healthy adult participants were recruited and exposed to two experimental conditions: scented and unscented.

After each exposure, participants completed a self-assessment survey to evaluate arousal and pleasure levels on a 5-point Likert scale, forming the two-dimensional arousal–pleasure plane. EEG recordings were performed using the Bitbrain Versatile-32 system, with 23 active electrodes selected based on the 10–20 system. Signals were sampled at 256 Hz in a controlled, quiet environment to minimize motion artifacts. All procedures were approved by the institutional ethics committee and informed consent was obtained from all participants. Self-reported arousal and pleasure ratings on the 5-point Likert scale were averaged across all participants and trials per condition. Scores above the median were labeled as 'high' (1), and those below were labeled as 'low' (0). This procedure yielded balanced binary labels for both arousal and pleasure dimensions.

B. EEG Preprocessing

The preprocessing pipeline followed our previous study [16] with adaptations for heatmap generation: 1- Bandpass filtering from 0.5–45 Hz to remove DC drift and high-frequency noise. 2- Independent Component Analysis (ICA) to identify and remove ocular and muscle artifacts. 3- Segmentation into one-second non-overlapping epochs, with five segments per condition per participant retained for further analysis. This procedure yielded clean artifact-free segments suitable for spectral analysis and image-based classification.

C. Spectral Feature Extraction and Heatmap Generation

For every EEG segment, Welch's approach, a strong method for examining non-stationary signals [2], was used to calculate Power Spectral Density (PSD). Five canonical EEG frequency bands were extracted: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–14 Hz), Beta (14–30 Hz), Gamma ($>$ 30 Hz). Using electrode locations, band-specific power levels were projected onto 2D scalp topographies to generate spectral heatmaps (Fig. 1). Each heatmap represents the spatial distribution of power in a single frequency band, emphasizing localized neural responses to olfactory stimuli.

D. CNN-Based Classification of EEG Heatmaps

The classification of olfactory EEG responses in this study was conducted using a convolutional neural network (CNN) implemented entirely in MATLAB R2023a. The workflow followed a supervised learning framework similar to our previous work [16], with modifications to incorporate spectral-domain topographic heatmaps as an input. Heatmap images generated from the spectral power of each EEG segment under scented and unscented conditions were loaded and labeled into two classes, resized to 128×128 pixels, and converted into single-precision RGB arrays. A total dataset of 570 (57 participants $\times$ 2 conditions $\times$ 5 segments) images (comprising both classes) was randomly partitioned into training (80%) and testing (20%) subsets using MATLAB's `cvpartition` function with a fixed random seed for reproducibility.

The CNN was designed as a compact binary classifier with three convolutional blocks (each containing a 3×3 convolution, batch normalization, ReLU activation, and 2×2 max-pooling), followed by a fully connected layer, softmax, and classification layer. In schematic form, the flow can be represented as: *Input heatmap* $\rightarrow$ *Conv1 (Conv–BN–ReLU–MaxPool)* $\rightarrow$ *Conv2 (Conv–BN–ReLU–MaxPool)* $\rightarrow$ *Conv3 (Conv–BN–ReLU–MaxPool)* $\rightarrow$ *FC* $\rightarrow$ *Softmax*.

The training process tracked accuracy and loss per iteration and epoch, with convergence confirmed by stable curves. The trained CNN classified the unseen test set, and performance was measured using accuracy (ACC), precision (PRE), recall (REC), and F1-score (F1), all derived from confusion matrix elements (TP, TN, FP, FN).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

Finally, iteration-wise and epoch-wise accuracy/loss trends were exported as CSV tables, and corresponding plots were generated to document training dynamics. This MATLAB-based end-to-end framework allowed reproducible training, testing, and evaluation of CNN classifiers for olfactory EEG heatmaps, consistent with the methodology validated in [16].

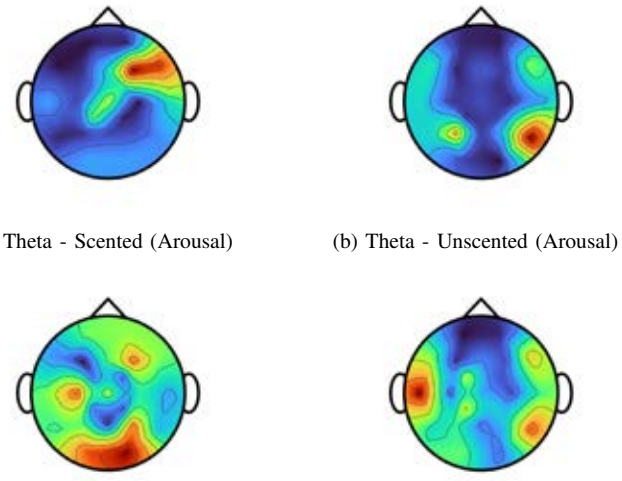
III. Results and Discussion

This section presents the classification results obtained using spectral topographic heatmaps as CNN inputs along with a discussion of model behavior, neurophysiological interpretations, and comparison with recent state-of-the-art methods.

A. Topographic Heatmap Analysis

Fig. 1 illustrates representative Theta-band heatmaps for both arousal and pleasure under scented and unscented conditions. The color distribution reflects the spectral power of Theta oscillations where red regions indicate higher activity and blue regions indicate lower activity.

Under scented conditions, Theta activity shows strong localized frontal–temporal increases—regions closely linked to limbic circuitry and olfactory emotional processing [3],[5]—whereas unscented trials exhibit weaker, more diffuse patterns, yielding clear heatmap separability that underpins the high CNN performance. Theta is emphasized in Fig. 1 because it produced the most discriminative *arousal* patterns (Table I); although both Theta and Beta achieved the best *pleasure* accuracies (Table II), Theta most clearly illustrates arousal separability.



(a) Theta - Scented (Arousal) (b) Theta - Unscented (Arousal)

(c) Theta - Scented (Pleasure) (d) Theta - Unscented (Pleasure)

Fig. 1: Examples of Theta-band EEG topographic heatmaps under scented and unscented conditions. Warmer colors correspond to higher spectral power, highlighting frontal and temporal activations during olfactory stimulation.

B. Classification Performance and Model Behavior

The CNN classifier achieved high recognition performance using single-band spectral heatmaps. The classification indicators for arousal and pleasure are presented in Tables I and II. For *arousal*, Theta dominated (98.39% accuracy, 96.15% precision, 100.00% recall, 98.04% F1), followed by Alpha (93.55%), while Beta, Delta, and Gamma were lower (88.71%, 85.48%, 80.65%). For *pleasure*, Theta and Beta both reached 96.55% with complementary error profiles (Theta: 100.00% precision, 91.67% recall; Beta: 92.31% precision, 100.00% recall), and Alpha remained strong at 94.83%. These outcomes align with Fig. 1, where Theta maps show clear frontal–temporal separability under olfactory stimulation.

TABLE I: Classification Performance for Arousal (Spectral Heatmaps)

Features	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Alpha	93.55	86.21	100.00	92.59
Beta	88.71	82.14	92.00	86.79
Delta	85.48	94.44	68.00	79.07
Gamma	80.65	88.24	60.00	71.43
Theta	**98.39**	**96.15**	**100.00**	**98.04**

TABLE II: Classification Performance for Pleasure (Spectral Heatmaps)

Features	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Alpha	94.83	95.65	91.67	93.62
Beta	**96.55**	92.31	100.00	96.00
Delta	89.66	84.62	91.67	88.00
Gamma	86.21	78.57	91.67	84.62
Theta	96.55	100.00	91.67	95.65

The confusion matrix for Theta-band arousal classification (Fig. 2) shows near-perfect separation, with only one misclassification. The CNN achieved 100% recall for scented and 97.3% for unscented confirming the reliability of Theta heatmaps.



Fig. 2: Confusion matrix for Theta-band arousal classification. The model achieved 98.39% accuracy with only one misclassified sample.

The training and validation loss curves in Fig. 3 indicate that the CNN converges rapidly, reaching near-zero loss within the first few epochs. Validation loss remains stable showing no signs of overfitting which is critical for small EEG datasets.

C. Neurophysiological Interpretation

The dominance of Theta in arousal recognition aligns with its role in limbic activity, attention, and emotional memory [3],[5]. Theta oscillations are highly responsive to olfactory stimulation, producing localized frontal-temporal activations that are observable in the heatmaps. The high Beta performance for pleasure is consistent with its role in active cognitive

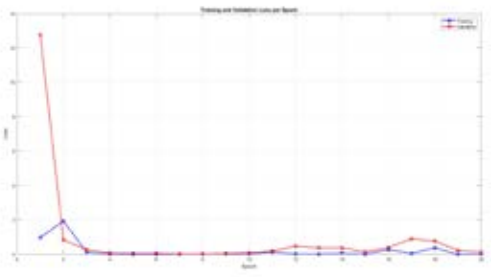


Fig. 3: Training and validation loss curves per epoch. The CNN shows fast convergence and stable generalization.

and sensory integration [1]. In contrast, Delta and Gamma contribute less due to their weaker involvement in short-latency olfactory processing.

D. Comparison with State-of-the-Art

Table III summarizes the performance of the proposed CNN alongside alternative lightweight CNNs evaluated on our dataset, and representative literature methods on different datasets. External approaches (ERENet [12], DSSA-Net [13], TSGCN [14]) were originally assessed on DEAP/SEED and are shown for context rather than direct comparison. The proposed CNN achieved the highest accuracy while using fewer parameters and converging faster.

TABLE III: Comparison of the proposed CNN and alternative CNNs on our dataset; literature methods on different datasets are shown for context only.

Method	Dataset / Features	Acc. (%)
Our dataset (57 subj., Spectral Heatmaps)		
Proposed CNN (Theta/Arousal)	Our dataset / Heatmaps	**98.39**
Proposed CNN (Beta/Pleasure)	Our dataset / Heatmaps	**96.55**
MobileNetV2	Our dataset / Heatmaps	95.2
SqueezeNet	Our dataset / Heatmaps	93.8
External literature (different datasets)		
ERENet [12]	DEAP / Topography	95
DSSA-Net [13]	SEED/DEAP / Spec.-Spat.	94–96
TSGCN [14]	SEED/DEAP / Temp.-Spec.	95–96

These findings confirm that spatially-informed Theta and Beta heatmaps, combined with a lightweight CNN, achieve or surpass state-of-the-art accuracy without complex multi-domain features or graph models, supporting their use in real-time olfactory emotion recognition and Brain–Computer Interface (BCI) applications.

IV. Conclusion

This study presented a lightweight framework for classifying olfactory-induced emotional responses using EEG spectral topographic heatmaps and convolutional neural networks. By focusing on single-band spectral information, the method avoids complex multi-domain feature extraction while maintaining high discriminative power.

Experiments showed that Theta was the most reliable band for arousal classification (98.39% accuracy, 98.04% F1), while both Theta and Beta were highly effective for pleasure recognition (96.55%). Alpha contributed moderately, whereas Delta and Gamma were less informative. Topographic heatmaps revealed clear differences between scented and unscented conditions, particularly in frontal and temporal regions linked to olfactory processing.

Compared with alternative lightweight CNNs on our dataset and state-of-the-art models such as ERENet [12], DSSA-Net [13], and TSGCN [14], the proposed CNN achieved similar or higher accuracy with a simpler architecture. Future work will focus on cross-subject generalization and temporal modeling to enable real-time applications in neuromarketing, affective computing, and brain–computer interface systems.

References

[1] S. D. Pollak, L. A. Camras, and P. M. Cole, "Progress in understanding the emergence of human emotion," *Developmental Psychology*, vol. 55, no. 9, pp. 1801–1811, 2019.
[2] X. Wang *et al.*, "Deep learning-based EEG emotion recognition: Current trends and future perspectives," *Frontiers in Psychology*, vol. 14, p. 1126994, 2023.
[3] H. Hou, X. Zhang, and Q. Meng, "Odor-induced emotion recognition based on average frequency band division of EEG signals," *Journal of Neuroscience Methods*, vol. 334, p. 108599, 2020.
[4] S. Pehlivan, B. Akbugday, A. Akan, and R. Sadighzadeh, "Detection of olfactory stimulus from EEG signals for neuromarketing applications," in *Proc. 30th SIU*, IEEE, 2022.
[5] A. Lanata *et al.*, "Automatic recognition of pleasant content of odours through EEG activity analysis," in *Proc. IEEE EMBC*, 2016, pp. 4519–4522.
[6] M. A. Becerra *et al.*, "Odor pleasantness classification from EEG signals and emotional states," in *Colombian Conference on Computing*, Springer, 2018.
[7] S. Yildirim, "The effect of ambient scent on customers' perception and behavioral responses: A field study in airport terminal," *International Journal of Marketing Studies*, vol. 9, no. 4, pp. 99–108, 2017.
[8] F. Basso, O. Droulers, T. Giraud, and C. Hillaire, "Odors in the store: EEG and self-reported measures of pleasantness and arousal," *Journal of Retailing and Consumer Services*, vol. 58, p. 102288, 2021.
[9] M. Seker and M. S. Ozerdem, "EEG-based emotional response to olfactory stimuli: A pilot study," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, no. 1, pp. 200–208, 2017.
[10] D. Gabriel, M. Grassi, and M. T. Arredondo, "Emotional evaluation of scented cosmetic products using EEG signals," *Frontiers in Psychology*, vol. 12, p. 675297, 2021.
[11] L. Wang, J. Huang, and X. Zhang, "Using EEG to examine emotional responses to advertisements: A meta-analysis," *Journal of Consumer Psychology*, vol. 32, no. 1, pp. 146–161, 2022.
[12] Y. Ren *et al.*, "ERENet: Multi-band EEG topology map learning for emotion recognition," *Applied Sciences*, vol. 12, no. 20, p. 10273, 2022.
[13] Z. Zhang *et al.*, "DSSA-Net: Directional spatial-spectral attention network for EEG-based emotion recognition," *Frontiers in Neurorobotics*, vol. 19, 2025.
[14] L. Li *et al.*, "Temporal-spectral graph convolutional networks for EEG emotion recognition," *Brain Informatics*, vol. 11, 2024.
[15] B. Akbugday, S. Pehlivan Akbugday, R. Sadikzade, A. Akan, and S. Unal, "Detection of Olfactory Stimulus in EEG Signals Using Machine and Deep Learning Methods," *Electrica*, vol. 24, no. 1, pp. 175–182, Jan. 2024.
[16] S. Pehlivan Akbugday *et al.*, "Decoding Olfactory EEG Signals Using Multi-Domain Features and Machine Learning," in *Proc. TIPTEKNO*, IEEE, 2024.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

A Multi-Chatbot Evaluation Framework for Knee MRI Diagnosis Assistance

Serhat Uzunbayir[1], Beyza Altuner[2], Deniz Eren Terziler[1], Cinar Gedizlioglu[2]
[1]*Software Engineering Department*
[2]*Computer Engineering Department*
Izmir University of Economics
Izmir, Türkiye

***Abstract*—Knee injuries, and in particular abnormalities of the Anterior Cruciate Ligament (ACL) and the meniscus, are diagnosed frequently using MRI scans. Although MRI interpretations typically require expert knowledge, that expertise may not always be accessible. Recently, researchers have begun using Large Language Models (LLMs) in the medical domain, applied to assist with diagnostic interpretative tasks. Here, we investigate the potential for LLM-based chatbots to assist and augment the reasoned diagnostic interpretation of knee MRI images. Specifically, we report our comparisons across chatbot diagnostics including ChatGPT-4o, Gemini 2.5 Flash, and Claude Sonnet 4, to see if they can annotate ACL injury, meniscal tear, and abnormality of any type. Using visual MRI slices as input, we evaluated the interpretations produced by multimodal capable chatbots against the ground truth data labelled by professional radiologists. Our findings illustrate each chatbot model's relative strengths and weaknesses in medical imaging analysis that contribute evidence towards supporting the development of AI-augmented workflows for medical imaging and radiology.**

***Keywords*—medical decision-making, ai, chatbot, healthcare, medical imaging, large language models.**

I. Introduction

Magnetic Resonance Imaging (MRI) is the gold standard diagnostic tool for soft tissue structures of the knee joint, including the Anterior Cruciate Ligament (ACL) and meniscus. Timely and accurate interpretations of knee MRI scans can lead to injuries that will limit somebody's ability to move, quality of life, and ultimately function of the joint in the long term. However, radiological resources may not always be available, particularly in low-resource contexts. This has created excitement to develop and implement Artificial Intelligence (AI) tools to assist with clinical decision making.

Large Language Models (LLMs) and multimodal AI chatbots have grown increasingly powerful over the past few years. They are able to process text and visual information simultaneously. Models like GPT-4o, Deepseek, and Kimi-K2 have demonstrated some success in various fields, including health care [1]. These technology have the potential to help augment diagnostic activities in radiology by interpreting medical images with clinical text; however, their performance has yet to be fully explored in narrower applications (e.g., musculoskeletal imaging).

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

The objective of this study is to assess and compare the performance of various leading AI chatbot models to interpret knee MRI images on three separate diagnostic criteria: the detection of general abnormalities, identification of ACL injuries, and identification of meniscal tears.

We will carefully examine how and if each chatbot interprets standardized MRI inputs, and will specifically look to answer two research questions:

> **RQ1:** Which AI chatbot provides the most accurate assessment of knee MRI images for detecting general abnormalities, ACL injuries, and meniscal tears?
> **RQ2:** How do AI chatbot diagnostics compare with models trained to handle MRI images?

We contribute a unified evaluation framework that feeds standardized knee MRI images into chatbot systems and benchmark their performance across three core diagnostic tasks. Our comparative analysis covers three state-of-the-art chatbots -ChatGPT-4o, Gemini 2.5 Flash, and Claude Sonnet 4- and a CNN baseline. The results reveal the current strengths and limitations of LLMs in clinical imaging tasks, along with broader implications for ethical deployment in real-world workflows.

II. Related Work

In this section we summarize the literature on artificial intelligence and its applications, particularly deep learning and large language models, in interpreting knee MRI images.

One popularly reported method has been using Convolutional Neural Networks (CNN) trained on knee MRI datasets, like the MRNet dataset [2], and annotated by expert radiologists [3]. These models can identify specific forms of abnormality like ACL tears and meniscal injuries fairly accurately and in some cases are even superior. CNNs are able to learn very specific, localized patterns and although are suitable to segmentation tasks like the one we discussed, CNNs are limited in the generalizability, interpretability, and contextual reasoning for specific clinical cases that constitute our gaps. This has led researchers to evaluate alternatives and more flexible forms of computational AI models.

LLMs increasingly being used in healthcare particularly destructive analysis, clinical decision support, and medical

imaging [4] [5]. With the necessary adjustments made for radiology, these foundations demonstrated versatility beyond natural language processing, with inclinations towards radiology report generation, protocol recommendation, and even multimodal image interpretation [4] [6]. In recent literature, vision-language foundation models have helped connect imaging findings and clinical context. This offers promise for improved levels of diagnostic reasoning and is an active area for future study [7].

For example, [6] assessed ChatGPT and similar models in image-based radiologic interpretation tasks, rating them on instruction-following ability, diagnostic relevance, and consistency. Their results show advantages for structured output generation, but also demonstrate lingering concerns like the frequent occurrence of hallucinations and a lack of robustness in novel situations. In related work, [8] evaluated a number of LLMs (Claude, Bing, and Perplexity) for radiologic decision-making for imaging pulmonary embolism, relative to the American College of Radiology Appropriateness Criteria.

Comparative analyses such as those in [9] and [10] show notable differences in LLM performance depending on task design, medical complexity, and contextual detail. While Claude and Gemini have shown strong contextual understanding, ChatGPT often excels in structured output and consistency. These underscore the need for targeted evaluation frameworks.

Furthermore, useful literature reviews have looked at larger implications of LLMs in clinical workflows. For instance, [11] reviews how LLMs integrated into workflow in radiology, identifying advantages of images LLMs including automated structured reporting and potential workflow efficiencies. In a similar vein, [12] survey a scalable data-driven approaches, including low-rank approximations, to improve medical imaging pipeline efficiency.

Even with these advancements, there have been limited, comparisons unique to multimodal LLMs for interpreting MRI images, and knee injuries or conditions like ACL and meniscus tears. There have been prior evaluations of multimodal LLMs, but they were limited primarily to chest images, or general radiology tasks [4] [7]. Therefore, the number of studies related to applications associated with musculoskeletal imaging is limited. This research provides related contribution by evaluating three LLMs with high-quality visual input capabilities, ChatGPT-4o, Gemini 2.5 Flash, and Claude Sonnet 4, and their ability to examine knee MRI images and demonstrate abnormalities. The findings provide insights into their interpretive accuracy, diagnostic confidence, and overall feasibility for integration into AI-augmented radiology workflows.

III. Methodology

In this section, we outline the methodology that we developed to assess the performance of different multimodal chatbot models on knee MRI interpretations.

A. Dataset

We used the publicly available **MRNet dataset** by Stanford University, which is comprised of knee MRI scans obtained from clinical exams. The dataset contains **1,370 MRI studies** from **1,088 patients** and each study consists of three standard MRI planes: *sagittal*, *coronal*, and *axial*. Each exam is annotated with three binary labels indicating if there was a title presence or absence of ACL tears, meniscal tears, and general abnormalities.

Selectively, we chose a representative subset of 100 studies from the MRNet dataset, with an equal split between positive and negative cases across the three diagnostic categories, and to prepare the data for chatbot evaluation, we aggregated all slices from each anatomical plane (sagittal, coronal, axial) per study. Thus, for each case we created three composite images and a composite image for each plane. Each of these composite images were utilized as the input.

B. Baseline Model for Comparison

To provide a basis for comparison with the performance of chatbots empowered by LLM, we constructed a baseline Deep Learning model using a Convolutional Neural Network (CNN). The architecture for the baseline followed the MRNet approach, consisting of 3 separate 2D CNNs which are trained separately on each MRI plane and then cohort late-fusion model for the final prediction across the three target labels.

Each per-plane network comprises five convolutional blocks with increasing channel dimensions ($8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128$). All convolutional layers use a kernel size of 3×3, except the first layer which uses a 5×5 kernel.

In each block, we applied ReLU activation functions and 2×2 max-pooling, which was ultimately followed by an adaptive max-pooling that transformed the output to a 1×1 spatial map. A final fully connected layer was used to map the resulting 128-dimensional feature vector to 3 binary outputs corresponding to our diagnostic targets; general abnormality, ACL tear, and meniscal tear.

The CNN model was trained using training split of the dataset. Performance was evaluated using the same diagnostic metrics and samples as the chatbot models.

C. Chatbot Models

We examined the diagnostic performance of three current, general-purpose multimodal chatbot models: ChatGPT-4o (OpenAI) [13], Gemini 2.5 Flash (Google DeepMind) [14], and Claude Sonnet 4 (Anthropic) [15]. These models were selected based on ease of access, recent launch into general circulation and broad public accessibility. Each is capable of processing both images and text, making them suitable candidates for evaluating performance on medical imaging tasks without domain-specific fine-tuning. Our aim was to explore how general-purpose multimodal models, without domain-specific medical fine-tuning, perform on specialized tasks such as knee MRI interpretation.

Each model was provided with three composite images per case patient case. The models were prompted to deliver diagnostic predictions for **ACL tears**, **meniscal tears**, and **general abnormalities**. In addition to the predicted diagnoses, we

requested **textual explanations** supporting their conclusions. This protocol enables a multi-dimensional assessment of each model's diagnostic capability, interpretability, and reliability in the medical imaging domain.

D. Prompt Engineering and Input Formatting

To standardize the evaluation across models, we applied a uniform prompt format that includes a brief clinical context and a directive to provide a diagnosis. Visual inputs were uploaded to the chatbot interfaces where image capabilities were available. For each case, the prompt structure was as follows:

> **You are a radiologist specializing in musculoskeletal MRI interpretation.** Here are three MRI slices of a human knee in different planes: *Sagittal view*, *Coronal view*, *Axial view*. Please analyze these images and respond to the following clinical classification questions. Use the exact format shown below.
>
> 1) Do you see any abnormality overall? Answer yes/no and explain. What abnormalities (if any) are visible?
> 2) Do you see any signs of an ACL tear? Answer yes/no and explain.
> 3) Do you see any signs of a meniscal tear? Answer yes/no and explain.
>
> **Example Format:**
>
> - Abnormal: ___
> - ACL Tear: ___
> - Meniscal Tear: ___
>
> 4) Briefly summarize the imaging features that support your conclusions.

This structure ensured consistency in model queries, minimized ambiguity, and allowed fair comparative analysis.

E. Evaluation Protocol

We developed an comprehensive **evaluation protocol** for assessing comprehensive chatbot model performance, which will consider:

- **Diagnostic Accuracy:** Comparison of chatbot output against the ground truth labelled outputs for ACL, meniscus, and aberration detection which will include performance metrics including *precision*, *recall*, *F1-score*, and *accuracy*.
- **Comparative Analysis:** A cross-model analysis to find which chatbot represents the best trade-off between diagnostic accuracy, reliability of confidence, and quality of explanation.

IV. Results

For our analysis of diagnostic performance, we examined the mean classification scores for the three tasks: ACL tear, meniscal tear, and general abnormality detection. The comparisons were between the CNN model and the three AI chatbots: ChatGPT-4o, Gemini 2.5 Flash, Claude Sonnet 4 using independent-samples t-tests. Reported p-values reflect two-tailed significance testing with a threshold of $p < 0.05$.

A. Total Diagnostic Performance

The CNN model achieved the highest overall diagnostic performance ($M = 2.39$, $SD = 0.57$). ChatGPT-4o ($M = 1.71$, $SD = 0.92$; $p < 0.001$), Gemini ($M = 1.55$, $SD = 0.89$; $p < 0.001$), and Claude ($M = 1.39$, $SD = 0.91$; $p < 0.001$) all performed significantly worse than the CNN.

There was no statistically significant difference between ChatGPT-4o and Gemini or between Gemini and Claude. However, ChatGPT-4o significantly outperformed Claude ($p = 0.014$).

Table 1 shows the performance metrics supported these findings. The CNN achieved the highest overall accuracy (0.80), recall (0.78), precision (0.77), and F1-score (0.77). ChatGPT-4o followed by balanced recall (0.79) and F1 (0.62) performed better than Gemini (F1 = 0.54) and Claude (F1 = 0.59).

B. ACL Tear Detection

The CNN again led in ACL detection performance ($M = 0.87$, $SD = 0.34$). ChatGPT-4o ($M = 0.48$, $SD = 0.50$; $p < 0.001$), Gemini ($M = 0.48$, $SD = 0.50$; $p < 0.001$), and Claude ($M = 0.25$, $SD = 0.43$; $p < 0.001$) all performed significantly worse than the CNN.

In terms of inter-LLM analysis, both ChatGPT-4o and Gemini significantly outperformed Claude ($p < 0.001$).

In terms of metrics, Claude had higher recall (0.92) than Gemini (0.67) and ChatGPT-4o (0.75), but lower precision (GPT: 0.16; Gemini: 0.14; Claude: 0.13) which resulted in comparatively low F1-scores (GPT: 0.26; Gemini: 0.24; Claude: 0.23). The CNN had the worst recall (0.08) but the best precision (0.33).

C. Meniscal Tear Detection

The CNN once more produced performance metrics with F1 = 0.61. GPT (F1 =0.60) and Claude (F1=0.56) performed closely behind while Gemini slightly lagged behind (F1=0.51). Claude had the highest recall (0.86) but the lowest precision (0.41).

D. General Abnormality Detection

For abnormality detection, the CNN performed best ($M = 0.83$, $SD = 0.38$). GPT ($M = 0.69$, $SD = 0.47$; $p = 0.02$), Gemini ($M = 0.57$, $SD = 0.50$; $p < 0.001$), and Claude ($M = 0.71$, $SD = 0.46$; $p = 0.044$) all scored lower.

Differences between GPT and other chatbots were not statistically significant: GPT vs. Gemini ($p = 0.08$), GPT vs. Claude ($p = 0.76$). However, Claude underperformed Gemini ($p = 0.039$).

The F1-scores for Claude were highest (0.83), followed closely by GPT (0.80) and then Gemini (0.70). The best F1 overall was the CNN (0.90). GPT had good balance across recall (0.78) and precision (0.83). Both Gemini and Claude had slightly higher precision (0.79), but Claude had the best recall (Claude = 0.86 and Gemini = 0.63).

TABLE I: PERFORMANCE COMPARISON OF MODELS ACROSS ACL, MENISCUS, AND ABNORMAL CATEGORIES

Model	ACL				MENISCUS				ABNORMAL				AVERAGE			
	Acc	Recall	Precision	F1	Acc	Recall	Precision	F1	Acc	Recall	Precision	F1	Acc	Recall	Precision	F1
ChatGPT-4o	0.48	0.75	0.16	0.26	0.54	0.83	0.47	0.60	0.69	0.78	0.83	0.80	0.57	0.79	0.51	0.62
Gemini 2.5 Flash	0.48	0.67	0.14	0.24	0.50	0.62	0.43	0.51	0.57	0.63	0.79	0.70	0.52	0.63	0.47	0.54
Claude Sonnet 4	0.25	0.92	0.13	0.23	0.43	0.86	0.41	0.56	0.71	0.86	0.79	0.83	0.46	0.87	0.45	0.59
CNN Model	0.87	0.08	0.33	0.13	0.69	0.57	0.65	0.61	0.83	0.99	0.83	0.90	0.80	0.78	0.77	0.77

V. Discussion

This research investigated the differential diagnostic abilities of ChatGPT-4o, Gemini 2.5 Flash, and Claude Sonnet 4 vs a CNN which was trained on knee MRI data. The evaluation focused on their performance across multiple diagnostic tasks involving musculoskeletal imaging. The findings answer two previously asked **RQs**:

- **Answer to RQ1:** ChatGPT-4o showed the overall best consistent performance among the LLMs.
- **Answer to RQ2:** All of the chatbots performed at a significantly lower level than the CNN for all tasks.

ChatGPT-4o demonstrated the most consistent and higher performance across recall and precision particularly with general abnormality detection and therefore, it may be the most reliable LLM for general screening purposes.

Claude's high recall in the ACL and meniscus tasks indicates it takes an aggressive positive classification stance, but its low precision must be qualified by the potential for high false positivity rates. This may be useful for triage systems potentially but would need to be followed by human verification to reduce the risk of overdiagnosis.

Currently, as found in the LMArena [1] vision benchmark results, GPT has a better understanding of vision than the Gemini 2.5 Flash and Claude Sonnet 4 models. The results of our findings are consistent with the benchmark results. This alignment reinforces the credibility of both the benchmark and our experimental methodology.

The results also indicate that LLMs are not yet reliable as primary diagnostic systems but may assist in supportive or preliminary roles in clinical workflows. Furthermore, systems tailored to processing MRI images appear to offer clear advantages over general-purpose LLMs.

The use of AI chatbots in clinical care can speed up and standardize diagnoses, but it also brings ethical risks. The biased training data can yield treatment bias and a loss of patient trust, and many AI models have "black-box" natures that obscure how decisions are reached making it difficult to assign accountability or gain a genuinely informed consent [10]. To address this, we must mandate that (1) data sets were diverse and fair, (2) models were assessed for bias and accuracy before use, and (3) AI approaches produce documents demonstrating how decisions are made, i.e., provide an explanation. In public facing examples, the MRNet dataset provides decomposed MRIs for knee images, 1,370 anonymous knees from Stanford each with available for research, which allows researchers compare models simultaneously and transparently. With solid data rules and respect for intellectual property, we can use AI in healthcare in a fair and clear way.

VI. Conclusion

LLMs offer a flexible, multimodal interface and language reasoning capabilities. GPT's balanced metrics highlight their utility in broad diagnostic screening. Nevertheless, LLMs still exhibit sensitivity to calibration and hallucinations. For example, although the Claude had high legibility control (ACL) recall, it exhibited low precision resulting in a high false positive rate which may be problematic in clinical workflows.

In sum, LLMs may be useful in a triage, education or decision support role - particularly in resource poor or volume rich environments - but should be applied with caution.

References

[1] LM Arena, "Leaderboard overview." https://lmarena.ai/leaderboard/, 2025. [Online; accessed 4-August-2025].
[2] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, B. N. Patel, K. W. Yeom, *et al.*, "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet," *PLoS medicine*, vol. 15, no. 11, 2018.
[3] A. Botnari, M. Kadar, and J. M. Patrascu, "A comprehensive evaluation of deep learning models on knee mris for the diagnosis and classification of meniscal tears: A systematic review and meta-analysis," *Diagnostics (Basel)*, vol. 14, p. 1090, May 2024.
[4] Z. Liu, T. Zhong, and et al., "Radllm: A comprehensive healthcare benchmark of large language models for radiology," 2023.
[5] L. Pan, Z. Zhao, Y. Lu, K. Tang, L. Fu, Q. Liang, and S. Peng, "Opportunities and challenges in the application of large artificial intelligence models in radiology," *Meta-Radiology*, vol. 2, no. 2, p. 100080, 2024.
[6] T. A. D'Antonoli, A. Stanzione, C. Bluethgen, F. Vernuccio, L. Ugga, M. E. Klontzas, R. Cuocolo, R. Cannella, and B. Koçak, "Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions," *Diagnostic and Interventional Radiology*, vol. 30, no. 2, p. 80, 2024.
[7] S. Lee, W. J. Kim, J. Chang, and J. C. Ye, "Llm-cxr: instruction-finetuned llm for cxr image understanding and generation," *arXiv preprint arXiv:2305.11490*, 2023.
[8] V. K. Uppalapati and D. S. Nag, "A comparative analysis of ai models in complex medical decision-making scenarios: evaluating chatgpt, claude ai, bard, and perplexity," *Cureus*, vol. 16, no. 1, 2024.
[9] P. K. Sarangi, S. Datta, M. S. Swarup, S. Panda, D. S. K. Nayak, A. Malik, A. Datta, and H. Mondal, "Radiologic decision-making for imaging in pulmonary embolism: accuracy and reliability of large language models—bing, claude, chatgpt, and perplexity," *Indian Journal of Radiology and Imaging*, vol. 34, no. 04, pp. 653–660, 2024.
[10] S. Srivastav, R. Chandrakar, S. Gupta, V. Babhulkar, S. Agrawal, A. Jaiswal, R. Prasad, M. B. Wanjari, V. P. Babhulkar, S. Agarwal, *et al.*, "Chatgpt in radiology: the advantages and limitations of artificial intelligence for medical imaging diagnosis," *Cureus*, vol. 15, no. 7, 2023.
[11] Z. Liu, T. Zhong, and et al., "Evaluating large language models for radiology natural language processing," 2023.
[12] S. Hamlomo, M. Atemkeng, Y. Brima, C. Nunhokee, and J. Baxter, "A systematic review of low-rank and local low-rank matrix approximation in big data medical imaging," *Neural Computing and Applications*, pp. 1–56, 2025.
[13] OpenAI, J. Achiam, and et al., "Gpt-4 technical report," 2024.
[14] G. Comanici, E. Bieber, and et al., "Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities," 2025.
[15] Anthropic, "Introducing claude 4." https://www.anthropic.com/news/claude-4, May 2025.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Mekânsal Kısıtlar Altında Hodgkin–Huxley Nöron Modelinin İyon Bağımlı Davranışı
Ion-Dependent Behavior of the Hodgkin–Huxley Neuron Model under Spatial Constraints

Zübeyir Özcan
Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Türkiye
Emails: zubeyirozcan@ktu.edu.tr

***Özetçe* —Bu çalışmada, Hodgkin–Huxley nöron modelinin mekânsal varsayımları ve bu varsayımların model üzerindeki kısıtlayıcı etkileri incelenmiştir. Model, hücre içi ve dışı iyon derişimlerinin zamana bağlı olarak değişimini hesaba katacak şekilde yeniden yapılandırılmıştır. Farklı hücre yarıçaplarına sahip yapılarda gerçekleştirilen simülasyonlarla, klasik Hodgkin–Huxley modelinin geçerliliği sorgulanmış ve geometrik sınırlamaların zar potansiyeli üzerindeki etkileri değerlendirilmiştir. Ayrıca, nörona uygulanan iyon bazlı dış akımların (Na^+, K^+, Cl^-) zar dinamiklerini nasıl etkilediği araştırılmıştır. Farklı iyon türlerine özgü akım kenetleme uygulamaları karşılaştırmalı olarak ele alınmış ve her birinin nöral dinamiklerde ayırt edici etkileri ortaya konmuştur. Elde edilen bulgular, küçük hacimli yapılar için Hodgkin–Huxley modelinin sınırlılıklarını ortaya koymakta ve iyon kaynaklı uyarımların biyofiziksel etkilerine dair önemli sonuçlar sunmaktadır.**

Anahtar Kelimeler*—*Hodgkin–Huxley modeli, biyofiziksel modelleme, iyon derişimi, zar potansiyeli

***Abstract*—In this study, the spatial assumptions of the Hodgkin–Huxley neuron model and their limiting effects on the model were examined. The model was reconstructed to incorporate temporal changes in intra- and extracellular ion concentrations. Through simulations conducted on structures with different cell radii, the validity of the classical Hodgkin–Huxley model was questioned, and the effects of spatial constraints on membrane potential were evaluated. Furthermore, the influence of ion-based external currents (Na^+, K^+, Cl^-) applied to the neuron on membrane dynamics was investigated. Current-clamp applications specific to different ion types were comparatively analyzed, and the distinctive effects of each on neural dynamics were identified. The findings highlight the limitations of the Hodgkin–Huxley model for small-volume structures and provide significant insights into the biophysical effects of ion-based stimulations.**

Keywords*—*Hodgkin–Huxley model; biophysical modeling; ion concentration; membrane potential

I. GİRİŞ

Hodgkin–Huxley (HH) modeli, hesaplamalı sinirbilimde temel modellerden biri olmakla beraber birçok yapay ve biyolojik nöron modelinin referansı konumundadır [1]. Günümüzde hem tekil nöronlar hem de nöronal ağların modellenmesinde yaygın bir şekilde kullanılmaktadır. Ancak modelin temel sınırlaması, hücre içi ve dışı iyon derişimlerinin zamana bağlı olarak sabit kaldığı varsayımıdır. Büyük hacimli hücrelerde bu kabul makul olsa da küçük hücreler veya sınırlı hacimli yapılarda zar akımları bu derişimleri anlamlı düzeyde değiştirme potansiyeline sahiptir. Bu durum, klasik HH modelinde biyofiziksel geçerliliğin sorgulanmasına neden olmaktadır [2].

Klasik HH modelinde, hücre içi ve dışı iyon derişimlerinin sabit olduğu kabulü, büyük hacimli hücreler veya açık sistemler için geçerli olsa da, küçük hacimli yapılar veya yüksek yüzey alanı/hacim oranına sahip sistemlerde geçerliliğini yitirmektedir [3]. Bu tür yapılarda zar üzerinden geçen iyon akımları, kısa sürede derişim değişimlerine yol açarak zar potansiyelinde geri beslemeli sapmalara neden olabilmektedir [4]. Ancak literatürde bu tür geometrik sınırlamalar altında HH modelinin geçerliliği nadiren sorgulanmış; iyon geri beslemesinin etkisi çoğu zaman ihmal edilmiştir. Bu nedenle, söz konusu varsayımın sınırlarının belirlenmesiyle literatürdeki bu boşluğun doldurulması mümkündür.

Öte yandan klasik HH modelinde, kullanılan akım kaynakları genelde idealize edilerek iyon türünden bağımsız "genel enjektör" gibi modellenmektedir. Ancak gerçek hayatta uygulanan akımın hangi iyon türü aracılığıyla taşındığı zar potansiyeli ve kanal dinamikleri üzerinde doğrudan belirleyicidir. Her iyon türünün denge potansiyeli, kanal dinamikleri ve zar tepkisi farklıdır. Bu doğrultuda, gerilime duyarlı iyon kanallarının iyon türüne özgü davranışlarının, nöronal ateşleme özelliklerini belirgin biçimde etkilediği literatürde gösterilmiştir [5]. Literatürde, iyon türüne özgü akım enjeksiyonlarının spesifik etkileri hâlâ yeterince detaylı şekilde analiz edilmemiştir. Bu boşluk, HH modelinin fiziksel olarak belirli koşullar altında yeniden değerlendirilmesi ve genişletilmesini gerektirmektedir.

Bu çalışma, klasik HH modelinin geçerlilik sınırlarını ortaya koyarak, özellikle mikro ölçekte iyon dinamiklerinin ihmal edilmesinin neden olabileceği model hatalarına dikkat çekmektedir. Literatürde genellikle sabit iyon derişimi varsayımı altında sürdürülen modellemelere karşılık, bu çalışmada derişimlerin zamanla değişebileceği ve bunun elektriksel aktiviteyi doğrudan etkileyebileceği gösterilmiştir. Özellikle hücre geometrisinin iyon akımları üzerindeki etkisi analiz edilmiştir. Ayrıca, uygulanan akımın taşıdığı iyon türüne bağlı olarak nöronal yanıtın farklılık göstermesi, klasik yaklaşımlardan ayrılan özgün bir bakış açısı sunmaktadır.

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

II. YÖNTEM

A. Hodgkin–Huxley Nöron Modeli

HH nöron modelinde, bir nöron zarındaki elektriksel dinamikler açıklanırken, zar potansiyelinin zamana bağlı değişimi ve iyon akımları göz önünde bulundurulmaktadır. Modelin çıkış noktası, zarın kapasitif davranışı ile iyon kanallarından geçen akımlar arasındaki dengedir. Bu ilişki Denklem (1)'de özetlenmiştir. Burada; C_{zar} zar kapasitansını, φ zar potansiyelini, I_{ext} dışarıdan uygulanan akımı ve I_i ise "i" iyonunun toplam kanal akımını temsil etmektedir.

$$C_{\text{zar}} \frac{d\varphi}{dt} = I_{\text{ext}} - \sum_i I_i \tag{1}$$

$$I_i = g_i(\varphi, t)\left(\varphi - E_i\right) \tag{2}$$

$$E_i = \frac{RT}{z_i F} \ln\left(\frac{[X]_{\text{dış}}}{[X]_{\text{iç}}}\right) \tag{3}$$

$$g_i(\varphi, t) = \bar{g}_i \, m^p(\varphi, t) \, n^q(\varphi, t) \, h^r(\varphi, t) \tag{4}$$

Her bir iyon için akım, ilgili kanalın iletkenliği ve potansiyel farkına bağlı olarak Denklem (2)'de ifade edilmektedir. Buradaki $g_i(\varphi, t)$, kanalın zaman ve gerilim bağımlı iletkenliğini; E_i ise iyonun denge potansiyelini göstermektedir. E_i, iyonun hücre içi ve dışı derişimlerinin göz önünde bulundurulduğu ve Denklem (3) ile verilen Nernst denklemiyle elde edilmektedir. Kanal iletkenliği ise Denklem (4)'te görüldüğü üzere gerilime duyarlı kapı değişkenleri ile tanımlanmaktadır. Burada $\bar{g}_i$, maksimum kanal iletkenliğini; m, n, h ise iyon kanalının açık/kapalı oranlarını temsil eden kapı değişkenleridir.

B. İyon Hareketleri

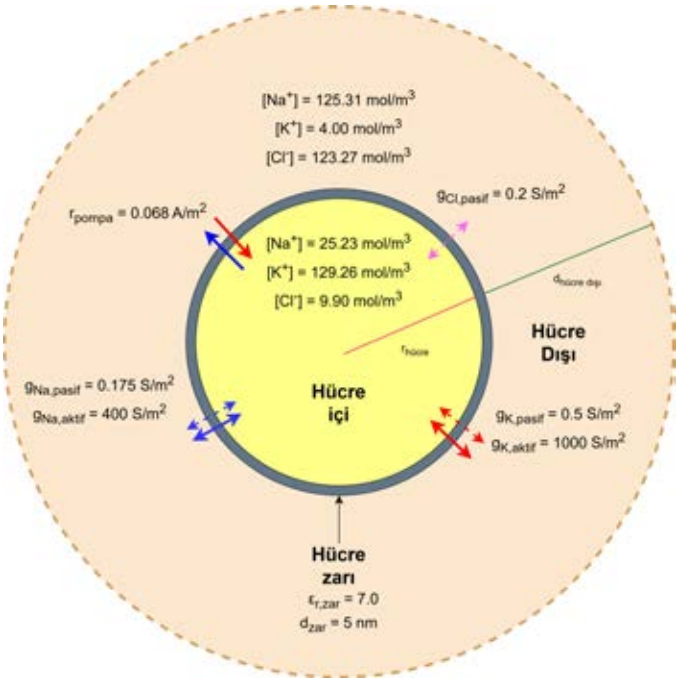
HH nöron modelinde temsil edilen iyon akımları üzerinden zar üzerinden akan iyon akısının belirlenmesi mümkündür. Buna göre, Denklem (5), birim zamanda birim alandan geçmesi beklenen iyon miktarının mol cinsinden hesabını göstermektedir.

$$J_i = \frac{I_i}{z_i F} \tag{5}$$

$$\frac{d[X]_{\text{iç}}}{dt} = -\frac{A_{\text{zar}}}{V_{\text{iç}}} J_i \tag{6}$$

$$\frac{d[X]_{\text{dış}}}{dt} = \frac{A_{\text{zar}}}{V_{\text{dış}}} J_i \tag{7}$$

Bu iyon akısı ile, hücre içi ve dışı hacimlerdeki derişim değişimleri belirlenmektedir. Hücre içi ve dışı derişimleri için zar akısının etkisi Denklem (6) ve Denklem (7)'de verildiği gibidir. Burada; A_{zar}, hücre zarı alanını; $V_{\text{iç}}$ ve $V_{\text{dış}}$ ise sırasıyla hücre içi ve dışı sıvı hacimlerini temsil etmektedir. Bu denklemler, iyon derişimi temelli geri besleme mekanizmalarının simülasyonunu mümkün kılmaktadır.



Şekil 1: Geometrik tasarım ve kullanılan model parametreleri

C. Model Varsayımları

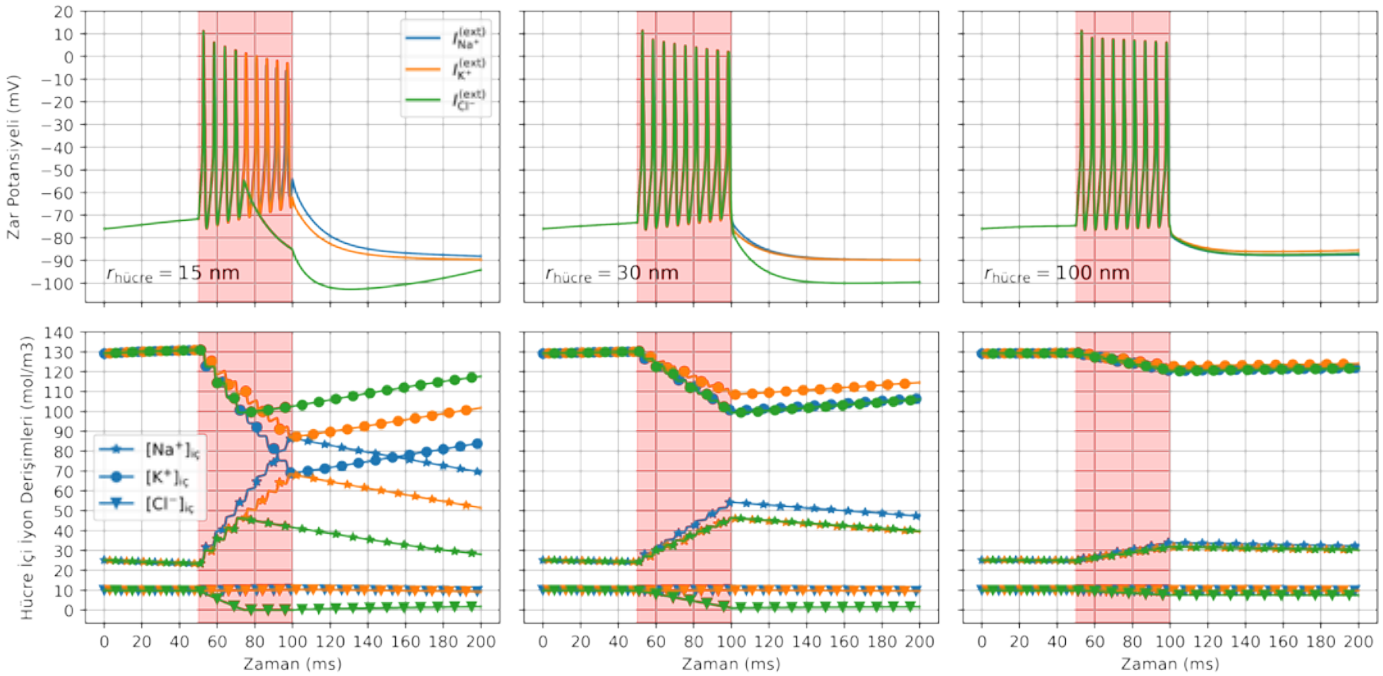
Klasik HH modelinde, iyon akışlarının hücre içi ve dışı iyon derişimlerinde herhangi bir değişime neden olmadığı varsayılmaktadır. Bu çerçevede, Nernst denklemi ile hesaplanması beklenen denge potansiyelleri sabit kabul edilmekte ve zamana bağlı olarak güncellenmemektedir. Benzer şekilde, zar üzerinden geçen iyon akılarının, hücre içi ve dışındaki iyon derişimlerini etkilemediği varsayılmaktadır. Bu yaklaşım, Denklem (6) ve Denklem (7)'den de anlaşılacağı üzere yalnızca yüzey alanı/hacim oranının çok düşük olduğu sistemler için geçerlidir.

Ancak, gerçek nöronların özellikle ince ve uzun dendritlerinde, yüzey alanı/hacim oranı bu varsayımları geçersiz kılabilecek kadar büyük olabilmektedir. Bu durumda, sabit derişim varsayımı modelin biyofiziksel doğruluğunu önemli ölçüde azaltmakta ve elde edilen sonuçların geçerliliğini sınırlamaktadır.

D. Çalışma Modeli

Bu çalışmada kullanılan model, klasik HH yaklaşımını genişleterek zamanla değişen iyon derişimlerini ve aktif/pasif iyon taşıyıcılarını dikkate almaktadır. Modelin geometrik ve biyofiziksel özellikleri Şekil 1'de sunulmuştur. Buna göre, küresel bir nöron yapısı ile onu çevreleyen küre kabuğu biçimindeki hücre dışı ortam arasında, Hübel ve Dahlem'in [4] çalışmasında ayrıntılı olarak tanımlanan iyon mekanizmalarına dayalı bir hücre zarı modeli oluşturulmuştur. Bu doğrultuda, modelde hücre zarında Na^+ ve K^+ iyonları için hem aktif hem de pasif iyon kanalları; Cl^- iyonu için ise yalnızca pasif iyon kanalları yer almaktadır.

İyon kanallarına ek olarak, zar üzerinde homeostatik dengeyi sağlamak amacıyla sodyum–potasyum ATPaz pompası



Şekil 2: Hücre yarıçapı ve iyon akımlarına göre zar potansiyeli ve hücre içi derişim değişimi

da modellenmiştir. Bu pompa, hücre içi ve dışındaki Na^+ ve K^+ derişimlerini izleyerek gerektiğinde her üç Na^+ iyonunu dış ortama çıkarırken, iki K^+ iyonunu hücre içine almaktadır. Pompanın çalışması, modelde tanımlanan bir pompa parametresi aracılığıyla kontrol edilmektedir.

Modelde yer alan bu iyon kanalları ve ATPaz pompası aracılığıyla zar üzerinden geçen iyon akıları, her iyon türü için Denklem (8), Denklem (9) ve Denklem (10)'da gösterildiği üzere ayrı ayrı hesaplanmaktadır. Bu hesaplamalarda, toplam iyon akısı yalnızca pasif ve aktif kanal akımlarını değil, aynı zamanda ATPaz pompasının katkısını ve dışarıdan uygulanan iyon bazlı uyarımları da içermektedir.

$$J_{Na^+} = \frac{I^{(aktif)}_{Na^+} + I^{(pasif)}_{Na^+} + 3I_p - I^{(ext)}_{Na^+}}{F z_{Na^+}} \tag{8}$$

$$J_{K^+} = \frac{I^{(aktif)}_{K^+} + I^{(pasif)}_{K^+} - 2I_p - I^{(ext)}_{K^+}}{F z_{K^+}} \tag{9}$$

$$J_{Cl^-} = \frac{I^{(pasif)}_{Cl^-} - I^{(ext)}_{Cl^-}}{F z_{Cl^-}} \tag{10}$$

Modelde, hesaplanan iyon akıları hücre içi derişimlerin güncellenmesinde doğrudan kullanılmaktadır. Ancak herhangi bir iyon türü için bu güncelleme sonucunda derişimin fiziksel olarak negatif bir değere düşmesi söz konusuysa, ilgili iyon akısı sınırlandırılarak bu durum önlenmiştir. Bu sayede modelin biyofiziksel tutarlılığı korunmuş ve simülasyonun gerçekçiliği artırılmıştır.

III. Bulgular ve Tartışma

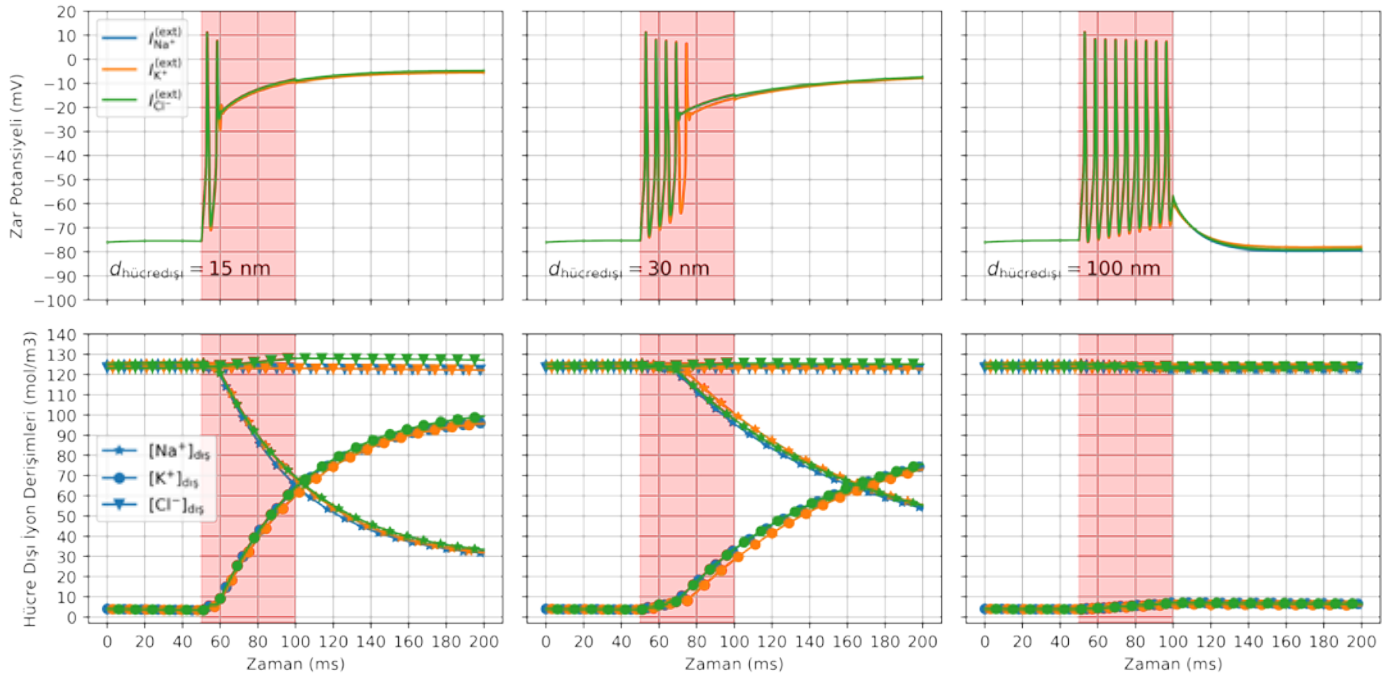
A. Hücre Boyutlarının Etkisi

Şekil 2'de, yarıçapları 15 nm, 30 nm ve 100 nm olan hücreler için zar potansiyeli ve hücre içi iyon derişimlerinin zamana bağlı değişimi gösterilmiştir. Hücre dışı derinlik 1 µm, akım uyarımları ise 50–100 ms aralığında her iyon için ayrı ayrı 150 A/m^2 olarak uygulanmıştır.

Şekil 2, hücre boyutu küçüldükçe iyon akımlarına karşı hücre içi derişimlerin duyarlılığının arttığını göstermektedir. Özellikle 15 nm yarıçaplı hücrede, uyarım anında iyon derişimleri büyük değişim göstermiş; Cl^- tabanlı akımlarda ise Cl^- derişimi tükenerek minimum seviyede kalmıştır. Daha büyük hücrelerde bu etki azalarak sürmüştür.

Hücre içi derişimlerdeki yarıçapa bağlı değişim, zar potansiyellerini de önemli ölçüde etkilemiştir. 100 nm yarıçaplı hücrede, iyon derişimleri büyük oranda sabit kaldığı için zar potansiyeli klasik HH modeline benzer bir seyir izlemiş ve akım uyarımının iyon bağımlılığına duyarlılığı azalmıştır. Buna karşılık, daha küçük yarıçaplarda iyon derişimlerindeki belirgin değişiklikler, denge potansiyellerinin kaymasına ve dolayısıyla zar potansiyelinde ciddi sapmalara yol açmıştır. Uyarımın başlangıcında tüm hücrelerde benzer zar potansiyelleri gözlenirken, zamanla iyon derişimlerinin farklılaşmasıyla zar potansiyelleri de birbirinden ayrılmıştır.

Akım uyarımının iyon türüne bağlı etkileri, öncelikle iyon derişimlerinde farklılık yaratmış, ardından bu fark zar potansiyelinin iyonlara özgü duyarlılığını artırmıştır. Örneğin, 15 nm yarıçaplı hücrede Cl^- derişimi minimuma indiğinde aksiyon potansiyeli üretimi durmuştur. Yarıçap büyüdükçe bu etkinin aksiyon potansiyeli üzerindeki etkisi zayıflasa da, zar potansiyelinde izlenebilir düzeyde kalmıştır. Tüm hücrelerde,



Şekil 3: Hücre dışı ortam derinliğinin, zar potansiyeli ve hücre dışı iyon derişimleri üzerindeki etkisi

akım uyarımı sona erdiğinde zar potansiyeli yeniden denge durumuna dönme eğilimi göstermiştir.

B. Dış Ortam Derinliğinin Etkisi

Şekil 3, yarıçapı 1 µm olan bir hücrenin bulunduğu ortamda farklı ortam derinliklerine ait sonuçları göstermektedir. Buna göre, ortam derinliğinin azalması hücre dışı iyon derişimlerinde belirgin sapmalara yol açmıştır. Ortam derinliği yeterince büyük olduğunda model, klasik HH modeliyle uyumlu sonuçlar vermektedir; ancak derinliğin azalmasıyla bu uyum giderek bozulmaktadır.

Hücre dışı derişimlerde gözlemlenen sapmalar, benzer şekilde zar potansiyelini de etkilemiştir. Özellikle düşük derinlik değerlerinde, hücre dışı Na^+ ve K^+ derişimleri akım uygulamaları sonucunda geri döndürülemez seviyelere ulaşmıştır. Yüksek derinlikte, uyarım sonrası hücre zar potansiyeli eski denge değerine dönebilirken, derinliğin azalması bu dengeye dönüşü engellemiş ve hücre başlangıç durumunu sürdüremez hâle gelmiştir. Bu durum, hücrenin homeostatik dengesini kaybetmesi şeklinde yorumlanabilmektedir.

Denge bozulmasının nedeni ATPaz pompasının çalışma dinamiklerinde saklıdır. Bu modelde, hücre hacminin yeterince büyük olması nedeniyle molar iyon akıları hücre içi derişimlerde belirgin değişikliklere yol açmamakta; dolayısıyla hücre içi Na^+ derişimi büyük ölçüde sabit kalmaktadır. Bu durum, ATPaz pompasının aktive olmasını gereksiz kılmaktadır. Ancak bu aynı zamanda, hücre dışındaki K^+ derişimindeki önemli değişimlere karşı da etkili bir yanıt üretilememesi anlamına gelmektedir. Bu nedenle, özellikle hücre dışı ortam derinliğinin düşük olduğu durumlarda homeostatik dengenin kaybolması olası hâle gelmektedir.

IV. Sonuç ve Öneriler

Bu çalışmada, klasik Hodgkin–Huxley modeline, iyon derişimlerinin zamana bağlı değişimini ve iyon türüne özgü akım bileşenlerini içeren genişletilmiş bir yaklaşım sunulmuştur. Simülasyon sonuçları, özellikle küçük hacimli yapılarda klasik modelin yetersiz kaldığını göstermiştir. Ayrıca, akımın hangi iyon üzerinden uygulandığının zar dinamikleri üzerinde belirleyici olduğu ortaya konmuştur. Gelecek çalışmalarda, farklı taşıyıcı mekanizmaların da modele entegre edilmesi önerilmektedir.

Kaynakça

[1] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, p. 500, 1952.

[2] L. Øyehaug, "Slow ion concentration oscillations and multiple states in neuron–glia interaction—insights gained from reduced mathematical models," *Frontiers in Network Physiology*, vol. 3, p. 1189118, 2023.

[3] M. J. Sætra, G. T. Einevoll, and G. Halnes, "An electrodiffusive, ion conserving Pinsky-Rinzel model with homeostatic mechanisms," *PLOS Computational Biology*, vol. 16, no. 4, p. e1007661, 2020.

[4] N. Hübel and M. A. Dahlem, "Dynamics from seconds to hours in Hodgkin-Huxley model with time-dependent ion concentrations and buffer reservoirs," *PLOS Computational Biology*, vol. 10, no. 12, p. e1003941, 2014.

[5] W. Teka, D. Stockton, and F. Santamaria, "Power-law dynamics of membrane conductances increase spiking diversity in a Hodgkin-Huxley model," *PLOS Computational Biology*, vol. 12, no. 3, p. e1004776, 2016.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Kortikal L5PC6 Piramidal Nöronların Alt Bölge Kombinasyonları Üzerinden Tümör Sınıflandırması
Tumor Classification via Subregion Combinations of Cortical L5PC6 Pyramidal Neurons

Zübeyir Özcan[1], İlknur Kayıkçıoğlu Bozkır[2], and Temel Kayıkçıoğlu[1]
[1]Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Türkiye
[2]Department of Software Engineering, Gümüşhane University, Gümüşhane, Türkiye
Emails: zubeyirozcan@ktu.edu.tr, ilknurkayikcioglu@gumushane.edu.tr, tkayikci@ktu.edu.tr

***Özetçe* —Bu çalışmada, benign ve malignant tümör örneklerinin, bir kortikal L5PC6 piramidal nöronun farklı alt bölgeleri kullanılarak sınıflandırılması amaçlanmıştır. Gerçek veri seti üzerinde spike tabanlı bir öğrenme yöntemi uygulanmış, her sınıf için ayrı bir nöron eğitilerek aksiyon potansiyeli üretimine dayalı bir karar mekanizması oluşturulmuştur. Eğitim süreci boyunca farklı sinaptik bağlantı türleri ve spike kodlama yöntemleri değerlendirilmiş, piramidal nöron için en yüksek performans yoğunluk kodlamayla elde edilmiştir. Bölgesel analizler, özellikle bazal bölgenin sınıflandırma başarısında öne çıktığını göstermiştir. Elde edilen sonuçlar, piramidal nöronların morfolojik ve biyofiziksel çeşitliliğinin sınıflandırma görevlerinde anlamlı bir avantaj sağladığını ortaya koymaktadır.**

Anahtar Kelimeler—piramidal nöron; L5PC6; alt bölge uyarımı; tümör sınıflandırması

***Abstract*—In this study, it was aimed to classify benign and malignant tumor samples by utilizing different subregions of a cortical L5PC6 pyramidal neuron. A spike-based learning method was applied to a real dataset, and a decision mechanism based on action potential generation was established by training a separate neuron for each class. During the training process, various types of synaptic connections and spike encoding schemes were evaluated and intensity coding yielded the highest performance for the pyramidal neuron. Regional analyses revealed that the basilar region particularly stood out in terms of classification success. The results demonstrate that the morphological and biophysical diversity of pyramidal neurons provides a significant advantage in classification tasks.**

Keywords—pyramidal neuron; L5PC6; subregion stimulation; tumor classification

I. GİRİŞ

Kortikal piramidal nöronlar, beyinde sıkça rastlanan ve yapısal olarak çok bölmeli bir mimariye sahip hücrelerdir. Karmaşık morfolojik yapıları ve biyofiziksel çeşitlilikleri sayesinde, farklı kortikal katmanlara uzanabilen geniş dendritik ağlara sahiptirler. Bu özellikleri, hem uzak kortikal bölgelerden bilgi alabilmelerine hem de işledikleri bilgiyi uzak bölgelere iletebilmelerine olanak tanımaktadır. Apikal ve bazal dendritik bölgeleri aracılığıyla gelen uyarımları mekânsal olarak ayrıştırarak işleyebilen bu hücrelerin alt bölgeleri, ulaştıkları kortikal katmanlar ve taşıdıkları biyofiziksel özellikler bakımından birbirinden ayrışmaktadır. Literatürde, bu alt bölgelerin gelen uyarımlara farklı tepkiler verdiği ve nöronal çıktılar üzerinde etkilerinin olduğu çeşitli çalışmalarla gösterilmiştir [1]. Ancak, söz konusu bölgelerin bilgi işleme ve sınıflandırma gibi görevlerdeki katkısı hâlâ detaylı araştırmalara ihtiyaç duymaktadır. Bu nedenle, piramidal nöronlar, karar verme süreçlerine entegre edilebilecek potansiyel hesaplamalı birimler olarak dikkat çekmektedir.

Yapılan çalışmalar, apikal ve bazal dendritlerin sinaptik uyarımları farklı biçimlerde işlediğini göstermektedir. Polsky ve arkadaşları, bazal dendritlerin sinaptik uyarımları doğrusal şekilde işlerken, apikal dendritlerin doğrusal olmayan işlemler gerçekleştirdiğini ortaya koymuştur [1]. Spruston ise bu bölgelerdeki iyon kanal dağılımı ve iletim özelliklerinin nöronal çıktılar üzerinde önemli etkiler yarattığını vurgulamıştır [2]. Ayrıca, Branco ve Häusser, farklı bölgelerden gelen uyarımların zamansal özelliklerine olan duyarlılıkların değiştiğini göstermiştir [3].

Moldwin ve Segev, yaptıkları çalışmada bir piramidal nöronun eğitilerek iki sınıflı bir sınıflandırma problemini başarıyla çözebileceğini göstermiştir [4]. Çalışmalarında kullandıkları piramidal nöron modelinin alt bölge performanslarını da ele almış olsalar da, sınıflandırma için kullanılan veri setinin yapay olması, yalnızca uyarıcı (eksitatör) sinaptik bağlantıların dikkate alınması ve sinaptik ağırlıkların yalnızca pozitif değerler alabileceği yönündeki kısıtlar, yöntemin potansiyel performansını sınırlayabilecek unsurlar olarak dikkat çekmektedir. Bu eksikliklere Kayıkçıoğlu Bozkır ve arkadaşlarının çalışmasında değinilmiştir [5]. Bir EKG veri seti kullanılarak piramidal nöronların sınıflandırma performansı incelenmiş ve önerilen aynalama yaklaşımı sayesinde ağırlıkların pozitif olma kısıtının yol açtığı performans kayıplarının önüne geçildiği gösterilmiştir. Ancak bu çalışmada da baskılayıcı (inhibitör) sinaptik bağlantıların sınıflandırma üzerindeki etkisi araştırılmamıştır. Özcan ve arkadaşları ise benzer bir yaklaşımı baskılayıcı sinapsları dahil ederek uygulamışlardır [6]. Çeşitli veri setleri kullanarak basit nöron modellerinin performansını değerlendirmişlerdir. Ancak bu çalışmada piramidal nöron modeline yer verilmemesi, konuyla ilgili araştırmalarda önemli bir boşluk bırakmıştır.

Bu çalışmada, yukarıda belirtilen sınırlılıkların aşılması

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

ve piramidal nöronların alt bölge düzeyinde bilgi işleme potansiyelinin daha gerçekçi koşullar altında değerlendirilmesi amaçlanmaktadır. Buna göre, L5PC6 tipi bir kortikal piramidal nöron modeli kullanılarak iki sınıflı bir tümör sınıflandırma problemi ele alınmıştır. Çalışmada, tümör veri setinin öznitelik vektörleri, daha önce literatürde önerilmiş bir öznitelik–spike dönüşüm yöntemiyle spike örüntülerine dönüştürülmüştür. Bu örüntüler, nöronun farklı dendritik alt bölgelerine sinaptik giriş olarak uygulanmıştır. Apikal ve bazal dendrit bölgelerine ek olarak soma hem bireysel hem de kombinasyon halinde uyarılırken oluşan somatik aksiyon potansiyelleri üzerinden sınıflandırma performansları karşılaştırılmıştır. Ayrıca, bu performans sonuçları yaygın olarak kullanılan HH, IZH ve LIF gibi basitleştirilmiş nöron modelleriyle karşılaştırılarak, piramidal morfolojinin sınıflandırma üzerindeki etkisi ortaya konulmuştur.

II. MATERYAL VE YÖNTEM

A. Veri Seti

Bu çalışmada *Breast Cancer Wisconsin* veri seti kullanılmıştır [7]. Toplam 569 örnek içeren veri setinde, her örnek bir hastaya ait olup tümörün iyi huylu (benign) veya kötü huylu (malignant) olduğunu gösteren etiket bilgisiyle birlikte 30 sayısal öznitelik barındırmaktadır. Bu öznitelikler, hücre çekirdeğine ait morfolojik özelliklerin ortalama, en büyük ve standart sapma değerlerini içermektedir. Veri seti kullanılırken; rastgele karıştırıldıktan sonra yarısı eğitim, yarısı test olmak üzere iki alt kümeye ayrılmıştır. Bu sayede çalışma için eğitim ve test başarımları ayrı ayrı değerlendirilmiştir.

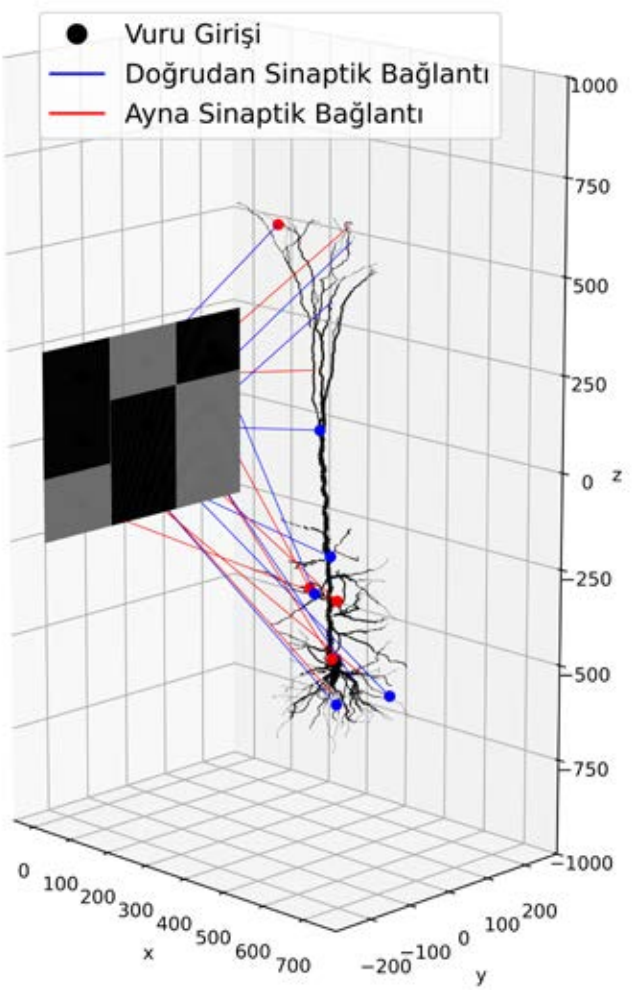
B. Nöron Modelleri

Kullanılan piramidal nöron modeli, Hay ve arkadaşları tarafından sunulan üç farklı kortikal piramidal nöron modelinden biridir [8]. Bu çalışma kapsamında yalnızca Şekil 1'de gösterilen ve L5PC6 tipi bir piramidal hücreye ait olan model esas alınmıştır. Modelin iyon kanal dağılımı ve diğer biyofiziksel parametreler, Shai ve arkadaşlarının çalışmasındaki düzenlemelere uygun biçimde güncellenmiştir [9].

Elde edilen sonuçları değerlendirmek ve karşılaştırmak amacıyla, Hodgkin–Huxley (HH), Izhikevich (IZH) ve Leaky Integrate-and-Fire (LIF) nöron modelleri ayrıca çalışmaya dahil edilmiştir. Bu modeller, sınıflandırma performansları yönünden referans olarak ele alınmış ve alt bölge bazlı değerlendirmelerle karşılaştırılmıştır.

C. Öznitelik–Spike Dönüşümü

Bu çalışmada, sayısal öznitelik vektörlerinin zamansal spike örüntülerine dönüştürülmesi için SPIGEN yöntemi kullanılmıştır [6]. Yöntem üç temel adımdan oluşmaktadır. İlk adımda, öznitelik değerleri yedi bitlik kuantalama işlemine tabi tutulmuştur. İkinci adımda, kuantal aralıklar ikili, yoğunluk veya Gray kodlama olmak üzere üç farklı yöntemle ikili koda dönüştürülmüştür. Elde edilen ikili kodlarda "1" değeri ilgili sinapsta giriş olduğuna, "0" ise giriş olmayacağı anlamına gelmektedir. Tüm örnekler için sinaptik girişler, elde edilen bu kodlar üzerinden yapılmıştır.



Şekil 1: Piramidal nöron morfolojisi ve girdi entegrasyonu

D. Aynalama Yaklaşımı

Bu çalışmaya, Kayıkçıoğlu Bozkır ve arkadaşlarının önerdiği aynalama yaklaşımı ayrıca dahil edilmiştir [5]. Aynalama yaklaşımında, her spike giriş örüntüsü bit düzeyinde 1'ler 0'a, 0'lar 1'e olmak üzere terslenmiş ve genişletilmiş bir giriş vektörü elde edilmiştir. Bu şekilde, her giriş hem orijinal haliyle hem de aynalanmış biçimiyle nöron modeline sunulmuştur. Böylece, Moldwin ve Segev'in çalışmalarında değindikleri sinaptik ağırlıkların pozitif olma kısıtı aşılmıştır [4].

E. Öğrenme Yöntemi

Çalışmanın öğrenme yönteminde üç temel unsur bulunmaktadır.

- **Sinaptik girişler:** Özniteliklerin spike örüntülerine dönüştürülmesiyle elde edilmiştir ve ikili değerler almaktadır. Burada "1", sinaptik girişin bir spike içerdiğini; "0" ise spike bulunmadığını ifade etmektedir.
- **Nöronal çıktı:** Nöronun verilen spike örüntüsüne karşılık olarak aksiyon potansiyeli üretip üretmediği

değerlendirilmiştir. Hiç aksiyon potansiyeli oluşmazsa çıktı "0", en az bir aksiyon potansiyeli oluşursa çıktı "1" kabul edilmiştir.

- **Sinaptik bağlantı tipi:** Bu çalışmada iki uyarıcı sinaps (AMPA, NMDA) ve iki baskılayıcı sinaps ($GABA_A$, $GABA_B$) kullanılmıştır. Eğitim süreci boyunca sinaptik güncellemeler her sinaps tipi için ayrı ayrı yapılmıştır. Ayrıca, her öznitelik girişi dört farklı sinaptik tip üzerinden nörona bağlanmıştır.

$$\Delta w_i = \eta \times (d - \text{çıktı}) \times x_i \tag{1}$$

$$w_i^{\text{Yeni}} \leftarrow \begin{cases} \max\left(0,\, w_i^{\text{Eski}} + \Delta w_i\right), & \text{Eksitatör} \\ \max\left(0,\, w_i^{\text{Eski}} - \Delta w_i\right), & \text{İnhibitör} \end{cases} \tag{2}$$

Bu çalışmada uyarlanmış bir perceptron öğrenme kuralı kullanılmıştır. Bu kurala göre ağırlık güncellemeleri tüm sinaps tipleri için ayrı ayrı ve negatif değer almayacak şekilde düzenlenmiştir. İlgili ifadeler Denklem (1) ve Denklem (2)'de gösterilmektedir.

Öncelikle nöronal çıktı ile hedef çıktı arasındaki fark hesaplanmış, elde edilen hata değeri belirli bir öğrenme oranıyla her bir sinaptik girişin değeriyle çarpılarak ağırlık güncelleme miktarı belirlenmiştir. Bu değişim, sinapsın tipi dikkate alınarak uygulanmıştır.

F. Eğitim ve Test

Çalışmada ele alınan sınıflandırma problemi iki sınıflıdır. Bu doğrultuda, her bir sınıf için ayrı bir piramidal nöron modeli kullanılmıştır. Eğitim sürecinde, benign sınıfı temsil eden nöron yalnızca benign örnekler için aksiyon potansiyeli üretmeye, malignant örnekler karşısında ise sessiz kalmaya yönelik olarak eğitilmiştir. Benzer süreç malignant sınıfı için de uygulanmıştır. Böylece her nöron yalnızca kendi sınıfına ait örnekleri seçici olarak temsil edecek şekilde özelleştirilmiştir.

Eğitim süreci, önceden belirlenen spike örüntüleri ve sinaptik bağlantı tipleri üzerinden yürütülmüştür. Toplam 20 epok boyunca tüm eğitim verileri her iki nöron için ayrı ayrı işlenmiştir. Her epok sonunda nöronal çıktılar ile hedef çıktılar arasındaki farklar üzerinden ortalama kare hata (MSE) değeri hesaplanmış ve öğrenme eğilimi takip edilmiştir.

Test aşamasında hem piramidal nöron modeli hem de karşılaştırmalı olarak kullanılan üç farklı basit nöron modeli için sınıflandırma performansı değerlendirilmiştir. Basit nöron modellerinde, farklı spike kodlama yöntemlerinin sınıflandırma başarısına etkisi ayrıca analiz edilmiştir. Elde edilen sonuçlara göre en iyi performansı sağlayan yoğunluk kodlama yöntemi, tüm piramidal nöron deneylerinde standart olarak kullanılmıştır.

Basit nöron modellerinde, iki sınıfı temsil eden iki ayrı nöronun çıktıları değerlendirilerek sınıf etiketine karar verilmiştir. Eğer yalnızca bir nöron aksiyon potansiyeli üretmişse, örnek o sınıfa atanmıştır. Her iki nöron da çıktı üretirse, en erken aksiyon potansiyelini üreten nöronun sınıfı esas alınmıştır. Hiçbiri ateşleme gerçekleştirmemişse, zar potansiyeli en yüksek seviyeye ulaşan nöronun sınıfı tercih edilmiştir.

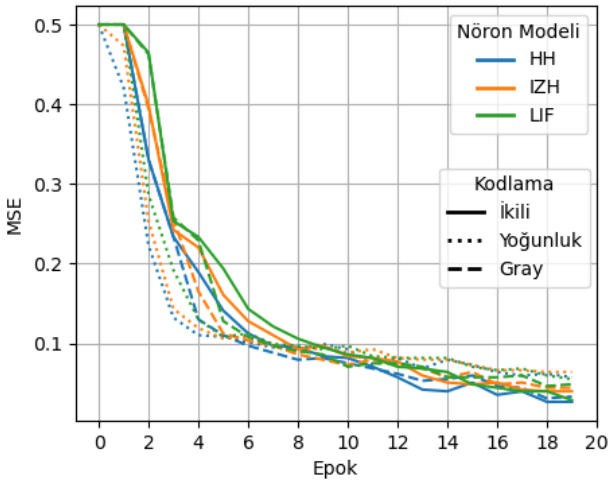
Piramidal nöron için sınıflandırma başarımı iki düzeyde değerlendirilmiştir. İlk olarak, benign ve malignant sınıflarını temsil eden iki piramidal nöronun her birinin kendi sınıfına ait örnekleri ne ölçüde doğru sınıflandırdığı analiz edilmiştir. Ardından, yukarıda açıklanan bütüncül karar mekanizması piramidal nöronlar için de uygulanarak genel sınıflandırma başarımı elde edilmiştir.

Son olarak, modelin genellenebilirliğini değerlendirmek amacıyla eğitim ve test verileri bağımsız şekilde ele alınmış, tüm deneylerde performans F1-skor ölçütü ile değerlendirilmiştir.

III. BULGULAR VE TARTIŞMA

A. Eğitim Hataları

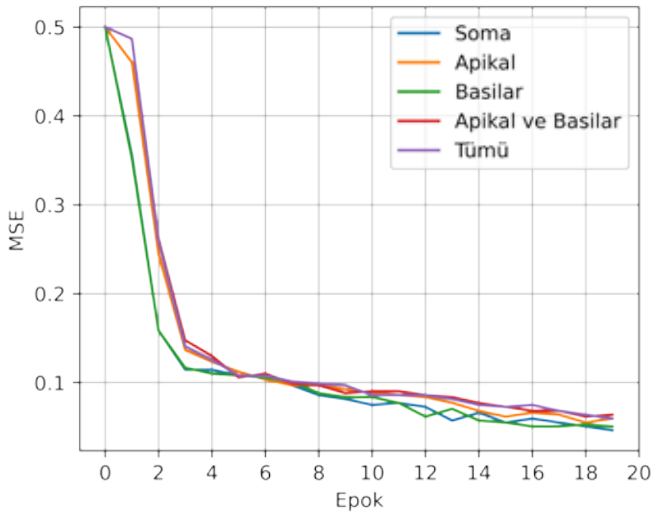
Şekil 2 ve Şekil 3'te eğitim aşamalarında elde edilen hataların epoklara göre seyirleri gösterilmektedir. Tüm modellerde, epok sayısı arttıkça MSE değerlerinde belirgin bir azalma gözlenmiştir. Basit nöronlarda, özellikle yoğunluk kodlama, erken epoklardan itibaren daha hızlı ve kararlı bir öğrenme eğrisi sunarak diğer kodlama yöntemlerine göre daha düşük hata seviyelerine ulaşmıştır. Piramidal nöron için, eğitim süreci boyunca bazal bölge üzerinden yapılan eğitim en düşük hata ile sonuçlanırken, apikal bölgenin tek başına kullanımı diğerlerine kıyasla daha yüksek hata değerleri üretmiştir. Genel olarak sonuçlar incelendiğinde ise piramidal nöronlar için alt bölgelerin eğitiminde küçük farklılıklar mevcuttur.



Şekil 2: Basit nöron modelleri için eğitim hatalarının seyri

B. Basit Nöron Modellerinin Sınıflandırma Performansları

Tablo I'de, HH, IZH ve LIF nöron modelleri için farklı spike kodlama yöntemleri kullanılarak elde edilen eğitim ve test başarı oranları karşılaştırılmaktadır. Eğitim aşamasında en yüksek doğruluk, HH modeli ile ikili kodlama yöntemi kullanıldığında %99,56 olarak elde edilmiştir. Tüm modellerde ikili ve Gray kodlamalar benzer başarılar sağlarken, yoğunluk kodlama yöntemi eğitim aşamasında diğer yöntemlere kıyasla biraz daha düşük performans göstermiştir.



Şekil 3: Piramidal nöron için eğitim hatalarının seyri

Aşama	Kodlama	HH	IZH	LIF
Eğitim	İkili	99,56%	98,68%	98,68%
	Yoğunluk	97,36%	97,35%	96,49%
	Gray	98,68%	98,24%	98,68%
Test	İkili	92,64%	92,65%	92,65%
	Yoğunluk	97,09%	96,49%	97,38%
	Gray	95,29%	95,60%	95,31%

Tablo I: Basit nöron modelleri için performans bulguları

Ancak test aşamasında bu durum tersine dönmüştür. Yoğunluk kodlama, üç modelde de en yüksek test başarılarını sağlamıştır: HH için %97,09, IZH için %96,49 ve LIF için %97,38. Bu sonuçlar, yoğunluk tabanlı spike kodlamanın genellenebilirlik açısından daha güçlü bir temsil sunduğunu göstermektedir. Özellikle HH ve LIF modellerinde test başarısının yüksek oluşu, bu modellerin karmaşık örüntülere karşı daha iyi öğrenme gerçekleştirdiğini ortaya koymaktadır.

C. Piramidal Nöronun Sınıflandırma Performansı

Tablo II'de, piramidal nöron modeli kullanılarak elde edilen eğitim ve test başarıları hem sınıf bazlı hem de bütüncül olarak verilmiştir. Eğitim verileri üzerinden değerlendirildiğinde, bazal bölge ile yapılan sınıflandırmalar, malignant ve benign örneklerde dengeli bir başarı sergilemiştir. Benzer şekilde, yalnızca soma bölgesi kullanıldığında da birleşik başarı oranı %97,80'e ulaşmıştır. Apikal ve diğer bölge kombinasyonları ise biraz daha düşük başarımlar göstermiştir.

Test aşamasında ise apikal ve apikal + bazal bölgeler, birleşik başarımda %97,38 ile en yüksek sınıflandırma performansını göstermiştir. Özellikle benign sınıflar için tüm bölgelerde yüksek doğruluklar elde edilmiş, ancak malignant sınıf için başarı oranları göreceli olarak daha düşük kalmıştır. Bu durum, benign örneklerin daha kolay sınıflandırılabildiğini, malignant sınıfın ise daha karmaşık nöronal yanıtlar gerektirdiğini düşündürmektedir.

Aşama	Bölge	Malignant	Benign	Birleşik
Eğitim	Soma	95,18%	96,93%	97,80%
	Apikal	95,18%	94,29%	96,93%
	Bazal	95,62%	95,58%	97,80%
	Apikal ve Bazal	94,34%	95,15%	96,93%
	Tümü	94,32%	95,16%	96,93%
Test	Soma	93,58%	97,07%	96,79%
	Apikal	92,14%	96,20%	97,38%
	Bazal	93,02%	96,77%	97,09%
	Apikal ve Bazal	91,89%	96,47%	97,38%
	Tümü	92,18%	96,47%	97,09%

Tablo II: Piramidal nöron için performans bulguları

Genel olarak, piramidal nöronun farklı alt bölgeleri üzerinden yapılan değerlendirmeler hem tekil hem de bütüncül bilgi işlemeyi destekleyecek şekilde yüksek sınıflandırma başarıları sunmuş ve bölgesel seçiciliğin sınıf ayrımı üzerindeki etkisini ortaya koymuştur.

IV. SONUÇ VE ÖNERİLER

Bu çalışmada, piramidal nöronun alt bölgeleri kullanılarak iki sınıflı bir tümör verisi üzerinde spike tabanlı bir sınıflandırma sistemi geliştirilmiştir. Bulgular, özellikle bazal bölgenin sınıflandırma başarısında öne çıktığını ve yoğunluk kodlamanın genellenebilirlik açısından avantaj sağladığını göstermiştir. Basit nöron modelleriyle karşılaştırıldığında, piramidal nöronun morfolojik çeşitliliği karar doğruluğunu artırıcı bir etki yaratmıştır. Gelecek çalışmalarda, inhibisyonun daha ayrıntılı ele alınması ve çok sınıflı yapılarla modelin genişletilmesi önerilmektedir.

KAYNAKÇA

[1] A. Polsky, B. W. Mel, and J. Schiller, "Computational subunits in thin dendrites of pyramidal cells," *Nature Neuroscience*, vol. 7, no. 6, pp. 621–627, 2004.

[2] N. Spruston, "Pyramidal neurons: dendritic structure and synaptic integration," *Nature Reviews Neuroscience*, vol. 9, no. 3, pp. 206–221, 2008.

[3] T. Branco and M. Häusser, "The single dendritic branch as a fundamental functional unit in the nervous system," *Current Opinion in Neurobiology*, vol. 20, no. 4, pp. 494–502, 2010.

[4] T. Moldwin and I. Segev, "Perceptron learning and classification in a modeled cortical pyramidal cell," *Frontiers in Computational Neuroscience*, vol. 14, p. 33, 2020.

[5] I. Kayikcioglu Bozkir, Z. Ozcan, C. Kose, T. Kayikcioglu, and A. E. Cetin, "Improving a cortical pyramidal neuron model's classification performance on a real-world ECG dataset by extending inputs," *Journal of Computational Neuroscience*, vol. 51, no. 3, pp. 329–341, 2023.

[6] Z. Özcan, İ. K. Bozkır, and T. Kayıkçıoğlu, "SPIGEN: A Configurable Population-Coded Spike Pattern Generator for Single-Spike Learning in Spiking Neural Networks," unpublished, 2025.

[7] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast Cancer Wisconsin (Diagnostic) [Dataset]," *UCI Machine Learning Repository*, 1993.

[8] E. Hay, S. Hill, F. Schürmann, H. Markram, and I. Segev, "Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties," *PLoS Computational Biology*, vol. 7, no. 7, p. e1002107, 2011.

[9] A. S. Shai, C. A. Anastassiou, M. E. Larkum, and C. Koch, "Physiology of layer 5 pyramidal neurons in mouse primary visual cortex: coincidence detection through bursting," *PLoS Computational Biology*, vol. 11, no. 3, p. e1004090, 2015.

A SYSTEM DESIGN FOR SCOLIOSIS DETECTION

Cansel Ficici
Elektical and Electronics Enginnering, Ankara University,
ogretmenoglu@ankara.edu.tr

Kutay Yıldırım
Elektical and Electronics Enginnering, Ankara University,
kutayildirimm@gmail.com

İsmail Can Yılmaz
Elektical and Electronics Enginnering, Ankara University,
can0786881@gmail.com

Arif Koçoğlu
Biomedical Device Technologies Department, Baskent University,
arif@baskent.edu.tr

Onur Koçak
Computer & Biomedical Engineering, Baskent University,
okocak@baskent.edu.tr

Ziya Telatar
Biomedical Engineering, Baskent University,
ztelatar@baskent.edu.tr

***Abstract*— Scoliosis, characterized by lateral curvature of the spine, can lead to significant health complications if not detected early. Conventional diagnostic methods, such as radiographic imaging, require exposure to ionizing radiation and may not be accessible in resource-limited settings. This study presents a non-invasive and portable scoliosis screening system that integrates ADXL345 triaxial accelerometers into a wearable corset. The system performs three established physical assessments—Shoulder Level Test, Adam's Forward Bend Test, and Spinal Angulation Test—by collecting posture data and estimating spinal curvature angles. Sensor outputs were processed using exponential moving average filtering and smoothing techniques to minimize noise. Estimated Cobb angles and other posture metrics are visualized through a user interface developed in C# with the .NET framework. Validation against measurements obtained with a conventional scoliometer demonstrated good agreement, demonstrating the system's potential utility as a radiation-free alternative for scoliosis screening.**

Keywords*— *scoliosis; cobb angle; accelerometer; posture analysis; non-invasive screening

I. INTRODUCTION

Scoliosis is defined as a lateral curvature of the spine greater than 10°, as measured on radiographs [1]. The condition is characterized by a sideways curvature and vertebral rotation, affecting approximately 2% of the global population [2]. The risk is significantly higher in individuals with a family history of scoliosis, with estimates of up to 20%, suggesting a hereditary component [3]. Clinical manifestations can include chronic back pain, shoulder and lumbar instability, and in severe cases, reduced thoracic volume that impairs pulmonary function [4].

Treatment strategies vary depending on curve severity, progression potential, and patient age. Observation is often sufficient in mild cases, bracing is typically recommended for moderate curvatures, and surgery may be required in progressive or severe deformities.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Current diagnostic systems, however, present notable shortcomings. Standard radiographic methods expose patients to ionizing radiation, which is a concern in children and adolescents who require frequent follow-ups. Moreover, radiographs are costly and not always available in rural or resource-limited settings. These limitations highlight the urgent need for alternative diagnostic approaches that are safe, cost-effective, and widely accessible.

The types of scoliosis can be classified as congenital, neuromuscular, or idiopathic [5]. The most common form, idiopathic scoliosis, has no specific cause identified to date [5]. The International Society on Scoliosis Orthopedic and Rehabilitation Treatment (SOSORT) has proposed three main classification types for idiopathic scoliosis: chronological, angular, and topographic [5].

Idiopathic scoliosis (IS) can also be classified according to age of onset: infantile (birth to two years), juvenile (three to nine years), and adolescent (10 years and older) [6]. Adolescent idiopathic scoliosis (AIS) is the most common pediatric spinal deformity [7].

In one study conducted in India, involving 119 patients diagnosed with scoliosis between December 2006 and December 2012, thoracolumbar curvature was the most common type, observed in 43.6% of cases [8]. Bunnell proposed a simple handheld device called a scoliometer, which is frequently used to measure and assess the angle of trunk rotation during the Adams Forward Bend Test and for screening purposes [9].

For examination with the scoliometer, the patient is asked to bend forward until a curve appears in the thoracic region. The scoliometer is then placed and moved along the spine to measure the highest point of the thoracic curve. The patient is subsequently asked to bend forward until the lower back (lumbar) curve is visible, and again the scoliometer is placed along the spine to measure the highest point of the curve [10].

Bunnell defined the following screening cut-off criteria [9] :

$0^o \leq ATR \leq 3^o$ Normal
$4^o \leq ATR \leq 6^o$ the trunk rotation is intermediate
$7^o \leq ATR$ high possibility of having scoliosis

The scoliometer-measured ATR has proven to be a reproducible and repeatable examination. In this test, an ATR of 7º or higher has shown a sensitivity of 83.3% and specificity of 86.8% [11].

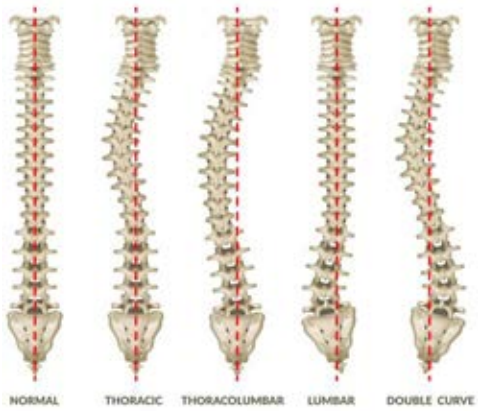


Figure 1. Types of Scoliosis of Spine [12]

The severity of spinal curvature can be determined by measuring the Cobb angle, which is defined as the angle between two perpendicular lines traced from the upper edge of the uppermost involved vertebra and the lower edge of the lowermost involved vertebra.

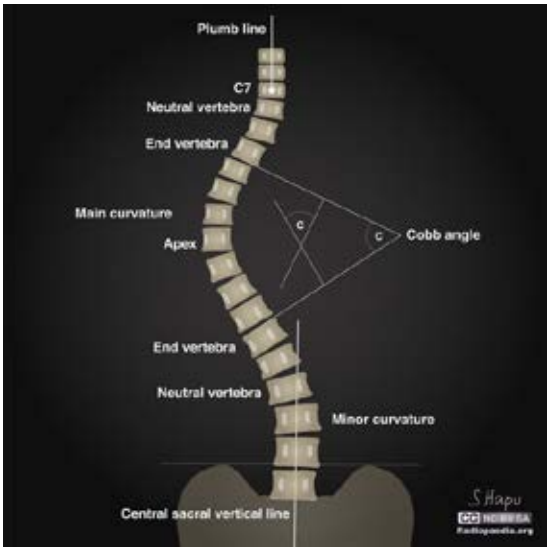


Figure 2. Illustration of the Cobb angle used for scoliosis assessment. [13]

The aim of this project is to design a sensor-based system to support the early diagnosis of scoliosis. The system focuses on obtaining accurate and stable posture data to assess spinal alignment and angulation. Three specific tests were performed within the scope of this project: Shoulder Level Test, Adam's Forward Bend Test, and Spinal Angulation Test. In addition to sensor-based data collection, a user-friendly interface was developed using C# and the .NET framework to effectively visualize real-time data. The interface displays the results of scoliosis diagnostic tests in tabular form and provides visual feedback by coloring the back, spine, and shoulder areas on a human figure according to the measurement results.

Consequently, this project aims to provide a practical, portable, and cost-effective alternative to existing scoliosis screening methods.

II. MATERIALS AND METHODS

3.1 Materials

The system developed in this study to support the early diagnosis of scoliosis consists of both hardware and software components. For the hardware, an Arduino Uno microcontroller board was used to collect and process sensor data. A total of nine ADXL345 triaxial accelerometers were used for posture analysis. These sensors were integrated into a corset corresponding to specific anatomical points on the patient's back. The connections between the Arduino board and the sensors were established using jumper cables. Plastic clamps were used to secure the cable connections and partially prevent tangling. The wearable corset ensured accurate positioning of the sensors while minimizing measurement errors due to movement.

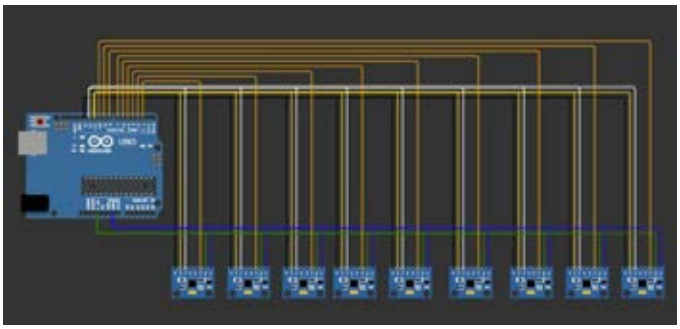
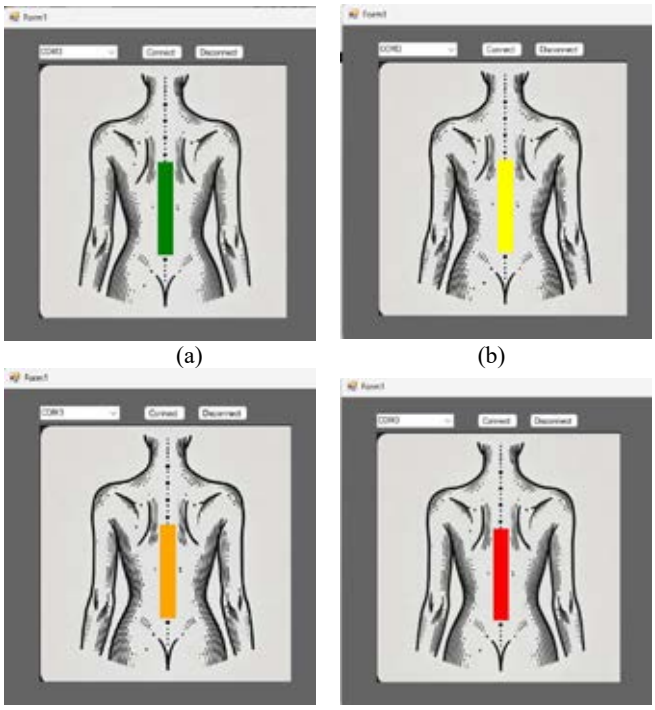


Figure 3. Connection of Arduino and ADXL345

For the software development environment, Microsoft Visual Studio was used for the development of the system interface. For programming and uploading the microcontroller code, the Arduino IDE was employed. Interface development and data processing were performed using the C# programming language and the .NET Framework.

The developed interface was designed to effectively visualize scoliosis diagnostic measurements and provide a user-friendly experience. The results of the Shoulder Level Height Test, Adam's Forward Bend Test, and Spinal Angulation Test are presented in tabular format on the interface. The Cobb angle is calculated from sensor data and displayed on the interface. Dynamic color coding is applied to a representative human figure based on the measurement results.

For Cobb angle measurement, 0–10° is displayed in green, 10–25° in yellow, 25–45° in orange, and values above 45° in red. For the Shoulder Level Test, the higher shoulder side is displayed in blue. The developed interface is shown in Figure 4.



(a) (b) (c) (d)

Figure 4 (a) : Cobb Angle < 10 , (b): 10 < Cobb Angle < 25, (c): 25 < Cobb Angle < 45 , (d): 45 < Cobb Angle

3.2 Data Collection and Processing

The Cobb angle was estimated from angular data obtained from sensors instead of conventional radiographic images. Specifically, differences in inclination between sensors placed on the upper and lower parts of the spine were analysed to derive values representing the Cobb angle. This approach integrates traditional measurement principles with sensor technology.

In addition to the Cobb angle, two other clinical tests were implemented using sensor data. The Shoulder Level Test was performed by calculating the difference in orientation between sensors placed on the left and right scapulae. The angular discrepancy in the sagittal plane reflects shoulder level asymmetry, a common indicator of scoliosis.

The Cobb angle, widely used in scoliosis assessment, was estimated from accelerometer data using the following formula:

$$Cobb\ Angle = (\arctan(^{x_2}/_{y_2}) - \arctan(^{x_1}/_{y_1})) + (\arctan(^{x_4}/_{y_4}) - \arctan(^{x_5}/_{y_5}))$$

Here, (x_i, y_i) represent the projections of the orientation vector from different sensor positions along the X and Y axes. The angle differences are combined to reflect the spinal curvature, corresponding to the Cobb angle.

In this test, two sensors are placed over the left and right scapulae to measure their respective orientation angles in the sagittal (YZ) plane. Each angle is calculated using the arctangent of the ratio between the vertical (Y) and dorsal (Z) acceleration components

$$\theta = \arctan(\frac{y}{z})$$

The difference between these two angles provides an objective measure of scapular asymmetry:

$$\Delta\theta = \arctan(^{y_6}/_{z_6}) - \arctan(^{y_7}/_{z_7})$$

A greater absolute value of Δθ indicates a higher degree of scapular imbalance, which may correspond to underlying spinal deformity such as scoliosis. Conversely, when Δθ approaches zero, it suggests a more symmetric and upright posture. This method enables a quantifiable, sensor-based assessment of scapular alignment, complementing clinical observations and supporting early detection of postural deviations.

Additionally, the Y-axis values from two upper sensors (y_8 and y_9) are compared directly to detect possible asymmetry or displacement. This comparison provides qualitative data on alignment differences, which may support other scoliosis indicators.

3.3 System Testing and Comparative Measurements

To evaluate the accuracy and reliability of the developed system, practical tests were conducted on multiple participants. First, measurements were obtained from each of the 19 participants using the sensor-based system, and the posture data and estimated Cobb angles were recorded in the system interface. Manual scoliosis measurements were then performed on the same participants using a scoliometer, a method widely used in clinical applications. Finally, the data obtained from both methods were compared, and the agreement between the system's measurement results and the traditional method was evaluated.

This experimental study was carried out with three voluntary participants. No interventions posing electrical safety risks were performed on the subjects, and they were duly informed regarding this issue. Written informed consent was obtained from all participants. Ethical approval will be secured for the forthcoming clinical validation phase of the study.

III. RESULTS AND FINDINGS

This section presents the findings from scoliosis detection tests conducted with the developed sensor-based system on multiple individuals under standardized conditions to evaluate its accuracy, functionality, and consistency. Three different physical test scenarios—Shoulder Level Test, Adam's Forward Bend Test, and Spinal Angulation Test—were applied to each participant, and the corresponding sensor data were collected using a custom-designed corset equipped with ADXL345 accelerometers.

For each test, posture deviations and spinal angulations were analyzed, and the estimated Cobb angles were calculated from the sensor data. The results were compared with manual

measurements obtained using a scoliometer to assess the system's measurement reliability. The summarized findings are presented in the tables below, illustrating the performance of the proposed method across different participants and test types. In addition, Figure 5 shows the final version of the system design. As shown in Figure 5, the system was worn by the participants during measurement acquisition.

TABLE I. RESULTS OF SPINAL ANGULATION TEST

Person	*Cobb Angle*	*Scoliometer*
Person-1	6.80	3
Person-2	4.79	2
Person-3	6.61	2

TABLE II. RESULTS OF FORWARD BENDING TEST

Person	*Cobb Angle*
Person-1	4.14
Person-2	6.94
Person-3	10.49

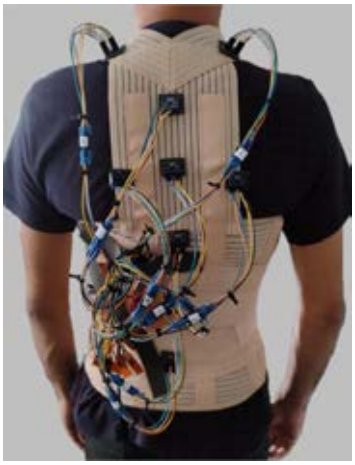


Figure 5. The System Final Design

In the bending test performed in thin individuals, healthier data were obtained compared to the tests performed in overweight individuals.

In a study of 300 healthy adults aged 18–85, the Scapular Balance Angle (SBA) averaged $2.505° \pm 2.340°$. Values above 7.185° were defined as abnormal. Since SBA quantifies the angle between the inferior angles of the scapula and the spine, it corresponds directly to shoulder blade asymmetry measured in our system [14].

III. CONCLUSION

In this study, a prototype scoliosis assessment system based on accelerometer sensors was designed and implemented. The system, comprising both hardware and software components, employs triaxial accelerometers positioned at specific anatomical points on the back to estimate the Cobb angle and evaluate posture. A custom-developed interface was used to present sensor data and provide visual feedback.

Tests conducted on multiple participants allowed comparison between the sensor-based results and those obtained using a standard scoliometer. The findings demonstrate that the system can produce measurements comparable to conventional methods. This approach may offer a practical, radiation-free alternative for scoliosis screening in settings where radiographic imaging is unavailable or not preferred.

Future improvements may focus on reducing the hardware size and expanding the diversity of the participant sample to further validate and enhance system performance. By ensuring compliance with industrial usability standards, the developed system is envisioned as a practical pre-diagnostic tool that may complement the radiological imaging techniques currently regarded as the gold standard for scoliosis diagnosis. In the present study, subjects remained stationary during the experimental data acquisition process, and all potential sources of electronic noise were eliminated. Future work will focus on trials conducted outside the laboratory environment to further enhance and validate the performance of the system.

REFERENCES

[1] V. Giannoglou and E. Stylianidis, "Review of advances in Cobb angle calculation and image-based modelling techniques for spinal deformities," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-5, pp. 129–136, 2016, doi: 10.5194/isprsannals-III-5-129-2016.

[2] S. Adibatti, K. R. Sudhindra, and J. M. Shivaram, "Design and performance analysis of Cobb angle measurement from X-ray images," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3530–3540, 2023.

[3] H.-H. Ma, C.-L. Tai, L.-H. Chen, C.-C. Niu, W.-J. Chen, and P.-L. Lai, "Application of two-parameter scoliometer values for predicting scoliotic Cobb angle," *Biomedical Engineering Online*, vol. 16, no. 1, p. 136, Dec. 2017, doi: 10.1186/s12938-017-0427-7.

[4] C. Chen, *et al.*, "Automating Cobb angle measurement for adolescent idiopathic scoliosis using instance segmentation," 2022. [Online]. Available: https://arxiv.org/pdf/2211.14122. Accessed: Aug. 10, 2025.

[5] I. Karpiel, A. Ziębiński, M. Kluszczyński, and D. Feige, "A survey of methods and technologies used for diagnosis of scoliosis," *Sensors*, vol. 21, no. 24, p. 8410, 2021, doi: 10.3390/s21248410.

[6] F. G. Ü. Demir, "Skolyoz sınıflaması," *JAMER*, vol. 8, no. 1, pp. 1–13, 2023.

[7] Y. Ishikawa, T. Kokabu, K. Yamada, Y. Abe, H. Tachi, H. Suzuki, T. Ohnishi, T. Endo, D. Ukeba, K. Ura, M. Takahata, N. Iwasaki, and H. Sudo, "Prediction of Cobb angle using deep learning algorithm with three-dimensional depth sensor considering the influence of garment in idiopathic scoliosis," *Journal of Clinical Medicine*, vol. 12, no. 2, p. 499, 2023.

[8] N. Gupta, R. S., B. G., and A. Shetty, "Vertebral and intraspinal anomalies in Indian population with congenital scoliosis: A study of 119 consecutive patients," *Asian Spine Journal*, vol. 10, no. 2, pp. 276–281, Apr. 2016, doi: 10.4184/asj.2016.10.2.276.

[9] Bunnell, W. P. (1984). An objective criterion for scoliosis screening. *JBJS*, 66(9), 1381-1387.

[10] T. Kuru Çolak, "Adolesan idiopatik skolyozda Schroth yöntemine göre 3 boyutlu egzersizlerin etkinliği," M.S. thesis, Sağlık Bilimleri Enstitüsü, İstanbul Üniversitesi, İstanbul, Türkiye, 2020. [Online]. Available: https://nek.istanbul.edu.tr/ekos/TEZ/49328.pdf. Accessed: Aug. 10, 2025.

[11] D. M. Coelho, G. H. Bonagamba, and A. S. Oliveira, "Scoliometer measurements of patients with idiopathic scoliosis," *Brazilian Journal of Physical Therapy*, vol. 17, no. 2, pp. 179–184, Mar.–Apr. 2013, doi: 10.1590/S1413-35552012005000081.

[12] "Skolyoz (omurga eğriliği)," *Terapi Room*. [Online]. Available: https://terapiroom.com/en/skolyoz-omurga-egriligi/. Accessed: Aug. 10, 2025.

[13] "Cobb angle," *Radiopaedia*. [Online]. Available: https://radiopaedia.org/articles/cobb-angle. Accessed: Aug. 10, 2025.

[14] J. Contreras, D. Gil, J. de Dios Errázuriz, P. Ruiz, C. Díaz, P. Águila, and F. Soza, "Valores de referencia del ángulo de balance escapular en población sana," *Revista Española de Cirugía Ortopédica y Traumatología*, vol. 58, no. 1, pp. 24–30, 2014, doi: 10.1016/j.recot.2013.09.009.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Real-time 3D Human Pose Estimation Model for Physiotherapy Exercises on Mobile Devices

Hamdi Yalın Yalıç, Alaettin Uçan, Ali Yaşar Yiğit, Adem Ali Yılmaz
Department of Research and Development
Tiga Information Technologies
Ankara, Türkiye
{yalin.yalic, alaettin.ucan, ali, adem}@tigahealth.com

***Abstract*—We present a lightweight, real-time 3D human-pose estimation model specific to mobile devices that can be used for physiotherapy exercises. Our two-stage training strategy first learns 2D joint coordinates via an EfficientNetB2-based backbone, then freezes those weights and adds a compact regressor for the third (depth) dimension. Deployed in ONNX format, the model delivers full-body 3D poses with 30 joints, which is more comprehensive than common 17-point joints, while achieving accuracy on par with state-of-the-art methods. Extensive experiments on the public Human3.6M dataset and the TheraPose physiotherapy dataset demonstrate real-time inference (less than 15 ms on ARM CPUs) and mean per-joint position errors of less than 40 mm, making it ideal for mobile health and exercise-monitoring applications.**

***Keywords*—3D human pose estimation, physiotherapy exercises, lightweight model optimization.**

I. Introduction

3D human pose estimation has become more popular due to the rise of mobile health and remote exercise apps. To track full-body motion, study biomechanics, and give real-time feedback during exercise routines, you need accurate 3D pose information. This is especially important in physiotherapy, where even small changes in joint angles or posture can have big effects on how well the treatment works. 2D pose estimation works well for basic posture tracking, but it doesn't give us the depth information we need to measure range of motion, limb orientation, and spinal curvature, which are all important parts of therapeutic assessments and rehabilitation exercises.

Using 3D pose estimation models on mobile devices could make physiotherapy more accessible to everyone by allowing low-cost, real-time analysis without the need for special hardware. But getting to this vision is not without its own set of problems. Mobile and embedded platforms are limited in their capabilities because they do not have enough processing power, memory, or battery life. Most of the time, state-of-the-art (SOTA) 3D pose estimation models are big, require a lot of processing power, and are designed for desktop GPUs, which means they cannot be used directly on mobile devices.

To address these limitations, we present a real-time 3D pose estimation model explicitly optimized for mobile devices. Our method uses a two-stage training strategy: we first train a 2D joint coordinate regressor using an EfficientNetB2 backbone, then freeze its weights and train a compact depth regressor to estimate the Z-coordinates. This modular approach makes optimization easier, reduces the cost of computation, and improves generalization in different poses. The model gives 30 joint coordinates, which are visualized in Figure 1, in 3D, providing more detailed information than most 18-joint COCO-based systems. This long skeleton has four spine-related joints that are very important to check and fix posture, balance, and alignment during physiotherapy exercises. This level of detail makes it easier to monitor areas such as lumbar extension, thoracic flexion, and cervical adjustments, which are areas that traditional skeleton layouts often overlook.

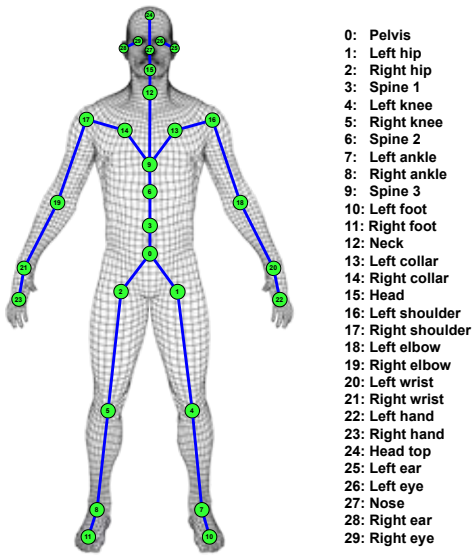


Fig. 1: Human pose model joint points

We use a number of optimization methods, such as post-training quantization and exporting to the ONNX format [1], to ensure that edge devices work in real time. Our last model is less than 10 MB in size and can process frames in less than 15 milliseconds on mid-range ARM CPUs. It also has comparable accuracy on 3D pose benchmarks.

In summary, our key contributions are:

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

- A two-stage 2D→3D training pipeline that decouples depth learning, improving efficiency and robustness.
- A lightweight EfficientNetB2-based architecture with ONNX deployment support, under 10 MB in size.
- A 30-joint pose representation specifically designed for physiotherapy, with enhanced spinal joint coverage.
- Real-time inference on commodity mobile hardware with competitive accuracy compared to existing 3D pose estimation models.

II. Related Work

There have been a number of attempts to estimate poses on platforms with limited resources, with a focus on real-time 2D pose estimation. BlazePose by Bazarevsky et al. [2] was the first landmark model made for mobile phones. It ran at over 30 FPS on a Pixel 2 device. It predicts 33 full-body landmarks using a regression head without a heatmap, which makes it very useful for real-time uses like fitness tracking. Zhang et al.'s MobiPose [3] also aims to improve multi-person 2D pose estimation on smartphones by combining motion tracking, lightweight convolutional architectures, and parallel scheduling for mobile CPUs. These works show excellent real-time 2D performance, but they do not estimate 3D joint positions, which are very important for understanding depth and posture in physiotherapy situations. Another work has explored two-stage architectures that decouple 2D estimation and 3D lifting. Zhao et al. [4] proposed a transformer-based 3D regressor that leverages frozen 2D features and contextual image cues, improving depth accuracy while maintaining modularity.

Researchers have explored methods to compress and deploy models to bring 3D pose estimation to mobile devices. Hwang et al.'s MoVNect [5] uses knowledge distillation to train a MobileNetV2-based model to figure out the 3D pose of one person. MoVNect runs at more than 20 frames per second on mobile CPUs when used with joint-angle filters for smoothing over time. In the same way, Choi et al.'s MobileHumanPose [6] uses a U-Net-like structure with skip connections and parametric activations to get 3D estimates in less than 13 ms per frame. In TransSMPL [7], a compressed transformer-based SMPL model, other methods like pruning and quantization have been looked into. TransSMPL uses early-exit logic and dynamic quantization to significantly reduce memory use while still accurately estimating the shape and pose of a 3D body.

Recent works have begun customizing pose estimation systems for clinical and rehabilitation use. Mercadal-Baudart et al. [8] demonstrated that a lightweight transformer model trained on exercise data could achieve high accuracy (less than 10 degrees error) in estimating joint angles critical to physiotherapy, such as pelvic tilt and knee flexion. They use 3D keypoints estimated from monocular RGB to measure full-body movement during rehabilitation exercises. Lee et al. [9] did another study that was based on physiotherapy. They used a 2D-to-3D lifting pipeline to find forward head posture, a common musculoskeletal problem, and got promising classification results using only upper-body joint data. These studies show how useful high-resolution joint tracking is for clinical feedback, but most of them use pre-trained general-purpose skeletons (like 17 joints), which might miss important spinal or therapy-specific areas.

III. Methodology

A. Model Architecture

Our model features a two-stage architecture designed to accelerate and enhance the accuracy of estimating 3D human poses. We use an EfficientNet-B2 backbone as the basis for our method. This backbone is a lightweight but powerful feature extractor. Two task-specific heads come after this backbone: a 2D pose estimation head and a depth regression head. Model architecture is illustrated in Figure 2.

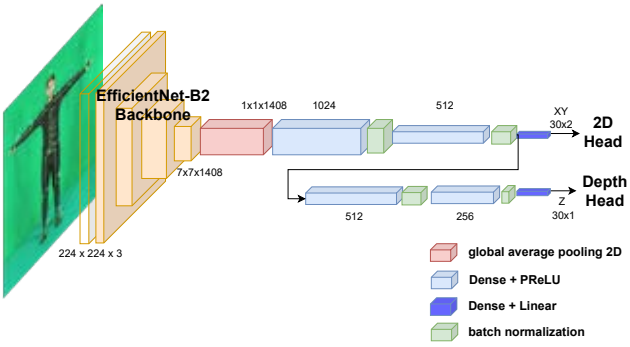


Fig. 2: 3D Pose Estimation Model Architecture

1) 2D Head: After the convolutional layers of EfficientNet-B2, the extracted feature maps are globally averaged and passed through two fully connected layers, each followed by batch normalization and a Parametric ReLU (PReLU) activation function. The last layer of output is a dense linear layer with 60 units, which are the (x, y) coordinates of 30 joints. We chose PReLU as the activation function because it learns the negative slope of the activation on its own. This makes it easier to generalize across different poses and improves gradient flow in small models, which is especially useful for mobile devices that have limited computational resources.

2) Depth Head: After training and freezing the 2D backbone and prediction layers, a second head is added to the 2D output. This head gets the flattened 2D joint coordinates as input and sends them through two dense layers, each of which is followed by batch normalization and PReLU activation. The last linear layer gives the Z-coordinates of the 30 joints, which offers complete 3D spatial information. This modular design allows for efficient fine-tuning of the depth regressor while preserving 2D accuracy.

B. Loss Functions

1) 2D Pose Loss: During the first stage, the model is trained to minimize the mean squared 2D distance between predicted and ground truth keypoints. Let $\hat{Y}_{2D} = \{(\hat{x}_i, \hat{y}_i)\}_{i=1}^{30}$ and

$Y_{2D} = \{(x_i, y_i)\}_{i=1}^{30}$ be the predicted and true 2D coordinates, respectively. The loss is defined as:

$$\mathcal{L}_{2D} = \frac{1}{30} \sum_{i=1}^{30} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

This function encourages the network to minimize the Euclidean distance between joint predictions in image space.

2) 3D Fine-tuning Loss: In the second stage, we train the Z-coordinate head while keeping the 2D model frozen. The loss function is a combination of MSE on Z and a small-weighted consistency term with the already learned 2D outputs. Let $\hat{Z} = \{\hat{z}_i\}_{i=1}^{30}$, $Z = \{z_i\}_{i=1}^{30}$, and reuse $\hat{Y}_{2D}, Y_{2D}$ as before. The loss becomes:

$$\mathcal{L}_{3D} = \underbrace{\frac{1}{30} \sum_{i=1}^{30} (z_i - \hat{z}_i)^2}_{\text{Z loss}} + \lambda \cdot \underbrace{\frac{1}{30} \sum_{i=1}^{30} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]}_{\text{Reprojection consistency}}$$

Here, λ is a small scalar (e.g., 1×10^{-6}) to preserve the previously learned 2D accuracy while focusing training on the Z dimension.

C. Data Preparation

Our model is trained on a curated subset of the TheraPose [10] dataset, which provides synchronized RGB frames and 3D joint annotations from an IMU-based motion capture system. The full dataset contains millions of frames covering over 120 physiotherapy exercises; we selected approximately 177,000 representative samples covering a wide range of body movements, including those critical to spinal alignment, balance, and posture. This subset was manually reviewed and corrected to ensure annotation quality, particularly for spine and limb joints.

We applied data augmentation using the Albumentations [11] library to improve model robustness. Changes in hue and saturation, random changes in brightness and contrast, and limited-angle rotations are all part of the transformations. These additions ensure spatial consistency between images and labels by considering keypoints. All images are resized to 224×224, and the keypoints are also scaled to fit this size. Before being fed into the network, the RGB inputs are brought down to the range of [0,1]. This standardization makes sure that it will work the same way on all devices and input resolutions.

D. ONNX Export and Mobile Optimization

We used the `tf2onnx` library to convert our trained TensorFlow model into ONNX format for real-time inference on mobile devices. Post-training quantization improved the model even more by lowering memory and computing overhead without losing much accuracy. We used the XNNPACK delegate with ONNX Runtime for inference. This is the best option for ARM CPU architectures. This setup can process frames in less than 15 milliseconds on mid-range mobile devices, which shows that it is possible to run full 3D pose estimation pipelines in real time, even in edge devices where GPUs are not available.

IV. Experiments

We trained the proposed model in two phases, using the Adam optimizer with a starting learning rate of 1×10^{-3} for 2D training and a lower rate of 1×10^{-4} for the Z-coordinate regressor. We set the batch size to 32 and the input resolution to 224×224. Training was performed on an NVIDIA RTX A6000 ADA GPU with 48GB of VRAM, enabling the processing of large batches and high-resolution data efficiently. We used early stopping with a patience of 40 epochs during training to keep from overfitting and speed up convergence, in a total of 200 epochs of training. In addition, a learning rate scheduler was used to reduce the learning rate by half if the validation loss remained unchanged, which made the model more stable during fine-tuning. Both phases were monitored using validation loss, and the best-performing weights were saved for deployment. An example output of the model on a single RGB frame is visualized in Figure 3, which contains both the pose model for the RGB input frame drawn on top of it, and a slight side-angle representation of the pose within the 3D grid.

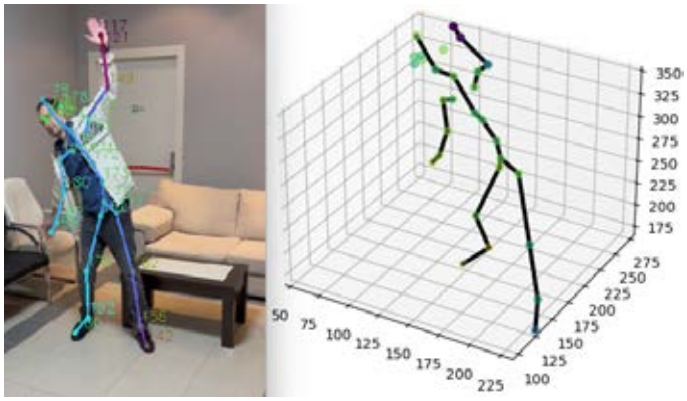


Fig. 3: Model output visualization overlayed on the RGB image and inside the 3D grid

We evaluated multiple backbone architectures to identify the optimal trade-off between accuracy and efficiency on our curated physiotherapy dataset. We used the MPJPE (Mean Per Joint Position Error) metric, which is defined as the average Euclidean distance between the predicted and ground truth 3D joint positions across all joints and frames. Among the tested variants, as shown in Table I, EfficientNet-B2 achieved the lowest 3D error (45.4 mm), outperforming both smaller (B0, B1) and larger (B3) EfficientNet versions. While EfficientNet-B3 has a higher parameter count, it surprisingly produced the worst accuracy, suggesting overfitting or reduced generalization in our scenario. MobileNet-v2 offered a good balance between compactness and accuracy, whereas MobileNet-v3Small yielded the smallest ONNX model size, but at the cost of higher error. As a result, EfficientNet-B2 was selected for the final model due to its favorable accuracy-to-size trade-off.

To enable a fair comparison with prior work, we retrained our model on the Human3.6M dataset using its standard 16-joint layout. The results are given in Table II. Our model

TABLE I: Training results using various backbones

Backbone	3D Error (MPJPE)	# Params (M)	ONNX Model size (MB)
EfficientNet-B0	50.3	5.8	6.2
EfficientNet-B1	48.7	8.3	8.8
EfficientNet-B2	**45.4**	9.7	9.8
EfficientNet-B3	83.4	12.5	13.0
MobileNet-v2	49.5	4.2	4.5
MobileNet-v3Small	61.2	**3.1**	**3.3**
MobileNet-v3Large	70.5	5.9	6.3

achieved a competitive 3D error of 39.8 mm MPJPE, on par with SOTA methods like PoseMamba [12], while maintaining real-time inference speed (<15 ms per frame) on mobile devices. Most lightweight models in the literature either omit runtime benchmarks on mobile platforms or do not report evaluation results on Human3.6M. For example, models like P-STMO [13] and Tiny-HourglassNet [14] do not share latency metrics, making real-time deployment viability unclear. In contrast, our model is one of the few that delivers both high accuracy and verified mobile efficiency, highlighting its practical value for edge-based physiotherapy applications.

TABLE II: Lightweight models comparison on Human3.6M dataset

Models	3D Error (MPJPE)	# Params (M)	Inference time (ms/frame) on mobile device
PoseMamba	**38.1**	6.7	-
P-STMO	42.8	6.2	-
MobileHumanPose	51.4	5.6	∼30–40
MoVNect	97.3	**1.3**	∼22
Tiny-Hourglassnet	64.9	-	-
Ours	39.8	9.8	**< 15**

V. Conclusion

In this work, we introduced a lightweight and mobile-optimized 3D human pose estimation model tailored for physiotherapy applications. By adopting a two-stage training strategy and selecting an EfficientNet-B2 backbone, we achieved high accuracy with real-time inference performance on mobile devices. Our model offers full-body 30-joint coverage, including critical spinal landmarks, and demonstrates competitive results compared to SOTA lightweight methods. In future work, we aim to enhance robustness under occlusion by incorporating temporal priors and visibility-aware modules, and explore partial body models that can estimate joint confidence scores when certain limbs or body parts are not visible.

References

[1] *ONNX — open neural network exchange*, https://onnx.ai/, [Accessed 29-07-2025].

[2] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020.

[3] J. Zhang *et al.*, "Mobipose: Real-time multi-person pose estimation on mobile devices," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 136–149.

[4] Q. Zhao, C. Zheng, M. Liu, and C. Chen, "A single 2d pose with context is worth hundreds for 3d human pose estimation," *Advances in Neural Information Processing Systems*, vol. 36, pp. 27 394–27 413, 2023.

[5] D.-H. Hwang, S. Kim, N. Monet, H. Koike, and S. Bae, "Lightweight 3d human pose estimation network training using teacher-student learning," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, pp. 479–488.

[6] S. Choi, S. Choi, and C. Kim, "Mobilehumanpose: Toward real-time 3d human pose estimation in mobile devices," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2328–2338.

[7] Y. Kim, H. Yoo, J.-H. Ryu, S. Lee, J. H. Lee, and J. Kim, "Transsmpl: Efficient human pose estimation with pruned and quantized transformer networks," *Electronics*, vol. 13, no. 24, p. 4980, 2024.

[8] C. Mercadal-Baudart *et al.*, "Exercise quantification from single camera view markerless 3d pose estimation," *Heliyon*, vol. 10, no. 6, 2024.

[9] H. Lee, B. Oh, S.-C. Kim, *et al.*, "Recognition of forward head posture through 3d human pose estimation with a graph convolutional network: Development and feasibility study," *JMIR Formative Research*, vol. 8, no. 1, e55476, 2024.

[10] H. Y. Yalıç, A. E. Usta, Ü. Atila, A. Uçan, A. A. Yılmaz, and A. Y. Yiğit, "Therapose: A large video dataset for physiotherapy exercises," in *In Proceedings of 15th Turkish Congress of Medical Informatics*, 2024, pp. 88–97.

[11] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.

[12] Y. Huang, J. Liu, K. Xian, and R. C. Qiu, "Posemamba: Monocular 3d human pose estimation with bidirectional global-local spatio-temporal state space model," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 3842–3850.

[13] W. Shan, Z. Liu, X. Zhang, S. Wang, S. Ma, and W. Gao, "P-stmo: Pre-trained spatial temporal many-to-one model for 3d human pose estimation," in *European Conference on Computer Vision*, Springer, 2022, pp. 461–478.

[14] B. Shi *et al.*, "Tiny-hourglassnet: An efficient design for 3d human pose estimation," in *2020 IEEE international conference on image processing (ICIP)*, IEEE, 2020, pp. 1491–1495.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Enhanced Tumor Classification via Radiomics and Deep Feature Generation using Neural Network

Integrating Radiomics with CNN-Based Deep Features

Dilek Betül SARIDEDE
Biomedical Engineering Department, Faculty of Engineering and Natural Sciences, İstanbul Atlas University, İstanbul, Türkiye
dilek.saridede@atlas.edu.tr

Rukiye Nur KAÇMAZ
Software Engineering Department Faculty of Engineering, Erciyes University, Kayseri, Türkiye
rukiyekacmaz@erciyes.edu.tr

Sevim CENGİZ
College of Technological Innovation, Zayed University, Dubai, United Arab Emirates
sevim.Cengiz@zu.ac.ae

***Abstract*—Glioblastoma (GBM) is the most aggressive and prevalent malignant brain tumor in adults. Accurate preoperative differentiation between GBM and other gliomas remains a critical challenge due to imaging overlaps and tumor heterogeneity. Radiomics has emerged as a powerful tool to extract quantitative imaging features; however, most existing approaches rely on traditional radiomic features derived from limited modalities or tumor regions. In this study, we have proposed a novel hybrid pipeline that integrates traditional radiomic features with deep features extracted using a convolutional neural network (CNN). We used multimodal magnetic resonance images and four pre-segmented tumor subregions (necrosis, edema, enhancing tumor, and whole tumor) from 185 brain tumor patients publicly available in The Cancer Imaging Archive (141 GBM, 44 non-GBM). A total of 1544 radiomic features were extracted using the PyRadiomics library for each patient. After applying FeatureWiz for feature selection, a Support Vector Machine classifier was trained using only selected traditional radiomic features and both selected traditional radiomic features and CNN-based deep features. Our results demonstrate that combining traditional radiomic and deep features significantly improves classification performance. Validation accuracy increased from 0.838 to 0.948, F1-score from 0.558 to 0.948, and recall from 0.436 to 0.950. These improvements indicate better generalization and robustness of the hybrid model. This study highlights the potential of integrating deep and traditional radiomic features for glioblastoma classification. Our hybrid pipeline addresses limitations of existing radiomics-only approaches and provides a promising direction for non-invasive tumor characterization in neuro-oncology.**

Keywords: Radiomic Features; Machine Learning; Medical Imaging; Feature Engineering

I. INTRODUCTION

Gliomas are the most common primary tumors of the central nervous system, arising from glial cells in the brain or spinal cord. Among them, glioblastoma (GBM) is the most aggressive and prevalent malignant subtype in adults, classified as a World Health Organization (WHO) grade IV glioma [1]. Despite significant advances in neuro-oncology, accurate preoperative differentiation between glioblastoma and other gliomas such as lower-grade gliomas or astrocytomas remains a clinical challenge due to overlapping imaging features and tumor heterogeneity [2].

Radiomics has emerged as a promising approach in computational oncology, enabling the extraction of high-dimensional quantitative features from medical imaging data to detect subtle patterns of tumor phenotype and heterogeneity [3]. It has shown considerable potential in differentiating glioblastoma from lower-grade gliomas and in survival prediction tasks. Nonetheless, many prior studies have relied exclusively on traditional features extracted from either a single MRI modality or non-specific tumor masks [4–6]. Akbari et al. [5] addressed this limitation by incorporating both intratumoral and peritumoral regions in feature extraction, demonstrating that spatial context can enhance predictive performance. However, their study remained limited to classical radiomics-based approaches. More recently, Wan et al. [4] evaluated survival prediction in recurrent glioblastoma patients by comparing traditional radiomic signatures with an end-to-end CNN model, but without integrating traditional and deep features. Similarly, Tan et al. [2] explored the impact of different volume-of-interest fusion strategies for glioma grading, yet their approach was also restricted to traditional features. Hassan et al. [7] proposed a multi-modal integration strategy about traditional radiomic features and deep features generated via a CNN; however, their study focused on dementia subtyping using multi-omics data and structural neuroimaging. It did not address glioblastoma classification. Similarly, Kobayashi et al. [8] and Zhou et al. [9] investigated the integration of deep learning and radiomics for glioblastoma classification, but their CNN components were primarily designed either as full classification models or to enhance interpretability rather than as dedicated feature extractors. Therefore, advancing glioblastoma-specific classification models that combine traditional and learned features from multimodal, multi-region MRI inputs remains an important research direction.

To address this limitation, we propose a novel hybrid pipeline that integrates traditional radiomic features extracted from four magnetic resonance imaging (MRI) modalities and four tumor subregions (necrosis, edema, enhancing tumor and whole tumor) with deep features generated via a CNN. By combining both data-driven and domain-specific features, our

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

approach aims to improve the diagnostic performance in distinguishing glioblastoma from non-glioblastoma tumors.

II. MATERIALS AND METHODS

A. Subjects and MRI Acquisition

Clinical information and MR images of 185 brain tumor patients, comprising 141 confirmed and glioblastoma and 44 non-glioblastoma tumors were obtained from The Cancer Imaging Archive [https://www.cancerimagingarchive.net/collection/mu-glioma-post/] [10].

Clinical information used in this study included age and sex information of the patients (Table 1). T1-weighted, T1-weighted post-contrast, T2-weighted and Fluid-attenuated inversion recovery (FLAIR) MR images were used. The dataset provides manually pre-segmented tumor subregions, including necrosis, edema, and enhancing tumor regions.

Table 1. Demographic information of the patients.

Patient Group	Age	Gender (F:M)
Patient with GBM tumors	63.19±12.14	55:68
Patient with non-glioblastoma tumors	41.27±15.95	23:21

B. Radiomic Feature Extraction

Radiomic features were extracted using a custom Python script utilizing the PyRadiomics 2.2.0 library [https://github.com/Radiomics/pyradiomics]. Feature categories included: 3D shape features (n = 17), first-order statistics (n = 19), and higher-order texture features (n = 75). Due to high collinearity among certain shape and first-order features, a reduced set of features was used in the analysis. Specifically, compactness, compactness 2, and spherical disproportion were removed from the shape group, and standard deviation was removed from the first-order group due to its strong correlation with variance. Ultimately, 14 shape features and 18 first-order statistics features were retained. Shape features, being contrast-invariant, were extracted once per segmented compartment per patient. All other features were extracted across 4 MRI modalities (T1, T1-contrast, T2, and FLAIR) and 4 tumor subregions (necrosis, edema, enhancing tumor and whole tumor).Thus, the total feature number of each patient equals 1544. Fig. 1 shows that multimodal MR images with manually segmented tumor subregions (edema, enhancing tumor, necrosis/non-enhancing tumor) were used to extract shape, first-order, and texture features. The extracted features were combined into a structured tabular format, where each row corresponds to a patient and each column to a specific radiomic feature. This table served as the input for the classification.

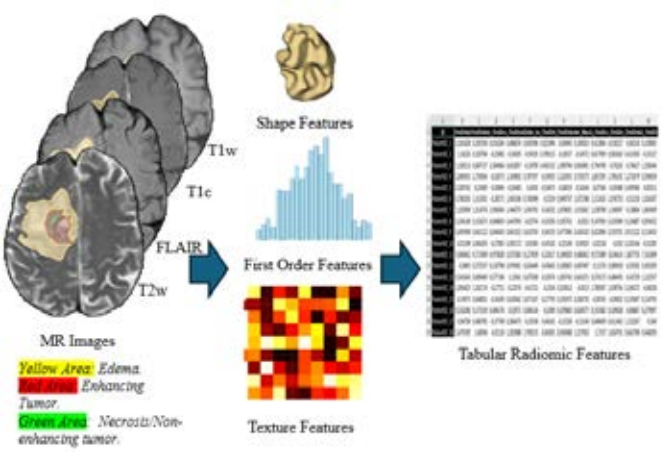


Fig. 1. The pipeline of the Radiomic Feature Extraction.

C. Classification

Fig. 2 shows the overview of the classification pipeline. Selected traditional radiomic features were directly used as input for machine learning algorithms (a). In the proposed model (b), the same features were passed through a CNN consisting of convolution, pooling, and flattening layers to extract deep features. These CNN-based deep features were then concatenated with the original radiomic features to form a combined feature set. After applying feature selection, the refined feature sets were used to train machine learning model for tumor classification.

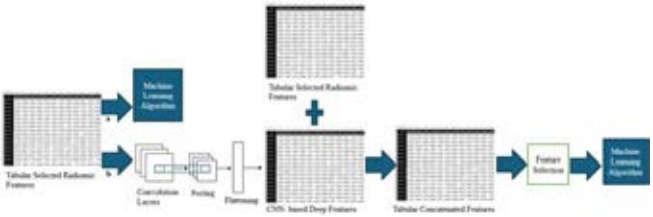


Fig. 2. The pipeline for the classification.

Firstly, the extracted features were preprocessed by replacing missing values using median imputation within each class. Featurewiz was applied for automated feature selection (correlation threshold = 0.85). Class imbalance was addressed with Synthetic Minority Over-sampling Technique (SMOTE) oversampling in the training dataset. In small datasets, particularly common in medical research, applying SMOTE effectively improves model performance by mitigating class imbalance [11]. Support Vector Machine (SVM) Learning Model was employed. SVM is frequently highlighted as a preferred classifier in SMOTE-augmented, small-sample applications due to their robustness and generalization capabilities [12]. Performance metrics included train accuracy, validation accuracy, F1-score, Receiver-operating characteristic curve (ROC) Area under the curve (AUC), precision, recall (sensitivity) and specificity were evaluated. In

the proposed model, to enhance the feature space, Deep Feature Generation (DFG) was proposed. DFG employed CNN as a feature extractor. After the application of Featurewizz, the selected traditional radiomic features were scaled and reshaped to fit a CNN which was constructed with two convolutional layers followed by global average pooling. A dense (fully connected) layer with 64 units was used to represent deep features: Input → Conv1D(32) → Dropout → Conv1D(16) → GlobalAvgPooling → Dense(64). These CNN-based deep features were then concatenated with the traditional radiomic features. The combined feature set was reprocessed using featurewiz and the same ML classifier was re-applied. Then, performance metrics were evaluated. As with the classical pipeline, oversampling (SMOTE), normalization, and inclusion of demographic variables (sex) were carefully handled during training.

III. RESULTS

There was a clear increase in performance metrics including validation accuracy, F1 score, ROC AUC, precision, and recall when concatenated features (Selected Traditional Radiomic Features + CNN-based Deep Features) were used, compared with using either Selected Traditional Radiomic Features or CNN-based Deep Features alone (Table 2). Specifically, when Concatenated Features (Selected Traditional Radiomic Features + CNN-based Deep Features) were compared with Selected Traditional Radiomic Features alone, F1 score improved from 0.558 to 0.948, recall from 0.436 to 0.950, and ROC AUC from 0.860 to 0.981. Although train accuracy rose only slightly (from 0.992 to 0.998), validation accuracy showed a more meaningful increase (from 0.838 to 0.948). On the other hand, using CNN-based Deep Features alone resulted in much lower performance (validation accuracy 0.503, F1 0.505, ROC AUC 0.379).

Table 2. Performance Comparison of SVM Classifier Using Selected Traditional Features vs. Concatenated Features (Selected Traditional Features + CNN-based Deep Features).

Performance Metrics	Selected Traditional Features	CNN-based Deep Features	Concatenated Features (Selected Traditional Features + CNN-based Deep Features)
Train Accuracy	0.992	0.527	0.998
Validation Accuracy	0.838	0.503	0.948
F1 Score	0.558	0.505	0.948
ROC AUC	0.860	0.379	0.981
Precision	0.797	0.584	0.948
Recall (Sensitivity)	0.436	0.364	0.950
Specificity	0.965	0.646	0.946

Fig. 3 shows that the learning curve shows the differences in training and validation accuracy across all 5 folds when using the concatenated feature set. In each fold, both training and validation accuracy remain consistently high.



Fig. 3. Differences in accuracy of training and testing applied by SVM with Concatenated Features (Selected Features + DFG).

IV. DISCUSSION

In this study, we propose a hybrid classification pipeline that integrates traditional radiomic features with deep features generated via a CNN to distinguish glioblastoma from non-glioblastoma brain tumors. Our results demonstrate that combining traditional radiomic features with data-driven deep features significantly enhances classification performance across multiple evaluation metrics.

Our results demonstrate that combining radiomic features with deep generation features significantly enhances classification performance across multiple evaluation metrics. While training accuracy remained consistently high (0.992 vs. 0.998), the validation accuracy improved markedly from 0.838 to 0.948 when using the concatenated feature set, suggesting enhanced generalizability and reduced overfitting. This finding aligns with prior work highlighting the complementary value of integrating deep learning and traditional radiomics for tumor classification. This finding aligns with prior work highlighting the complementary value of integrating deep learning and radiomics for tumor classification [8,9]. However, unlike these studies, where CNNs were primarily used either as the main classification model or for interpretability purposes, our approach employs a CNN as a dedicated feature extractor operating on multi-region, multi-modality traditional radiomic features. The resulting deep features are systematically fused with pre-selected traditional radiomic features, followed by an additional feature selection step to enhance generalizability and reduce overfitting. The result of this study shows that the relatively small gap between training and validation curves, which suggests that the model generalizes well to unseen data. This stability indicates that integrating CNN-based deep features with the selected features contributes positively to model robustness and predictive performance. Furthermore, our findings resonate with the conceptual basis of radiomics defined by Lambin et al. [3], who underlined the value of quantitative imaging biomarkers, and with Akbari et al. [5], who highlighted the importance of spatial context in radiomic feature design. Similarly, Tan et al. [2] and Wan et al. [4] showed that combining multi-region or multi-modal features

improves grading and survival prediction tasks, though their studies lacked deep learning integration. By incorporating a CNN-based deep feature generator into the radiomic pipeline, our model bridges this gap and complements previous traditional-only studies like Chaddad et al. [6].

The proposed model used FeatureWiz-selected radiomic features (about 300–400) instead of raw 3D MRI images. Because CNN operated only on this small, pre-filtered table derived from the traditional radiomic features, the computational cost remained very low: it required only a few hundred thousand parameters, trained within minutes on a standard GPU, and needed far less memory than image-based CNNs. The CNN-based deep features were simply new hidden representations learned from these selected radiomics, not a second round of image processing. Our results showed that using only Selected Traditional Features yielded about 84% validation accuracy, using CNN-based deep features alone yielded about 62%, while combining both raised validation accuracy to about 95%. Similar findings in other study have reported that direct 3D CNNs often perform worse on small medical datasets [7]. Therefore, radiomic features were enriched with deep learning rather than extracting deep features directly from images.

To the best of our knowledge, dataset used in this study was most recently updated on March 21, 2025 and has not yet been cited or analyzed in any published studies. Therefore, no directly comparable classification results are currently available. [10]

Despite these promising results, there are several limitations including small dataset size and the use of neural network architecture which may restrict generalizability and highlight the need for further validation across diverse cohorts.

V. CONCLUSION

This study presents a novel classification framework that combines traditional radiomic features and CNN-based deep features, achieving notable improvements in diagnostic performance for distinguishing glioblastoma from non-glioblastoma tumors. The results support our hypothesis that integrating domain-driven and data-driven features enhances the representational power of the model and improves generalization, particularly in small, imbalanced medical datasets. Future work may explore external validation cohorts and the extension of this framework to other tumor types or prognostic tasks, further advancing precision diagnostics in neuro-oncology.

REFERENCES

[1] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization classification of tumors of the central nervous system: A summary," Acta Neuropathologica, vol. 131, no. 6, pp. 803–820, 2016, doi: 10.1007/s00401-016-1545-1.

[2] R. Tan, C. Sui, C. Wang, and T. Zhu, "MRI-based intratumoral and peritumoral radiomics for preoperative prediction of glioma grade: A multicenter study," Frontiers in Oncology, vol. 14, p. 1401977, 2024, doi: 10.3389/fonc.2024.1401977.

[3] P. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," European Journal of Cancer, vol. 48, no. 4, pp. 441–446, 2012, doi: 10.1016/j.ejca.2011.11.036.

[4] Q. Wan et al., "Comparative analysis of deep learning and radiomic signatures for overall survival prediction in recurrent high-grade glioma treated with immunotherapy," Cancer Imaging, vol. 25, no. 5, 2025, doi: 10.1186/s40644-024-00818-0.

[5] H. Akbari et al., "Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma," Neurosurgery, vol. 78, no. 4, pp. 572–580, 2016, doi: 10.1227/NEU.0000000000001102.

[6] A. Chaddad, C. Desrosiers, and M. Toews, "Radiomic analysis of multi-contrast brain MRI for the prediction of survival in patients with glioblastoma," IEEE Access, vol. 6, pp. 73031–73041, 2018, doi: 10.1109/ACCESS.2018.2880542.

[7] S. Hassan et al., "MINDSETS: Multi-omics integration with neuroimaging for dementia subtyping and effective temporal study," Scientific Reports, vol. 15, no. 15835, 2025, doi: 10.1038/s41598-025-97674-0.

[8] Z. Hu et al., "A deep learning model with radiomics analysis integration for glioblastoma post-resection survival prediction," arXiv preprint, arXiv:2203.05891, 2022.

[9] Z. A. Shboul, M. Alam, L. Vidyaratne, L. Pei, M. I. Elbakary, and K. M. Iftekharuddin, "Feature-Guided Deep Radiomics for Glioblastoma Patient Survival Prediction," Frontiers in Neuroscience, vol. 13, p. 966, 2019, doi: 10.3389/fnins.2019.00966.

[10] D. Yaseen et al., "University of Missouri post-operative glioma dataset (MU-Glioma-Post) (Version 1) [Data set]," The Cancer Imaging Archive, 2025, doi: 10.7937/7K9K-3C83.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[12] Q. Wang, Z. Luo, J. Huang, Y. Feng, and Z. Liu, "A novel ensemble method for imbalanced data learning: Bagging of extrapolation-SMOTE SVM," Computational Intelligence and Neuroscience, vol. 2017, Article ID 1827016, 2017, doi: 10.1155/2017/1827016.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

An AI-Supported Hospital Laundry Processing System (AIHLPS)

Ibrahim Zincir, Kaan Kurtel, Ufuk Celikkan
Izmir University of Economics, Izmir, Turkiye

***Abstract*—To ensure a hygienic and safe healthcare environment, all operational processes, including the collection, cleaning, and storage of patient linens, are strictly regulated in accordance with standards and policies. Personnel must clearly understand and implement these standards and policies to effectively prevent contamination. Healthcare facilities maintain these regulations written in natural language, in many cases containing ambiguous information. They also must contend with local, national, and international variations in these regulations. This further exacerbates ambiguities. Artificial Intelligence (AI) and natural language processing serve as a bridge between human language and ICT-based applications. They can provide faster, more precise, and more adaptable solutions to ever-changing conditions. This study proposes an artificial intelligence supported laundry processing system (AIHLPS) to do risk analysis tailored to the regulatory and operational risk factors associated with hospital laundry management. The framework processes various text documents using a Large Language Model (LLM) and classifies laundry items via its Decision Model to determine washing parameters.**

Keywords—medical laundry, hospital laundry management, large language model, artifical intelligence

I. Introduction

Pinpointing the exact origin of hospital-acquired infections can be challenging due to a wide range of possible transmission routes within hospital settings—one of which is patient laundry. It is reasonable to consider hospital linens as potential carriers of infectious pathogens. These textiles commonly include bedding, towels, and clothing used by both patients and staff, such as scrubs, lab coats, uniforms, patient gowns, and pajamas, as well as surgical drapes and curtains. In response to growing concerns, hospital and healthcare facility administrators are placing increased emphasis on the role of nurses and staff in ensuring proper management and handling of medical laundry. Staff must understand and follow detailed protocols for the transportation, cleaning, and storage of linens to avoid cross-contamination. It is also essential for them to distinguish between hygienically clean and fully sanitized laundry.

Several factors influence the hospital laundry cleaning processes. These include reducing human error, complying with diverse regulations, assessing linen conditions, responding quickly to changing circumstances, and integrating hospital-and healthcare systems. To emphasize the importance of quick reaction to changing conditions, consider a case where the hospital management needs to change washing and drying policies for healthcare textiles due to the challenges created by Monkeypox, which is spreading rapidly today, as it did with Covid-19 in the past. Also, consider that hospital laundry items have different cleaning needs due to their different contexts, and healthcare facilities follow different policies regarding reusing, washing, or disposing of these items. Additionally, be reminded that there are often conflicting local, national, or international regulations. International health organizations have published numerous technical guidelines defining the requirements for regulating laundry hygiene in hospitals. Despite some minor differences, these international standards generally share a common narrative. However, in cases of rapidly developing epidemics and migration events, they announce additional sub-rule making it difficult for hospitals to adapt. A similar situation applies to country-based domestic legislation and hospital-specific regulations.

These standards identify risk levels based on the condition of hospital linens. However, the overall system is sensitive to factors such as geographic variations, specific outbreak types, and patient volumes. This means that policies and guidelines for linen sterilization and disinfection can be affected by these factors, leaving healthcare facilities vulnerable to uncontrolled (or delayed) risks. This shows that risk levels must be adaptable to changing conditions.

Traditional manual classification is prone to human error, leading to inefficiencies and potential infection breaches. Our goal in this study is to address the aforementioned issues by proposing a framework called the Risk-Based AI Analysis Module which processes human-generated text to accurately classify the risk level of laundry items and set cleaning parameters accordingly using a decision model. We have primarily used four sources of information as input to the framework. 1) Numerous technical guidelines and regulations defining the requirements for hospital laundry hygiene published by international health organizations (i.e. the World Health Organization-WHO and the European Centre for Disease Prevention and Control-ECDC). 2) country-based domestic regulations (i.e. the US Centers for Disease Control and Prevention-CDC). 3) hospital-based policies, and 4) a rational risk classification of patient care instruments and items. The proposed framework uses large language models, with ChatGPT being a popular example.

II. Literature Review

Hospital laundry cleaning policy, rules, and guidelines for the reduction of risks related to laundry against potential or actual contamination with an infectious agent can be found in [1]-[7]. The effective management of hospital laundry is crucial in infection prevention, as underscored by numerous studies.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

According to [7], inadequate textile management directly correlates with increased healthcare-associated infections, emphasizing the criticality of proper laundry handling protocols. Moreover, Lee et al. stated that machine development models, which utilize continuous programming advancements, can help people accurately process their waste. They also noted that the speed and quantity of waste depletion will facilitate the effective increase of waste separation policies [8].

GenAI (Generative AI), a subset of artificial intelligence where artificial intelligence learns from existing data and generates new content based on the patterns and structures it learns [9], [10], is also revolutionizing medical informatics. Chopra [11] and Chakraborty [12] examined the transformative potential of ChatGPT, its development in generative AI, and its applications in healthcare. They highlighted the far-reaching impact of using these technologies in medical decision-making. Furthermore, identified that AI-based systems, especially those utilizing Large Language Models (LLMs), significantly enhance infection control measures by automating real-time data processing and improving compliance with hygiene protocols.

Baker et al. [13] discusses the overall suitability for an IoT healthcare system while evaluating the strengths and weaknesses of the system. Similarly, under the title Artificial Intelligence in Healthcare, [14], [15] discusses the opportunities, challenges and ethical issues in the application of AI to healthcare. Kumar et al. propose an Artificial Intelligence (AI)-based automated solution for COVID-related medical waste type classification and provide data-driven decisions for recycling [16]. Zainordi et al. demonstrated that integration of IoT (Internet of Things) and AI (Artificial Intelligence) technologies in hospital laundry systems leads to improved operational efficiencies [17].

As discussed before, hospitals produce vast quantities of contaminated linen daily, including patient gowns, bedding, surgical drapes, and staff uniforms. Contaminated laundry poses a significant risk for HAIs, potentially leading to severe patient health outcomes increased operational costs and the need for stringent handling protocols [1], [6]. Abney et al. define several factors affecting bacteria, viruses, and mold growth in laundry as fabric composition, storage conditions, usage, season, age of clothing, type of detergent, and several others. They also categorized pathogens detected in textiles before washing as bacteria, fungi, viruses, helminths and protozoa [18]. Considering that these categories also have subcategories, it becomes clear how difficult manual sorting is.

III. Hospital Laundry Processes

This section describes the conventional operations of hospital laundry based on [18] and [20], which helps us in understanding the basic requirements of an ICT-based solution. Laundry operations in a hospital are quite different from those in other domains such as hotel and hospitality. Even within a hospital, different units operate in different contexts, requiring varying cleaning parameters (for example, infectious diseases and maternity wards may use different pH and temperature values). In the light of these and other specific requirements, we propose a Hospital Laundry Management Process consisting of five major activities as illustrated in Fig. 1.

Laundry Collection, Pick-up and Sorting (LC): During the collection process, the staff collects laundry from hospital units and sort the items using the associated laundry information. Article [18] proposes a system that uses washable RFID tags to store this information and assist with item sorting.

Data Collection (DC): During the Data Collection activity all necessary information is gathered to enable the Risk Analysis activity to assign a risk level. This activity may involve the use of physical hardware such as RFID, for data acquisition, in addition to manual data provisioning. Medical staff may visually examine the items and manually enter information about each item's condition into a data repository (e.g., a database). Additionally, after visual inspection staff may make an immediate judgement, assign a risk level to the laundry item and dispose of it in appropriate color-coded collection bins.

Risk Analysis (RA): The primary function of Risk Analysis process is to assign a risk level to each item so that a decision can be made on how the item is cleaned or whether the item must be destroyed. Risk Analysis process uses artificial intelligence to assess the risk level of fabrics. This paper focuses on this process and discusses its details in the following sections.

User Interface (UI): This process facilitates interaction between human users and the rest of the system. Medical staff enter data described in the Data Collection activity using a user interface device, such as a mobile phone or a terminal. If a manual risk assignment is needed it is also performed by this process.

Cleaning (CP): The details of the cleaning activity are outside the scope of this paper. This study focuses on the Risk Analysis aspect of the hospital laundry handling processes.

IV. AI-Based Risk Analysis Module

AI-based risk analysis module learns, infers and makes decisions to ensure that the system follows the correct washing procedures. To achieve this the module must classify the linens using a defined classification scheme.

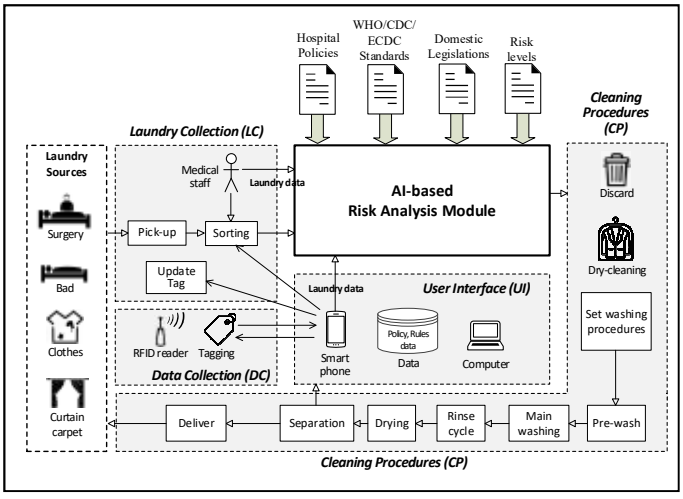


Fig. 1. Hospital Laundry Management Activities

TABLE I. SAMPLE DATASET

Sample	Class
Linens issued but returned unopened	Unused (I)
Blanket from a non-infectious orthopedic patient	Low Risk (II)
Gown from patient suspected of tuberculosis	Medium Risk (III)
Bed cover soaked with blood from emergency trauma unit	High Risk (IV)
Bed sheet returned from a patient in the oncology ward who had nosebleeds and was confirmed COVID-19 positive.	Discard (V)

Spaulding proposed the conventional risk classification in healthcare facilities approximately 65 years ago. Spaulding developed a rational approach to patient care instruments and items, by introducing a clear and logical classification consisting of three main categories: *Critical*, *Semi-critical*, and *Non-critical* [21]. Many organizations [22], [23] used this classification as the de facto standard. After analyzing a number of laundry handling policies we have concluded that the risk specifications given in various standards documents and Spaulding's work needed some update. For example, it is written in some policy documents that laundry was discarded without being cleaned due to its condition [4, p. 4] [23, pp. 10, 16] and in others it was returned to the store without being opened [4, p. 3]. Although rare compared to other types of contamination, any instance of radioactive contamination must be handled using entirely different procedures. For these reasons, a new risk categorization is needed. We leveraged Large Language Models (LLMs) to create five distinct risk classes optimized according to the guidelines from [3]-[7], [22]-[24] These new risk classes are *Unused (I), Low Risk (II), Medium Risk (III), High Risk (IV)* and *Discard (V)*.

The model uses the new five-level risk classes and the standards to train the dataset. A sample dataset is presented in Table I. As conditions and other environmental factors change over time, the dataset is updated and reused to update the generated model. This leads us to a scalable and adaptable model that enables the automatic classification of laundry in real time into five distinct risk-based classes. The generated LLM model ensures compliance with [3]-[7], [22]-[24] and processes real-time data to classify laundry accurately.

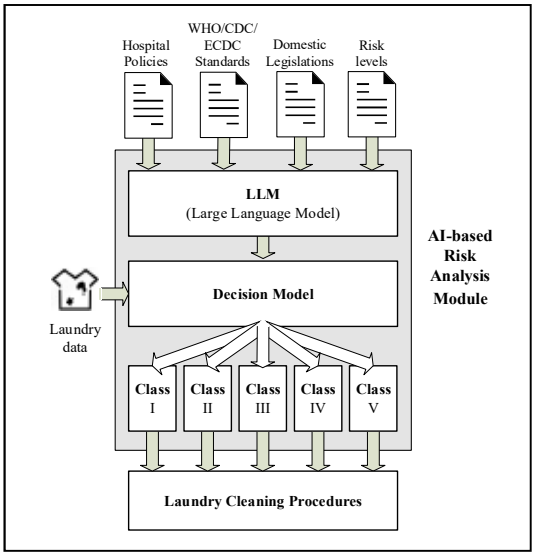


Fig. 2. Technical framework

AI-Based Risk Analysis Module works as follows: LLM model is first trained using risk levels defined by the specific hospital in question, the hospital's own policies, relevant domestic legislations and regulations, and finally standards from WHO, CDC, and ECDC. This training generates a decision model, and both the input data and the final model are stored inside a database. Then new laundry data (that inherits all the necessary information with respect to i.e. from where this laundry comes from, to whom it belongs to, at what time it is used and for how long as well as for what purpose, and finally when it is last disinfected/cleaned), is presented to the generated decision model via reading the RFID tag implemented on it. After this, the decision model makes a suggestion with respect to which class this laundry belongs to and then presents the necessary laundry cleaning procedures according to that specific class. Fig. 2 demonstrates this AI-Based Risk Analysis Module, presented inside Fig 1, in detail.

TABLE II. SAMPLE WORKFLOW SCENARIOS

- LLM Input Example: "*Unopened surgical gown returned from storage.*"
- Class Assignment: I (Unused)
- Workflow:
  i. Scan RFID tag to confirm unopened status.
  ii. LLM Verification: AI checks EHR for any contamination alerts.
  iii. If cleared, return to clean storage.
  iv. If contamination is suspected (e.g., torn packaging), reroute for rewashing.
- Automation: LLM logs the item's status in the inventory system.

- LLM Input Example: "*Blanket from non-infectious orthopedic patient.*"
- Class Assignment: II (Low Risk)
- Workflow:
  i. Place in white linen bags (automated dispensing system).
  ii. LLM Confirmation: AI verifies patient's infection status via EHR.
  iii. Wash at ≥60°C with standard detergent.
  iv. Store in designated clean linen area (tracked via barcode).
- Automation: AI updates linen usage logs in real time.

- LLM Input Example: "*Gown from patient with persistent cough awaiting TB results.*"
- Class Assignment: III (Medium Risk)
- Workflow:
  i. Place in yellow biohazard bags (color-coded).
  ii. LLM Alert: Flags for separate washing at ≥70°C with disinfectant.
  iii. Post-wash inspection using AI-powered visual scanning for stains.
  iv. If cleared, return to circulation; else, escalate to Class IV.
- Automation: AI documents handling compliance for audits.

- LLM Input Example: "Blood-soaked bed cover from trauma unit."
- Class Assignment: IV (High Risk)
- Workflow:
  i. Place in red leak-proof biohazard bags.
  ii. Staff must wear PPE (gloves, gowns, masks).
  iii. LLM-controlled Washing:
    - Main washing: High temperature -- 70°C - 90°C (158°F - 194°F), special bleaches, 10 minutes.
    - Optional UV treatment for additional sterilization.
  iv. Final AI Verification: Ensures no residual contamination.
- Automation: AI generates disposal logs for regulatory compliance.

- LLM Input Example: "*Single-use chemotherapy drape.*"
- Class Assignment: V (Discard)
- Workflow:
  i. Barcode-tagged for hazardous waste tracking.
  ii. LLM Confirmation: Ensures no accidental reintroduction into the system.
  iii. Disposed of via certified medical waste contractors.
  iv. AI generates compliance reports for audits.
- Automation: Real-time alerts if improper handling detected.

The LLM generates a ruleset to classify the laundry items. A sample ruleset is given below:

1. IF unused, unopened, and shows no contamination, THEN Class I.
2. IF used by non-infectious patients or staff and shows no biological contamination, THEN Class II.
3. IF from a ward with patients under observation or suspected infection (e.g., isolation ward) but not confirmed, THEN Class III.
4. IF visibly soiled with blood, body fluids, or from a confirmed infectious patient, THEN Class IV.
5. IF single-use, irreversibly damaged, or contaminated with hazardous chemicals, THEN Class V.

V. Example Workflow Scenarios

The proposed module generates optimized workflow procedures for each of these five classes described above accordingly. Table II depicts sample scenarios and workflows. We generated a simulated dataset to test the proposed framework and achieved 100% accuracy.

VI. Conclusion

We have proposed a software system to address the needs of medical laundry. The handling of medical laundry must follow strict guidelines without exception. The rules, policies, and protocols for handling laundry are typically specified in written documents using natural language. These documents attempt to categorize linens and based on this categorization they specify different handling requirements. LLMs exhibit strong language understanding capabilities; therefore, using LLMs in a laundry setting can enable the automation of laundry processing and reduce staff errors caused by interpreting ambiguous documents. By integrating AI-Based Risk Analysis Module (in other words LLM-driven automation), any hospital can transform its laundry management into a much safer, more efficient, and fully traceable system. Furthermore, this model can be replicated in other healthcare settings to reduce HAIs, improve compliance, and optimize resource allocation.

The proposed system uses various documents such as hospital laundry documents, public health agency guidelines for handling medical laundry and legislative documents as input to the LLM and classifies laundry belonging to institutions that provide health services. Once the laundry is classified the washing parameters are determined, or the laundry is discarded. Future improvements could incorporate robotic sorting, blockchain technologies for secure tracking, and expanded sensor integration to support even greater reliability and security procedures. We are currently in discussions with the university hospital to explore a potential collaboration aimed at validating the ideas presented in this paper using real-world datasets.

References

[1] World Health Organization, "Infection prevention and control in the context of coronavirus disease (COVID-19): a living guideline", Geneva: WHO, Aug. 2023. Reference: WHO/2019-nCoV/IPC/guideline/2023.2.

[2] World Health Organization, "Global analysis of healthcare waste in the context of COVID-19: status, impacts and recommendations," Geneva: World Health Organization; 2022.

[3] European Centre for Disease Prevention and Control, "Surveillance of healthcare-associated infections and prevention indicators in European intensive care units," HAI-Net ICU protocol, version 2.2, 2015.

[4] Centers for Disease Control and Prevention, "G. Laundry and Bedding," in Environmental Infection Control Guidelines, CDC, Jan. 2024.

[5] NHS East London Foundation Trust, "Linen and Laundry Policy 3.0," Feb. 2024.

[6] NHS Leicestershire Partnership, "Linen and Laundry Management Policy," March 2025.

[7] W.A. Rutala, and D.J. Weber, "Infection control: The role of disinfection and sterilization," in Journal of Hospital Infection, 2019, vol. 103, no. 3, pp. 207-214.

[8] Z. Lee, Y.C. Wu, and W. Xukang, "Automated Machine Learning in Waste Classification: A Revolutionary Approach to Efficiency and Accuracy," in 2023 12th Int. Conf. on Computing and Pattern Recognition (ICCPR 2023), October 27-29, 2023, Qingdao, China. ACM.

[9] X. Wang et al., "The Internet of Things in the Era of Generative AI: Vision and Challenges," in IEEE Internet Comp., 2024, vol. 28, no. 5, pp. 57-64.

[10] C. Stryker, and M. Scapicchio, "What is generative AI?," IBM, 22 March 2024, https://www.ibm.com/think/topics/generative-ai.

[11] A. Chopra, et al., "Enhancing medical decision-making with ChatGPT and explainable AI," Int. Jour. of Surgery," 2024, vol. 110, no. 8, pp. 5167-5168.

[12] C. Chakraborty, et al., "ChatGPT or LLMs can provide treatment suggestions for critical patients with antibiotic-resistant infections: a next-generation revolution for medical science?," Int. Journal of Surgery, March 2024, vol. 110, no. 3, pp. 1829-1831.

[13] S.B. Baker, W. Xiang and I. Atkinson, "Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities," in IEEE Access, 2017, vol. 5, pp. 26521-26544.

[14] European Commision, "Artificial Intelligence in healthcare, Transforming the future of medicine," EC, Accessed: 8 Agu. 2025. [Online]. Available: https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_en.

[15] T. Davenport, and R. Kalakota, "The potential for artificial intelligence in healthcare," in Future Healthc J., Jun. 2019, vol. 6, no. 2, pp. 94-98.

[16] N.M. Kumar, et al., "Artificial intelligence-based solution for sorting COVID related medical waste streams and supporting data-driven decisions for smart circular economy practice," in Process Safety and Environmental Protection, 2021, vol. 152, pp. 482-494.

[17] Zainordi M., et al., "Smart Linen Hamper and Laundry Monitoring for Hospital Industry," Evolution In Electrical And Electronic Engineering, 2024, vol. 5, no. 1, pp. 596-602.

[18] S.E. Abney, M.K. Ijaz, J. McKinney, and C.P. Gerba, "Laundry Hygiene and Odor Control: State of the Science," Applied and Environmental Microbiology", 2021, vol. 87, no. 14.

[19] K. Kurtel, and U. Celikkan, "Application of Context Aware Laundry Management in Medical Settings," 2024 Medical Technologies Congress (TIPTEKNO), Mugla, Türkiye, 2024, pp. 1-4.

[20] U. Celikkan, and K. Kurtel, "A Context-Aware Architecture for the Management of Laundry Business Processes," 14th European Conf. on Knowledge Management - ECKM 2013, Lithuania, 2013, pp. 159-166.

[21] E.H. Spaulding, "Chemical disinfection of medical and surgical materials," in: Lawrence CA ,Block SS, eds. Disinfection, sterilization and preservation. Philadelphia, Lea & Febiger, 1968; pp: 517-531.

[22] W.A. Rutala, D.J. Weber, and the Healthcare Infection Control Practices Advisory Committee (HICPAC), "Guideline for Disinfection and Sterilization in Healthcare Facilities, 2008", CDC, Update: June 2024,

[23] World Health Organization and Pan American Health Organization, "Decontamination and Reprocessing of Medical Devices for Health-care Facilities," WHO and PAHO, 2016.

[24] European Centre for Disease Prevention and Control, "Infection prevention and control in the household management of people with suspected or confirmed coronavirus disease (COVID-19)," March 2020.

A Low-Cost Portable Vital Sign Monitoring System and Mobile Application Customized for Medical Device Designs

Arif Koçoğlu
Biomedical Device Technologies Department, Baskent University,
arif@baskent.edu.tr

Halid Akkuş
Electrical and Electronics Engineering, Ankara University,
halidakkus01@gmail.com

Sacid Erdoğan
Electrical and Electronics Engineering, Ankara University,
saciterd@gmail.com

Cansel Ficici
Electrical and Electronics Engineering, Ankara University,
ogretmenoglu@ankara.edu.tr

Onur Koçak
Computer & Biomedical Engineering, Baskent University,
okocak@baskent.edu.tr

Ziya Telatar
Biomedical Engineering, Baskent University,
ztelatar@baskent.edu.tr

***Abstract*— This paper presents the design and implementation of a vital signs monitoring system based on embedded electronics and wireless communication. The system aims to acquire real-time physiological data — specifically heart rate, oxygen saturation (SpO_2), and electrocardiogram (ECG) signals — using the MAX30100 and AD8232 biomedical sensors interfaced with an ESP32 microcontroller. The collected data are transmitted wirelessly to a mobile application developed using the Flutter framework and Dart programming language. The system operates over a local Wi-Fi network and employs UDP broadcast for automatic device discovery, eliminating the need for manual IP entry. The mobile application includes responsive charts, dynamic status messages, and ECG signal filtering with real-time R-peak detection. To reduce power consumption, energy-saving strategies were implemented on the microcontroller side. A comprehensive engineering approach was followed: sensor-level signal acquisition, embedded software design, mobile UI development, and real-time visualization. The system provides a low-cost, extensible platform suitable for both academic and practical health monitoring applications, particularly in home-care environments. Project reached accuracy between 95-98% for vital sign results compared with smart watch measurements.**

Keywords*— *ESP32, MAX30100, AD8232, ECG, SpO_2, heart rate, Flutter, Dart, UDP, vital signs monitoring, mobile health, signal processing, simple interface

I. INTRODUCTION

In recent years, the demand for portable and real-time health monitoring systems has increased significantly due to advancements in wireless communication, embedded electronics, and sensor technologies [1]. The rise of wearable devices and remote patient monitoring platforms has opened up new opportunities for personal health tracking, early diagnosis, and chronic disease management [2]. Particularly after the COVID-19 pandemic, the need for non-invasive, cost-effective, and user-friendly biomedical systems has become more prominent in both clinical and non-clinical environments [3].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Recent studies highlight IoT-based and AI-assisted wearables, including a cloud-connected ECG patch with clinical-grade accuracy [4], an IoT-enabled framework for real-time cardiovascular monitoring [5], and a machine learning system for early arrhythmia detection [6]. These developments illustrate the shift toward intelligent, connected, and patient-centered healthcare solutions.

Vital signs such as heart rate (HR), blood oxygen saturation (SpO_2), body temperature, blood pressure, and electrocardiogram (ECG) are critical physiological parameters that reflect the overall condition of the cardiovascular and respiratory systems [7-8]. Continuous and accurate monitoring of these parameters allows for timely detection of anomalies such as arrhythmia, hypoxemia, or cardiac arrest. Traditionally, such measurements require bulky, expensive, and hospital-based equipment, which limits accessibility and early intervention. To overcome these limitations, microcontroller-based wearable systems have emerged as compact alternatives for personal healthcare applications [9].

This paper presents the design and development of a portable, low-cost vital signs monitoring system based on the ESP32 microcontroller [10]. The system integrates the MAX30100 for pulse oximetry and heart rate and the AD8232 for ECG acquisition [11]. Sensor signals are processed and transmitted via Wi-Fi to a mobile device, where they are displayed through a Flutter-based Android application [12].

The system acquires SpO_2, heart rate, and ECG signals in real time, with digital filtering and R-peak detection ensuring accurate analysis. Lead-off detection alerts users to electrode disconnection, preserving data reliability.

Wireless transmission over a local network enables untethered operation and enhances user comfort. The mobile application provides a responsive interface with smooth

animations, theme switching, and automatic device discovery via UDP broadcast, simplifying deployment and improving usability.

This work contributes to mobile health by offering an extensible embedded platform that can be expanded with additional sensors or algorithms, making it suitable for academic research, student training, and personal health monitoring.

II. MATERIAL AND METHODS

The scope of this paper is confined to the design, development, and validation of a prototype-level biomedical monitoring system intended for non-clinical applications, including academic research, personal health tracking, and related investigative purposes. This project does not seek to replace clinical-grade medical devices but rather aims to demonstrate the following:

- The feasibility of integrating embedded systems with mobile platforms to enable continuous health monitoring.
- The implementation of real-time wireless communication between hardware components and software interfaces.
- The provision of both visual and signal-level insights into users' vital signs within a compact system architecture.

Additionally, this project employs a progressive development model, wherein iterative weekly stages of testing, analysis, and optimization were conducted. This approach accounts for practical constraints typical of embedded system design in real-world environments.

The system architecture is organized into five distinct layers to ensure modularity, real-time performance, and clear separation of functionality:

A. Sensor Layer - Biomedical Signal Acquisition

The MAX30100 sensor captures heart rate and peripheral oxygen saturation (SpO_2) using red and infrared photoplethysmography via the I^2C interface. Simultaneously, the AD8232 module provides analog electrocardiogram (ECG) signals through differential leads (RA, LA, RL), which are sampled by the ESP32's analog-to-digital converter (ADC) [13]. All components are interconnected on a breadboard to facilitate rapid prototyping [14].

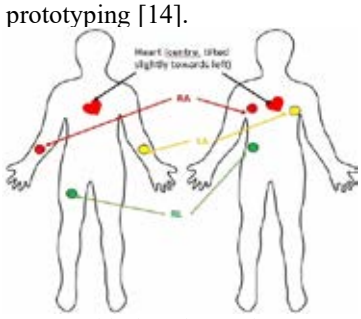


Figure 1. Sensor Positioning Diagram for ECG Measurement

B. Embedded Software Layer – ESP32 Firmware

Firmware development was performed in the Arduino IDE using C++. Key libraries utilized include PulseOximeter, ESP Async Web Server, and WiFi Udp. A state machine orchestrates the measurement sequence: finger detection, stabilization (4 seconds), active measurement (10 seconds), and reset. ECG signals are sampled at approximately 250 Hz, with lead-off detection implemented via LO+/LO– inputs. Sensor data are formatted as JSON and delivered via HTTP, while the ESP32 broadcasts its IP address every 3 seconds over UDP for network discovery.

C. Communication Layer – Wi-Fi Networking

UDP broadcast is employed to enable automatic discovery of the ESP32 by the mobile application. Data retrieval is performed through HTTP polling at intervals of approximately 40 ms, ensuring low-latency acquisition of heart rate, SpO_2, ECG data, and system status to support real-time user interaction. The web server component is implemented on the ESP32 [15-16].

D. Processing Layer

ECG signals are initially processed offline in MATLAB, where raw ADC values are converted to millivolts, then bandpass filtered between 0.5 and 40 Hz. R-peak detection is accomplished using a 5-point sliding window with a 0.5 mV threshold. This detection algorithm was subsequently ported and optimized for real-time execution in Dart [17]. Detected R-peaks are marked on the mobile ECG interface, enabling accurate heart rate visualization under stringent computational constraints.



Figure 2. Bioinstrumentation Module with ECG Leads

E. Mobile Application Layer – Real-Time Visualization

The mobile application is developed using Flutter and Dart frameworks. It supports automatic connection to the ESP32 via UDP discovery and provides real-time graphical displays of heart rate and SpO_2, including animated feedback. The app implements digital ECG filtering within 0.5–40 Hz, real-time R-peak detection, lead-off alerts, and offers dynamic theme switching between dark and beige modes. The interface includes two primary tabs: one dedicated to ECG waveform visualization and another for vital signs such as heart rate and SpO_2.

The design consists of a microcontroller, an AD8232 module for acquiring ECG signals, a MAX30100 sensor for pulse signal detection, and Android applications for visualizing the acquired signals. The critical design block diagram of the system is presented in Figure 4. The electronic system, incorporating an ESP-series microcontroller, was integrated within the Proteus

simulation environment, and the on-board implementation was successfully completed.



Figure 3. Mobile Application for Android



Figure 4. System Block Diagram

III. RESULTS AND DISCUSSION

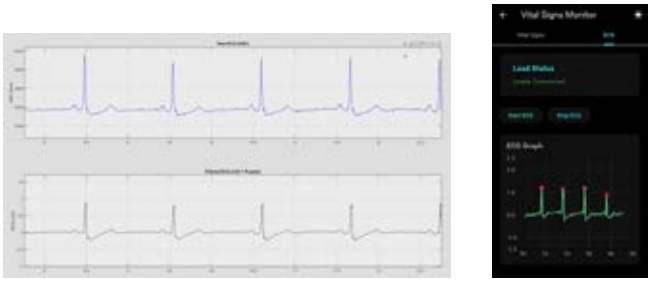
The processing of electrocardiogram (ECG) signals is a fundamental component of physiological monitoring systems, enabling the detection and analysis of the heart's electrical activity. In this study, ECG signal processing was initially prototyped using MATLAB to develop and validate detection algorithms. Subsequently, the implementation was adapted for real-time operation within a mobile application developed using Flutter. This dual-platform methodology ensured the reliability of signal quality assessment and algorithmic performance prior to deployment.

Analog-to-digital converter (ADC) values were acquired from the ESP32 via HTTP at an approximate sampling rate of 250 Hz. A lightweight high-pass and low-pass filter approximation was applied to the incoming data, and the resulting signal was scaled to enhance the visibility of the ECG waveform [16]. A five-point local maxima detection algorithm was employed to identify R-peaks, which were visually indicated as red markers on the real-time ECG trace. The waveform was rendered using the *fl_chart* library. To optimize performance and memory usage, signal buffers were cleared every five seconds. All processing operations were performed asynchronously to maintain smooth user interface updates. The mobile application environment was designed with minimal computational overhead and visual optimization in mind [18].

The two-stage development approach—initially employing MATLAB for algorithm prototyping and validation, followed by implementation in Flutter for real-time mobile deployment—facilitated a robust and systematic design process for ECG signal processing. The current system demonstrates effective heartbeat detection and smooth waveform visualization, establishing a strong foundation for future enhancements, including beat-per-minute (BPM) estimation and arrhythmia classification.

Table I. Comparative Summary of MATLAB and Flutter Implementations

Feature	*MATLAB*	*Flutter (Mobile)*
Filtering Method	Butterworth Bandpass (0.5–40 Hz)	Simple IIR-style approximation
R-Peak Detection	Local maxima + threshold	Same logic
Visualization	Dual real-time plots	Single graph + peak markers



a) (b)

Figure 5. (a) MATLAB Real-Time Processing ECG Signal Plot (b) Mobile App: ECG Tab – ECG Signal Visualization

The Connection Page is the first screen displayed to the user. Its purpose is to automatically detect and connect to the ESP32 device by broadcasting its IP via UDP. Network discovery and connection procedures are managed within the *IPInputPage* module. Incoming UDP datagrams are decoded using the String.fromCharCodes(datagram.data) method and subsequently parsed by splitting the resulting string at the delimiter ":". This design eliminates the need for manual static IP configuration, thereby streamlining system setup. The *Vital Signs* tab, integrated into the *HomePage* interface, is dedicated to displaying real-time heart rate and oxygen saturation values acquired from the ESP32 [16, 19].

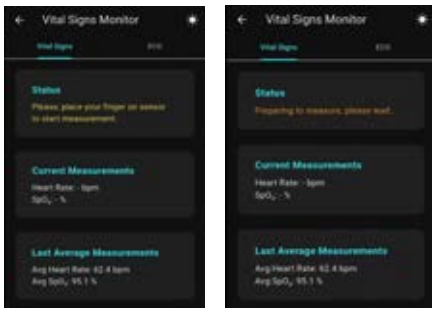


Figure 6. Mobile App: Vital Signs Tab – Status Card pt1

The ECG tab provides real-time visualization of the analog ECG signal processed from the AD8232 module. It performs filtering and basic R-peak detection on the mobile side.

In this study, heart rate and oxygen saturation (SpO_2) values were measured using the *Lifenet Medical*™ pulse oximeter available in our laboratory. In addition, measurements were

also obtained from a *Huawei™ Watch D2* smartwatch, and the results were compared accordingly. The acquired ECG signal is presented in Figure 5. This signal was further compared with the 3-channel ECG measurement of a bedside monitor in our laboratory on the same volunteer subject. The ECG waveforms were consistent, and for validation purposes, the heart rate (R-peaks) was counted and compared with the one-minute reference measurement, demonstrating identical results.

SpO_2 validation was performed across three volunteer subjects. Based on one-minute averaged recordings, the smartwatch measurements exhibited a 98% similarity, while the pulse oximeter demonstrated a 95% similarity. Furthermore, R-peak detection results showed 99% agreement. The novelty of this validation process lies in demonstrating that cost-effective wearable devices can achieve high agreement with clinical-grade monitors under experimental conditions, highlighting their potential use in practical biomedical applications.

IV. CONCLUSION AND FUTURE WORKS

The system successfully fulfilled its intended functionalities by measuring heart rate and oxygen saturation using the MAX30100 sensor, acquiring ECG signals from the AD8232 analog front-end module, broadcasting the ESP32's IP address via UDP for automatic network discovery, serving physiological data through an HTTP server in JSON format, and visualizing heart rate (HR), oxygen saturation (SpO_2), and ECG waveforms in real time within a mobile application. Each subsystem, encompassing sensor data acquisition, signal processing, data transmission, and visualization, was seamlessly integrated and operated reliably under typical test conditions.

The MAX30100 sensor consistently provided accurate heart rate and SpO_2 measurements under proper finger placement and stable contact conditions, with a four-second stabilization delay implemented prior to recording to minimize noise. The ECG waveform exhibited clearly distinguishable QRS complexes, particularly when motion was minimized and electrodes were properly positioned, although visual signal quality decreased with hand movement as expected from a non-diagnostic, low-cost sensor. The mobile application, polling the ESP32 every 40 milliseconds, delivered real-time feedback without performance degradation, while UDP-based IP discovery reliably completed within 1–3 seconds after device startup, eliminating the need for manual IP entry. The Flutter-based mobile interface was responsive and user-friendly, supporting both dark and beige/light themes, with a theme toggle feature to enhance readability under varying lighting conditions.

In conclusion, this study presents a technical roadmap for designing a cost-effective vital monitor by integrating embedded electronics, biomedical sensors, and cross-platform mobile software. The system combines real-time physiological feedback with a user-friendly interface, offering both practical healthcare applications and educational value. The novelty of this work lies in providing a structured framework that translates recent electronic advancements into practical strategies for vital monitoring.

REFERENCES

[1] How2Electronics, "ECG monitoring with AD8232 ECG sensor and Arduino," 2021. [Online]. Available: https://how2electronics.com/ecg-monitoring-with-ad8232-ecg-sensor-arduino/. [Accessed: Aug. 10, 2025].
[2] K. Tan, A. R. Kavsaoğlu, O. Koçak, and C. Akbay, "Remote monitoring system for incubator data," in *Proc. National Congress on Medical Technologies (TIPTEKNO)*, 2018, pp. 1–4, doi: 10.1109/TIPTEKNO.2018.8597156.
[3] E. Elif and O. Koçak, "A low cost design non-invasive glucose monitoring experimental setup," in *Proc. National Congress on Medical Technologies (TIPTEKNO)*, 2018.
[4] J. Lopez, M. Navarro, S. De la Torre, and P. Casaseca, "Wearable patch for continuous electrocardiogram and respiration monitoring: Design, characterization, and clinical evaluation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 3124–3134, Nov. 2020, doi: 10.1109/JBHI.2020.2991749.
[5] S. M. R. Islam, A. T. Ghoneim, and K. S. Kwak, "Smart healthcare monitoring framework for IoT-based wearable devices," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15645–15655, Nov. 2021, doi: 10.1109/JIOT.2021.3071564.
[6] H. T. Nguyen, Y. Zhang, and A. M. Khan, "Early arrhythmia detection using wearable sensors and machine learning: A mobile health approach," *Sensors*, vol. 23, no. 5, p. 2651, 2023, doi: 10.3390/s23052651.
[7] W. Q. Mok, W. Wang, and S. Y. Liaw, "Vital signs monitoring to detect patient deterioration: An integrative literature review," *International Journal of Nursing Practice*, vol. 21, no. S2, pp. 91–98, 2015.
[8] U. U. Tarhan, D. Cengiz, and O. Koçak, "Bilgisayar kontrollü damar yeri tespiti cihazı," in *Proc. BPS*, 2014.
[9] A. T. Tuncer, O. Koçak, and A. Koçoğlu, "PIC microcontroller based RF wireless system design for patient," in *Proc. 7th Int. Advanced Technologies Symp. (IATS)*, 2013.
[10] Espressif Systems, *ESP32 Datasheet: Wi-Fi & Bluetooth SoC*, 2020. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf. [Accessed: Aug. 10, 2025].
[11] Analog Devices, *MAX30100 Pulse Oximeter and Heart-Rate Sensor for Wearable Health, Datasheet*, 2015. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/MAX30100.pdf. [Accessed: Aug. 10, 2025].
[12] Google, *Dart Programming Language Overview*, 2024. [Online]. Available: https://dart.dev/. [Accessed: Aug. 10, 2025].
[13] Circuit Digest, "Interfacing MAX30100 pulse oximeter with Arduino for heart rate & SpO_2 monitoring," 2020. [Online]. Available: https://circuitdigest.com/microcontroller-projects/interfacing-max30100-pulse-oximeter-with-arduino. [Accessed: Aug. 10, 2025].
[14] Analog Devices, *AD8232: Single-Lead, Heart Rate Monitor Front End, Datasheet*, 2016. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/ad8232.pdf. [Accessed: Aug. 10, 2025].
[15] GitHub, *ESPAsyncWebServer – Asynchronous Web Server for ESP8266 and ESP32*, 2024. [Online]. Available: https://github.com/me-no-dev/ESPAsyncWebServer. [Accessed: Aug. 10, 2025].
[16] J. Postel, "User Datagram Protocol," *RFC 768*, Internet Engineering Task Force (IETF), 1980. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc768. [Accessed: Aug. 10, 2025].
[17] Google, *Flutter Documentation: Build Beautiful Apps*, 2024. [Online]. Available: https://docs.flutter.dev. [Accessed: Aug. 10, 2025].
[18] MathWorks, *Signal Processing Toolbox: Bandpass Filtering and Peak Detection in Real-Time*, 2024. [Online]. Available: https://www.mathworks.com/products/signal.html. [Accessed: Aug. 10, 2025].
[19] GitHub, *WiFiUDP – Arduino Core for ESP32, Wi-Fi UDP Library for ESP32 Boards*, 2024. [Online]. Available: https://github.com/espressif/arduino-esp32. [Accessed: Aug. 10, 2025].

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Comparison of Leading Deep Learning Architectures on Skin Lesion Classification Using the ISIC 2019 Dataset

Anil Tan Aktan, Alaa Almouradi, Erchan Aptoula and Berrin Yanikoglu
Sabanci University, Istanbul, 34956 Türkiye
{*tan.aktan,alaaalmouradi,erchan.aptoula,berrin*}*@sabanciuniv.edu*

***Abstract*—This paper addresses the problem of automatic classification of dermoscopic skin lesion images into one of eight diagnoses that include multiple cancer types. The goal of this research was to evaluate methods to address problems of class imbalances and image variation in skin lesion classification, as well as to propose an ensemble model that performs well on multi-class metrics. The widely used dataset developed for the International Skin Imaging Collaboration (ISIC) challenge, ISIC 2019, was chosen to test the methods proposed in this paper. Methods evaluated in this paper include modern architectures such as Vision Transformer, Convolutional Neural Networks, and Vision Mamba, as well as training strategies such as data augmentation, test-time augmentation, and specialized loss functions addressing class imbalance. The best performance was obtained using an ensemble approach, achieving an overall accuracy of 74.80% and a macro F1 score of 74.30%. These results compare favorably to the state-of-the-art, where drops in performance with multi-class metrics have typically been observed.**

***Index Terms*—Skin cancer, class imbalance, test-time augmentation, heterogeneous ensemble.**

I. Introduction

Skin cancer is one of the most widespread variants, posing significant threats to public health and burdening healthcare systems. Melanoma, in particular, presents high metastatic potential and rapid progression, making early and accurate detection crucial for successful treatment [1]. Dermoscopic examination performed by dermatologists has been the clinical standard for lesion diagnosis. However, accurate diagnosis is a challenge even for clinicians, as they typically achieve around 60% accuracy in lesion classification [2]. This diagnostic uncertainty, combined with limited availability and unequal geographic distribution of dermatologists, has motivated research towards robust and accurate automated diagnostic methods.

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), transformers, and hybrid models, have substantially improved performance in medical image analysis [3]. However, despite these advances, critical challenges persist. First, medical datasets frequently exhibit significant class imbalance. This imbalance can bias model training and undermine performance on minority classes [4]. Second, due to the introduction of many hyperparameters in ensembling techniques and complex models, standard training methodologies have often overlooked comprehensive hyperparameter optimization, limiting model generalization [5]. Additionally, techniques performing well in other problems are underexplored in the skin lesion problem. Traditional approaches have rarely leveraged test-time augmentation (TTA) fully, resulting in missed opportunities to improve robustness, a critical requirement for this task given the variations commonly encountered in clinical practice [6]. Furthermore, while individual model architectures have been extensively studied, ensembles combining diverse architectures have remained underexplored [7].

The official leaderboard for the International Skin Imaging Collaboration (ISIC) 2019 challenge reveals the uneven performance of top teams across different metrics (Table II). While accuracy measures overall correctness, it can be misleading in imbalanced datasets. Balanced accuracy accounts for class imbalance by averaging recall across classes and was the official evaluation metric of the challenge. F1-score, on the other hand, emphasizes the harmonic mean of precision and recall, offering a more sensitive measure of performance in the presence of skewed class distributions. These evaluation inconsistencies and the underlying performance gaps motivate this research to study methods that deliver balanced and generalizable results across all key metrics.

The contributions of the paper can be summarized as follows:

- Evaluating different methods that address class imbalance and model robustness in the area of skin lesion classification.
- Proposing an ensemble model for skin lesion classification with high performance on multi-class metrics.

II. Related Works

Skin lesion classification methodologies generally fall into two architecture categories: deep convolutional neural networks and transformer-based architectures. Hybrid approaches and additional architecture components have also been proposed to improve generalization and robustness.

Deep Learning Architectures: CNNs have become widespread in automated skin lesion analysis [3]. Architectures such as EfficientNet [8] have demonstrated superior performance, effectively capturing the local spatial features critical for lesion recognition tasks. However, CNNs have primarily

This study was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under Project No. 124N417.
979-8-3315-5565-8/25/$31.00 ©2025 IEEE

focused on localized patterns, often lacking sufficient global context, limiting their performance on visually ambiguous lesion categories [9].

To address this, transformer-based models like Vision Transformer (ViT) [9] have introduced self-attention mechanisms that explicitly model long-range dependencies. This global modeling capability complements the local pattern extraction of CNNs and is particularly beneficial in dermatology, in which lesion differentiation may rely on subtle global contextual clues.

More recently, hybrid vision architectures such as MambaVision [10] aim to balance contextual awareness with computational efficiency by integrating ViT-style attention and Mamba's linear-time sequence modeling. While such hybrid architectures have shown promising results in broader medical imaging domains, their application in dermatology remains limited and underexplored.

Strategies for Class Imbalance and Robustness: Addressing class imbalance has been a significant research area, with methods such as Weighted Cross-Entropy and Focal Loss [4], [11] widely applied to balance model attention towards minority classes. Data augmentation methods have also been extensively studied as a means to improve generalization performance [12]. Test-time augmentation, although effective in other domains, has rarely been optimized in the context of dermatology [6]. Thus, systematic exploration and evaluation of augmentation techniques, combined with robust loss functions, have remained limited in dermoscopic image tasks.

Ensemble Learning: Ensemble methods have proven highly effective in various machine learning tasks by leveraging model diversity to enhance predictive performance and robustness [7]. However, heterogeneous ensembles combining fundamentally different architectures (CNNs, transformers, hybrids) have remained under-explored in dermatological image classification, motivating the explicit exploration present in this paper.

In summary, while significant advancements have been made, critical aspects such as class imbalance, robust augmentation strategies, and heterogeneous ensemble modeling have remained insufficiently addressed. This paper explicitly targets these gaps, systematically integrating and evaluating these strategies for robust, accurate, and generalizable skin lesion classification.

III. Proposed Approach

The primary goal of the proposed methodology is to achieve robust and generalizable multiclass image classification by addressing several critical challenges, including high class imbalance, limited data for underrepresented classes, sensitivity to image variability, and combining the strengths of different deep learning architectures. In this section, each design choice is explicitly justified by detailing its expected contribution towards solving these specific challenges.

Loss Functions for Class Imbalance: To improve model performance on skewed class distributions, the proposed method explicitly incorporates loss functions designed to address class imbalance, including standard cross-entropy, Weighted cross-entropy, and focal loss [4] functions. Focal Loss was selected due to its focus on difficult and minority-class samples.

Test-Time Augmentation (TTA): Another issue is to obtain robustness with different viewpoints and variations in appearance caused by lighting, orientation, or scale. Test-time augmentation is implemented at inference time to improve the robustness of predictions by averaging model outputs across augmented versions of each input image. The augmentation strategies selected include horizontal flips, rotations at increments of 90°, and additional small-degree rotations, color jittering, and resizing.

The TTA approach aims to reduce prediction variance caused by alterations in images obtained in a clinical scenario. Indeed, producing predictions from multiple augmented views improves the output stability and model accuracy, as demonstrated in our experiments. The enhanced reliability and generalizability are important in real-world clinical deployment scenarios.

Hyperparameter Optimization: Model performance and generalization capability critically depend on careful hyperparameter tuning [5]. To systematically address this challenge, hyperparameters were optimized by comparing validation macro-F1 scores using Optuna, a state-of-the-art optimization framework that efficiently explores parameter spaces via Bayesian optimization methods [5]. Key hyperparameters that were tuned include: learning rate, weight decay, and focal loss gamma.

Learning rate impacts convergence behavior and final accuracy; weight decay controls regularization and prevents overfitting; and focal loss gamma determines the trade-off between focusing on challenging examples and ignoring easily classified ones.

Heterogeneous Ensemble Learning: Different deep learning architectures inherently possess distinct internal biases, capturing complementary visual features [7]. Therefore, a heterogeneous ensemble comprising three fundamentally different architectures (EfficientNet-B5 (CNN-based) [8], Vision Transformer (ViT-base-patch16-224, transformer-based) [9], and MambaVision (ViT–Mamba hybrid) [10] is proposed in this work. Aggregating predictions from these complementary architectures was implemented through simple mean ensembling, i.e., averaging their class probability outputs, which enhances overall accuracy, robustness, and generalizability by leveraging their diverse strengths.

EfficientNet-B5 was chosen due to its parameter-efficient scaling, enabling effective processing of high-resolution images critical for capturing detailed lesion features [8]. ViT was selected for its capability to model global context and dependencies between image regions, aiding recognition of overall lesion structures and patterns [9]. MambaVision, as a hybrid architecture, was selected because it combines the global context modeling of Vision Transformers with the efficient sequential state-space modeling of Mamba layers,

effectively capturing both global dependencies and structured representations [10].

IV. Experiments and Results

Experiments are conducted to evaluate different methods that address challenges such as class imbalance, sensitivity to image variability, and other previously mentioned challenges in the skin lesion classification problem. These methods include: different architectures, data stratification, data augmentation, test-time augmentation, and different losses. All experiments are conducted using the ISIC 2019 dataset.

Dataset: The ISIC challenge was organized to benchmark automated classification algorithms using the ISIC 2019 dataset [13]–[15] - a publicly available collection of dermoscopic skin lesion images. ISIC 2019 was selected since it is the most established multi-class benchmark, unlike newer ISIC datasets that are larger but framed differently (e.g., binary classification in 2020). The challenge leaderboard uses multi-class balanced accuracy as the metric with which to compare algorithms. The ISIC 2019 challenge has an official training and test split, consisting of 25,331 images for training and 8,238 images for testing. There is no official split for validation.

The class distribution of the ISIC 2019 training set is as follows: Melanocytic nevus (NV) 50.8%, Melanoma (MEL) 17.9%, Basal cell carcinoma (BCC) 13.1%, Benign keratosis (BKL) 10.4%, Actinic keratosis (AK) 3.4%, Squamous cell carcinoma (SCC) 2.5%, Vascular lesion (VASC) 1.0%, and Dermatofibroma (DF) 0.9%. As can be seen, the distribution is highly biased; NV makes up over half of all samples, while DF and VASC contribute about 1% each. This resembles the real-life distribution of these skin lesions, which makes the ISIC 2019 dataset appropriate to evaluate methods that address class imbalance. To maintain class distribution during training and validation, a stratified splitting was used, specifically, 80% for training and 20% for validation. As for the final evaluation, the official test set was used.

All images in the dataset are color dermoscopic photographs in three-channel RGB format. There is no fixed aspect ratio; the majority of images fall within the resolution range of 576×768 pixels to 1024×1024 pixels. This variability reflects the diversity of capture devices and clinical settings encountered in real-world practice.

Data Augmentation and Transformations: Data augmentation was applied to improve model generalization given the class imbalance and variations typically encountered in clinical imaging scenarios. A series of stochastic and geometric transformations was applied during training only. The transformations included:

- Random rotation up to $45°$
- Horizontal and vertical flips
- Random affine transformations (translation up to 10%, scaling between 1.0 and 1.05)
- Color jitter with brightness and contrast adjustments (0.1)
- Normalization using ImageNet statistics (`mean` = [0.485, 0.456, 0.406], `std` = [0.229, 0.224, 0.225])
- Center square crop: Crops the central square region of the image, to preserve aspect ratio.

For validation and testing, only deterministic transformations were applied (center crop, resize, and normalization) to ensure consistent evaluation conditions.

Model Training: Models were initialized with ImageNet pre-trained weights, then trained for 25 epochs on the ISIC 2019 dataset using AdamW. As found by hyperparameter tuning on EfficientNet-B5, ViT, and MambaVision, the learning rate was set to (8.4E-05, 1.00E-05, 4.56E-05) and weight decay to (8.6E-03, 8.6E-03, 4.9E-05). Training and validation accuracies, losses, and macro F1-scores were monitored. The best checkpoints were selected by the highest validation macro F1.

The loss function and TTA configuration used for the final evaluations were selected based on the results of the experiments conducted on different loss functions and TTA configurations, as explained as follows, with the gamma hyperparameter for focal loss being further tuned to 2.41.

Loss Functions: Experiments were conducted on 3 loss functions: standard cross-entropy, weighted cross-entropy, and focal loss, all conducted on the ViT model. The gamma parameter for focal loss was set to the default (2) [4]. Other parameters for training were set as mentioned previously. As seen in Table I, focal loss markedly outperformed both standard and weighted cross-entropy, achieving up to a 1.75% point gain in weighted F1 score on the class-imbalanced dataset.

Test-Time Augmentation (TTA): To evaluate the applicability of TTA, 4 configurations were tested with 8 and 16 views, using stochastic transformations of flips, 90°rotations, additional 10-15°rotations, and color jitter of 10%. All experiments were conducted with focal loss, which gave the best results in initial experiments. Test-time augmentation yielded a 0.79% point improvement in F1 compared to no augmentation, highlighting the value of inference-time transformations in enhancing model robustness.

TABLE I
Performance of ViT Model with Different Settings

Loss Function	Test F1 (%)	Acc (%)
Focal Loss (gamma=2)	**72.15**	**72.41**
Weighted Cross-Entropy	71.21	71.54
Cross-Entropy	70.40	71.72
Test-time Augmentation	**Test F1 (%)**	**Acc (%)**
8x (Flip + Rotate90 + Rot10)	**72.15**	**72.41**
8x (Flip + Rotate90 + Rot15)	72.12	72.38
16x (Flip + Rotate90 + Rot15 + ColorJitter)	72.00	72.30
No test-time augmentation	71.36	72.01

Ensemble Results: As can be seen from Table II, a heterogeneous ensemble of three architectures resulted in a 1.67% and 1.02% point increase in F1 score and accuracy, respectively, compared to our best-performing single model. Furthermore, the confusion matrix in Table III confirms

these improvements by demonstrating accurate classification of major classes and consistent performance on minority classes. Moreover, the proposed ensemble represents notable improvements over leading literature and leaderboard methods, with the highest F1-score (0.743) and AUC (0.942). Overall, the ensemble method effectively leverages the complementary strengths of different architectures, demonstrating enhanced robustness and generalization performance in skin lesion classification tasks.

TABLE II
COMPARISON OF ISIC 2019 PERFORMANCE: LEADERBOARD, LITERATURE, AND PROPOSED METHODS (BEST IN BOLD, SECOND BEST UNDERLINED)

Method	Weighted F1	AUC	Bal. Accuracy
ISIC2019 Leaderboard:			
DaisyLAB [16]	0.685	<u>0.913</u>	**0.636**
DysionAI [17]	0.703	0.780	<u>0.607</u>
AImageLab -PRHLT [18]	0.709	0.886	0.593
Literature:			
Doctors' Average [2]	0.597	—	—
Skin-Net [19]	<u>0.713</u>	—	—
Kassem *et al.* [20]	0.575	—	—
This work:			
EfficientNet-B5	0.736	0.925	0.598
ViT	0.722	0.912	0.587
MambaVision	0.723	0.918	0.591
Proposed Ensemble	**0.743**	**0.942**	<u>0.607</u>

TABLE III
CONFUSION MATRIX FOR THE PROPOSED ENSEMBLE MODEL.

	MEL	NV	BCC	AK	BKL	DF	VAS	SCC
MEL	**918**	225	67	40	69	0	7	1
NV	119	**2216**	77	21	52	1	3	6
BCC	27	29	**828**	46	26	3	5	11
AK	20	11	84	**168**	87	0	0	4
BKL	83	94	60	70	**333**	10	0	10
DF	1	15	16	1	3	**53**	0	2
VASC	11	8	24	0	0	1	**54**	6
SCC	9	4	62	15	13	1	0	**61**

V. CONCLUSION

Our work presents the performance of several deep learning architectures optimized to address class imbalance and improve robustness in multi-class skin lesion classification. The results demonstrate the value of focal loss, systematic hyperparameter optimization, and an optimized test-time augmentation strategy, each contributing measurable gains in generalization.

This study also introduces an ensemble model that achieves strong performance across multi-class metrics. The proposed ensemble model achieved the highest F1-score (0.743) and AUC (0.942) among all evaluated methods, outperforming both the leading literature and the official ISIC 2019 leaderboard submissions. Future work will extend this framework to larger and more recent datasets and explore the use of metadata and external data sources to further improve the recognition of underrepresented classes.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2021," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
[2] P. Tschandl, C. Rosendahl, and H. Kittler, "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," *The Lancet Oncology*, vol. 20, no. 7, pp. 938–947, Jul. 2019.
[3] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE TPAMI*, vol. 42, no. 2, pp. 318–327, 2020.
[5] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the ACM SIGKDD*, 2019, pp. 2623–2631.
[6] D. Shanmugam, D. Blalock, G. Balakrishnan, and J. Guttag, "When and why test-time augmentation works," *arXiv:2011.11156*, 2021.
[7] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.
[8] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of ICML*, 2019, pp. 6105–6114.
[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of ICLR*, 2021.
[10] A. Hatamizadeh and J. Kautz, "Mambavision: A hybrid mamba-transformer vision backbone," in *Proceedings of IEEE CVPR*, 2025, pp. 25 261–25 270.
[11] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of IEEE CVPR*, 2019, pp. 9268–9277.
[12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
[13] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018.
[14] N. C. F. Codella, D. A. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. K. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC)," *CoRR*, vol. abs/1710.05006, 2017.
[15] C. Hernández-Pérez, M. Combalia, S. Podlipnik, N. C. Codella, V. Rotemberg, A. C. Halpern, O. Reiter, C. Carrera, A. Barreiro, B. Helba, S. Puig, V. Vilaplana, and J. Malvehy, "Bcn20000: Dermoscopic lesions in the wild," *Scientific Data*, vol. 11, no. 1, p. 641, 2024.
[16] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," *MethodsX*, vol. 7, p. 100864, 2020.
[17] S. Zhou, Y. Zhuang, and R. Meng, "Multi-category skin lesion diagnosis using dermoscopy images and deep cnn ensembles," DysionAI, Tech. Rep., 2019.
[18] F. Pollastri, J. Maronas, M. Parreno, F. Bolelli, R. Paredes, C. Grana, and A. Albiol, "AImageLab-PRHLT at ISIC challenge 2019," Università degli Studi di Modena e Reggio Emilia and Universitat Politècnica de València, Tech. Rep., 2019.
[19] Y. S. Alsahafi, M. A. Kassem, and K. M. Hosny, "Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier," *Journal of Big Data*, vol. 10, p. 105, 2023.
[20] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114 822–114 832, 2020.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Karar Destekli Akıllı İnfüzyon Pompası Tasarımı
Design of a Smart Infusion Pump With Decision Support

Melahat KIZIL ve Nursun ÜSTÜNKARLI
Sağlık Hizmetleri Meslek Yüksekokulu
İzmir Ekonomi Üniversitesi
İzmir, Türkiye
{melahat.kizil,nursun.ustunkarli}@ieu.edu.tr

Berkay İlber TUNAKAN ve Aydın AKAN
Elektrik-Elektronik Mühendisliği Bölümü
İzmir Ekonomi Üniversitesi
İzmir, Türkiye
{berkay.tunakan,akan.aydin}@ieu.edu.tr

***Özetçe* —Acil sağlık hizmetlerinde hızlı ve doğru ilaç uygulaması, hasta yaşamını doğrudan etkileyen kritik bir süreçtir. Bu çalışmada; hastanın tıbbi geçmişi, yaşamsal bulguları ve klinik durumu dikkate alınarak, paramediklerin karar verme süreçlerini destekleyen akıllı bir infüzyon pompası tasarlanmıştır. Tasarım süreci; literatür taramaları, 112 ekipleriyle yapılan görüşmeler, öğrenci staj gözlemleri ve protokol bazlı doz hatalarına dair geri bildirimler temelinde yürütülmüştür. Pompa, QR kod veya entegre sistemlerle hasta bilgilerini tanıyarak algoritma destekli doz önerileri sunmakta; görsel/sesli yönlendirmeler sağlamaktadır. "Hızlı Mod" özelliği ile kardiyak arrest, anaflaksi ve hipoglisemi gibi durumlara yönelik protokoller içermektedir. Sistem, riskli ilaçlara karşı uyarı verip doğru ilaç-doz-hız kombinasyonlarını önermekte; uygulama zamanlarını izleyerek otomatik uyarılar oluşturabilmektedir. Kullanıcı dostu arayüzü ve düşük ışıkta çalışabilen LCD ekranı ile sahada güvenli kullanım sunar. Cihaz, hasta verilerini kullanarak doğru doz hesaplamalarını desteklemektedir.**

***Anahtar Kelimeler*—Akıllı infüzyon pompası; acil sağlık hizmetleri; ilaç uygulama; karar destek sistemi; paramedik; hastane öncesi bakım.**

***Abstract*—Rapid and accurate drug administration is a critical process that directly affects patient outcomes in emergency medical services. In this study, a smart infusion pump was designed to support paramedics' decision-making by considering the patient's medical history, vital signs, and clinical condition. The design process was guided by literature reviews, interviews with ambulance teams, student internship feedback, and observations of protocol-based dosing errors. The pump identifies patient data via QR codes or integrated systems and offers algorithm-supported dosage suggestions with visual and audio guidance. The "Quick Mode" includes predefined protocols for common emergencies such as cardiac arrest, anaphylaxis, and hypoglycemia. The system can alert users about high-risk medications, recommend appropriate drug-dose-rate combinations, and issue automatic reminders for administration timing. Its user-friendly interface and LCD screen, operable in low light, ensure safe and effective use in the field. The device aims to support accurate dosage calculations using patient-specific data such as age, weight, height, and medical history.**

***Keywords*—Smart infusion pump; emergency medical services; drug administration; decision support; paramedics; prehospital care.**

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE

I. GIRIŞ

Acil sağlık hizmetleri (112), hastane öncesi dönemde hasta ve yaralılara olay yerinde veya ambulansta müdahale eden sağlık sisteminin önemli bir alanıdır. Bu alanda görev yapan paramedikler, çok çeşitli sağlık problemleriyle karşılaşabileceklerinden, her türlü zorluğa karşı hazırlıklı olmalıdırlar. Hastane öncesi ortam, hastane koşullarından belirgin şekilde farklıdır; genellikle hareketli, dar, öngörülemez ve sınırlı bir alan ve ekipmanla hasta ve yaralılara müdahale gerektiren bir yapıya sahiptir.

Paramediklerin yalnızca güçlü bir mesleki eğitim almaları yeterli değildir; aynı zamanda kullandıkları tıbbi cihaz ve ekipmanlara da son derece hâkim olmaları gerekir. Bu ekipmanlar arasında defibrilatörler, vantilatörler, sedyeler, atel sistemleri ve intravenöz sıvı tedavisi uygulamalarında kullanılan infüzyon pompaları v.s yer alır. Günümüzde ambulanslarda kullanılmaya başlanan infüzyon pompaları, hastaya damar yolu açarak sıvı ve ilaç verilmesinde önemli rol oynamaktadır. İlaç uygulamalarında doz hesaplamalarının belirli formüllere göre yapılması gerektiğinden, acil durumun stresi altında paramedikler yanlış doz uygulayabilmekte ve bu da hastanın klinik durumunu olumsuz etkileyebilmektedir. Eğer infüzyon pompası kullanılmıyorsa, sıvı uygulaması sırasında damla sayımı ile manuel hesaplamalar yapılmakta; bu durum zaman kaybına neden olmakta ve hata riskini artırmaktadır. Ayrıca kullanılan serum setlerindeki kıskaçlar, ufak bir hareketle açıldığında hastaya istenilenden fazla sıvı geçişine neden olabilmekte, bu durum özellikle bebek hastalar için hayati risk oluşturabilmektedir. Bu ortamlarda yapılan ilaç hataları, hastaların sağlığını ciddi biçimde tehdit etmektedir. Nitekim 2023 yılında yayımlanan karma yöntemli bir sistematik derlemede, hastane öncesinde ilaç hatasına yol açan faktörlerin çok katmanlı olduğu ve bu hataların sıklığının önemli düzeyde olduğu raporlanmıştır [1]. İsrail'deki bir üniversite hastanesinin acil servisine sevk edilen yetişkin hastalar üzerinde yapılan retrospektif bir araştırmada verilen ilaç sayısının fazlalığının ve uzun transport sürelerinin daha yüksek oranda ilaç hatalarına olduğu anlaşılmıştır [2].

Tüm bu sorunlar göz önünde bulundurulduğunda, hastane öncesi acil sağlık hizmetlerinde kullanılmak üzere; kompakt, yerçekimine ihtiyaç duymadan sıvı akışını sağlayabilen, gerektiğinde hastanın üzerine veya koluna monte edilebilen,

ilaç dozlarını otomatik olarak hesaplayabilen, sağlam yapılı, şiddetli sıcaklık değişimlerine uyumlu, hareket ve elektriksel parazitlere dayanıklı, taşınabilir infüzyon pompası tasarımlarına ihtiyaç duyulmaktadır. İdeal olarak bu cihazlar, en az 24 saat kesintisiz çalışabilecek kapasitede, şarj edilebilir ve uzun ömürlü pillerle donatılmalıdır.

2013-2015 yılları arasında Almanya, Bad Belzig acil tıp hizmet merkezinin hastane öncesi acil servis doktorları tarafından tedavi edilen 708 hastayı retrospektif olarak değerlendirilmiş ve özellikle yaşlı hastalar ilaç hataları açısından yüksek riskli olarak tanımlanmıştır [3]. Aynı şekilde pediatrik olgularda durum daha da endişe vericidir. Paramedikler hasta ağırlıklarıyla karşılaştırıldığında sıklıkla uygun doz aralığının dışında uygulama gerçekleştirdkleri anlaşılmıştır [4].

Hastane ortamında yaygın olarak kullanılan akıllı infüzyon pompaları programlanabilir doz sınırları, ve alarm sistemleri sayesinde hataları azaltma potansiyeline sahiptir. Ancak mevcut pompaların çoğu, ambulans veya olay yeri gibi dinamik koşullara uyum sağlayacak biçimde tasarlanmamıştır. Karmaşık menüler, küçük ekranlar ve uzun kurulum adımları, yüksek tempolu acil senaryolarda cihazların kullanımını kısıtlamaktadır. Bununla birlikte, literatürde ambulanslarda kullanılan infüzyon pompalarının hasta güvenliğini artırma ve sağlık personelinin iş yükünü azaltma potansiyeline sahip olduğu vurgulanmaktadır [5]. (Bu bağlamda, hastanın tıbbi geçmişini (alerjiler, kronik hastalıklar, daha önceki ilaç intoleransları vb.) işleyerek karar veren ve paramedikleri, protokolere dayalı olarak yönlendiren akıllı ve taşınabilir bir infüzyon sistemi gereksinimi ortaya çıkmaktadır. Infüzyon pompaları genellikle büyük hacimli olup, yerçekimine ve herhangi bir elektrik veya pil desteğine ihtiyaç duymadan çalışamazlar (Şekil 1). Bu nedenle bu çalışmada, halihazırda kullanılan infüzyon pompolarının hantal olması, çok yer kaplaması ve taşıma sırasında ortaya çıkan zorlukları gidermek amacıyla daha pratik, yer ve zamandan tasarruf edebilecek bir pompa tasarlanması amaçlanmıştır.



Şekil 1: Geleneksel (sol) ve Taşınabilir (sağ) infüzyon sistemi

II. Malzeme ve Yöntem

Bu araştırmada, hastane öncesi acil sağlık hizmetlerinde görev yapan paramediklerin ilaç uygulama süreçlerinde karşılaştıkları zorluklara çözüm sunmak amacıyla karar destekli, taşınabilir ve kullanıcı dostu bir akıllı infüzyon pompası prototipi geliştirilmiştir. Çalışma üç aşamada yürütülmüştür:



Şekil 2: Önerilen infüzyon pompası tasarımı

1. İhtiyaç Analizi: Araştırmanın ilk aşamasında, İzmir Ekonomi Üniversitesi İlk ve Acil Yardım Programı öğrencilerinin yaz stajı değerlendirmesi sonucu sahada karşılaşılan problemler toplanmıştır. Görüşmelerde özellikle ilaç doz hesaplamalarındaki hatalar, sıvı kesintileri, ekipman hantallığı ve transport sırasında cihazların ergonomik zorlukları ön plana çıkmıştır.

2. Literatür ve Mevcut Sistem Analizi: İlaç hataları, infüzyon pompalarının kullanımındaki sınırlılıklar ve hasta güvenliği ile ilgili ulusal ve uluslararası literatür taranarak sistemin ihtiyaç duyulan özellikleri belirlendi. Mevcut infüzyon pompa sistemlerinin teknik yetersizlikleri analiz edildi ve bunlara alternatif oluşturabilecek fonksiyonel kriterler çıkarılmıştır.

3. Prototip Tasarımı ve Sistem Akışı: Geliştirilen akıllı infüzyon pompası, hastanın yaş, boy, kilo, medikal geçmiş ve ilaç kontrendikasyon bilgilerini kullanıcıdan etkileşimli olarak alan bir arayüzle tasarlandı. Mikrodenetleyici kontrollü sistem, algoritma destekli doz hesaplaması yapmakta, ilaç-hasta uyumsuzluklarında sesli ve görsel uyarılar vermektedir. Ayrıca cihaz, "Hızlı Mod" özelliğiyle sık karşılaşılan acil durumlar için (kardiyak arrest, anaflaksi, hipoglisemi vb.) önceden tanımlı protokollerle hızlı ilaç uygulamasına olanak sağlamaktadır. Cihazın küçük boyutlu ve kola monte edilebilir yapıda olması, sahada taşınabilirliğini ve kullanım kolaylığını artırmaktadır. Bu çalışmada, acil sağlık hizmetlerinde görev yapan sağlık profesyonellerinin ilaç doz hesaplama ve uygulama süreçlerini desteklemek amacıyla akıllı bir infüzyon pompası geliştirilmiştir. Sistem yalnızca hastanın yaş, boy ve kilo bilgilerini değil, aynı zamanda tıbbi geçmişe dair kritik verileri de etkileşimli sorular yoluyla dikkate alır. Kullanıcıya doz hesaplama, kontrendikasyon kontrolü ve motor seçimi gibi karar noktalarında destek sunar.

III. Cihaz Tasarımı

A. Kullanıcı İş Akışı

Kullanıcı iş akışı, cihazın başlatılmasından infüzyon işleminin tamamlanmasına kadar olan süreci kapsar. İşlem cihazın açılması ve ilaç kutusunun hazneye yerleştirilmesiyle başlar. Borular peristaltik pompaya düzgün yerleştirildikten sonra kullanıcıdan hasta hakkında bilgi alınır. Eğer hasta bilinci yerindeyse, sistem yaş, boy, kilo gibi verileri ister ve ardından ilaç seçimi yapılır. Kontrendikasyon soruları kullanıcıya yöneltilir. Sonuçlara göre ilaç verilebilirlik durumu belirlenir.

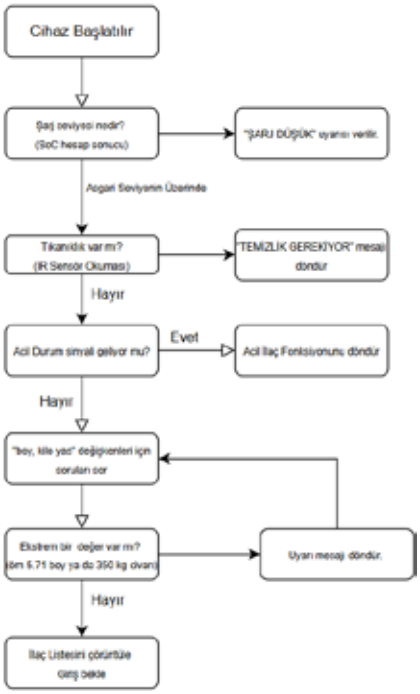
Eğer hasta bilinci kapalıysa, kullanıcı acil durum modunu aktif eder. Bu modda ilaç ve doz bilgileri manuel olarak girilir ve sistem doğrudan infüzyon işlemini başlatır. Kullanıcı ayrıca zaman kontrollü infüzyon seçeneğini de tercih edebilir.



Şekil 3: Kullanıcı İş Akışı Diyagramı

B. Teknik İş Akışı

Cihazın Başlatılması ve Hazırlık Aşaması: Cihaz ilk çalıştırıldığında sistem otomatik olarak açılış kontrol döngüsünü başlatır. Öncelikle batarya seviyesi kontrol edilir ve State of Charge (SoC) değeri değerlendirilir [6]. Eğer bu değer belirlenen eşiğin altında ise, kullanıcıya *"ŞARJ DÜŞÜK"* uyarısı sesli ve görsel yollarla iletilir. Ardından hortum hattının uygunluğu kontrol edilir. Bu aşamada IR sensörü kullanılarak hortum içerisindeki sıvı geçişi takip edilir. Eğer sensör, ışığın kırılma indeksindeki sabitliğe dayanarak sıvı akışı olmadığını tespit ederse sistem, hattın tıkalı olduğunu varsayar [7]. Tıkanıklık ya da kirlenme gibi durumlar söz konusuysa, kullanıcıya *"TEMİZLİK GEREKİYOR"* uyarısı gösterilir. Bu kontrollerin ardından kullanıcıdan yaş, boy ve kilo bilgileri istenir. Giriş verileri, tanımlı eşik değerlerin dışında ise kullanıcıya uyarı verilir ve düzeltme istenir. Doğrulanan veriler üzerinden uygun ilaçlar listelenir ve sistem bir sonraki adıma geçer.



Şekil 4: Başlatma ve Hazırlık Aşaması

Acil Modun Aktifleşmesi ve Süreç Akışı: Acil mod, hastanın bilinçsiz, iletişime kapalı ya da yaşamsal riski yüksek olduğu senaryolarda devreye alınmak üzere tasarlanmıştır. Bu mod, kullanıcıdan bilişsel geri bildirim alınamayan durumlarda vakit kaybını önlemek amacıyla sistem yönlendirmeleri devre dışı bırakılarak çalışır. Paramedik, acil mod aktif hale getirildiğinde kütüphanede tanımlı ilaç bilgisi ve dozajlarından birini seçebilir. Ya da doğrudan ilaç seçimi ve doz bilgilerini elle girer. Kullanıcı tarafından girilen doz bilgisi, sistemde tanımlı maksimum güvenli doz eşikleriyle karşılaştırılır. Eğer girilen değer bu sınırı aşıyorsa, sistem görsel ve sesli uyarılarla kullanıcıdan onay ister. Onay alınmadan infüzyon işlemi başlatılmaz. Onay verildiği takdirde sistem iki seçenek sunar: Doğrudan uygulama veya belirli bir zaman diliminde kademeli uygulama. Seçime göre motor kontrol birimi gerekli gerilim ve zamanlama parametrelerini alarak infüzyon sürecini başlatır. Bu sayede acil müdahale sırasında sistem, hızla devreye girerek kritik zaman kazanımı sağlar ve hayat kurtarıcı rol üstlenir.

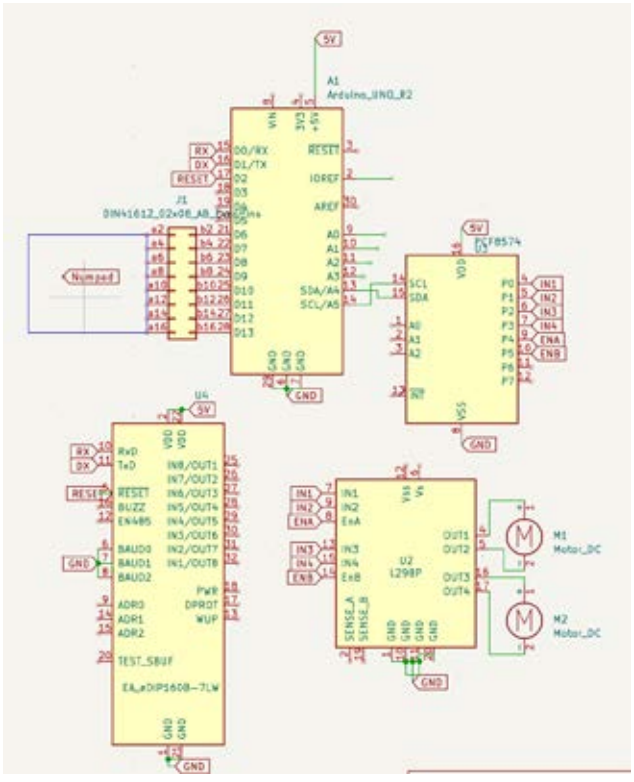
Kontrendikasyon Kontrollü Otomatik Doz Hesaplama Süreci: Normal modda kullanıcı, sistem tarafından sunulan kontrendikasyon sorularına yanıt verir. Bu yanıtlar ikili olarak kodlanır ve her bir ilaç için tanımlı kurallar dizisine göre değerlendirilir. Eğer kullanıcı yanıtları belirli bir risk düzeyini işaret ederse, sistem ilacın uygulanabilirliğine dair karar verir. Kısıtlı doz önerilen senaryolarda sistem kullanıcıya azami doz miktarını bildirir; uygun koşullarda ise otomatik doz hesaplama algoritması devreye girer.

Tasarlanan elektronik kartın şematik devresi Şekil 6'da görülmektedir. Hesaplanan doz doğrultusunda sistem, peristaltik pompa ya da hacme uygun başka bir motoru seçerek sürücü devresine sinyal gönderir. İlaç infüzyonu



Şekil 5: Acil Modun Başlatılması ve Manuel Süreç

başlatıldığında, IR sensör yalnızca tıkanıklık tespiti amacıyla değil, aynı zamanda sıvı geçişindeki sürekli değişimleri okuyarak gerçek zamanlı akış kontrolü sağlar [8]. Bu sensör sayesinde sıvı hızındaki ani azalmalar, mikro tıkanıklıklar ya da beklenmedik pompa arızaları erken tespit edilir. Böylece sistem yalnızca doğru dozda değil, doğru hızda ve kesintisiz infüzyon akışı sağlayacak şekilde çalışır. Bu geri bildirim mekanizması, infüzyon güvenliğini artıran hayati bir güvenlik katmanı işlevi görür.



Şekil 6: Tasarlanan Elektronik Kart Şematik Devresi

IV. TARTIŞMA VE SONUÇ

Bu çalışmada geliştirilen karar destekli akıllı infüzyon pompası, hastane öncesi acil sağlık hizmetlerinde ilaç uygulama hatalarını azaltmayı ve paramediklerin karar verme süreçlerini desteklemeyi amaçlamaktadır. Literatürde bu tür hataların hasta güvenliğini ciddi şekilde tehdit ettiği bilinmektedir [9], [10]. Geliştirilen cihaz, kola monte edilebilen kompakt yapısı, kullanıcı dostu arayüzü, düşük ışıkta çalışabilen ekranı ve hasta verilerine dayalı doz hesaplama algoritmasıyla öne çıkmaktadır. "Hızlı Mod" özelliği sayesinde kardiyak arrest, anaflaksi veya hipoglisemi gibi durumlarda önceden tanımlı protokoller hızla uygulanabilmektedir. Ayrıca kontrendikasyonlara dayalı otomatik doz önerisi ve IR sensörle sıvı akış kontrolü gibi özellikler, hem doğruluk hem de güvenlik açısından önemli avantajlar sağlamaktadır. Sonuç olarak, bu akıllı infüzyon pompası, paramediklerin iş yükünü azaltabilecek ve hasta güvenliğini artırabilecek önemli bir teknolojik yenilik sunmaktadır. Klinik testlerle desteklendiğinde, hastane öncesi bakımda yaygın şekilde kullanılabilecek potansiyele sahiptir.

KAYNAKÇA

[1] Walker D, Moloney C, SueSee B, Sharples R, Blackman R, Long D, Hou XY. Factors Influencing Medication Errors in the Prehospital Paramedic Environment: A Mixed Method Systematic Review. Prehosp Emerg Care. 2023;27(5):669-686. doi: 10.1080/10903127.2022.2068089. Epub 2022 Jun 27. PMID: 35579544.

[2] Lifshitz AE, Goldstein LH, Sharist M, Strugo R, Asulin E, Bar Haim S, Feigenberg Z, Berkovitch M, Kozer E. Medication prescribing errors in the prehospital setting and in the ED. Am J Emerg Med. 2012 Jun;30(5):726-31. doi: 10.1016/j.ajem.2011.04.023. Epub 2011 Jul 8. PMID: 21741787.

[3] Ramadanov N, Klein R, Schumann U, Aguilar ADV, Behringer W. Factors, influencing medication errors in prehospital care: A retrospective observational study. Medicine (Baltimore). 2019 Dec;98(49):e18200. doi: 10.1097/MD.0000000000018200. PMID: 31804342; PMCID: PMC6919435.

[4] Hoyle JD, Davis AT, Putman KK, Trytko JA, Fales WD. Medication dosing errors in pediatric patients treated by emergency medical services. Prehosp Emerg Care. 2012 Jan-Mar;16(1):59-66. doi: 10.3109/10903127.2011.614043. Epub 2011 Oct 14. PMID: 21999707.

[5] Atanda O, West J, Stables T, Johnson C, Merrifield R, Kinross J. Flow rate accuracy of infusion devices within healthcare settings: a systematic review. Ther Adv Drug Saf. 2023 Jul 21;14:20420986231188602. doi: 10.1177/20420986231188602. PMID: 37492690; PMCID: PMC10363896.

[6] A. Bayraktar and O. Aydın, "Determination of Battery State of Charge with a Data-Driven Ensemble Learning Model for Electric Vehicle Applications," *Frontiers in Energy Research*, vol. 13, p. 1573972, 2025. doi: 10.3389/fenrg.2025.1573972.

[7] Y. Saputro and R. S. Nugroho, "Design and Prototype of Infusion Monitoring System Based on Infrared Sensor and Microcontroller," *Jurnal Teknologi dan Sistem Komputer*, vol. 15, no. 1, pp. 1–6, 2022. doi: 10.35882/teknokes.v15i1.1.

[8] M. F. A. Basher, S. A. Vasantharaj, A. A. Elattar, and H. M. A. Fahmy, "Smart Intravenous (IV) Infusion Monitoring System," *Inventions*, vol. 9, no. 4, p. 89, 2021. doi: 10.3390/inventions9040089.

[9] Poranen A, Kouvonen A, Nordquist H. Human errors in emergency medical services: a qualitative analysis of contributing factors. Scand J Trauma Resusc Emerg Med. 2024 Aug 30;32(1):78. doi: 10.1186/s13049-024-01253-7. PMID: 39215372; PMCID: PMC11363522.

[10] Takhtinejad NJ, Stewart D, Nazar Z, Hamad A, Hadi MA. Identifying factors influencing clinicians' reporting of medication errors: a systematic review and qualitative evidence synthesis using the theoretical domains framework. Expert Opin Drug Saf. 2024 Oct;23(10):1271-1282. doi: 10.1080/14740338.2024.2396397. Epub 2024 Sep 16. PMID: 39192820.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Navigating the Healthcare Maze through Natural Language: HAAN An AI-Based Health Tool

Tareq M F Abdullah AlHammoodi, Kaan Özer, Hulki Enes Uysal, Kutluhan Erol, İlker Korkmaz
Faculty of Engineering
Izmir University of Economics
Izmir, Türkiye
{tareq.alhammoodi, kaan.ozer, enes.uysal}@std.ieu.edu.tr, {kutluhan.erol, ilker.korkmaz}@ieu.edu.tr

***Abstract*—Many people have a hard time navigating the healthcare systems, especially young people with little previous experience, or the elderly, who might get lost in the maze of mobile and web applications. The proposed system aims to provide user-friendly guidance through a chatbot platform utilizing pretrained Large Language Models (LLMs) combined with a Retrieval-Augmented Generation (RAG) in order to deliver reliable answers with references to their direct sources.**

Whether someone wants to know about a certain procedure, or learn about health insurance coverage, the system is designed to make the process easier and more understandable for everyone. This supportive approach not only empowers individuals to take better care of their health, but also helps reduce the burden on healthcare providers. Ultimately, the project seeks to close the information gap and offer a reliable source of assistance for people navigating everyday health concerns, particularly within the context of the Turkish healthcare system.

***Keywords*—digital health tool, retrieval augmented generation (RAG), large language models (LLMs), e-health chatbots.**

I. Introduction

Although modern healthcare systems have made significant advances in diagnosis and treatment, the process of actually using their services can be daunting for many people, including the young and the elderly, nonnative speakers, or others with limited digital literacy. For example, someone might want to get an eye exam, but does not know where to start. Should they go to a hospital, or a clinic? Are they allowed to see a specialist directly, or do they need a referral from their primary care physician? Do they need an appointment? What costs are covered by their insurance versus out-of-pocket expenses? How do they track the ever changing rules and regulations?

In Türkiye, as in many places, these challenges are compounded by a decentralized healthcare environment involving public hospitals, private clinics, family physicians, and specialized care centers, each one of them is governed by different procedures, insurance policies and appointment systems. While the Turkish Ministry of Health provides digital services such as MHRS (Centralized Hospital Appointment System) [1], e-Nabız (Personal Health Record) [2], and the Health Ministry Communication Center (SABİM 184) [3], these platforms often assume a level of familiarity that many users do not possess. This information gap can lead to unnecessary emergency visits or confusion.

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

This paper introduces HAAN (Health-Aware AI Assistant for Needs), an AI-powered chatbot that utilizes a Retrieval-Augmented Generation (RAG) framework to provide guidance for everyday health-related questions. The system is built on a Large Language Model (LLM) and trained with reliable health documents, ministry guidelines, and hospital protocols. HAAN provides users with answers related to appointments, tests, referrals, insurance procedures, and more. It also incorporates a feedback mechanism where users can rate and review responses, improving the accuracy and maintainability.

By bridging the gap between individuals and the services they need, HAAN is a step toward a more inclusive, efficient, and supportive healthcare environment in alignment with several UN Sustainable Development Goals (SDG) [4]. It advances SDG 3 (Good Health and Well-being) by improving access to healthcare knowledge, supports SDG 10 (Reduced Inequalities) by empowering underserved communities, and promotes SDG 16 (Peace, Justice, and Strong Institutions) by improving transparency in healthcare interactions.

II. Literature Review

In recent years, the integration of Artificial Intelligence (AI) into healthcare has gained significant momentum in improving patient care and clinical decision-making, as well as in expanding access to medical services. While AI applications in healthcare have historically focused on diagnosis and data analytics, patient-facing chatbots now represent a growing area of innovation. WHO has recognized the role of AI chatbots in increasing health awareness and early intervention, especially in underserved or digitally inexperienced populations [10]. These tools are also being incorporated into chronic disease management, vaccination scheduling, and mental health support, expanding their value beyond initial triage. They are especially valuable in facilitating communication between patients and healthcare systems, reducing the burden on professionals, and increasing health literacy [5].

Despite widespread adoption of electronic health records, usability and infrastructure challenges remain [6]. Several studies suggest that AI chatbots with conversational interfaces and structured guidance, such as Ada Health and Babylon Health, can successfully help bridge this gap [7, 8]. These systems use natural language processing (NLP) for symptom

assessments and recommendations, though many rely on static knowledge bases with infrequent updates. To address this, our proposed chatbot leverages a RAG architecture, combining vector search and LLMs to provide contextually accurate, up-to-date responses. This design, as supported by Amugongo et al. [9], reduces hallucinations and improves user trust in clinical settings.

Crowdsourced evaluation has also emerged as a valuable method for refining AI chatbot performance. Amershi et al. [11] emphasize the importance of crowdsourced evaluation in improving interactive AI systems, especially in real-world, user-centered applications. This approach invites users to rate chatbot responses based on usefulness or relevance, and optionally explain any issues they faced. These ratings are stored and analyzed to prioritize the most helpful answers in future conversations. As a result, each interaction contributes to improving the system's performance over time, creating a continuous feedback loop between users and AI.

Ultimately, AI chatbots based on RAG can advance digital health equity and offer scalable, cost-effective communication tools as the healthcare sector continues to digitize.

III. Methodology

This section presents the overall approach used to design and develop the system. The methodology focused on selecting effective AI techniques, building a relevant knowledge base, implementing a modular architecture, and validating the system through testing and user feedback.

After exploring various methods for building AI systems that deliver accurate and relevant responses, we chose RAG, as it integrates document retrieval with text generation, enabling AI to leverage a curated knowledge base. By grounding responses in retrieved documents, we aimed to reduce the risk of hallucinations and improve the factual reliability.

A. Data Collection and Knowledge Base Creation

In order to create a reliable knowledge base, we manually gathered hundreds of documents from trusted sources like the Ministry of Health, insurers, and hospitals and organized them into smaller, structured sections optimized for retrieval. To enable semantic search, we used an embedding model to convert each text chunk into a high-dimensional vector that captures its meaning as shown in Fig. 1. This allows the system to find relevant content based on semantics rather than just keywords, using models like nomic-embed-text or OpenAI's text-embedding-ada-002.

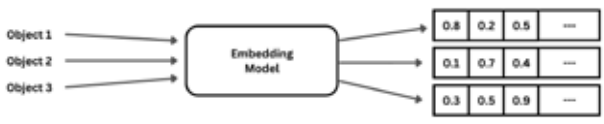


Fig. 1: Embedding representation.

These embeddings were then stored in a Vector Database (VDB), allowing efficient similarity search during the retrieval phase. As represented in Fig. 2, when a user asks a question, the system uses the embedding of the question to find the most semantically similar document chunks from the VDB.

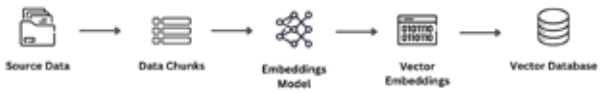


Fig. 2: Knowledge base preparation.

This is done by using a mathematical approach known as cosine similarity to compare the stored embeddings. Cosine similarity measures the cosine of the angle between two vectors in a multi-dimensional space, which reflects how similar they are in terms of direction rather than magnitude [12]. A cosine similarity score close to 1 indicates that the vectors are pointing in nearly the same direction, suggesting a high degree of similarity between the corresponding data points as shown in Fig. 3.

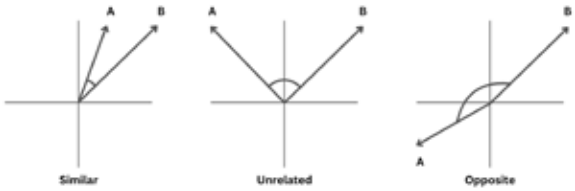


Fig. 3: Cosine similarity representation.

This makes cosine similarity particularly effective for comparing high-dimensional data like text embeddings, where the actual values may differ but the overall meaning or context is preserved in their directional alignment. This approach enables the chatbot to understand user queries even when they are phrased differently from the original documents.

B. System Architecture

The system was designed to offer a smooth and secure user experience. The front-end, built with React.js and styled using Tailwind CSS [13], provides an interactive chat interface where users can ask questions, see real-time updates, and leave feedback. It communicates with the backend securely using Axios. On the backend, Node.js with Express.js handles user authentication (via Google OAuth), session management, and chat storage using MongoDB [14]. A separate AI service, developed with Python and FastAPI, manages message processing and response generation by integrating tools like Ollama [15], Chroma DB [16], and LangChain [17]. The architecture follows a modular structure to ensure clarity, maintainability, and scalability across both the application and AI services.

The chatbot's core AI was built using Ollama, an open-source platform for running LLMs locally. Ollama was chosen for its support of retrieval-augmented workflows, efficient hardware usage, and easy model customization. We tested several mid-size models (e.g., Mistral 7B, LLaMA 3.2 derivatives)

and selected one optimized for conversational performance and low hallucination.

The answer generation process uses a RAG architecture with two main stages: retrieval and generation as shown in Fig. 4. In the document retrieval phase, the system first embeds the user's question into a dense vector using an embedding model. This vector is then used to perform a semantic search against a VDB to retrieve the top-k most relevant document chunks. These retrieved chunks are concatenated and formatted to build the context needed for answering.

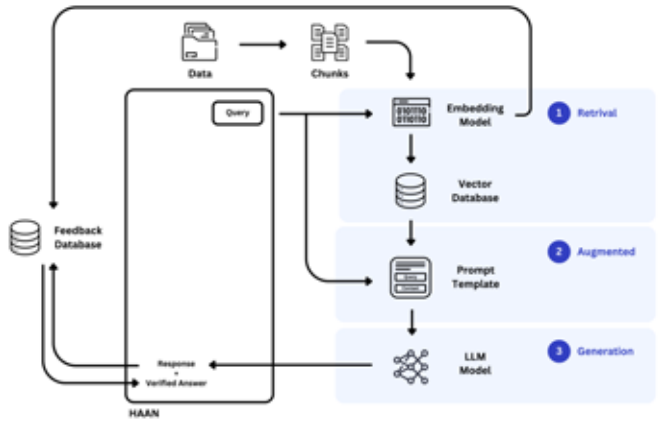


Fig. 4: RAG data flow diagram.

In the contextual generation phase, this context is passed to the Ollama model. A structured prompt is created that includes a system message, the retrieved documents, and the original user's question. This prompt is sent to the Ollama API for inference, where the model is configured with a low temperature (e.g., 0.2, meaning minimal randomness), which generates an answer based strictly on the provided content and reduces deviations from the source material. Fig. 5 depicts an example of the system prompt.

```
System Message: "You are a helpful assistant that only answers
based on the following reference materials. If information is
not available, respond with ' I'm sorry, but I don't have
enough knowledge to answer this question.'"
Reference Materials:
[Data Chunk 1]
[Data Chunk 2]
[Data Chunk 3]
User Question:
[User's Input]
```

Fig. 5: A sample prompt structure.

This approach helps maintain answer reliability and minimize hallucination (i.e., the model making up facts). By grounding the AI's outputs in official documents, we make sure that users receive contextually appropriate advice tailored to their needs.

HAAN was designed with a focus on user experience. The design philosophy was rooted in simplicity and clarity. It is important to provide a system that is easy to interact with users from different backgrounds. Therefore, the interface was deliberately kept focused on the following essentials: ask, get guidance, and act as seen in Fig. 6. It supports responsive design across various devices, aligning with the goal of simplifying access to healthcare services.



Fig. 6: HAAN user interface.

C. User Feedback and Analysis Mechanism

Recognizing that AI performance cannot be fully accurate, we implemented a user feedback system. After each response, users can indicate if it was helpful. Negative feedback prompts a brief explanation (e.g., "Incomplete," "Not relevant"). Satisfactory responses along with their query, with user consent, are stored as a `QUESTION` object in the database as a semantic embedding vector. Embeddings allow us to compare meaning rather than just text, so similar questions with different wording will have close vectors. Using cosine similarity, new queries are matched against past ones. If similarity exceeds a threshold value (e.g., greater than 0.90), the system retrieves past answers that are positively rated, anonymized and stripped of any user-identifiable information, then displays them alongside a freshly generated AI response. All questions and responses are stored whether they are voted or not, using a unified schema. This supports analysis of common issues, content improvement, and better future responses without needing to retrain the model. The feedback loop enables a self-improving system driven by real user input.

IV. Discussion

The HAAN system demonstrated that RAG has the potential to assist users in complex healthcare tasks. By using embedding-based semantic search over a curated medical knowledge base, the chatbot consistently provided contextually relevant and factually grounded answers, and showed flexibility in handling different phrasings of similar questions in our preliminary tests.

HAAN combines information retrieval and generative AI to reduce hallucination by grounding answers in real sources, a critical factor in healthcare applications. The use of vector-based embeddings and cosine similarity effectively bridged the gap between user queries and source content, corroborating existing literature on the role of vector databases in NLP.

By offering accurate and understandable health information in a conversational format, HAAN has the potential to make healthcare more accessible. Its adaptable nature also makes it useful for institutions managing communication during public health events. The built-in feedback system promotes trust and ensures continuous improvement based on real user input.

Although the system performed well in early scenarios, certain limitations remain. Some legal and bureaucratic topics are highly dynamic and require constant updates, which place a burden on maintaining the knowledge base. Furthermore, while the feedback mechanism is powerful, it relies on users being proactive in rating responses, which may not always occur in reality. In addition, users might shy away from giving consent to have their questions and responses utilized by the system. To address these challenges integration with government databases or APIs in real time, enhancing anonymization and scrubbing techniques could prove useful. Other improvements could include multilingual support and adaptive summarization for complex documents. The system should also be tested more extensively in cases involving rare legal situations or scenarios with incomplete or ambiguous user input to see its performance in diverse real-world contexts.

Despite these limitations, HAAN lays a meaningful foundation for accessible, AI-driven health-care support, and with future improvements, it has the potential to become an even more reliable and inclusive tool for those who need it the most.

V. Conclusion

In this project, we developed HAAN, a web-based chatbot designed to help users navigate the healthcare system by providing accurate, document-backed answers using a combination of LLMs and RAG. Built on a structured knowledge base from trusted sources and featuring a user-friendly interface with semantic search and feedback capabilities, HAAN demonstrated the viability of delivering relevant, grounded responses. While the system as it is currently implemented has several aforementioned limitations including language support and occasional gaps in completeness, future improvements and further testing can address those. Future improvements should focus on expanding content coverage, adding multilingual and accessibility features, and strengthening data protection. Scalability and security remains another critical area of improvement. Overall, HAAN highlights how AI can improve healthcare access when designed with reliability, usability, and real-world needs in mind.

References

[1] T.C. Sağlık Bakanlığı, "Merkezi Hekim Randevu Sistemi (MHRS)," https://www.mhrs.gov.tr, (accessed on July 25, 2025).
[2] T.C. Sağlık Bakanlığı, "E-Nabız Kişisel Sağlık Sistemi," https://enabiz.gov.tr, (accessed on July 25, 2025).
[3] Sağlık Bakanlığı İletişim Merkezi (SABİM), https://sabim.gov.tr, (accessed on August 5, 2025).
[4] United Nations Sustainable Development Goals, https://sdgs.un.org/goals, (accessed on May 10, 2025).
[5] L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, and A.Y.S. Lau, "Conversational agents in healthcare: a systematic review," Journal of the American Medical Informatics Association, vol. 25, no. 9, pp. 1248–1258, Sep. 2018, doi: 10.1093/jamia/ocy072.
[6] İ. Köse, J. Rayner, Ş. Birinci, M.M. Ulgu, İ. Yılmaz, and Ş. Güner, HIMSS Analytics Team, and MoH Team, "Adoption rates of electronic health records in Turkish hospitals and the relation with hospital sizes," BMC Health Services Research, vol. 20, article no. 967, 2020, doi:10.1186/s12913-020-05767-5.
[7] S.M. Jungmann, T. Klan, S. Kuhn, and F. Jungmann, "Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: Comparative case study with lay and expert users," JMIR Formative Research, vol. 4, no. 4, e13863, Oct. 2019.
[8] S. Razzaki, A. Baker, Y. Perov, K. Middleton, J. Baxter, D. Mullarkey, D. Sangar, M. Taliercio, M. Butt, A. Majeed, A. DoRosario, M. Mahoney, and S. Johri, "A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis," arXiv, Jun. 2018, https://arxiv.org/abs/1806.10698, (accessed on May 10, 2025).
[9] L.M. Amugongo, P. Mascheroni, S. Brooks, S. Doering, and J. Seidel, "Retrieval augmented generation for large language models in healthcare: A systematic review," Nature Medicine, 2025, Article ID 12157099.
[10] World Health Organization, "Ethics and governance of artificial intelligence for health," WHO, Geneva, 2021, https://www.who.int/publications/i/item/9789240029200, (accessed on August 5, 2025).
[11] S. Amershi, M. Cakmak, W.B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," AI Magazine, vol. 35, no. 4, pp. 105–120, 2014. doi:10.1609/aimag.v35i4.2513.
[12] "Cosine similarity," ScienceDirect Topics, Elsevier, https://www.sciencedirect.com/topics/computer-science/cosine-similarity, (accessed on May 10, 2025).
[13] Meta Platforms, Inc., "React – A JavaScript library for building user interfaces," https://reactjs.org, (accessed on August 5, 2025).
[14] MongoDB Inc., "MongoDB: The developer data platform," https://www.mongodb.com, (accessed on August 5, 2025).
[15] Ollama Inc., "Ollama: Run LLMs locally," https://ollama.com, (accessed on August 5, 2025).
[16] Chroma, "Chroma: The AI-native open-source vector database," https://www.trychroma.com, (accessed on August 5, 2025).
[17] LangChain, "LangChain: Building applications with LLMs through composability," https://www.langchain.com, (accessed on August 5, 2025).

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Artificial Intelligence Models for the Detection of Breast Cancer in Medical Imaging

Ece Akaydın[1], Ayşe Rüya Arslan[2] and Gülay Çiçek[3]
[1,2]Department of Software Engineering, [3]Department of Biomedical Engineering
Istanbul Beykent University, Faculty of Engineering and Architecture, Sariyer, Istanbul, Turkey
[1]2103013295@student.beykent.edu.tr, [2]2203013050@student.beykent.edu.tr
[3]gulaycicek@beykent.edu.tr

***Abstract*—This study compares the performance of deep learning (ResNet50, VGG16) and traditional machine learning (Random Forest, SVM) algorithms in breast cancer diagnosis. Convolutional neural networks (CNN) were used for feature extraction, and the impact of Principal Component Analysis (PCA) and dimensionality reduction on model performance was analyzed. In experiments on the BreakHis dataset, the effect of PCA on AUC values was evaluated, and comparisons were made between models. The results show that the ResNet50 model using PCA achieved the highest performance, reaching 99.21% AUC. While PCA improved performance in some models, it negatively impacted performance in others. This study is unique in that it compares deep learning and traditional machine learning methods under the same dataset and experimental conditions, with and without PCA. This study allows for the integration of the obtained results not only into academic but also clinical decision support systems. The comprehensive analysis provides valuable findings for both researchers and healthcare professionals...**

***Keywords*—Breast Cancer Diagnosis, Deep Learning, Machine Learning, PCA, BreakHis Dataset**

I. Introduction

Breast cancer is the most common malignant tumor in women, accounting for approximately 30% of all female cancer cases. Early diagnosis is critical to treatment success. Imaging methods such as mammography, ultrasound, and MRI are widely used in the diagnostic process; however, due to limitations such as cost, access, and margin of error, these methods can be inadequate in some cases. Manual evaluations by pathologists are also time-consuming and error-prone. In this context, the use of artificial intelligence, and particularly deep learning methods, has great potential to increase accuracy and speed up the process of breast cancer diagnosis. Image processing techniques and CNN-based structures can detect abnormal tissues at an early stage. Significant artificial intelligence and deep learning-based research on breast cancer conducted in recent years is summarized. The reviewed studies were evaluated in terms of the datasets used, methods, results obtained, shortcomings, and future contributions. Dembrower et al. (2023) reported that artificial intelligence was not as effective as a second radiologist in a study of 55,581 data points, indicating that the technology is still developing. [1]. İbrahim et al. (2020) reported that the CNN model they developed on mammography images obtained from different hospitals demonstrated higher AUC performance than doctors; however, the lack of dataset size limits the generalizability of the study. [2].Alom et al. (2019) reported accuracies of 97.95% and 99.05% using the IRRCNN model with the Kaggle and BreakHis datasets, respectively; however, the study did not include machine learning methods, which is limiting in terms of methodological diversity [3].

The unique contributions of this study to the literature are as follows:

- While many studies in the literature focus solely on deep learning or machine learning methods, this study considers both approaches together and systematically examines the effects of PCA implementation/non-implementation on both classical algorithms and deep network architectures.
- By evaluating the performance of different model groups on breast cancer images under the same dataset and conditions, the study provides a fair and comparable analysis. In this respect, it serves as an important reference for model selection in the field of medical image processing.
- Designing the findings in a way that can be integrated not only into academic settings but also into clinical applications adds practical value to the study and increases its applicability for decision support systems.

II. *Method*

A. *Dataset and Preprocessing*

This study used the BreakHis dataset, obtained from the Kaggle platform, which contains histopathological images for breast cancer detection. [4]. The dataset comprises a total of 7,909 images collected from 82 patients, with 2,480 benign and 5,429 malignant samples, highlighting a class imbalance that could potentially affect model training. Images were captured at magnification levels of 40X, 100X, 200X, and 400X, each with a resolution of 700×460 pixels and a 3-channel RGB format. During preprocessing, the accuracy of the labels was checked and corrected if necessary to ensure data quality. The dataset was then divided using a 5-fold stratified cross-validation strategy, allocating approximately 70% of the data for training and 30% for testing in each fold. This systematic preprocessing and splitting approach ensures that both training and evaluation phases are robust and reproducible.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

B. Feature Extraction

In this study, Convolutional Neural Networks (CNN) will be used for feature extraction. In CNN, image pixels are processed with convolutional layers to extract features.

C. Feature Selection

Features extracted from visual data are often high-dimensional, which can increase model complexity and negatively impact the learning process. Therefore, dimensionality reduction techniques were employed to improve classification performance and reduce the risk of overfitting. In this study, PCA was used to retain only the most informative features, providing more efficient input to the classification models.

D. Model Selection

- Machine Learning Algorithms: Machine Learning is a process that utilizes mathematical models to enable a computer to learn without requiring explicit instructions. In recent years, machine learning methods have become frequently preferred because they have significantly increased classification accuracy. Two machine learning algorithms, Random Forest and Support Vector Machines, were implemented in this study.
  - Random Forest: A Random Forest is an algorithm consisting of multiple decision trees. It trains each decision tree with different pieces of data. This way, each tree provides a different perspective, and the overall accuracy of the model increases [5]
  - Support Vector Machines (SVM): This algorithm is generally used for data classification. It is used to classify both linearly separable and non-linearly separable data [6].
- Deep Learning Architectures: Deep learning is a more advanced version of artificial neural networks. It is used specifically for classification, recognition, and detection [7]. The fundamental architecture of Deep Learning is considered the CNN architecture.
  - ResNet: ResNet is a specialized CNN architecture consisting only of convolutional layers. Essentially, it's a structure that allows layer blocks to be skipped via shortcuts. Unlike previous architectures, it's a deeper architecture designed with 152 layers. It also won the 2015 ImageNet competition with a 3.6% error rate.
  - VGGNet: VGG is a CNN architecture developed for image processing on large datasets. It consists of a structure consisting of 11, 16, and 19 layers, called VGG-11, VGG-16, and VGG-19, respectively.
  - Convolutional Neural Networks (CNNs) were chosen for model selection due to their success, particularly in image processing and classification. CNNs automatically extract features from images, reducing the need for manual feature engineering. Furthermore, when used in conjunction with techniques such as data augmentation, they reduce the risk of overfitting and increase the model's generalization ability. This increases the model's accuracy and reliability.

E. Model Training

Dropout was applied after fully connected layers in model training to prevent overfitting. [8]. Training continued for 50 epochs, and hyperparameters such as learning rate and batch size were determined to increase model performance [9]. Model performance was evaluated using metrics such as accuracy, sensitivity, specificity, precision, and F-score. [10]. These metrics are calculated based on the TP, TN, FP, and FN values in the complexity matrix. Additionally, the model's classification performance is visualized with the ROC curve, and the overall success is measured with the AUC (area under the curve). The AUC value reflects the model's ability to distinguish between positive and negative classes.

III. Experimental Studies

This section presents experimental studies conducted with different deep learning and machine learning models for the classification of breast cancer images. PCA was applied to the obtained features for dimensionality reduction, and their classification performance was evaluated. All models were tested on the BreakHis dataset and analyzed using metrics such as accuracy, precision, sensitivity, and F-measure.

Confusion Matrix: It is one of the main metrics used to evaluate the performance of classification algorithms. This matrix, whose rows represent the classes predicted by the model and whose columns represent the true classes, numerically represents the model's correct and incorrect predictions. Each cell represents the number of examples at the intersection of a given class prediction and the true class.

A. PCA On - Off Applications and Metrics

This section evaluates the performance impact of applying PCA to different classification models. Table I presents a comparison of the accuracy, precision, sensitivity, and F-measure values obtained for Support Vector Machines (SVM), Random Forest (RF), ResNet50, and VGG16 models before (OFF) and after (ON) PCA. This examines the contribution of dimensionality reduction to the classification performance of deep learning-based feature extractors, along with traditional machine learning algorithms. The graph below provides a comparative graph of the Machine and Deep Learning models. Table I presents a class-by-class comparison of the performance of four different models before and after PCA application in terms of metrics such as accuracy, precision, sensitivity, and F1 score. This table allows us to numerically observe the impact of PCA on each model,

and clearly demonstrates which models experience performance gains or losses.

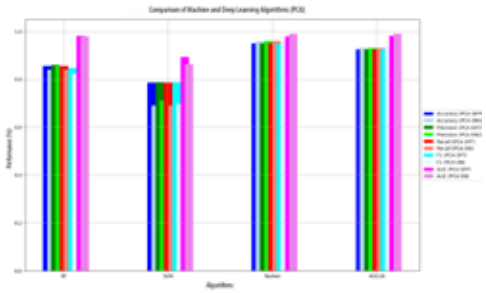


Fig. 1: Comparison of Machine and Deep Learning Models

When the results presented in Table I are evaluated in general, the performance differences between the models before and after PCA can be summarized as follows:
Analyzing the obtained results, the accuracy rate for the VGG16 model before PCA was applied was calculated as 92.58%. The area under the ROC curve (AUC) value for this structure was determined as 0.9693. With PCA applied, the accuracy rate increased to 93.16%, and the AUC value reached 0.9781. Small improvements were also observed in metrics such as sensitivity, specificity, and F1-measure. The accuracy rate achieved without PCA in the Random Forest (RF) algorithm was 85.57%. In this case, the AUC value was measured as 0.9315. However, when PCA was applied, the accuracy dropped to 83.99%, and the AUC value was recorded as 0.9307.The reason for this negative effect is that the RF algorithm creates decision trees based on raw features. When PCA loses some discriminative features, the model loses some of the distinction between classes and performance declines. Similar decreases were observed in metrics such as macro-average and F1-score. These data reveal that PCA has a negative impact on the RF model in terms of both overall accuracy and discrimination power. The ResNet model achieved an accuracy rate of 94.98% without PCA, and the AUC value for this structure was calculated as 0.9889, a very high level. With PCA applied, the accuracy increased to 95.85%, and the AUC value increased even further, reaching 0.9921. The reason for this decline is that PCA loses some of the discriminative information from the high-dimensional CNN features. In particular, the fine details that distinguish malignant and benign cells are not sufficiently represented after dimensionality reduction, causing the model performance to decrease somewhat. Significant improvements were also observed in the precision, recall, and F1-score metrics. These results show that PCA provides a significant performance increase over the ResNet architecture, and that the model can distinguish the positive class with higher accuracy. The accuracy rate of the SVM algorithm before PCA was applied was determined to be 78.77%. The AUC value for this structure was 0.8354. However, when PCA was applied, the accuracy rate decreased to 69.04%, and the AUC value decreased to 0.7274. Significant decreases were also observed in the precision, recall, and F1-score metrics. This indicates that PCA significantly negatively impacted the classification success of the SVM model. This negative effect can be explained by the fact that SVM is highly sensitive to the distribution and scaling of features. PCA reduces dimensionality by selecting principal components that explain the most variance, but some features that are discriminative for SVM classification may be discarded during this process. Overall, while PCA increased the classification success and discrimination ability (AUC) in the ResNet and VGG16 models, it decreased the performance in terms of both accuracy and area under the ROC curve in the Random Forest and SVM algorithms. However, the sensitivity (recall) values for the malignant class were generally high in all models, indicating that the models were generally successful in detecting malignant cases.
Figure 2 compares the average AUC values of four different models (VGG16, SVM, ResNet50, and Random Forest) without (PCA OFF) and after (PCA ON). The results show that PCA provides small improvements in AUC values for some models (e.g., VGG16 and ResNet50), but performance decreases for the SVM model.

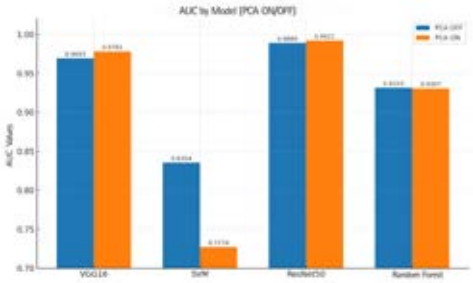


Fig. 2: Comparison of Machine and Deep Learning Models

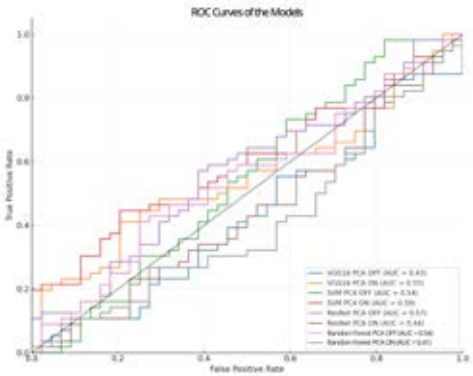


Fig. 3: Comparison of Machine and Deep Learning Models

Figure 3 presents the ROC curves obtained for all models with PCA ON and PCA OFF. The ROC curves visually demonstrate the classification performance of each model at different threshold values and are supported by AUC

TABLE I: Experimental Results of Classification Algorithms (PCA)

Algorithm	PCA	Accuracy	Precision		Recall		F-Skor		AUC
			Class 0	Class 1	Class 0	Class 1	Class 0	Class 1	
VGG16	OFF	0.93	0.90	0.94	0.86	0.96	0.88	0.95	0.9693
	ON	0.93	0.90	0.94	0.88	0.96	0.89	0.95	0.9781
RF	OFF	0.86	0.89	0.85	0.62	0.97	0.73	0.90	0.9315
	ON	0.84	0.94	0.82	0.52	0.98	0.67	0.89	0.9307
ResNet	OFF	0.95	0.92	0.96	0.92	0.97	0.92	0.96	0.9889
	ON	0.96	0.95	0.96	0.92	0.98	0.93	0.97	0.9921
SVM	OFF	0.79	0.66	0.85	0.67	0.84	0.66	0.84	0.8354
	ON	0.69	0.51	0.81	0.62	0.72	0.56	0.76	0.7274

scores. These graphs provide a visual comparison of the models' accuracy-performance balance.

IV. *Discussion*

The effectiveness of Principal Component Analysis (PCA) in machine learning varies significantly based on the model. In traditional models like Random Forest and SVM, applying PCA generally led to a decrease in performance. For Random Forest, accuracy dropped from 85.57% to 83.99%, while the drop was more substantial for SVM (from 78.77% to 69.04%), suggesting PCA can be detrimental to some tree-based and non-linear models. Conversely, PCA had a positive effect on deep learning models. For VGG16, accuracy saw a slight increase from 92.58% to 93.16%, and the improvement was more pronounced with ResNet, where accuracy rose from 94.98% to 95.85%, along with significant gains in other metrics. This indicates that PCA can be a beneficial preprocessing step in deep learning architectures, especially with complex, high-dimensional data. Ultimately, these findings show that dimensionality reduction techniques like PCA are not universally effective; their success is directly tied to the compatibility between the specific algorithm and the preprocessing technique used.

V. Conclusion

This study systematically evaluated the effect of Principal Component Analysis (PCA) on both traditional machine learning and deep learning models for breast cancer detection. The findings reveal that PCA's impact is model-dependent: it improved performance in deep learning architectures, achieving the highest accuracy of 95.85% with ResNet and also boosting VGG16. Conversely, PCA resulted in a decrease in metrics for traditional algorithms like Random Forest and SVM. This disparity underscores that dimensionality reduction techniques are not universally effective. The results clearly demonstrate that PCA can be particularly beneficial for deep learning models, highlighting the significant potential of AI-based systems for accelerating and improving the accuracy of early breast cancer diagnosis. The joint, systematic analysis of both model families under the influence of PCA constitutes a unique contribution to medical image processing research. Future work should focus on enhancing model generalization using larger, balanced datasets and exploring next-generation Transformer-based approaches.

References

[1] Karin Dembrower, Alessio Crippa, Eugenia Colón, Martin Eklund, Fredrik Strand, and ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit. Health*, 5(10):e703–e711, October 2023.

[2] Asmaa Ibrahim, Paul Gamble, Ronnachai Jaroensri, Mohammed M Abdelsamea, Craig H Mermel, Po-Hsuan Cameron Chen, and Emad A Rakha. Artificial intelligence in digital breast pathology: Techniques and applications. *Breast*, 49:267–273, February 2020.

[3] Md Zahangir Alom, Chris Yakopcic, Mst Shamima Nasrin, Tarek M Taha, and Vijayan K Asari. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *J. Digit. Imaging*, 32(4):605–617, August 2019.

[4] BreaKHis Dataset. BreaKHis – Breast Cancer Histopathological Image Classification. https://www.kaggle.com/datasets/ambarish/breakhis/data, 2025. Access date: August 8, 2025.

[5] Özlem Akar and Oğuz Güngör. Classification of multispectral images using random forest algorithm. *Journal of Geodesy and Geoinformation*, (106):139–146, 2012. [in Turkish].

[6] Özlem Akar and Esra Tunç Görmüş. Land use maps produced from göktürk-2 and hyperion eo-1 satellite images using random forest classifier and support vector machines. *Geomatics*, 4(1):68–81, 2019. [in Turkish].

[7] Ferdi Doğan and İbrahim Türkoğlu. Comparison of deep learning algorithms' success in leaf classification. *Sakarya University Journal of Computer and Information Sciences*, 1(1):10–21, 2018. [in Turkish].

[8] Numan Halit Güldemir and Ahmet Alkan. Classification of optical coherence tomography images using deep learning. *Firat University Journal of Engineering Sciences*, 33(2):607–615, 2021. [in Turkish].

[9] Murat Karakoyun and Mehmet Hacıbeyoğlu. Statistical comparison of machine learning classification algorithms with biomedical datasets. *Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering*, 16(48):30–42, 2014. [in Turkish].

[10] Gülay Gürgen and Soydan Serttaş. Performance comparison of machine learning algorithms for early diagnosis of heart failure. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 10(28):165–174, 2023. [in Turkish].

TIPTEKNO'25
MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Enhancing Fundus Image Quality for Improved Age-Related Macular Degeneration Detection using Deep Learning

Ceren Yilmaz, Sinem Yesil, Elif Ilkay Ozkan, Erdem Okur
Izmir University of Economics, Izmir, Turkiye

***Abstract*—Age-related Macular Degeneration (AMD) is a leading cause of vision loss in the elderly, where early detection plays a critical role in slowing disease progression. Deep learning approaches have shown strong potential for automated AMD diagnosis from retinal images; however, their performance can be hindered by image quality variations, illumination inconsistencies, and artifacts. In this study, we propose a custom contrast enhancement mask to improve lesion visibility in fundus images prior to classification. Using the publicly available ADAM challenge dataset, three deep learning architectures—YOLOv8n-cls, InceptionV3, and a modified U-Net encoder—were trained and evaluated on both the original and enhanced datasets. Experimental results demonstrate that the enhancement method substantially improves classification performance across all models, with YOLOv8n-cls achieving the highest accuracy with 91.70%, and specificity with 96.76%. These findings highlight the importance of preprocessing in medical image analysis and suggest that lightweight models, when combined with effective enhancement techniques, can achieve high accuracy suitable for clinical deployment.**

***Keywords*—AMD, Age-related Macular Degeneration, deep learning, image processing, neural networks**

I. Introduction

Age-related Macular Degeneration (AMD) is one of the leading causes of irreversible vision loss among elderly people globally, with a significant impact on quality of life. This neurodegenerative macular disorder of the retina, which is responsible for sharp and detailed vision, has a tendency to advance insidiously in the early phase. Early and accurate diagnosis of AMD is essential for early treatment, which can slow down the progression of the disease and preserve residual vision [1]. Traditional diagnostic methods have a significant reliance on ophthalmic expertise and manual examination of fundus images, which can be subjective, labor-intensive, and prone to inter-observer variability. The increasing prevalence of AMD due to a globally aging population further strains healthcare resources, echoing an urgent need for efficient, objective, and scalable diagnostic solutions.

In the last several years, the emergence of deep learning (DL) and computer vision has presented promising prospects for the automation of detection and categorization of various retinal diseases from fundus images [2]. Convolutional Neural Networks (CNNs), in particular, have demonstrated superior performance in feature extraction and pattern recognition, making them highly suitable for medical image analysis [3]. A number of studies have explored the application of CNNs to AMD detection, with different architectures and methods being employed to classify fundus images according to characteristic features [4][5][6].

Other than fundus images there are different datasets available with other type of images for instance, OCT images, as well as various approaches from several aspects. Serener and Serte aimed to classify dry and wet AMD types using OCT images [6]. The dataset they have used consisted of 83484 images categorized into four classes; healthy, diabetic macular edema (DME), dry AMD (drusen), and wet AMD (CNV). AlexNet and ResNet-18 architectures were compared, and ResNet demonstrated particularly high performance in dry AMD detection [7][8]. This study highlighted the effectiveness of transfer learning-based deep learning models, especially in the classification of dry AMD.

Ahnaf et al. conducted classification on 723 AMD and 1407 normal images from the Duke SD-OCT dataset using ResNet50 and VGG16 architectures [9][10][11][12]. The novelty of this study lied in its use of the Grad-CAM method to provide explainability in the classification process [13]. It was observed that the ONL-ISM (Outer Nuclear Layer to Inner Segment Myeloid) layer played a decisive role in both AMD and normal class decisions. This approach enhances the transparency and reliability of model decisions for clinical applications.

Lastly, in the study conducted by Singh et. al., the goal was to automatically classify ocular diseases such as diabetic retinopathy, glaucoma, AMD, and cataract [14]. Working on the ODIR dataset [15], which consists of images from 5000 patients, the researchers compared VGG19 and Inception V3 architectures. Interestingly, VGG19 outperformed InceptionV3 [16][17]. This success was attributed to VGG19's ability to capture fine details through its narrower receptive fields and its generalization capability across different demographics.

Despite the progress, however, there are still challenges to achieving robust and highly accurate automatic AMD detection from fundus images. Fundus images typically possess variable qualities, non-uniform illumination, and artifacts, which could potentially impair the performance of even state-of-the-art deep learning models. To address these limitations, our study proposes an enhanced framework for automated AMD detection with a particular focus on the critical role

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

played by image preprocessing. We hypothesize that by improving the quality and consistency of the fundus images prior to classification, we can boost the performance of off-the-shelf deep learning models.

In this work, we present a novel approach to the automatic AMD detection task, which integrates a preprocessing method for quality improvement of fundus images with their subsequent classification by well-known convolutional neural network models. We would like to demonstrate the impact of careful preprocessing on the accuracy and reliability of AMD detection, leading to a more stable and practical solution for clinical application. The remainder of this paper is organized as follows: Section II explains our new preprocessing approach and the deep learning architectures employed as well as the reasons why they are chosen. Section III presents the experimental setup and results. Finally, Section IV concludes the paper and addresses future work.

II. Methodology and Proposed Enhancement

The methodology to test whether the contrast mask proposed here affects the classification is simple. Three deep learning architectures —YOLOv8n-cls, InceptionV3, and a modified U-Net encoder classifier— were trained and evaluated for binary classification of retinal fundus images into Age-related Macular Degeneration (AMD) and Normal (NO) categories [18][17][19]. All models were trained on a uniformly preprocessed dataset and images resized to appropriate size for the input layer of models augmented with horizontal flips and random rotations to overcome class imbalance. They were also trained and evaluated with another dataset which is essentially the same but preprocessed with proposed enhancement mask which constitutes the novel step of this work. It is a custom contrast enhancement mask to improve lesion visibility for the network models.

The network models are chosen for their well-known structures and versions readily available to use due to popularity. Moreover, they were selected over other architectures due to certain unique properties. For instance, YOLOv8n-cls prioritizes computational efficiency and real-time capability. The InceptionV3 leverages a custom classification head for deep semantic learning. Lastly, U-Net retains its encoder for spatially rich feature extraction. In the end, they are trained, tested for both datasets and compared against each other.

Following subsections will explain the dataset used in the study and preprocessing step which includes the contrast enhancement mask generation in detail.

A. The Dataset

The Automatic Detection challenge on Age-related Macular Degeneration (ADAM), introduced at ISBI 2020, is the first international competition designed to benchmark algorithms for AMD detection and characterization [20]. Addressing the lack of large, standardized datasets and comparable evaluation protocols in medical image analysis, ADAM builds upon the successes of the REFUGE and AGE challenges. The dataset provided by the challenge now consists of over 1500 retinal images categorized as AMD or NO (normal). All images were obtained from CT Angiography (BTA), not traditional frontal fundus photography. This distinction is crucial because BTAderived images are often easier to acquire in clinical workflows and are less dependent on ophthalmology-specific imaging hardware. Their standardization across patient records also enhances model transferability. Specifically two devices -a Zeiss Visucam 500, producing at a resolution of $2124x2056$ pixels, and a Canon CR-2, yielding at $1444x1444$ pixels- are used to collect the images. All images were taken in strict adherence to the standard operating procedures for each camera.

B. Preprocessing

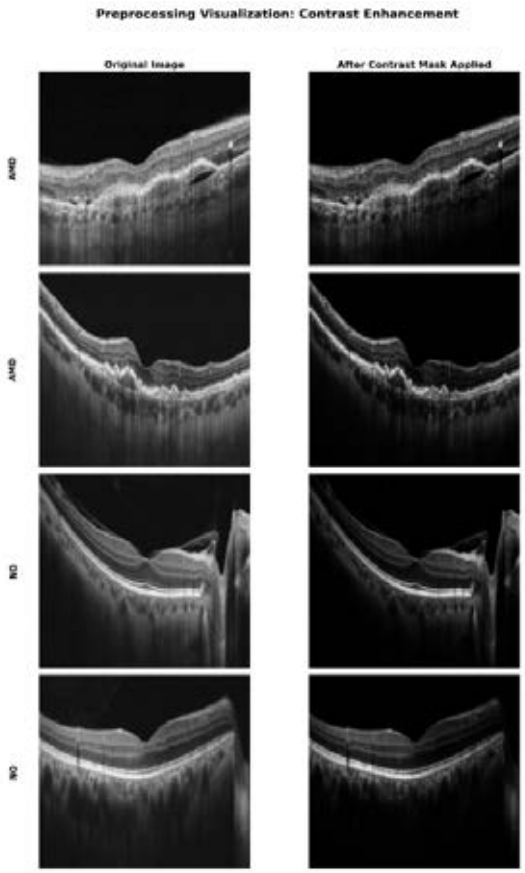


Fig. 1: After contrast enhancement preprocessing step the images as they are shown on right side. They look darker to the eye but the pixel value variation between tissue layers are higher now for detection.

In this study a similar image enhancement mask that is described by Okur and Turkan is developed but applied on regular BT images [21]. The mentioned study utilizes three different masks and their combinations to detect best result yielding enhancement mask. This study creates a mask similar to brightness mask of theirs and utilizes only this one due to the characteristics of the dataset images. The developed mask first transforms the raw images from the dataset to an RGB

image. Then calculates the mask as the absolute difference between its value and the mean of the color channels at the same spatial location. A feature map is generated for each color channel and linearly normalized to the [0, 1] range. When combined into an image, this process produces an RGB contrast-enhancement mask, where each pixel value functions as a weight. These weights derived from explained calculation then, applied on images by direct per-pixel weighting method. This mask ensures that the pixel value difference between each tissue that can be seen in the image is high and more separable. It is hypothesized that the applied mask will make chosen network models understand AMD related features better. Fig. 1 shows original images and images after the mask application.

The preprocessing step ends with resizing of dataset images to the appropriate resolution for the network model to be used. The resolutions are 320 x 320 for UNet, 299 x 299 for YOLO and Inception.

III. Experimental Setup and Results

All experiments were conducted on an MSI Alpha 15 A3DDK-034 laptop equipped with an AMD Ryzen 7 3750H processor and an AMD Radeon RX 5500M GPU. The system was configured with 16 GB DDR4 memory and the operating environment consisted of Python 3.10 with all experiments executed under consistent software and hardware conditions to ensure reproducibility.

TABLE I: Performance comparison of models on regular and enhanced datasets. The U-Net result on the regular dataset is omitted as it performed below chance level.

Model	Dataset	Accuracy	Recall (Sensitivity)	Specificity	NPV	Precision	F1-score
U-Net	Enhanced	86.60%	79.81%	91.62%	91.37%	80.03%	79.92%
YOLOv8	Enhanced	**91.70%**	84.88%	96.76%	92.99%	89.43%	86.88%
InceptionV3	Enhanced	89.36%	83.03%	94.05%	92.55%	84.57%	83.76%
YOLOv8	Regular	66.38%	78.28%	57.57%	99.53%	69.10%	64.28%
InceptionV3	Regular	78.72%	50.00%	100.00%	78.72%	39.36%	44.05%

Table I presents the classification performance of U-Net, YOLOv8, and InceptionV3 on both the enhanced and regular retinal image datasets. The results clearly show that the proposed enhancement method significantly improved performance across all models, suggesting that better lesion visibility and contrast contribute to more discriminative feature learning. Among the tested architectures, YOLOv8 trained on the enhanced dataset achieved the best overall metrics, with an accuracy of 91.70%, specificity of 96.76%, and F1-score of 86.88%. This combination of high specificity and strong F1-score indicates that YOLOv8 not only excels at correctly identifying non-AMD cases but also maintains strong detection capability for AMD, avoiding the common pitfall of biasing toward the majority class. InceptionV3 with the enhanced dataset followed closely, achieving 89.36% accuracy and a well-balanced sensitivity–specificity profile, reflecting its capacity to capture subtle pathological features despite its larger model size and longer training requirements.

In stark contrast, the models trained on the regular dataset displayed lower overall accuracy and substantially reduced F1-scores. YOLOv8's performance dropped to 66.38% accuracy, primarily due to a steep decline in specificity with 57.57%, suggesting difficulty in rejecting non-AMD cases without enhanced image features. InceptionV3 trained on the regular dataset exhibited perfect specificity but very low sensitivity, indicating severe bias toward predicting the majority non-AMD class—a clear sign of imbalance exploitation rather than true discriminative learning. The omission of U-Net's regular dataset result was warranted, as it performed below chance level, further reinforcing that raw, unenhanced fundus images provide insufficient discriminative signal for reliable AMD detection.

Overall, these results confirm that the proposed enhancement method improves both generalization and class balance, particularly benefiting lightweight, real-time models like YOLOv8, which demonstrated the most favorable trade-off between sensitivity and specificity in this work.

IV. Conclusion

This work introduced a contrast-based enhancement mask designed to improve lesion visibility in retinal fundus images for AMD detection. The method was evaluated using three distinct deep learning architectures on the ADAM challenge dataset in both regular and enhanced forms. Results consistently showed that the enhancement step improved classification metrics, with YOLOv8n-cls delivering the best overall performance in terms of accuracy, specificity, and F1-score. InceptionV3 and U-Net also benefited from the preprocessing, demonstrating improved generalization and reduced class imbalance effects. These outcomes confirm that targeted image preprocessing can significantly boost the performance of deep learning models in medical imaging tasks. This work serves as a preliminary step, confirming that the enhancement method can successfully boost the performance of well-known architectures. Resonating on that, future work will explore integrating this enhancement method into end-to-end trainable pipelines and validating the approach on additional datasets to further assess its robustness and clinical applicability. If successful, it is considered to move on the multi-disease retinal datasets to eventually boost lightweight systems to be used as diagnostic aid tools.

References

[1] "Age-related macular degeneration (amd)," https://www.hopkinsmedicine.org/health/conditions-and-diseases/agerelated-macular-degeneration-amd, Accessed: 2025-07-22.

[2] Y. Jeong, Y-J. Hong, and J-H. Han, "Review of machine learning applications using retinal fundus images," *Diagnostics*, vol. 12, no. 1, 2022.

[3] P. Celard, E.L. Iglesias, J.M. Sorribes-Fdez, R. Romero, A.S. Vieira, and L. Borrajo, "A survey on deep learning applied to medical images: from simple artificial neural networks to generative models," *Neural. Comput. & Applic.*, vol. 35, pp. 2291–2323, 2023.

[4] A. Bhuiyan, T.Y. Wong, D.S.W. Ting, A. Govindaiah, E.H. Souied, and R.T. Smith, "Artificial intelligence to stratify severity of age-related macular degeneration (amd) and predict risk of progression to late amd," *Trans. Vis. Sci. & Tech.*, vol. 9, no. 2, pp. 25–25, 2020.

[5] M.A. Zapata, D. Royo-Fibla, O. Font, J.I. Vela, I. Marcantonio, E.U. Moya-Sánchez, A. Sánchez-Pérez, D. Garcia-Gasulla, U. Cortés, E. Ayguadé, and J.Labarta, "Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma.," *Clin. Ophthalmol.*, pp. 419–429, 2020.

[6] A. Serener and S. Serte, "Dry and wet age-related macular degeneration classification using oct images and deep learning," in *2019 EBBT*, 2019, pp. 1–4.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*. 2012, vol. 25, Curran Associates, Inc.

[8] Goldy Verma, "Efficient retinal disease detection with resnet-18: A deep learning approach," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, 2024, pp. 812–816.

[9] S. M. Azoad Ahnaf, G. M. Atiqur Rahaman, and Sajib Saha, "Understanding cnn's decision making on oct-based amd detection," in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, 2021, pp. 1–4.

[10] Stephanie J. Chiu, Joseph A. Izatt, Rachelle V. O'Connell, Katrina P. Winter, Cynthia A. Toth, and Sina Farsiu, "Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images," *Investigative Ophthalmology I& Visual Science*, vol. 53, no. 1, pp. 53–61, 01 2012.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[12] Shuying Liu and Weihong Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730–734.

[13] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," *CoRR*, vol. abs/1610.02391, 2016.

[14] Preet Singh, Taniya Hasija, and K. R. Ramkumar, "Optimizing ocular disease detection: A deep learning approach with vgg19 and inceptionv3," in *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, 2024, pp. 1–6.

[15] Yin jianing, "Odir - dataset," 2022.

[16] Ramyea Ravindaran, Kasthuri N, Preethi S, Adithya B, Gunaranjan Sp, Dharanidharan K, and Aravinth S, "Performance analysis of a vgg based deep learning model for classification of weeds and crops," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–5.

[17] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[18] Athraa Kh. Al-Zihairy and A.E. Abdelkareem, "Optimizing yolov8-cls: A step towards smarter edge environments," in *2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI)*, 2024, pp. 1–6.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[20] Huihui Fang, Fei Li, Huazhu Fu, Xu Sun, Xingxing Cao, Fengbin Lin, Jaemin Son, Sunho Kim, Gwenole Quellec, Sarah Matta, Sharath M. Shankaranarayana, Yi-Ting Chen, Chuen-Heng Wang, Nisarg A. Shah, Chia-Yen Lee, Chih-Chung Hsu, Hai Xie, Baiying Lei, Ujjwal Baid, Shubham Innani, Kang Dang, Wenxiu Shi, Ravi Kamble, Nitin Singhal, Ching-Wei Wang, Shih-Chang Lo, José Ignacio Orlando, Hrvoje Bogunović, Xiulan Zhang, and Yanwu Xu, "Adam challenge: Detecting age-related macular degeneration from fundus images," *IEEE Trans. Med. Imag.*, vol. 41, no. 10, pp. 2828–2847, 2022.

[21] Erdem Okur and Mehmet Turkan, "Patch enhancement for melanoma detection with bag of visual words," in *2022 Medical Technologies Congress (TIPTEKNO)*, 2022, pp. 1–4.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Analysis of Dementia EEG Signals Using Empirical Mode Decomposition Variants and Deep Learning

Yahya Oguzhan Senol, Aydin Akan
Dept. of Electrical and Electronics Engineering
Izmir University of Economics
Izmir, Turkey
oguzhansenol1334@gmail.com, akan.aydin@ieu.edu.tr

Ozlem Karabiber Cura
Dept. of Biomedical Engineering
Izmir Katip Celebi University
Izmir, Turkey
ozlem.karabiber@ikcu.edu.tr

***Abstract*— Dementia is among the most frequent neurological diseases that result in worsening cognitive functions and damaging consequences on quality of life. In this study, a novel method is proposed by using electroencephalogram (EEG) signals and signal decomposition methods to detect Alzheimer's disease (AD). Intrinsic Mode Functions (IMFs) are obtained by utilizing Empirical Mode Decomposition (EMD) and Ensemble Empirical Mode Decomposition (EEMD) methods. Five time-domain and five spectral-domain features are extracted from one-minute and five-second EEG segments, as well as from the corresponding IMF segments, using EEG recording. Topographic heat maps are generated for each feature. These feature maps give temporal and spatial information simultaneously on a single image. Feature map images are classified with two-dimensional convolutional neural network (2D-CNN). Three CNN architectures are used for classification and comparison between networks. The proposed method achieves promising results, with accuracy up to 96%.**

Keywords—Alzheimer's Dementia (AD); topographic heat map; 2D-CNN; EMD; EEMD

I. INTRODUCTION

Dementia refers to the loss of brain functions. AD is the most common type of dementia. AD accounts for 65% to 80% of all dementia cases and is known as one of the leading causes of death in adults over the age of 65 [1]. AD is an advancing neurological illness and widespread public health issue. Its effects include cognitive decline, memory loss, and behavioral abnormalities [2]. A patient's response to treatment and the diagnosis of AD can be made using a variety of medical methods. Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) are performed to detect symptoms associated with AD [3]. Also, EEG is another technique that uses electrodes applied to the scalp to record electrical impulses from the brain [4]. While techniques like MRI and PET offer high spatial resolution, they may miss certain abnormalities due to their limited temporal resolution. In contrast, EEG is superior in terms of temporal resolution, cost-effectiveness, and availability [5].

There are many studies to analyze EEG signals. In some of them, features from EEG signals like tsallis entropy [6], multivariate multiscale weighted permutation entropy [7], permutation Jaccard distance [8], and Fuzzy Entropy (FuzzyEn) [9] are used to distinguish patient and control group. Each of these studies emphasized how the brain complexity of AD patients was lower compared to the control group. Apart from calculating many features from EEG signals, signal decomposition approaches with the help of machine learning are also used for AD classification. Some of these approaches are EMD, and EEMD. EMD, and EEMD are used especially to analyze nonlinear and non-stationary signals like EEG. EMD decomposes signal into intrinsic mode functions (IMFs). Each IMF represents the part of the signal that contains a certain oscillation. When the IMFs are added together, the original signal is reconstructed. Sometimes, similar oscillations can be found in different IMFs, and different oscillations can be found in the same IMF. To prevent this issue, the EEMD method is utilized. In a study, segmented EEG signals were decomposed into the first seven IMFs with EMD and EEMD methods. Then, spectral and time-domain features were extracted from these IMFs and used to classify AD patients and healthy control subjects using machine learning methods such as Decision Tree (DT), k-Nearest Neighbors (kNN), Support Vector Machine (SVM), and Random Forest (RF) [10]. In that study, 91.8% classification accuracy was obtained using EMD, and 94.1% classification accuracy was obtained using EEMD. Another study benefited from discrete wavelet transform (DWT) and EMD for decomposing the signal, and SVM, k-NN, and Linear Discriminant Analysis (LDA) for classification [11]. Their study achieved 97.64% accuracy using k fold cross-validation and 81.08% accuracy using leave-one-subject-out (LOSO) validation. Even though these manual feature extraction and machine learning-based techniques have produced good classification results, they are still computationally demanding and challenging for doctors to understand. These issues can be resolved by the suggested mixed deep learning (DL) approaches, which are used with EEG time-series data and frequency domain representations, also known as time-frequency representations (TFRs) [12].

In this study, the classification of patients and healthy control subjects is performed using EEG signals. The first three

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

IMFs are extracted from the EEG signals using EMD and EEMD. Spectral and time-domain features are then derived from both the IMFs which are extracted with EMD and the original EEG signals. Only spectral features are derived from the IMFs which are obtained with EEMD. Topographic heat maps in RGB format are generated based on these features. Later, these RGB images are classified using 2D CNN architectures, including ResNet50, EfficientNet-b0, and ResNet-18. The classification process is implemented in the MATLAB environment.

II. Materials and Methods

A. Dataset

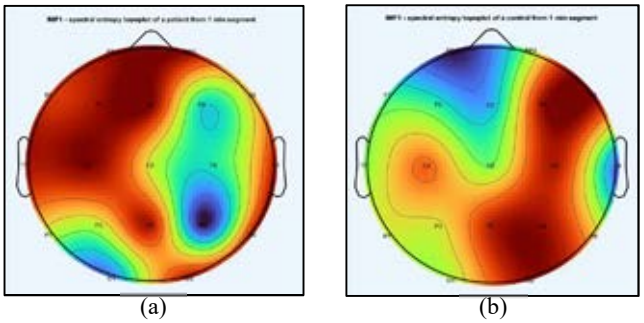
EEG data used in this study were obtained from patients evaluated at the dementia clinic of the Neurology Department, Izmir Katip Celebi University Faculty of Medicine. These patients were diagnosed with early-stage Alzheimer's disease by diagnostic tests and neuroimaging. EEG signals were recorded for 15 AD patients (30 minutes for each patient). Their average age is 64. 7 of them are males, and 8 of them are females. EEG recording was done for each patient from 19 channels at 200 Hz sampling frequency in the resting state with the eyes closed. EEG signals were recorded only, and different biological signals like ECG and EOG were not collected during the recording process. Recordings were obtained from 19 scalp electrodes placed according to the International 10–20 system at the following positions: Fp1, F7, T3, T5, O1, O2, T6, T4, F8, Fp2, F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4. For comparison, EEG data were also collected from 11 cognitively healthy, age-matched control subjects (CS) using the same recording system and electrode placement. 6 of them are males, and 5 of them are females. Their average age is 57. To collect these EEG signals for use in this study, ethical approval numbered **83** and dated **22.10.2020** is obtained from the Clinical Research Ethics Committee of Izmir Katip Celebi University.

B. Preprocessing and Feature Extraction

Firstly, at the preprocessing stage, a Butterworth band pass filter was used to all channels of patients and controls. Pass band of 0.5 - 40 Hz is applied to all channels to eliminate unwanted electrical noise, and some artifacts such as eye-blinking and muscle activity of individuals. Filtered EEG signals were divided into nonoverlapping 5 seconds and 1 minute segment durations separately. For each segment of 1 min, EMD and EEMD are applied. For each segment of 5s, only EMD is applied. 5 spectral-domain and 5 time-domain features are extracted from segments of EEG and IMFs, which are first three IMFs obtained using EMD and EEMD, for AD patients and CS. Spectral features are total power, higher order spectral (HOS) moments of order 1, 2, and 3, and spectral entropy. Time domain features are activity, mobility, complexity, skewness, and kurtosis.

C. Topographic Heat Map Generation

After the feature extraction process, topographic heat maps were generated for time and spectral features of 1 min and 5 s segments of EEG signals and IMFs. Matlab's eeglab toolbox was used for it. Topoplot function, defined in EEGLAB was used to plot topo maps. In order to get visually distinctive images, appropriate transformations are made on time-domain and spectral features. Log 10 transformation is applied for total power, activity, complexity, kurtosis, HOS moments of order 1, 2, and 3. Min-max normalization is applied for skewness. No transformation is applied for mobility and spectral entropy. Transformations are done per-image. Topographic heat maps give information about the spatial resolution of brain activity across EEG channels for a given feature. They allow for the visualization of which features are more dominant in specific channels for the corresponding signal (such as EEG or IMFs). Examples of topographic heat maps of spectral entropy of the first IMF obtained with EMD, extracted from 1 min segments of a control subject and an AD patient are given in Fig. 1.



(a) (b)

Figure 1: Examples topographic heat maps of spectral entropy for control and patient groups. a) AD patient b) CS

It can be observed that left temporal region is more active for a patient whereas right temporal region is more active for a control subject.

D. CNN Architectures for Classification

Topographic heat maps for spectral and time features from IMFs and EEG signals were used to classify patient and control with deep learning. 2D CNN was used for classification of images. ResNet-50, ResNet-18, and EfficientNet-b0 are the CNN architectures used in this study. These architectures are designed for classifying 1000 classes. In our study, suitable changes are made in layers, because there are two classes which are patient and control. Also, images are resized to 224 x 224 x 3, because these architectures expect 224 x 224 x 3 as an input image.

III. Results and Discussion

In this study, there are a total of 9 cases for 1-minute segment duration of spectral features topographic heat maps. Each CNN architecture has 3 cases. On the other hand, there are a total of 6 cases for 1-minute segment duration of time features topographic heat maps, and each CNN architecture has 2 cases. In addition, for the 5-second segment duration, there are a total of 6 cases for the time-domain features and another six cases for the spectral-domain features' topographic heat maps. For each case, data was split into 80% for training and 20% for validation. Data split was done using all patients and control subjects for each case. Training options are adjusted for all three CNN architectures. Maximum epoch is set to 10, and learning rate is set to 0.0001. Performance metrics which are validation accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), and F1 score are examined. Classification results for the 1-minute segment duration and the 5-second segment duration are given in Table 1 and Table 2, respectively.

TABLE I: Performance evaluation results of CS and AD EEG classification obtained using spectral and temporal features topographic heat maps for 1min segment duration.

Spectral Features						
CNN Architectures	**Method**	**Validation ACC**	**SEN**	**SPE**	**PRE**	**F1-Score**
ResNet-50	EEG	0.9474	0.9606	0.9292	0.9491	0.9548
	EMD	0.9271	0.9535	0.8910	0.9228	0.9379
	EEMD	**0.9560**	0.9569	0.9551	0.9550	0.9559
ResNet-18	EEG	0.9234	0.8967	0.9595	0.9676	0.9308
	EMD	0.9166	0.9303	0.8979	0.9255	0.9279
	EEMD	0.9228	0.9138	0.9316	0.9301	0.9219
EfficientNet-b0	EEG	**0.9598**	0.9800	0.9324	0.9515	0.9655
	EMD	0.9154	0.9210	0.9077	0.9315	0.9263
	EEMD	0.9366	0.9497	0.9235	0.9251	0.9372
Temporal Features						
CNN Architectures	**Method**	**Validation ACC**	**SEN**	**SPE**	**PRE**	**F1-Score**
ResNet-50	EEG	0.8336	0.8967	0.7754	0.7866	0.8380
	EMD	0.8	0.8185	0.7816	0.7885	0.8032
ResNet-18	EEG	0.8352	0.7900	0.8769	0.8556	0.8215
	EMD	0.7215	0.6182	0.8245	0.7784	0.6891
EfficientNet-b0	EEG	0.8240	0.8400	0.8092	0.8029	0.8208
	EMD	0.7494	0.7528	0.7459	0.7467	0.7497

TABLE II: Performance evaluation results of CS and AD EEG classification obtained using spectral and temporal features topographic heat maps for 5s segment duration.

Spectral Features						
CNN Architectures	**Method**	**Validation ACC**	**SEN**	**SPE**	**PRE**	**F1-Score**
ResNet-50	EEG	**0.9150**	0.9344	0.8960	0.8984	0.9160
	EMD	0.8402	0.8524	0.8273	0.8396	0.8459
ResNet-18	EEG	0.8687	0.9046	0.8333	0.8474	0.8724
	EMD	0.8002	0.8184	0.7808	0.7987	0.8085
EfficientNet-b0	EEG	0.8891	0.8790	0.8990	0.8955	0.8872
	EMD	0.7870	0.8432	0.7273	0.7666	0.8031
Temporal Features						
CNN Architectures	**Method**	**Validation ACC**	**SEN**	**SPE**	**PRE**	**F1-Score**
ResNet-50	EEG	0.7817	0.7764	0.7869	0.7820	0.7792
	EMD	0.6775	0.6943	0.6590	0.6923	0.6933
ResNet-18	EEG	0.7389	0.7713	0.7071	0.7217	0.7457
	EMD	0.5990	0.6781	0.5116	0.6054	0.6397
EfficientNet-b0	EEG	0.7013	0.5836	0.8172	0.7587	0.6597
	EMD	0.5960	0.4038	0.8084	0.6997	0.5121

From these tables, it can be observed that classification results of 1 min cases is better than 5 s cases. Also, spectral features topographic heat maps give better results than time features topographic heat maps. In addition, spectral features topographic heat maps for 1 min segment duration is best when we compare with other cases. Because of that, EEMD method is benefited for only spectral features topographic heat maps for 1 min segment duration to save computation time. The best accuracy value 95.98% is achieved with EEG from Efficient-b0 architecture. Although EEG gives higher performance from EMD and EEMD, we can observe that accuracy of EEMD is very close to EEG even EEMD accuracy (% 95.60) is higher than EEG for ResNet-50 network. So, performance of EMD is increased with EEMD which is a variant of EMD method. If we compare CNN architectures, Efficient-b0 and ResNet-50 architectures give better results than ResNet-18. This can be expected, as ResNet-18 is a more basic architecture compared to the others.

IV. CONCLUSION

Detection of AD is important in terms of starting treatment as soon as possible and raising an individual's awareness. In this study, detection of AD was done with signal decomposition methods which are EMD and its variant EEMD, and deep learning. It is aimed to help doctors' judgment as fast as possible. In this study, 2D CNN was used to classify control subjects and patients. Topographic heat maps were used as input images to CNN. EMD and EEMD were used to decompose EEG signals into IMFs. Spectral and temporal features were extracted from IMFs (obtained with EMD) and

original EEG signal for 1 min and 5 s cases. Only spectral features were extracted from IMFs (obtained with EEMD) for 1 min cases. These features were transformed into topographic heat maps. Topographic heat maps give information about the spatial resolution of brain activity across EEG channels for a given feature. They allow for the visualization of which features are more dominant in specific channels for the corresponding signal (such as EEG or IMFs). This spatial perspective adds interpretability to machine learning models by linking features to specific brain regions. These maps help show which brain areas work differently in different groups. Furthermore, researchers can learn more about how brain disorders work by looking at differences in heat maps for different features or conditions. In neurological conditions like AD, these visual comparisons are extremely helpful.

After generating topographic heat maps, classification of these topo maps was done with ResNet-50, ResNet-18 and EfficientNet-b0 architectures. Spectral features topo maps were classified better than time-domain features topo maps for 1 min and 5 s cases. In addition, EEG spectral topo maps were classified better than IMFs spectral topo maps. Classification performance of 1 min IMFs spectral topo maps was increased using EEMD. For all architectures, classification performance of EEMD is higher than EMD for 1 min IMFs spectral. For ResNet-50 architecture, performance of EEMD is also higher than EEG for 1 min spectral case. These results highlight the potential of EEMD as a robust decomposition method for increasing classification accuracy. 96% accuracy was achieved with 1 min EEG spectral topo maps as the highest accuracy rate in this study. This high accuracy demonstrates the reliability of spectral features and the effectiveness of the 2D CNN approach in EEG based AD detection.

In this study, because of large data, computation time is a constraint. It takes hours to do classification process with 2D CNN architectures. Later, to decrease computation time, another computer with powerful GPU and CPU can be used. Implementing parallel computing or using cloud-based platforms could also be considered to overcome these computational challenges.

For future work, feature selection can be done to improve classification performance of IMFs. Other signal decomposition methods like ITD, Multivariate Empirical Mode Decomposition (MEMD) can be obtained to compare with EMD and EEMD. Also, different features like Logarithmic band power (LBP), Approximate entropy (ApEn), Zero-crossing rate (ZCR), Norm (Norm) can be extracted from signal. These features can be compared with the features of the proposed study. In addition, the first three IMFs were used in this study. Different combinations of IMFs can be used to find out which IMFs are more successful in classification. As a result, IMF selection can be made to increase classification performance. After all comparisons between signal decomposition methods, extracted features, and IMFs selection processes the most effective approach can be learned to classify AD patients.

ACKNOWLEDGMENT

The authors would like to thank the physicians of Izmir Katip Celebi University Faculty of Medicine for their valuable contributions during the data acquisition process.

REFERENCES

[1] J. delEtoile and H. Adeli, "Graph theory and brain connectivity in Alzheimer's disease," The Neuroscientist, vol. 23, no. 6, pp. 616–626, April 2017.

[2] W.-Y. Yu, T.-H. Sun, K.-C. Hsu, C.-C. Wang, S.-Y. Chien, C.-H. Tsai, and Y.-W. Yang, "Comparative analysis of machine learning algorithms for Alzheimer's disease classification using EEG signals and genetic information," Computers in Biology and Medicine, vol. 176, p. 108621, May 2024.

[3] C. R. Jack, D. A. Bennett Jr., K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan, and N. Silverberg, "NIA-AA research framework: Toward a biological definition of alzheimer's disease," Alzheimer's & Dementia : the Journal of the Alzheimer's Association, vol. 14, no. 4, pp. 535–562, 2018.

[4] E. Perez-Valero, M. Á. Lopez-Gordo, C. M. Gutiérrez, I. Carrera-Muñoz, and R. M. Vílchez-Carrillo, "A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearable EEG," Computer Methods and Programs in Biomedicine, vol. 220, p. 106841, June 2022.

[5] U. Lal, Arjun Vinayak Chikkankod, and L. Longo, "A Comparative Study on Feature Extraction Techniques for the Discrimination of Frontotemporal Dementia and Alzheimer's Disease with Electroencephalography in Resting-State Adults," Brain sciences, vol. 14, no. 4, pp. 335–335, March 2024.

[6] K. D. Tzimourta, T. Afrantou, P. Ioannidis, M. Karatzikou, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, P. Angelidis, E. Glavas, N. Grigoriadis, D. G. Tsalikakis, and M. G. Tsipouras, "EEG Window Length Evaluation for the Detection of Alzheimer's Disease over Different Brain Regions," Brain Sciences, vol. 9, no. 4, p. 81, April 2019.

[7] B. Deng, L. Cai, S. Li, R. Wang, H. Yu, Y. Chen, and J. Wang, "Multivariate multi-scale weighted permutation entropy analysis of EEG complexity for Alzheimer's disease," Cognitive neurodynamics, vol. 11, no. 3, pp. 217–231, June 2017.

[8] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, and F. C. Morabito, "Permutation Jaccard Distance-Based Hierarchical Clustering to Estimate EEG Network Density Modifications in MCI Subjects," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 10, pp. 5122–5135, October 2018.

[9] S. Simons, P. Espino, and D. Abásolo, "Fuzzy Entropy Analysis of the Electroencephalogram in Patients with Alzheimer's Disease: Is the Method Superior to Sample Entropy?," Entropy, vol. 20, no. 1, p. 21, January 2018.

[10] O. K. Cura, A. Akan, G. C. Yilmaz, and H. Ture, "Detection of Alzheimer's Dementia by Using Signal Decomposition and Machine Learning Methods," International Journal of Neural Systems, vol. 32, no. 09, August 2022.

[11] M. S. Safi and S. M. M. Safi, "Early detection of Alzheimer's disease from EEG signals using Hjorth parameters," Biomedical Signal Processing and Control, vol. 65, p. 102338, March 2021.

[12] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, and P. Bramanti, "Deep convo lutional neural networks for classification of mild cognitive impaired and alzheimer's disease patients from scalp EEG recordings," in 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016, pp. 1–6.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Design of a Digital Content Framework with Gamification and Progress-Tracking for Mobile Assisted Diabetes Self-Education

Göksel Okandan
Dept. of Computer Eng.
Kocaeli University
İzmit, Turkey
210201058@kocaeli.edu.tr

Alpaslan Burak İnner
Dept. of Computer Eng.
Kocaeli University
İzmit, Turkey
binner@kocaeli.edu.tr

Adnan Kavak
Dept. of Computer Eng.
Kocaeli University
İzmit, Turkey
akavak@kocaeli.edu.tr

Sajjad A. Khan
Dept. of Software Eng.
Kocaeli University
İzmit, Turkey
sajjadahmad.khan@kocaeli.edu.tr

Ayaz Aktaş
Dept. of Computer Eng.
Kocaeli University
İzmit, Turkey
220202050@kocaeli.edu.tr

Alparslan Bozkurt
Dept. of Computer Eng.
Kocaeli University
İzmit, Turkey
220202081@kocaeli.edu.tr

Yeliz Demirhan
Diabetes Clinic
Kocaeli University Hospital
İzmit, Turkey
yelizerdem@gmail.com

Özgür Çakır
Dept. of Internal Medicine
Kocaeli University
İzmit, Turkey
ozgur.cakir@kocaeli.edu.tr

***Abstract*—Research indicates that personalized diabetes education increases engagement and long-term self-management among patients. This study presents a structured and personalized framework for mobile-assisted self-diabetes education. The framework was developed as a component of the larger AIDCare system, which is an AI-supported diabetes management platform. In this framework, educational content is developed using H5P and hosted on a WordPress platform. Gamification features, including achievements, badges, leaderboards, and user analytics are integrated to support engagement. Real-time progress tracking is used to continuously monitor and record the progress of each user. The framework enables healthcare personnel to add, edit, or update educational content as needed, while also facilitating ongoing customization and alignment with clinical workflows. In addition, this study discusses implementation considerations and describes a pathway for integration with routine care and the evaluation of patient outcomes.**

***Keywords*—Content Creation; Diabetes Education; Gamification; Learning Path; Real-time Progress Tracking**

I. Introduction

According to the World Health Organization (WHO), diabetes is a chronic disease that occurs when the pancreas does not produce enough insulin, a hormone that regulates blood glucose, or when the body cannot effectively use the insulin it produces [1]. It is a chronic and permanent disease that adversely affects the quality of life of patients if mismanaged. This can lead to a shortened lifespan. Over time however, with advances in technology, several means to support diabetes self-management have been developed, chiefly mHealth applications and sensors. AIDCare, an artificial intelligence-supported diabetes management system [2], provides a diabetes management platform that is easy to use and combines advanced healthcare technologies to meet the needs of people with diabetes.

Research on diabetes self-management and support has identified several factors that influence its complexity, ranging from socioeconomic, biological, and genetic determinants [3]. This diversity of factors calls for a personalized, structured education system that applies to everyone while having room to be personalized for each person (user) [4]. Additionally, numerous studies have also demonstrated the benefits of personalized diabetes self-management and education (DSMES) programs, including improved blood glucose control, lower risk of cardiovascular disorders, and lower risk of pregnancy failure [5] [6] [7]. Several studies have also found that the content of diabetes DSMES should be engaging for users to improve their memory retention and attendance at self-management education. Especially, the Gamification approach has been found to be effective by several studies [8] [9] [10].

To offer the benefits of structured, personalized and gamified diabetes education, the content of the AIDCARE Education Module has been developed and its progress has been tracked to meet the needs of diabetes patients today and onward. This study contributes a structured and scalable approach for personalized DSMES content integrated with gamification and a lightweight progress-tracking protocol, designed for seamless integration into clinical workflows via the AIDCare platform. This study is organized as follows: Section II describes the methodology for content generation, validation, and design of the gamification and analytics framework; Section III covers progress tracking and server integration; Section IV discusses the results, usability considerations, and clinical implications;

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

and Section V concludes with a summary and directions for future work.

II. Methodology

The methodology underpinning the AIDCare Education Module blends content engineering, expert validation, and a lightweight progress-tracking framework to deliver structured, personalized diabetes self-management education within routine care. This section outlines the end-to-end process for generating learning content, validating it with clinicians, and implementing the gamification and analytics components, with an emphasis on cross-platform delivery, scalability, and alignment with clinical workflows. This section also explains the content-generation workflow, validation procedures, the design of the gamification and analytics framework, and the progress-tracking system, followed by integration considerations and evaluation plans.

A. Generation of Educational Content for Diabetic Patients



Fig. 1. Unified graph of learning paths created by healthcare professionals.

1) Creating a Learning Path for Education Structure: To systematically create the educational content used in the module, a learning path based on the education received by a newly diagnosed diabetes patient was established with assistance from health professionals specializing in diabetes. The learning path consists of interrelated and non-sequential education modules, structured according to the user's health data and previously specified characteristics. The learning paths were designed to address the most common types of diabetes: type 1, type 2, and Gestational Diabetes. After the health professionals created the proper learning paths for each type, these paths were unified under one graph, as shown in Fig. 1, and the graph was technically implemented using the JSON format to offer the created content in a structured way.

2) Tools and Sources of Content Creation: Educational content was developed using H5P, a framework that does not require coding knowledge to add or modify content to facilitate future inclusion or alteration of content by health professionals. The WordPress platform was chosen to host the educational content of the module because of its ease of management and the ability to be used within the application by implementing WebView, without requiring the user to use an external application. Furthermore, the content used during the creation of education was based on the several sources listed below for each type of content:

- The text content used in the education was developed based on sources approved by guiding healthcare professionals, mainly including the guidelines written by the Turkish Diabetes Foundation [11] and the Turkey Endocrinology and Metabolism Association [12], along with the website of the Turkish Diabetes Society [13].
- The video content was taken from the YouTube channel of a health specialist working at Kocaeli University Hospital's Endocrinology Clinic [14], if available; otherwise, it was left blank at this stage of the project. In the next phase of the project, videos from approved and contracted parties were added to replace the blank videos.
- The questions in the education content were developed using various online sources [15] [16] [**?**] and Google's GEMINI artificial intelligence. A Word file containing these questions was sent to health professionals for feedback. Afterwards, the questions currently in the educational content were modified in line with their feedback.
- The educational games, along with the "Food Care Card Game" or "Insulin Injection Card Game" that are used by diabetes specialist nurses during their education for individuals with diabetes, have been created for the education module, alongside games like "Hypoglycemia Puzzle" for gamification purposes.

3) Validation of the Created Content: The created content was presented to the guiding healthcare professionals, and their feedback was considered. Their main issues were that the content created by using the "Course Presentation" type was "too boring" and that the alternative videos used during the development period could trigger copyright issues. Subsequently, the content was developed using alternative types to diversify content, and videos taken from external sources were changed with blank videos to be replaced by approved videos later in the development cycle.

B. Gamification and Profile Analytics

Educational research on the impact of gamification on education, specifically in diabetes self-management education, is mentioned in the "Introduction". Guided by these studies, this education framework incorporates a three-component gamification framework for memory retention of self-management information and retaining long-term user engagement:

- **Achievement / Badge System:** The achievement system uses an event-driven architecture based on Django signals, to separate gamification and core logic. The server-side `achievement` table defines six badges (for now) that reward learning progress: *First Step*, *You're Not Alone*, *Draw Your Path*, *Consistent Learner*, *Fast Learner* and *education chain*. The business logic of this system is as follows: Django signals monitor the relevant models and when a condition is met, the corresponding `UserAchievement` record is updated (`earned = True`), as shown in Fig. 2. The operation is atomic, thus it avoids race conditions. Several signal to badge can be found below.
  - `Consistent Learner`: opening at least one module every day for 30 consecutive days (`award_consistent_student_badge`).
  - `Fast Learner`: completing $\geq$ 10 lessons at 100% within the last seven days (`award_fast_student`).
  - `Education Champion` was awarded when all available lessons were completed (`award_champion_badge_for()`).
- **Leaderboard:** A PostgreSQL view named `v_leaderboard` ranks the top **six** users by completed lessons and total number of badges. The data were fetched via REST–Retrofit and rendered as `RecyclerView` cards on the home screen (Fig. 3).
- **Progress Analytics:** Records in `UsersProcess` and `UserModuleAccess` are summarized by a celery task triggered on every `completion rate` update. Three key indicators are stored in the `user_stats` view, which are *completed lessons* (`completion_rate = 100`), *total active time*, and *streak of consecutive days*.

This gamification framework boosts self-management motivation while the analytics layer delivers personalized feedback. As a result, users stay engaged longer and reach clinical goals more quickly.



Fig. 2. Gamification layer: earned badges and the top-six leaderboard rendered as `RecyclerView` cards.



Fig. 3. Progress-analytics dashboard showing *completed courses*, *total active time*, and *current streak*.

C. Progress Tracking and Server Integration

The *resume* and real-time completion-update features of H5P-based learning objects are implemented through a lightweight REST protocol with a minimal JSON load that connects the client and server. The approach delivers a consistent user experience to both web and Android clients, by allowing progress updates in real-time even without a stable internet connection, without requiring any additional plugins.

Client Layer – HTML Footer Script: A single JavaScript snippet, injected with the Code Snippets plugin into the footer of every H5P asset, fulfills two core duties: "*Resume*" and "*Progress Update*". During the education, the Javascript snippet The snippet listens to the H5P's native `xAPI` event. Every new "answered/completed/consumed" statement is then forwarded to the REST service in a compact JSON body containing only {`user`, `education`, `completion_rate`}. Therefore, network traffic is kept minimal and remains consistent across online and offline scenarios. When a user leaves and comes back to resume the the education, the most recent completion percentage for the user–content pair is fetched from the server as soon as the education page loads. When the H5P object emits the "loaded" signal, the timestamp is calculated by multiplying the stored percentage by the total duration T, allowing the learner to continue exactly where they are left off.

Android Layer – WebView Bridge: On a mobile device, the same HTML content is rendered in a local `WebView`. Three refinements are applied: (i) the JavaScript console output is redirected to Logcat for field debugging; (ii) an online `updateProgress(%)` bridge allows JavaScript to post percentages, and (iii) values are also cached in a Room database, so that progress is never lost even when connectivity is interrupted.

Server Layer – `UsersProcess` Endpoint: A single Django REST endpoint stores completion data: `GET` returns to the previous state; `POST` registers a newly opened lesson at 0 %; `PATCH` updates only the `completion_rate` field.

Because the schema is intentionally lean, clients do not need additional transformations.

The completion rate is dynamically calculated based on the number of interactive blocks (n) within each education, as shown in Table I. For example, in an education that has five interactive blocks (games, videos, text, etc.), each block would contribute 20% to the total progression rate. A user who has completed three of these blocks would have a completion rate of 60%. Static headings and images are only used for information purposes and are excluded from the calculation.

TABLE I
PER-BLOCK DYNAMIC COMPLETION CONTRIBUTION

H5P Item	Weight
Any interactive block	$100/n$
Static text/image	0

III. DISCUSSION

The prototype of this digital content framework, which includes the content creation system, gamification elements, and progress tracking, has been well-received by healthcare professionals at Kocaeli University Hospital's Endocrinology Clinic for its adaptability and ease of use. However, further testing is required to validate the reliability with different types of patients. Additionally, since there haven't been many tests of this framework or its gamification elements, there can only be assumptions regarding the effect of the gamification elements based on previous studies, such as improving user engagement or information retention. Overall, while the system has been well-received and deemed to have the potential to provide a solid foundation of content for digital DSMES programs, it needs to be validated by real-world testing regarding its impact on health metrics, such as HbA1C and blood glucose levels, and user engagement to demonstrate clinical effectiveness and reliability under different conditions.

IV. CONCLUSION

This study presented the design and development of a digital content framework that underpins the AIDCARE Education Module. We have proposed a structured and scalable approach for delivering personalized diabetes self-management education, integrated with gamification and a lightweight progress-tracking protocol, designed for integration into clinical workflows via the AIDCARE platform. In the future, we plan to strengthen security to meet established standards, replace placeholder videos with clinician-approved content, and expand both educational material and gamification features (for example, adding more achievements and refining the point distribution). Usability testing with individuals diagnosed with diabetes was also planned to assess the module's effectiveness and user experience. The results and insights from these tests will be used to further refine the educational experience of this module.

ACKNOWLEDGEMENT

This study is a part of the "AI Enabled Personalized Digital Assistant for Diabetes Self-Management (AIDCARE) project supported by Health Institutes of Türkiye (TUSEB), Turkish Ministry of Health under grant TUSEB 33987.

REFERENCES

[1] World Health Organization,"Diabetes: fact sheet," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes
[2] Ö.F.Üçer, A.Kavak, Y.Demirhan, M.Uzun, B.Savaş, A.B.İnner, et al. "Implementation of a decision support mechanism on a mobile platform using clinical evidence-based dynamic insulin dosage adjustment for artificial intelligence-enabled diabetes care (AIDCARE) system," MDPI Engineering Proceedings, vol. 104(27), 2025.
[3] S.R.Shrivastava, P.S.Shrivastava, J.Ramasamy, "Role of self-care in management of diabetes mellitus," Journal of Diabetes & Metabolic Disorders, vol. 12(14), 2013.
[4] J.Davis, A.H.Fischl, J.Beck, et al., "2022 National standards for diabetes self-management education and support," The Science of Diabetes Self-Management and Care, vol. 48(1), pp.44–59, 2022.
[5] Y.Zhou, G.Gan, "Impact of personalized education and supervision on pregnancy outcomes in women with gestational diabetes mellitus," Clinical and Experimental Obstetrics & Gynecology, vol. 50(4), pp.84, 2023.
[6] L.Huang, H.Guo, L.Xiu, H.Wang, "Efficacy of individualized education in patients with type 2 diabetes mellitus: a randomized clinical study protocol," Medicine, vol. 99(50), e23625, 2020.
[7] M.H.Fan, B.T.Huang, Y.C.Tang, X.H.Han, W.W.Dong, L.X.Wang, "Effect of individualized diabetes education for type 2 diabetes mellitus: a single-center randomized clinical trial," African Health Sciences, vol. 16(4), pp. 1157–1162, 2016.
[8] T.Kaihara, V.Intan-Goey, M.Scherrenberg, M.Falter, I.Frederix, Y.Akashi, P.Dendale, "Impact of gamification on glycaemic control among patients with type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials," European Heart Journal Open, vol. 1(3), oeab030, 2021.
[9] N.Luo, X.Meng, "Research on gamification of diabetes education," in Proc. 14th Int. Conf. Inf. Technol. Med. Educ. (ITME 2024), Guiyang, China, 2024, pp. 616–619.
[10] J.Zeng, D.Sun, C-K.Looi, and A.C.W.Fan, "Exploring the impact of gamification on students' academic performance: a comprehensive meta-analysis of studies from 2008 to 2023," British Journal of Educational Technology, vol. 55(6), pp. 2478-2502, 2024.
[11] Turkish Diabetes Foundation, "Diabetes diagnosis and treatment guide," [Online]. Available: https://www.turkdiab.org/admin/PICS/webfiles/DIYABET_TANI_ve_TEDAVI_REHBERI_16x23_2025.pdf
[12] Society of Endocrinology and Metabolism of Turkey, "Diabetes mellitus diagnosis, treatment and follow-up guide," [Online]. Available: https://file.temd.org.tr/Uploads/publications/guides/documents/diabetesmellitus2024.pdf
[13] Turkish Diabetes Association, "About diabetes," [Online]. Available: https://diyabetcemiyeti.org/diyabet-hakkinda.html
[14] Diyabetle Yaşam, [Online]. Available: https://www.youtube.com/@uzmandiyabetegitimhemsir-ib8lw/videos
[15] International Diabetes Federation, [Online]. Available: https://idf.org/
[16] Turkish Diabetes Foundation, "Everything about diabetes," [Online]. Available: https://www.turkdiab.org/diyabet-hakkinda-hersey.asp?lang=TR&id=46

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Transfer Learning Strategies for Leveraging Multi-Session and Multi-Subject Datasets to Improve sEMG Recognition

N. Sicard
Dept. of Electrical and Computer Engineering
Université du Québec à Trois-Rivières
Trois-Rivières, CANADA
nicolas.sicard@uqtr.ca

F. Nougarou
Dept. of Electrical and Computer Engineering
Université du Québec à Trois-Rivières
Trois-Rivières, CANADA
Francois.Nougarou@uqtr.ca

***Abstract*—Building movement detection models based on surface electromyography (sEMG) requires extensive data collection, which is time-consuming due to inter-session and inter-subject variability in the EMG signals. In fact, sensor placement can differ across sessions, and muscle physiology and user experience vary between subjects. The model-based approach of transfer learning offers a solution by transferring knowledge from an already acquired dataset into a model and deriving this pre-trained model to achieve good performance for a new session or a new subject with fewer data. However, how can this approach effectively exploit datasets from multiple sessions and multiple subjects? Compatibility issues arise between some pre-trained models and data from different sessions or subjects, affecting prediction accuracy. This study addresses these challenges and proposes strategies for combining and selecting transfer learning models from diverse datasets that significantly improve the prediction of eight forearm gestures with minimal training data. The proposed strategies produce promising results, with the combination strategy exceeding 90% precision and others between 85 and 90%.**

***Keywords*—EMG, gesture recognition, transfer learning strategies, inter-session, inter-subject.**

I. INTRODUCTION

Surface electro-myographic (sEMG) data persents high variability in the case of inter-session and inter-subject signals. The signals depend on the sensor placement relative to the arm and contact with the subject's skin and the individual muscle physiology that differs between subjects. This high variability makes training models for applications using these signals very long and straining on the subject. The use of transfer learning (TL) approach has gained traction as an extra step in hand recognition based on sEMG signals to add depth to the process[1-10]. Model-based transfer learning leverages knowledge from a pre-existing model trained from an initial dataset, which is then adjusted to achieve good performance for a new session or a new subject with less data. This improves the ease of model training, because the model only needs fine-tuning. This approach has been used within a variety of different models in sEMG recognition such as simple Neural Networks (NN) [1], Convolutional Neural Networks (CNN) [2-7], CNN combined with Long Short-Term Memory (LSTM) [8], Domain Generalization (DG) [9], and TL model ensemble decisional trees [10]. These studies show that recognition methods without transfer learning (TL) achieve high precision, and the TL improvement is often marginal but with fewer learning data[1-10].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

However, in the presence of several datasets from multiple sessions and multiple subjects, the strategies for exploiting these datasets using the TL approach are not trivial. Indeed, new data to be interpreted may not be compatible with some pre-trained models and this may affect the performance of the TL approach. This study highlights these compatibility challenges and explores strategies for combining and selecting multiple TL models from multiple datasets to improve motion prediction based on sEMG signals.

To do this, we conducted preliminary experiments in which EMG data from 8 sensors placed around the forearm were recorded on two subjects over several sessions with different electrode positions to detect 8 hand gestures. All these data enabled us (i) to evaluate the behavior of the pre-trained transfer learning models under different inter-session and inter-subject detection conditions, (ii) to observe the variability in the performance of these models and (iii) to propose transfer learning strategies for effectively exploiting multiple datasets in order to improve the gesture prediction while using a reduced number of training sEMG data.

II. MATERIALS AND METHODS

A. Data Acquisition and Features Extraction

The sEMG data is recorded using the MyoArmband® (Thalmic Labs, Canada), which consists of a bracelet with

TABLE I: Recorded hand gestures

Label	Image	Name	Label	Image	Name
M1		Rest	M5		Wrist ulnar flexion
M2		Hand closed	M6		Wrist extension
M3		Hand opened	M7		Wrist radial flexion
M4		Wrist flexion and fingers extension	M8		Thumb, index and little extension

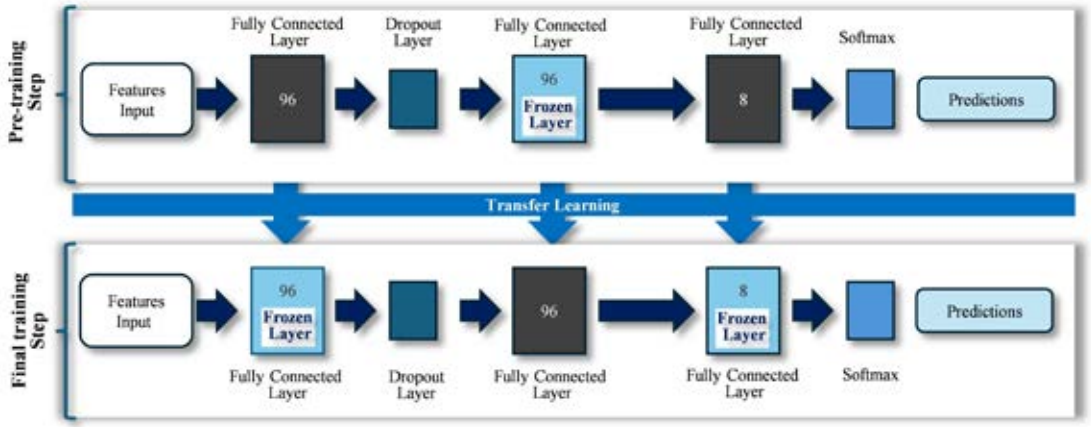


Fig. 1: Illustration of the training of the pre-trained model, the knowledge transfer and the training of the final model

8 electrodes that measure muscle activity around the forearm at a sampling rate of 200 Hz. Eight hand and wrist gestures (Table I) were recorded for training and testing using a custom Matlab/Python platform for target setting, sensor connection, and data collection. Two subjects were involved in this study to conduct a preliminary investigation on transfer learning based on multi-subject and multi-session datasets. Subject 1 completed three data collection sessions and Subject 2 completed four sessions. The bracelet is rotated after each session. Rotating the bracelet changes the electrode placement on the arm, simulating the variations that can occur between usage sessions of a given subject. Finally, a data collection session involves recording the sequences of each hand gesture three times, each lasting 30 seconds. A total of 48 size features were extracted from the EMG channels within a 100ms window, with a 50ms overlap: Mean Absolute Value (MAV), Modified MAV (MMAV), MAV Type 2 (MMAV2), Root Mean Square (RMS), Waveform Length (WL), and Mean Absolute Value Slope (MAVSLP) [11].

B. Training phases of the model-based transfer learning

1) Pre-training step: As shown in Fig. 1, the initial model is a neural network with three fully connected (FC) layers (96, 96, and 8 neurons), dropout, softmax, and cross-entropy. The second FC layer is frozen in the pre-training step and will only learn in the final training step. The purpose of freezing that layer is to teach the first and third layers to classify the position independent from the second layer. This pre-training step uses all data from the subject and session with a high number of epochs, allowing the model to capture data variability and specialize in hand gesture classification. Pre-training was done for all positions and both participants to obtain a model for each training condition.

2) Final training step: In model-based transfer learning, the pre-trained model serves as the initial state. The first and third FC layers are frozen with pre-trained weights, while the second FC layer is active (see Figure 1). The training dataset differs in bracelet position and/or subject from the pre-training datase. These datasets can be considered as belonging to two different domains. The model is retrained to map features and predict based on the input domain. The goal is to train the second FC layer to classify hand gestures despite domain changes, using minimal data. The resulting model is called the "TL model." The final training used all pre-trained models to create TL models for each domain and test inter-session and inter-subject TL.

C. Proposed combination and Selection strategies

With pre-trained models from multiple sessions and subjects, compatibility with new domain data varies. Some TL models perform well, others match no TL. Therefore, with multiple datasets, the strategies for using the TL approach are not trivial. Two strategies are explored for better prediction: combining and pre-selecting TL models (see Figure 2).

1) Combination strategy: Figure 2 (a) presents the proposed combination strategy based on the average of soft labels. For each sample, this approach averages the soft labels from all TL models and selects the gesture with the highest score. This technique uses maximum information during testing to ensure accuracy without selecting TL models. However, it requires

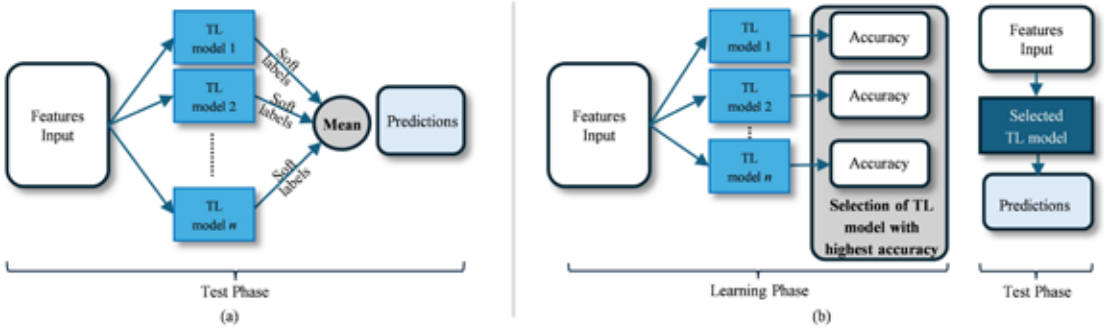


Fig. 2: Illustration of the proposed (a) combination strategy based on soft label average and the (b) selection strategy.

processing all TL models, with complexity increasing as their number grows.

2) Selection strategy: As shown in Figure 2 (b), during the learning phase, model selection considers all trained TL models and evaluates their accuracy using data that were not used for training. TL model accuracies are compared, and the highest-performing model is selected, which has the advantage of reducing computational complexity of the testing phase.

3) Selection & combination strategy: This approach combines the proposed selection and combination strategies, offering a compromise between computational efficiency and the amount of information used. Here, the two highest-performing models are first selected and then combined. This solution slightly increases the system's complexity but also enhances the amount of information leveraged for prediction. Additionally, it is possible to increase the number of TL models selected for averaging, further improving predictive accuracy at the cost of added complexity.

III. Results

A. Behavior of TL Models

As shown in Figure 3, the proposed strategies were evaluated using 5 to 25 training samples per gesture. All models aim to detect the 8 gestures from Table I using EMG data from a domain different from pre-training. The datasets used are detailed in Table II. Several observations can be made from Figure 3. First, the TL models, pre-trained on data from different domains, exhibit distinct behaviors. One model performs similarly to the model without TL (Model 6), most of them achieve better performance, and one model stands out significantly (Model 5). This indicates that TL models behave differently due to varying levels of compatibility with the new domain data. This variability in TL model behavior highlights the importance of developing strategies to effectively leverage multiple TL models. The model without TL underperforms compared to the TL models, and even with just 5 training samples per movement, the models can achieve over 90% accuracy to detect all gestures. In this scenario, the strategy combining all 6 models shows strong potential, reaching nearly 95% accuracy with just 5 samples per gesture and improving slightly with more data. The model selection strategy also delivers satisfactory results, the accuracy at 5 samples per movement is of 90%. Finally combining both strategies, which requires only two TL models, balances performance and complexity, showing accuracies between the other two strategies.

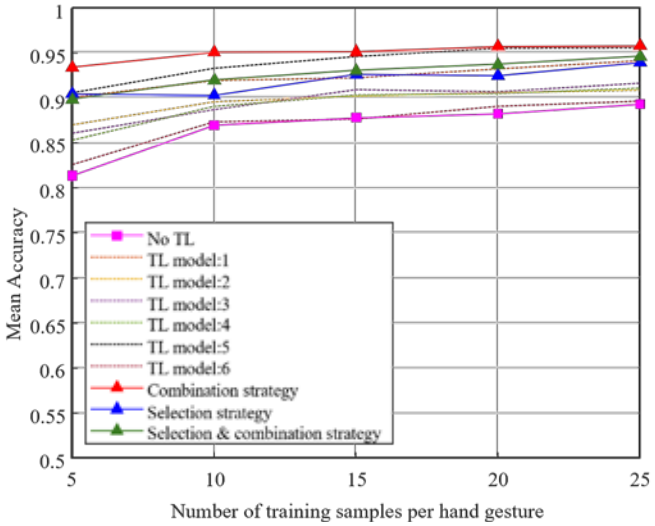


Fig. 3: Precision performances for available TL models and proposed strategies with varying training sample sizes.

TABLE II: TL data configurations used in Figure 3 for given participants and electrode position

TL Model	Pre-trained (Pos., Part.)	TL (Pos., Part.)
TL Model 1	Pos. 1, Part. 1	Pos. 2, Part. 2
TL Model 2	Pos. 2, Part. 1	Pos. 2, Part. 2
TL Model 3	Pos. 3, Part. 1	Pos. 2, Part. 2
TL Model 4	Pos. 1, Part. 2	Pos. 2, Part. 2
TL Model 5	Pos. 3, Part. 2	Pos. 2, Part. 2
TL Model 6	Pos. 4, Part. 2	Pos. 2, Part. 2

B. Evaluation of the combination and selection strategies

Accuracy results are presented in Figure 4 across three scenarios for a given subject *x*: inter-session (Cond.A*x*), separated inter-subject for each of the 2 subjects (Cond.B*x*), and inter-subject and inter-session combined (Cond.C*x*). Subject 1 is inexperienced, while Subject 2 is experienced in performing the gestures to acquire forearm EMG signals. To highlight the benefits of the TL approach with minimal training data, all simulations used only 5 samples per movement. The average precision, recall and F1-Score are also evaluated in Table III. Finally, Table IV presents the accuracy for each individual hand gesture. All results are averages over 20 cross-validation repetitions using all available TL model conditions. In each scenario, a reference model was trained from scratch using the same training and testing data as the TL models. The reference model for Subject 1 has an average accuracy of 70%, while Subject 2 reached 77%. As shown in Figure 4, across all tested training conditions, the proposed strategies consistently outperformed the reference model by approximately 10%.

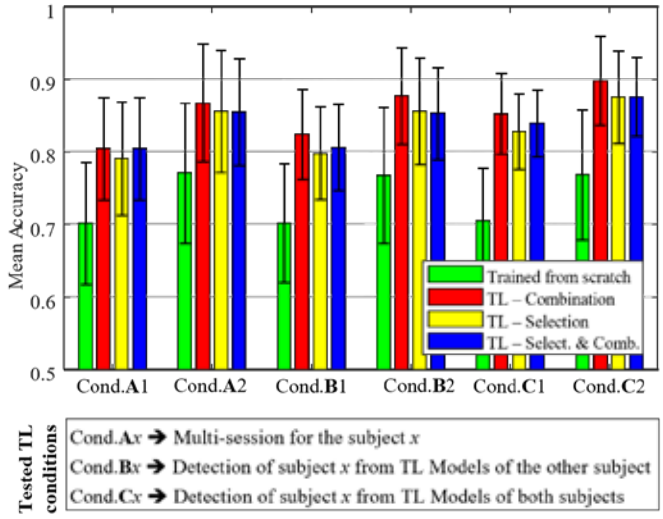


Fig. 4: Average accuracies of the model without TL and the proposed TL strategies in different TL conditions.

TABLE III: Performance metrics for CondC1 and CondC2 over 20 repetitions

Strategy	Precision	F1-Score	Recall
CondC1			
Scratch	69.36 ± 7.50	66.78 ± 8.81	70.96 ± 7.76
Selection	78.99 ± 6.58	76.85 ± 7.21	80.25 ± 6.18
Select. & Comb.	80.62 ± 5.80	78.73 ± 6.82	82.46 ± 4.88
Combination	**83.79 ± 6.33**	**82.21 ± 7.61**	**84.70 ± 5.54**
CondC2			
Scratch	80.08 ± 6.65	78.88 ± 6.91	82.57 ± 5.98
Selection	88.82 ± 5.77	88.16 ± 6.31	90.27 ± 5.32
Select. & Comb.	88.45 ± 5.12	87.70 ± 5.58	90.03 ± 4.56
Combination	**92.03 ± 4.77**	**91.66 ± 5.07**	**93.21 ± 4.11**

For both subjects, the proposed strategies show similar behaviors. The combination strategy shown in Figure 2 consistently provides the best performance under all conditions and training sample datasets. In the case of the inter-session with subject 1, the combination strategy and the selection & combination strategy give identical results, as only two TL models are available, meaning that both approaches use the same models. In most cases, the selection and combination strategy (section II.C.3) achieves the second best performance, and the selection strategy (section II.C.2) performs slightly worse. Thus, as previously mentioned, the combination strategy is the approach that makes the best use of the available TL models, although it requires all TL models during the testing phase. Selection-based strategies are slightly less effective but require fewer models for prediction.

When considering Cond. A*x* and Cond. B*x*, the proposed strategies effectively address the challenges of classification scenarios with minimal training data by leveraging the information of the available datasets. Note that Subject 2, being more experienced, consistently outperforms Subject 1. However, the results from Cond.B1 and Cond.B2 demonstrate effective knowledge transfer between subjects. For each subject, Cond.C*x* yields the best results, where the strategies use information from both inter-session and inter-subject data. The proposed strategies perform better with more information, indicating their effectiveness and great potential for achieving higher performance with fewer training data.

Table III gives a better perspective of the performances of the proposed strategies in CondC1 and CondC2. The bold values are the highest for each performance metric in each condition. For each metric of each condition, the combination strategy has the highest performance. For Subject 1, the combination strategy achieves over 80% average accuracy across all metrics, while Subject 2 exceeds 90%. Notably, models trained from scratch perform significantly worse.

Table IV details the results per movement for CondC2. It shows the effectiveness of the proposed methods. By leveraging multiple datasets, the detection of movements that are more difficult for the reference model improves.

TABLE IV: Accuracy separated by class using CondC2 over 20 repetitions

Strategy	M1	M2	M3	M4	M5	M6	M7	M8
Scratch	79.88	90.10	79.20	78.35	90.65	85.60	66.31	83.35
Selection	98.41	93.72	88.65	87.45	95.96	91.89	79.04	96.58
Select. & Comb.	96.13	91.05	86.42	81.96	94.47	91.85	80.33	98.91
Combination	98.63	95.36	89.20	87.89	98.02	94.99	83.52	98.31

IV. Conclusion

The goal of this preliminary study is to propose transfer learning (TL) strategies to exploit multi-session and multi-subject datasets to improve the prediction of eight hand gestures using reduced surface electromyography (sEMG) data. Three TL strategies are presented and proposed based on combination of TL models, selection of TL models, and selection & combination of TL models. The combination strategy is the most complex but most efficient approach to leverage the information from the available TL models. Selection-based strategies are slightly less effective but are significantly less complex. These three approaches outperform a model that doesn't go through TL, and only necessitates minimal training data due to its utilization of the information from the existing pre-trained models. Finally, it has been shown that the more these strategies leverage information, the better they perform. Thus, future work will involve a more in-depth evaluation of these proposals with a more complex recognition structures and a larger number of subjects and inter-session data.

ACKNOWLEDGMENT

Research supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] C. Andronache et al., "Towards extending real-time EMG-based gesture recognition system," 43rd International Conference on Telecommunications and Signal Processing (TSP), 2020, pp. 301-304.
[2] U. Côté-Allard et al., "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning," in IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 27, no. 4, pp. 760-771, April 2019.
[3] R. Deepa et al., "Revolutionizing Hand Gesture Recognition: A Transfer Learning Approach using Surface Electromyography and Convolutional Neural Networks," Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2023, pp. 70-75.
[4] J. Lu, P. Liang et al., "EffiE: Efficient Convolutional Neural Network for Real-Time EMG Pattern Recognition System on Edge Devices," 11th International IEEE/EMBS Conference on Neural Engineering (NER), Baltimore, MD, USA, 2023, pp. 1-5.
[5] S. S. Roy et al., "Hand Movement Recognition Using Cross Spectrum Image Analysis of EMG Signals-A Deep Learning Approach," Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA), 2020, pp. 1-5.
[6] H. Shi, X. Jiang, C. Dai and W. Chen, "EMG-based Multi-User Hand Gesture Classification via Unsupervised Transfer Learning Using Unknown Calibration Gestures," in IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1119-1131, 2024.
[7] C. Tang, W. Yi, J. M. Kim and L. G. Occhipinti, "EMG-Based Human Motion Analysis: A Novel Approach Using Towel Electrodes and Transfer Learning," IEEE SENSORS, 2023, pp. 1-4.
[8] X. Chen et al., "Hand Gesture Recognition based on Surface Electromyography using Convolutional Neural Network with Transfer Learning Method," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 4, pp. 1292-1304, April 2021.
[9] Y. Ye, Y. He, T. Pan, Q. Dong, J. Yuan and W. Zhou, "Cross-subject EMG hand gesture recognition based on dynamic domain generalization," 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2023, pp. 1-4.
[10] D. Wu, J. Yang and M. Sawan, "Transfer Learning on Electromyography (EMG) Tasks: Approaches and Beyond," in IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 3015-3034, 2023
[11] A. Phinyomark et al., "Evaluation of EMG feature extraction for hand movement recognition based on Euclidean distance and standard deviation," in Proc. ECTI-CON, May 2010, pp. 856–860.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

A Focused Survey on Patient Simulation with Large Language Models

Selin Katarci
Izmir Tinaztepe University
Private Buca Hospital
Izmir, Türkiye
selin.katarci@tinaztepe.com

Ayca Kumluca Topalli
Department of Electrical and Electronics Engineering
Izmir University of Economics
Izmir, Türkiye
ayca.topalli@ieu.edu.tr

***Abstract*—The rapid advancement of generative artificial intelligence technologies, particularly large language models such as ChatGPT, has opened new avenues in healthcare education. One of the most promising applications is the simulation of virtual patients to enhance clinical training. This study aims to critically analyze and compare ten recent peer-reviewed research articles that explore the integration of ChatGPT-based virtual or standardized patients into medical education. The selected studies vary in methodology, scope, and application, ranging from history taking and empathy training to gamified learning environments, emotion recognition, and structured simulation platforms. A qualitative synthesis was conducted, focusing on implementation strategies, learner engagement, user satisfaction, and educational outcomes. The review reveals that large language model-powered virtual patients show high potential in improving communication skills, diagnostic reasoning, and learner autonomy. Several studies report positive user feedback and notable cost and scalability advantages compared to traditional simulation methods. In addition, novel training techniques targeting the emotional awareness of artificial intelligence agents are emerging as an important direction for future development. However, limitations such as response accuracy, emotional realism, and the absence of non-verbal communication remain key challenges. Most authors agree that large language models should complement rather than replace human interaction in simulation-based learning. This paper contributes to a deeper understanding of how generative artificial intelligence can be effectively used in medical education, providing valuable insights for educators, curriculum designers, and technology developers seeking to integrate artificial intelligence tools into training programs.**

Keywords— Large Language Model; Virtual Patients; Healthcare Education; Generative Artificial Intelligence; Empathy; Emotion Recognition; Medical Simulation component

I. INTRODUCTION

The integration of artificial intelligence (AI) into medical education has gained rapid momentum in recent years, driven by the increasing capabilities of large language models (LLMs) such as ChatGPT. These generative AI tools are reshaping how educators design, deliver, and evaluate training in clinical settings.

Among the most innovative applications is the use of ChatGPT-powered virtual patients (VPs) and standardized patients (SPs), which offer learners safe, accessible, and cost-effective platforms to develop clinical reasoning, communication, and diagnostic skills [1, 2].

Traditional clinical education methods often rely on human actors or manikins to simulate patient encounters. While effective, these approaches are resource-intensive and constrained by logistical limitations, including personnel training, scalability, and cost [3, 4].

To address these challenges, researchers have begun exploring AI-based simulations that leverage the natural language understanding and generative capabilities of LLMs. This paradigm shift enables scalable, personalized, and repeatable training scenarios for students at all levels of medical education [5, 6].

Early studies show promising results regarding the realism, usability, and educational value of these AI-powered simulations. Vaughn et al. demonstrated that ChatGPT-generated scenarios could support faculty in developing credible and diverse case-based exercises, though limitations related to missing clinical detail and factual inaccuracies remain [4].

Similarly, Liu et al. [5] and Holderried et al. [2] confirmed that ChatGPT could simulate patient dialogues effectively, scoring high in realism and diagnostic plausibility when evaluated by medical experts. However, concerns persist regarding authenticity, especially in emotionally nuanced interactions and empathy-related contexts [7].

Another critical dimension of these tools is their impact on learner autonomy and skill development. Aster et al. found that students using ChatGPT for history-taking tasks experienced a heightened sense of autonomy and self-directed learning [7].

Meanwhile, the Korean medicine training platform developed by Ryu et al. emphasized structured competency development in specific domains such as suicide risk assessment, suggesting that these tools can be tailored for niche educational contexts [3].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Furthermore, the open-access platform described by Thesen et al. underscores the importance of transparency, scalability, and educator control in deploying generative AI tools for training purposes [8].

Despite their advantages, these systems are not without drawbacks. Issues such as occasional hallucination of facts, language biases, and lack of standardized pedagogical feedback limit their standalone use in high-stakes education [1, 5].

Additionally, most current implementations remain experimental or supplementary, highlighting the need for rigorous evaluation and integration into formal curricula [9].

This paper aims to provide a comprehensive literature-based analysis of ten recent studies exploring the use of ChatGPT and generative AI technologies as VPs in healthcare education. By comparing their methodologies, implementation strategies, evaluation metrics, and reported outcomes, the evolving role of AI in clinical training is illuminated, and both opportunities and limitations in this rapidly expanding field are identified.

II. LITERATURE REVIEW

The application of LLMs, particularly ChatGPT, in healthcare education has attracted considerable attention in recent years. This review evaluates ten studies published between 2021 and 2025 that explore the use of ChatGPT-powered VPs and SPs for clinical training. These studies vary in implementation methods, target learners, interaction style, and evaluation metrics.

Holderried et al. [2] designed a web-based chatbot using GPT-3.5-turbo to simulate a patient role in history-taking scenarios. Their findings indicate high realism and plausibility in responses, with 94% of chatbot outputs matching predefined scripts. Similarly, Liu et al. [5] implemented ChatGPT in ten clinical scenarios evaluated by senior physicians, confirming the model's ability to replicate realistic patient behavior.

Several studies explored structured or scripted VP systems. Aster et al. [7] evaluated the role of ChatGPT in simulated empathetic dialogue, noting improvements in learner autonomy but identifying limitations in emotional depth and non-verbal feedback. Pears et al. [6] introduced a gamified platform that integrates ChatGPT with video technologies, highlighting enhanced engagement but noting the lack of clinical depth due to the demo-based nature of the system.

Ryu et al. [3] proposed a simulation platform for Korean medicine education using a prompt-tuned GPT-4o. Their system achieved high student satisfaction, particularly in niche domains such as suicide risk assessment. Vaughn et al. [4] demonstrated a scenario generation tool using GPT-3.5 prompts to assist educators in case design. Though beneficial for rapid prototyping, expert revision was necessary to ensure accuracy.

In terms of deployment, Thesen et al. [8] developed an open-access web app using GPT-4o with human-authored case constraints. Their platform emphasized flexibility, multilingual access, and customization. Cologlu et al. [1] investigated intern physicians' interaction with an AI-powered VPs in clinical simulations, reporting high realism but identifying language glitches and lack of visual input.

Additional studies addressed backend architectures. Hsieh et al. [10] integrated multiple LLMs (ChatGPT, PaLM, ERNIE-4, Mistral) into a simulation engine with retrieval-augmented generation (RAG). This platform achieved an 85% match with human evaluation and improved realism in SP behavior.

Lastly, Tavarnesi et al. [9] employed manual prompt-based tools to design flexible cases for educators, offering speed and simplicity at the expense of accuracy and hallucination risk.

Overall, the literature highlights the rapid innovation and diversity of approaches in AI-powered medical training. These tools generally improve accessibility, scalability, and engagement, but often fall short in clinical depth, empathy modeling, and standardized assessment protocols.

III. METHODOLOGY

This paper employs a structured comparative literature review methodology to explore and synthesize recent research on the use of ChatGPT and related LLMs as VPs or SPs in healthcare education. This methodological approach allows for a rigorous, multi-dimensional examination of diverse implementations across different educational, clinical, and technological contexts. In particular, it enables a systematic mapping of how generative AI technologies are currently being designed, deployed, and evaluated in simulation-based medical training.

Structured literature reviews are particularly effective when comparing studies that apply varying tools and pedagogical strategies, as they enable the researcher to identify thematic patterns and methodological trends without relying solely on quantitative synthesis [9, 10]. This approach also facilitates a critical assessment of the pedagogical, technical, and evaluative dimensions of each study, ensuring that both strengths and limitations are explicitly addressed.

The reviewed literature consists of ten peer-reviewed articles published between 2021 and 2025. These studies were selected based on specific inclusion criteria: first, each study involved the use of ChatGPT or a comparable LLM such as GPT-3.5, GPT-4, or GPT-4o; second, the focus of the study was on medical or healthcare education, specifically in the simulation of patient communication, clinical reasoning, or diagnostic training; and third, the article included an evaluative component, such as expert reviews, learner feedback, or performance metrics.

All studies were published in English and represent a range of educational settings and clinical specialties, including general internal medicine [2], trauma [9], mental health [7], and Korean medicine [3]. The inclusion of diverse medical domains was intentional, as it allows the synthesis to capture both broad trends in AI-powered simulation and specialized applications tailored to niche training needs.

Following a full-text analysis of each study, a set of common analytical categories was inductively developed to ensure consistency and depth across the comparison. These categories include the type of simulation tool (e.g., chatbot, gamified simulation, prompt-based generator), the version and configuration of the LLM used (e.g., GPT-3.5, GPT-4, RAG-based architectures, LangChain), the target user group (e.g.,

medical students, interns, nursing educators), and the number and diversity of simulated clinical cases.

In addition to technical specifications, the studies were also coded for the pedagogical design principles applied, such as scenario complexity, adaptability to learner input, and the presence or absence of feedback mechanisms.

Furthermore, they were examined for reported advantages such as accessibility, realism, and learner autonomy [4, 5, 6], as well as for limitations including hallucination of facts, limited emotional depth, and lack of non-verbal communication [1, 7]. This dual focus on both benefits and challenges provides a balanced perspective on the educational potential of AI-powered simulation tools.

The extracted data were compiled into a comparative synthesis table (Table I), which enables side-by-side analysis of implementation strategies, evaluation methods, and outcomes. For example, Holderried et al. [2] evaluated ChatGPT's ability to generate realistic differential diagnoses across thirty clinical cases, while Aster et al. [7] focused on the model's performance in replicating empathic responses during history-taking scenarios. Ryu et al. [3] developed a structured simulation training platform for Korean medicine students, specifically targeting suicide risk assessment, illustrating how LLMs can be adapted to culturally specific educational frameworks. Similarly, Thesen et al. [8] presented an open-access generative AI tool built on GPT-4o that allowed educators to design and control clinical simulations, prioritizing transparency and adaptability. Tavarnesi et al. [9], on the other hand, used audiovisual storytelling and gamification to enhance engagement with AI-generated VPs, reflecting an emerging trend toward immersive, multimedia-enhanced simulation environments.

By analyzing these studies comparatively, it became possible to identify several emerging clusters in implementation, such as scripted versus open-ended simulations, faculty-controlled versus autonomous systems, and static versus adaptive feedback mechanisms. These patterns offer insights into how various design choices impact educational outcomes, learner engagement, and system scalability. For instance, scripted simulations tend to provide higher factual accuracy but lower flexibility, whereas open-ended models support richer learner interaction at the cost of consistency.

The thematic synthesis was guided by principles outlined in previous educational technology reviews, allowing for a rigorous yet flexible approach to data interpretation [9, 10].

While this methodology allowed for a rich, multi-dimensional comparison, certain limitations must be acknowledged. Differences in study design, inconsistent reporting of evaluation metrics, and a lack of transparency regarding AI configuration in some articles posed challenges to uniform analysis [4, 8].

Additionally, most implementations were still in pilot or early development stages, limiting the generalizability of their findings.

Another limitation concerns the absence of long-term outcome evaluations, as none of the reviewed studies provided longitudinal data on skill retention or transfer to clinical practice.

Nevertheless, the chosen methodology proved effective in mapping the current landscape of LLM-powered patient simulation tools in medical education and identifying critical success factors, educational potentials, and research gaps, laying the groundwork for more comprehensive future studies.

IV. CONCLUSION

This structured comparative review synthesized evidence from ten peer-reviewed studies investigating the application of ChatGPT and related LLMs as VPs and SPs in medical education.

By systematically comparing implementation methods, underlying model architectures, medical domains, participant demographics, and reported learning outcomes, this review established a detailed cross-study perspective on the state of AI-driven clinical simulation.

The analysis confirms that LLM-powered patient simulations can substantially enhance accessibility, adaptability, and learner autonomy, providing cost-effective, on-demand opportunities for clinical reasoning and communication training. Notably, implementations leveraging advanced prompting techniques, RAG, domain-specific fine-tuning, and interactive feedback loops demonstrated higher levels of contextual accuracy, clinical realism, and diagnostic plausibility.

Despite these promising developments, recurring limitations persist across the reviewed literature. Emotional depth remains limited, with AI systems often struggling to convey nuanced empathy or replicate the complexity of non-verbal communication.

Additionally, factual inaccuracies, or "hallucinations", occur with varying frequency, necessitating vigilant expert oversight in high-stakes scenarios. While some studies achieved successful integration of AI, based simulations into structured curricula and specialized training, such as Korean medicine, trauma management, and mental health assessment, others remained exploratory, lacking robust evaluation protocols or long-term outcome measurement. This disparity highlights the urgent need for standardized assessment frameworks, validated performance rubrics, and cross-institutional benchmarking to ensure educational reliability and clinical safety.

Looking forward, the continued evolution of multimodal LLMs, capable of processing and generating not only text but also speech, images, and potentially biometric data, presents an opportunity to address current limitations in empathy modeling and non-verbal interaction.

Incorporating adaptive feedback mechanisms, scenario branching logic, and integration with advanced clinical data simulation tools could further increase pedagogical value. Equally important are ethical and operational considerations, including bias mitigation, transparency in AI decision-making, data privacy safeguards, and educator training for effective integration of generative AI tools.

TABLE I. COMPARISON OF STUDIES

Title	Year	Tool Type	LLM Used	Diseases Covered	Participants	Advantages	Disadvantages
[7]	2025	Virtual Patient – text chatbot (empathy/history)	ChatGPT (GPT-3.5)	1 domain (Cardiology)	3rd-year medical students	Learner autonomy; empathic opportunities; repeatable practice	No non-verbal cues; variable empathy depth; text-only
[8]	2025	Open-access web app (educator-authored cases; formative feedback)	GPT-4o (OpenAI)	Multiple	NS	Open-access; multilingual; educator control; scalable	Feedback depth limited by rubric; case quality depends on uploads
[2]	2024	Simulated Patient chatbot (structured history-taking)	gpt-3.5-turbo (OpenAI)	1 case	Medical students	High plausibility (97.9%); 94.4% script-grounded answers; positive usability	Single case; occasional implausible/socially desirable answers; no non-verbal cues
[3]	2024	Simulation platform (CPX-style)	ChatGPT API (version NS)	1 domain (Suicide risk assessment)	KM students	Competency-focused; high satisfaction; standardized practice	Authenticity skepticism; niche scope; needs wider validation
[4]	2024	Scenario generation assistant (educator-facing)	ChatGPT (likely GPT-3.5)	5 scenarios	Peer reviewers	Rapid drafting; diverse cases; reduces authoring workload	Missing clinical details; accuracy requires expert curation
[6]	2024	Gamified platform (workshop)	ChatGPT-4 (OpenAI API)	Multiple	Healthcare learners	High engagement; cost-effective; supports remote/visual learning	Prototype scope; limited clinical depth; modest evaluation sample
[9]	2024	Virtual Patient system (interactive audiovisual + NLP)	NS (generative/NLP components; model not specified)	Multiple	Medical students	High realism via actor video; dialogic practice; end-of-visit feedback	Accuracy depends on NLP/generative layer; adaptability limited by authoring
[10]	2024	Standardized Patient engine (multi-LLM + RAG)	ChatGPT, PaLM, ERNIE-4, Mistral (ensemble)	Multiple	Students + expert evaluators	Higher factual consistency; improved realism via hybridization	Complexity/latency; compute cost; integration overhead
[5]	2023	Simulated Standardized Patient (text chat)	ChatGPT-3.5-turbo (OpenAI)	10 cases	Senior physicians	Rapid deployment; colloquial answers; high plausibility	Role-flip in long dialogs; repetitive phrasing; no non-verbal signals
[1]	2021	Virtual Patient (early AI VP prototype)	Non-LLM (classical NLP/rule-based)	Multiple	Medical trainees	Early low-cost VP; scalable beyond human SP constraints	Less natural language; narrower dialogue than LLM-based systems

In conclusion, while ChatGPT-powered VP/SP systems may not entirely replace traditional clinical education methods in the immediate future, they are likely to become a powerful and indispensable complement, especially in expanding access to realistic, personalized, and repeatable training experiences. With continued technological advancement and pedagogical refinement, these systems hold strong potential to shape the future of medical education in a positive and transformative way.

REFERENCES

[1] D. Cologlu, M. Ak, and O. E. Kurt, "An artificial intelligence powered virtual patient for clinical training: A preliminary study," Health Policy and Technology, vol. 10, no. 3, 2021, Art. no. 100534, doi: 10.1016/j.hlpt.2021.100534.

[2] F. Holderried, C. Stegemann-Philipps, L. Herschbach, J.-A. Moldt, A. Nevins, J. Griewatz, M. Holderried, A. Herrmann-Werner, T. Festl-Wietek, and M. Mahling, "A generative pretrained transformer (GPT)–powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study," JMIR Medical Education, vol. 10, p. e53961, 2024, doi: 10.2196/53961.

[3] J.-W. Ryu, C.-Y. Kwon, J.-S. Park, S.-R. Lim, H.-L. Jeon, H.-J. Kim, and S.-H. Kim, "Development and application of a ChatGPT-based simulation training platform for Korean medicine," Journal of Oriental Neuropsychiatry, vol. 35, no. 4, pp. 413–427, 2024, doi: 10.7231/jon.2024.35.4.413.

[4] J. Vaughn, S. H. Ford, M. Scott, C. Jones, and A. Lewinski, "Enhancing healthcare education: Leveraging ChatGPT for innovative simulation scenarios," Clinical Simulation in Nursing, vol. 87, 2024, Art. no. 101487, doi: 10.1016/j.ecns.2023.101487.

[5] X. Liu, C. Wu, R. Lai, H. Lin, Y. Xu, Y. Lin, and W. Zhang, "ChatGPT: when the artificial intelligence meets standardized patients in clinical training," Journal of Translational Medicine, vol. 21, no. 447, 2023, doi: 10.1186/s12967-023-04314-0.

[6] M. Pears, C. Poussa, and S. T. Konstantinidis, "Progressive healthcare pedagogy: An application merging ChatGPT and AI-video technologies for gamified and cost-effective scenario-based learning," in Smart Mobile Communication & Artificial Intelligence (IMCL 2023), Lecture Notes in Networks and Systems, vol. 937, Cham, Switzerland: Springer, 2024, pp. 103–116, doi: 10.1007/978-3-031-56075-0_10.

[7] A. Aster, S. V. Ragaller, T. Raupach, and A. Marx, "ChatGPT as a virtual patient: Written empathic expressions during medical history taking," Medical Science Educator, 2025, doi: 10.1007/s40670-025-02342-7.

[8] T. Thesen, N. A. Alilonu, and S. Stone, "AI patient actor: An open-access generative-AI app for communication training in health professions," Medical Science Educator, vol. 35, pp. 25–27, 2025, doi: 10.1007/s40670-024-02250-2.

[9] G. Tavarnesi, A. Laus, R. Mazza, L. Ambrosini, N. Catenazzi, S. Vanini, and D. Tuggener, "Learning with virtual patients in medical education: Agenerative AI–based approach," in 2024 IEEE Global Engineering Education Conference (EDUCON), pp. 1425–1432, 2024, doi: 10.1109/EDUCON60474.2024.10579571.

[10] H.-C. Hsieh, C.-H. Lee, H.-M. Huang, C.-H. Lin, and T.-Y. Kuo, "Integrating multiple LLMs with retrieval-augmented generation for standardized patient simulations," in 2024 International Conference on Advanced Learning Technologies (ICALT), pp. 256–260, 2024, doi: 10.1109/ICALT60092.2024.00052.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Tumor-Targeted Nanoparticle Delivery Prediction Using Machine Learning

Oguzhan Panatli [1,2], Açelya Yilmazer [1,3], Gazi Erkan Bostancı [2]

[1] Department of Biomedical Engineering, Ankara University, Gölbaşı, Ankara 06830, Türkiye
[2] Department of Computer Engineering, Ankara University, Gölbaşı, Ankara 06830, Türkiye
[3] Stem Cell Institute, Ankara University, Balgat, Ankara 06520, Türkiye
panatli@ankara.edu.tr, ayilmazer@ankara.edu.tr, ebostanci@ankara.edu.tr

***Abstract*—Because of its rising incidence and mortality rates, cancer is a major public health issue worldwide. Although one of the most common cancer treatments is chemotherapy, the method can cause serious side effects by damaging normal cells along with cancer cells. Nanomedicine is a promising field with significant potential to reduce the harmful effects of chemotherapy and increase the effectiveness of the treatment process. On the other hand, the effective delivery of nanoparticles to solid tumors continues to be a substantial obstacle in the clinical application of cancer nanomedicine. This study involved the development and comparison of six machine learning models to predict the tumor delivery efficiency of nanoparticles. The models were developed using the Nano-Tumor Database, comprising 376 samples and four target variables (DE_24, DE_168, DE_Max, DE_Tlast). After data preprocessing and hyperparameter tuning, the XGBoost model achieved the lowest error values on the test set, with root mean square error values of 2.87, 1.43, and 3.42 for the DE_24, DE_168, and DE_Max target variables, respectively. The Random Forest model achieved the lowest error values on the test set, with a root mean square error value of 2.83 for the DE_Tlast target variable. The findings suggested that machine learning approaches have significant potential in the field of cancer nanomedicine.**

Keywords—artificial intelligence; machine learning; nanomedicine; nanotechnology; drug delivery

I. Introduction

Cancer remains one of the leading global public health concerns, attributable to its high incidence and mortality rates [1]. According to global cancer statistics for 2022, there were approximately 20 million new cases and nearly 9.7 million cancer-related deaths worldwide [2]. Chemotherapy is a widely used cancer treatment method. Conventional chemotherapy can damage tumors and healthy cells, causing gastrointestinal reactions, bone marrow suppression, and hair loss. These side effects reduce treatment efficacy and dose tolerance, lowering quality of life [3]. Targeted drug delivery systems are being developed to diminish these obstacles. These systems target chemotherapeutic agents directly to the tumor site to reduce drug toxicity and improve treatment efficacy [4]. Liposomes, polymeric nanoparticles, dendrimers, and metal nanoparticles are commonly used in this field of nanotechnology [5, 6].

Nanomedicine is a valuable field in biomedical research in search of safer and more effective cancer chemotherapies [7, 8]. On the other hand, the limited delivery efficiency (DE) of nanoparticles (NPs) to solid tumors remains a significant obstacle to the clinical translation of cancer nanomedicine [9, 10]. The percentage of systemically administered nanoparticles that localize in tumor tissue is called DE. This metric quantifies the total accumulation of NPs in the tumor at a specific time point [11].

Wilhelm et al. [11] carried out a detailed literature review to assess the DE of NPs to solid tumors. They reported that a median of 0.7% of the injected dose (ID) of nanoparticles reached the tumor based on the literature analysis. Furthermore, they highlighted that DE varied depending on various factors, including physicochemical properties of the NP (type, core material, size, zeta potential, and shape), targeting strategy, tumor model, and cancer type. Recent studies have leveraged machine learning (ML) and artificial intelligence (AI)-assisted physiologically based pharmacokinetic (PBPK) approaches to predict nanoparticle delivery [12, 13], and our work extends these studies by comparing multiple ML models.

In the present study, we developed and compared different ML models to predict the tumor DE of NPs. These ML-based prediction models were based on diverse factors, including NPs' physicochemical properties, targeting strategy, and tumor-related parameters, using the Nano-Tumor Database, which comprises 376 samples. Our findings suggest that ML-based approaches, particularly when integrated with PBPK modeling, may provide valuable insights for advances in cancer nanomedicine.

II. Materials and Methods

A. Dataset and Data Preprocessing

Within the scope of this study, "Nano-Tumor Database" [9] was used to develop ML models. The dataset included 376 samples sourced from 200 studies, covering a diverse array of cancer nanomedicine research published between 2005 and 2018. The dataset comprised various parameters associated with physicochemical properties of NP, targeting strategy, cancer type, tumor size, tumor weight, tumor model, injected dose, and body weight of subject. In addition, the dataset comprises four target variables derived from the PBPK model [9]. More comprehensive information about the dataset is presented in Table 1.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

TABLE I. OVERVIEW OF VARIABLES AND THEIR CATEGORIES/VALUES USED IN THE MACHINE LEARNING MODELS

Feature	Unit	Category / Value Range
Nanoparticle type	–	Inorganic, Organic, Hybrid
Nanoparticle's core material	–	Polymeric, Gold, Other OM, Liposomes, Silica, Hydrogels, Dendrimers, Other IM, Iron Oxide, Silica-Liposomes, Iron Oxide-Polymeric, Other HM, Silica-PDA, Zinc Protein-Hydrogels
Nanoparticle shape	–	Spherical, Rod, Plate or Flake, Other
Hydrodynamic diameter	nm	[5, 456.6]
Zeta potential	mV	[-59.4, 71.3]
Targeting strategy	–	Passive, Active
Cancer type	–	Brain, Breast, Cervix, Colon, Liver, Lung, Ovary, Pancreas, Prostate, Skin, Stomach, Kidney, Other
Tumor model	–	Allograft Heterotopic, Allograft Orthotopic, Xenograft Heterotopic, Xenograft Orthotopic, Other
Tumor weight	g	[0.02, 5.09]
Tumor size	cm^3	[0.02, 1.8]
Administered dose	mg/kg	[0, 1220]
Subject Body weight	g	[16, 35]
DE at 24 hours	%ID	[0.008, 50.03]
DE at 168 hours	%ID	[0.003, 23.83]
Maximum DE	%ID	[0.008, 56.89]
DE at the last sampling time point	%ID	[0.009, 29.80]

In data preprocessing, to detect outliers, the interquartile range (IQR) method was used. After this step, the K-Nearest Neighbors (KNN) imputation method was employed to impute missing values. The distribution of rare categories was subsequently examined, and their target variable means were compared with those of other categories. Following this, some rare categories were merged. Afterwards, one-hot encoding was employed to transform categorical variables into numerical data. Following these steps, the dataset was split into train and test sets, allocating 80% for training and 20% for testing. Finally, z-score normalization, which is feature scaling, was applied to the numerical variables to rescale.

B. Model Development

A total of six ML algorithms were used to develop models that predict the DE of the NPs in this study. The ML algorithms included Linear Regression (LR) [14], KNN [15], Support Vector Machines (SVM) [16], Classification and Regression Trees (CART) [17], Random Forests (RF) [18], and eXtreme Gradient Boosting (XGBoost) [19]. Hyperparameter tuning was performed using the grid search method to develop an ML model that predicts the target variable with minimal error. Following the hyperparameter tuning, the top 25 features were selected based on their importance scores derived from RF models trained independently for each dependent variable. After feature selection, hyperparameter tuning was performed once more for each model, and the models were rebuilt.

C. Model Performance Evaluation

Two different performance metrics, root mean square error (RMSE) and coefficient of determination (R^2), were employed to assess and compare the developed models. The model's performance was evaluated using the training set, on which 5-fold cross-validation (CV) was applied as internal validation, and the test set for external validation.

D. Implementation Environment

Data preprocessing methods, development of ML models, hyperparameter tuning, and performance evaluation of the models were implemented in Python 3.12.7 in conjunction with the NumPy, Pandas, Scikit-Learn, XGBoost, GridSearchCV, SciPy, and scikit-posthocs libraries. Jupyter Notebook was used as an Integrated Development Environment (IDE). The computational setup consisted of a macOS operating system and 16 GB of RAM.

III. RESULTS

A. Model Development and Evaluation Overview

The Nano-Tumor Database [9] was used in this study. This dataset included 12 features, such as cancer treatment scenarios, NP characteristics, and tumor-related features. The study also included four target variables: DE_24, DE_168, DE_Max, and DE_Tlast. These target variables were estimated by Chen et al. in previous studies [9] using PBPK models based on the Nano-Tumor Database. Six ML algorithms were used for each target variable, and a total of 24 models were developed following data preprocessing, which involved detection of outliers, imputation of missing values, analysis of rare categories, feature selection, and scaling. The optimized hyperparameters for the ML models were determined by 5-fold CV. The final models were evaluated using both the training and test datasets. RMSE and R^2 were employed to evaluate the models' performance. The study workflow is illustrated in Figure 1.

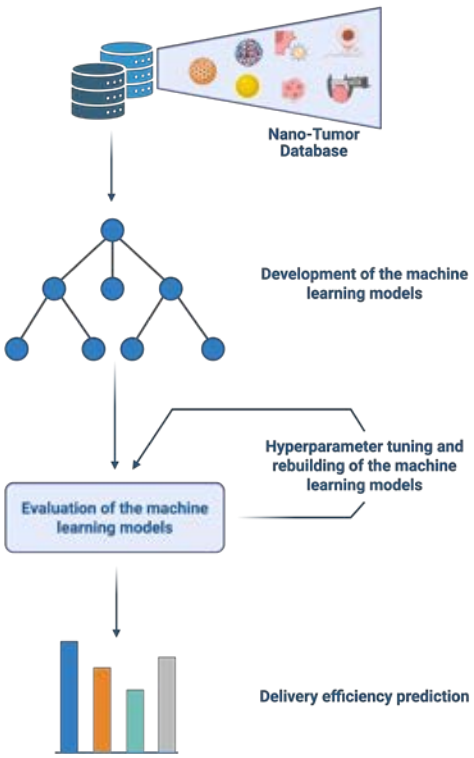


Fig. 1. Overview of the study framework to develop ML models to predict nanoparticle DE to the tumor site in mice. Created in https://BioRender.com.

B. *Training and Testing Results of Machine Learning Models*

Following data preprocessing, the ML algorithms were trained. The prediction performance of the ML models was evaluated utilizing R^2 and RMSE metrics. To assess the consistency of model performance, 5-fold CV was conducted on the training set. The models' final performance was determined according to the test set results, which reflect the model's generalization capability. Among the developed ML models, XGBoost achieved superior predictions with higher R^2 and lower RMSE values for the DE_24, DE_168, and DE_Max target variables, while the RF model achieved better predictions with higher R^2 and lower RMSE for the DE_Tlast target variable. In the test dataset, the R^2 and RMSE values for the XGBoost model ranged from 0.31 to 0.56 and from 1.43 to 3.42, respectively, across all target variables. In the test dataset, the R^2 and RMSE values for the RF model ranged from 0.43 to 0.52 and from 1.45 to 3.45, respectively, across all target variables. Conversely, the LR and CART models showed the poorest performance. In the test dataset, the R^2 and RMSE values for the LR model ranged from 0.02 to 0.25 and from 1.68 to 4.83, respectively, across all target variables. In the test dataset, the R^2 and RMSE values for the CART model ranged from 0.00 to 0.36 and from 1.72 to 4.08, respectively, across all target variables. The results of the performance metrics for these models are detailed in Table 2.

TABLE II. 5-FOLD CV AND TESTING RESULTS FOR TUMOR DE USING VARIOUS ML MODELS

	DE_24		DE_168		DE_Max		DE_Tlast	
Model	***5-fold CV***	***Test***	***5-fold CV***	***Test***	***5-fold CV***	***Test***	***5-fold CV***	***Test***
LR								
RMSE	2.65	4.02	1.49	1.68	4.23	4.83	2.99	3.73
σ	± 0.42		± 0.24		± 0.64		± 0.84	
R^2	0.08	0.13	0.33	0.25	0.13	0.03	0.08	0.02
KNN								
RMSE	2.38	3.82	1.50	1.57	3.85	3.73	2.96	3.52
σ	±0.33		± 0.28		± 0.65		± 0.87	
R^2	0.25	0.21	0.31	0.34	0.29	0.42	0.09	0.12
SVM								
RMSE	2.43	4.14	1.47	1.60	4.26	4.47	2.79	3.57
σ	± 0.35		± 0.25		± 0.86		± 1.02	
R^2	0.23	0.08	0.34	0.32	0.15	0.17	0.23	0.10
CART								
RMSE	2.61	4.08	1.55	1.72	4.21	3.92	3.08	3.76
σ	± 0.30		± 0.36		± 0.77		± 0.75	
R^2	0.09	0.10	0.27	0.20	0.15	0.36	0.00	0.00
RF								
RMSE	2.19	2.98	1.29	1.45	3.27	3.45	2.58	2.83
σ	± 0.36		± 0.24		± 0.65		± 0.92	
R^2	0.36	0.52	0.49	0.44	0.49	0.50	0.33	0.43
XGBoost								
RMSE	2.21	2.87	1.28	1.43	3.46	3.42	2.57	3.11
σ	± 0.31		± 0.24		± 0.69		± 0.87	
R^2	0.34	0.56	0.49	0.46	0.42	0.51	0.34	0.31

C. *Feature Importance*

The feature importance values were extracted from the RF model to evaluate the impact of input features on model predictions [18]. As depicted in Figure 2, zeta potential, NP size, injected dose, tumor weight, and tumor size had the greatest impact on all target variables. Even if the significance levels of these five variables varied across target variables, they contributed more to the final models than other variables. Overall, the cumulative predictive contribution of these variables to the target variables ranged from 52% to 61%.

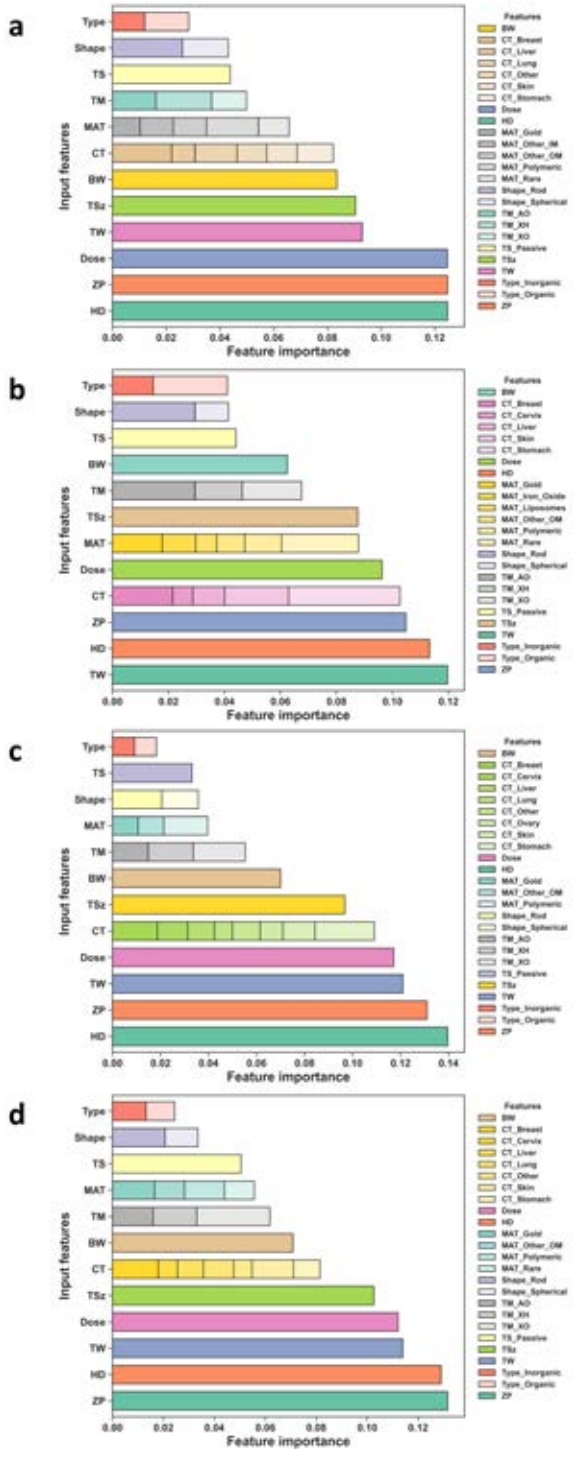


Fig. 2. Importance percentage in the RF model for each feature variable. (a–d) represent results for the feature variable of DE_24, DE_168 DE_Max, and DE_Tlast, respectively. Individual importance is represented by colors in the stacked bars. Abbreviations: ZP, zeta potential; Type, type of nanoparticles; TS, targeting strategy; TM, tumor model; HD, hydrodynamic size; Shape, shape of nanoparticles; MAT, core material of nanoparticles; CT, cancer type; AO, allograft orthotopic.; XH, xenograft heterotopic; XO, xenograft orthotopic; Dose, injected dose; TSz, tumor size; TW, tumor weight; BW, subject body weight.

IV. Discussion

This study presents a PBPK-based ML model that can predict the DE of NPs to tumors in mice, relying on various parameters, including the cancer type, the tumor model, and the physicochemical properties of NPs. The model's predictive performance has room for improvement. Nevertheless, the findings indicate that it has the potential to be an instrument that can help the development of novel NP-based drug formulations for cancer therapy. This tool could, in the future, be developed to avoid including nanoparticles with low tumor DE in preclinical studies, thus facilitating the advancement of nanomedicine research and reducing the need for animal experiments. This study additionally demonstrates the integration of ML approaches with PBPK modeling in the cancer nanomedicine field.

The enhanced delivery of NPs has the potential to mitigate adverse effects in non-target tissues and to reduce the total dose required for treatment [20]. However, the low DE of NPs to tumors poses a significant obstacle in cancer nanomedicine. Previous studies revealed the existence of critical factors affecting DE [11]. Studies were performed to ascertain the relationship between physicochemical properties of NP and tumor DE. Multiple regression analysis [11], ML, and artificial neural networks were used in these studies [12]. In the present study, we developed six ML models for each target variable and compared their predictive performance (Table 2). Among these, the RF and XGBoost models, both of which are tree-based algorithms, demonstrated better performance compared to the other models. These findings reveal that tree-based ensemble models can more successfully predict tumor DE based on physicochemical and tumor-related properties in this dataset. This superiority may be attributed to their ability to capture nonlinear relationships and model complex feature interactions more effectively than linear or single-tree methods; however, the performance and generalizability of the models remain constrained because of the relatively small sample size of the dataset, which may also introduce potential biases.

This study, in line with existing literature [12, 13], suggests that integrating ML approaches with PBPK modeling can significantly enhance nanomedicine research. Given the widespread use of PBPK modeling in various biomedical applications [12], including drug discovery and development, this study could serve as an example for similar integrative approaches. These findings also highlight a potential contribution to the clinical translation of nanomedicine through prioritizing nanoparticle formulations with higher tumor delivery efficiency.

V. Conclusion

In this study, the ML model was developed to predict the tumor DE of NPs. This study points out the potential of using ML-based approaches with PBPK models to enhance cancer nanomedicine research. In future studies, the dataset can be expanded to improve the model's generalization ability and achieve better results. Additionally, deep neural network models can be developed to capture subtle patterns that tree-based ensemble machine learning models might miss, thereby improving prediction performance. In conclusion, such multidisciplinary approaches can help determine which NP candidates are promising in the preclinical stage and improve design processes in nanomedicine applications. Thus, (i) the number of experimental studies can be reduced, (ii) animal experiments can be better prioritized, and (iii) the efficiency of translational research can be increased.

References

[1] W. Cao, K. Qin, F. Li, and W. Chen, "Comparative study of cancer profiles between 2020 and 2022 using global cancer statistics (GLOBOCAN)," Journal of the National Cancer Center, vol. 4, no. 2, pp. 128-134, 2024.

[2] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 74, no. 3, pp. 229-263, 2024.

[3] Y. Yao et al., "Nanoparticle-based drug delivery in cancer therapy and its role in overcoming drug resistance. Front Mol Biosci 2020; 7: 193," ed: Epub 2020/09/26. doi: 10.3389/fmolb. 2020.00193. PubMed PMID: 32974385, 2020.

[4] F. Danhier, O. Feron, and V. Préat, "To exploit the tumor microenvironment: Passive and active tumor targeting of nanocarriers for anti-cancer drug delivery," Journal of controlled release, vol. 148, no. 2, pp. 135-146, 2010.

[5] L. N. Nguyen et al., "The mechanisms of nanoparticle delivery to solid tumours," Nature Reviews Bioengineering, vol. 2, no. 3, pp. 201-213, 2024.

[6] A. Z. Wang, R. Langer, and O. C. Farokhzad, "Nanoparticle delivery of cancer drugs," Annual review of medicine, vol. 63, no. 1, pp. 185-198, 2012.

[7] D. Fan, Y. Cao, M. Cao, Y. Wang, Y. Cao, and T. Gong, "Nanomedicine in cancer therapy," Signal Transduction and Targeted Therapy, vol. 8, no. 1, p. 293, 2023.

[8] P. M. Giri, A. Banerjee, and B. Layek, "A recent review on cancer nanomedicine," Cancers, vol. 15, no. 8, p. 2256, 2023.

[9] Y.-H. Cheng, C. He, J. E. Riviere, N. A. Monteiro-Riviere, and Z. Lin, "Meta-analysis of nanoparticle delivery to tumors using a physiologically based pharmacokinetic modeling and simulation approach," ACS nano, vol. 14, no. 3, pp. 3075-3095, 2020.

[10] B. Ouyang et al., "The dose threshold for nanoparticle tumour delivery," Nature materials, vol. 19, no. 12, pp. 1362-1371, 2020.

[11] S. Wilhelm et al., "Analysis of nanoparticle delivery to tumours," Nature reviews materials, vol. 1, no. 5, pp. 1-12, 2016.

[12] Z. Lin, W.-C. Chou, Y.-H. Cheng, C. He, N. A. Monteiro-Riviere, and J. E. Riviere, "Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches," International journal of nanomedicine, pp. 1365-1379, 2022.

[13] W.-C. Chou et al., "An artificial intelligence-assisted physiologically-based pharmacokinetic model to predict nanoparticle delivery to tumors in mice," Journal of Controlled Release, vol. 361, pp. 53-63, 2023.

[14] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to linear regression analysis. John Wiley & Sons, 2021.

[15] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE transactions on information theory, vol. 13, no. 1, pp. 21-27, 1967.

[16] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," Advances in neural information processing systems, vol. 9, 1996.

[17] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, Classification and regression trees. Routledge, 2017.

[18] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5-32, 2001.

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785-794.

[20] C. O. Egwu et al., "Nanomaterials in drug delivery: strengths and opportunities in medicine," Molecules, vol. 29, no. 11, p. 2584, 2024.

Actuation and Sensor Issues: Portable CPR Machine

L.Canan Dülger,
İzmir University of Economics
Mechanical Eng. Dept.
İzmir, Türkiye
canan.dulger@ieu.edu.tr

Melahat Kızıl
İzmir University of Economics,
Vocational School of Health
Services, İzmir, Türkiye
melahat.kizil@ieu.edu.tr

B. Karaaslan,T.Gülek, D.Eroğlu,
M.G.Güneş, A.Uzşen
İzmir University of Economics
Mechanical Eng. Dept.
İzmir, Türkiye
bkaraaslan.kb@gmail.com

***Abstract*— A portable CPR (cardiopulmonary resuscitation) B machine is designed as a prototype. A nonlinear viscoelastic model is used to consider the human thoracic properties. Based on the mechanical movements and force-acceleration accruals associated with CPR deployment. The system model is studied for the chest response to CPR by considering the viscoelastic properties of the rib cage and intrathoracic organs. It is then numerically simulated in MATLAB/Simulink. Actuation and sensor integration details are discussed. A DC motor is applied for driving the slider crank mechanism.**

Keywords—CPR machine, Matlab-Simulink, Viscoelastic model, Chest model, slider crank mechanism, DC motors

I. INTRODUCTION

One of the leading causes of death is cardiac arrest. Those whose hearts have stopped may have a higher chance of survival using CPR. In an emergency, CPR is a first aid technique that can be immediately performed. The two most important factors, timing and the intervention quality, determine success in case of survival after cardiac arrest. Timing is the first, early identification and prompt treatment necessary. If the person's cardiac rhythm is detected, CPR or the use of a defibrillator should also be initiated. The Turkish Cardiology Association reports that there are 300,000 heart attacks and 125,000 heart attack-related deaths in Turkey annually. Between 20-30 % of heart attack victims die before they reach the hospital, while between 10-20% die subsequently. The death rate during a crisis's first 24 hours ranges from 30%-50% [1]. The American Heart Association and the European Resuscitation Council recommended that 30 chest compressions and two breaths should be administered initially during cardiopulmonary resuscitation (CPR). The chest should be checked at this point, and inflating it should take a few seconds [2]. Adult CPR requires chest compressions to be given at least 100–120 times per minute, and providing at least 5–6 cm deep [3].

CPR training is essential to provide immediate assistance during cardiac arrest. M. Kızıl has carried out a study on the effectiveness of CPR previously. First aid and basic life support have been performed at Izmir University of Economics. Seven different health programs were included with 153 volunteer students in 2017 [1]. The use of a training manikin is vital for CPR practice. Students can get hands-on practice by using such manikins. CPR training is given to project students from the Mechanical Engineering. Figure 1 shows the volunteer students during CPR training in İEU Laboratories. CPR manikins are used during training. Two hands on the lower half of the sternum are used for adults, teenagers, and children. For infants, the same compression rate will be performed using only two fingers on the lower half of the sternum. This helped to improve the CPR prototype design and its construction details during the project.

Based on the load and motion properties needed for the mechanism, a motor is selected with the required torque and cycle rates. The motor has to generate chest compressions with the corresponding 5 cm deep (approximately) and 100–120 compressions per minute (cpm) were the choice when prioritizing for the selection process. This criterion optimization ensured compatibility between load characteristics, actuation components, and the need for a linear motion, making it valid for mechanical structure. Relevant sensors were included in the project to offer a basis for future system enhancements. Suitable sensors are purchased for pulse and chest compression force. Furthermore, preliminary mathematical modeling studies were performed to understand the system behavior. The responses to different parameter variations are acquired. The simulated model is analyzed. The requirement is for a system to work with CPR scenarios involving adults or teenagers. For the phases of the study (like sensor integration, control algorithms enhancement, mechanical structure, and the prototype manufacturing-3D printing).

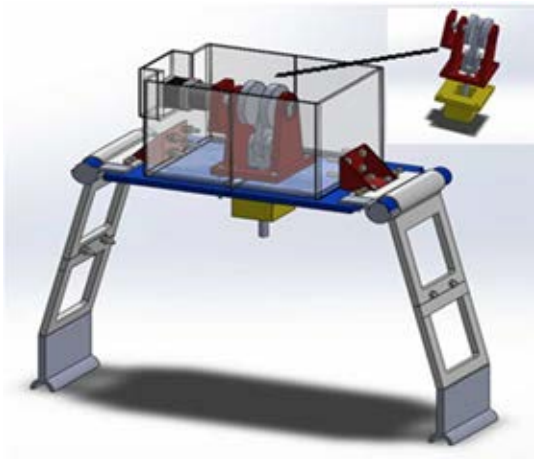


Fig.1 CPR Training-IEU Laboratories

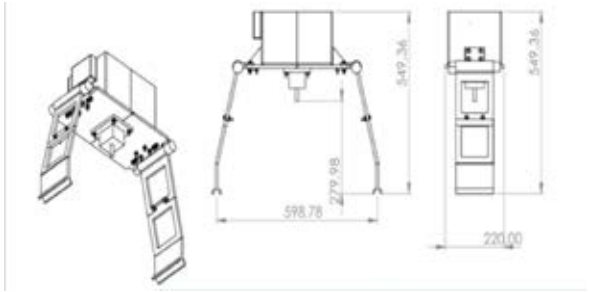
979-8-3315-5565-8/25/$31.00 ©2025

II Background On Proposed CPR

M. Kizil et al. [4] have presented a concept design based on active compression-decompression. Different piston-driven designs are then studied. L.C. Dülger et al. [5] have then offered a slider crank mechanism with detailed dimensions and simulations needed. A prototype has been printed in 3D, and its construction is completed. Figure 2(a) and (b) show the printed prototype using a slider crank mechanism with dimensions [5]. There have been numerous commercial attempts to address this requirement. The LUCAS™ and AutoPulse devices are examples of the two prominent types of MCPR systems on the market today [6]. M.J.Ghafoor et al. [7] and A. Athreya et al. [8] found that DC motors, particularly those that are gear-reduced, generate close human-like chest compressions when combined with slider-crank or cam-driven mechanisms. Kim et al. [9] demonstrated the feasibility of a system for CPR actuation that makes use of a 12V brushed DC motor and a 30:1 gearbox to produce the necessary torque.



(a)



(b)

Fig.2 Referred Prototype -Slider crank mechanism with dimensions [5]

II. Sensors For CPR Design

The development of automatic cardiopulmonary resuscitation (CPR) systems has prioritized the use of embedded sensor technologies in recent years. Human mistakes, exhaustion, or inexperience are the leading causes of issues like inadequate pressure, improper depth, and irregular rhythms that are commonly seen in manual CPR applications. Sensor-supported technologies that provide real-time feedback are designed to improve the quality of CPR and guarantee safety [10]. One of the most widely utilized force-measuring sensors in these systems is the load cell. By continuously and precisely measuring the force applied during compression, it aids in keeping the system within safe bounds. The measurements that load cells offer are crucial because too much pressure can injure the chest, and too little pressure can result in inefficient circulation. Ultrasonic sensors are used to measure the depth of chest compressions without contact. These sensors, which detect the distance between the piston and the chest surface with the help of sound waves, control whether the system provides the required 5–6 cm depth in each compression. [11]. Many CPR prototypes developed today use these three basic sensors together and process the data via a microcontroller. Sensor integration in CPR systems is a critical technology. Multi-sensor systems, where load cells, ultrasonic sensors, and accelerometers are used together, significantly increase the reliability and accuracy of CPR applications in a wide range from education to clinical settings. (Fig. 3)

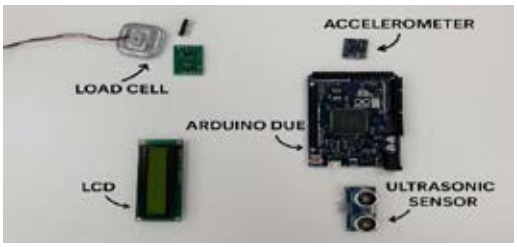


Fig 3. CPR Sensors Used in Our Study

III. Mathematical Model-Chest Compression

Design of CPR systems requires the approximation of thoracic behavior upon compressive load; otherwise, it will not correctly model the geometries and physiological dynamics in a medical compression. The goal is not simply to press on the chest but to simulate and deliver compressions in ways that mirror how the ribcage, sternum, lungs, and airflow systems interact during cardiac arrest scenarios. We review contributions that provide fundamental insights into the mechanical modeling of CPR that present data or methods to be used in our own nonlinear viscoelastic and energy-based formulations.

Ruiz de Gauna et al. [12] introduced a simplified viscoelastic model derived from real chest compressions collected via pre-hospital CPR monitors. Their work focuses on identifying mechanical parameters from actual resuscitation attempts. By computing chest displacement from force and acceleration data, each compression-recoil cycle was parameterized with an elastic stiffness constant k and separate damping coefficients for compression and recoil. J. Moradicheghamahi et al. [13] explored how thoracic geometry affects CPR outcomes using finite element simulations of anatomically variable ribcages. It then models the chest as a mass-spring-damper system. Kao et al. [14] have focused on CPR performance under controlled variations of thoracic stiffness. Their study used a multi-stiffness thorax manikin, in which subjects performed compressions, and the researchers determined depth, flow, and recoil.

A. Viscoelastic Modeling of Thoracic Mechanics for CPR

To accurately replicate the mechanical behavior of the human chest during cardiopulmonary resuscitation (CPR), we

introduced a dynamic model using a nonlinear viscoelastic model. This model is central to the design, simulation, and control of our portable CPR device, which uses a motor-driven slider-crank mechanism to generate cyclic compressions. Our goal was to simulate the thoracic response with sufficient detail to reflect stiffness, nonlinearity, energy loss, and real-time dynamics under the force and displacement cycles characteristic of CPR. A. Jalali et al [15] have performed a study to characterize the mechanical properties of a chest. The model implemented is built using Equation 1:

$$F = F_S + F_d + F_h + F_i \quad (1)$$

This total force F, acting on the sternum, is the total of elastic resistance (F_S), the viscous damping (F_d), the hysteresis damping (F_h), and the inertial response (F_i). Each of these terms reflects a distinct physical part that is important to the actual mechanics of CPR. They form a whole dynamic system that behaves realistically under simulation. The elastic force is defined as:

$$F_S = k_1 X + k_2 X^2 + k_3 X^3 + k_4 X^4 + k_5 X^5 \quad (2)$$

The fifth-order polynomial set above reflects the nonlinearity in thoracic stiffness. The lower-order terms k_1 and k_2 represent the initial compliance of the chest. The higher-order terms; k_3, k_4, and k_5 represent the geometric stiffness that occurs at greater depths. This is particularly important for compressions that are bigger than 3 cm. Because the chest will stiffen more strongly, both due to bone geometry and muscle tension. In the simulations, this equation produces realistic force curves and limits compression depth to the safe values. The viscous damping term is given by:

$$F_d = c_1 \dot{X} + c_2 \dot{X}^2 \quad (3)$$

Energy losses in the chest modeled by this equation depend on the speed. The linear damping, which is proportional to the compression velocity, is the first term that accounts for linear dissipation. The second one describes the nonlinear strain effects. This damping controls how smooth the actuator behaves. The hysteresis damping term is modeled as:

$$F_h = c_h \frac{\dot{X} sgn(\dot{X})}{w} \quad (4)$$

This term added internal energy dissipation in terms of not only velocity but also direction and frequency. The sign function makes sure that damping always takes place in both compression and release phases. The division by ω ensures that at faster rates, the damping weakens slightly, representing real biological hysteresis behavior. This lag is crucial to real-world CPR. It affects how much energy is transferred back into the chest and how well the blood flow is maintained over cycles. The inertial component is given by:

$$F_i = M\ddot{X} \quad (5)$$

Including this equation allows the simulation to correctly model force overshoot and lag during fast actuation. To generalize the model across patient sizes and simplify computational analysis, the model was non-dimensionalized [15]. Table 1 represents scale parameters. The variables were scaled using $x = \frac{X}{X_0}$, $\tau = \frac{t}{T}$, $m = \frac{M}{M_0}$, the system's frequency was similarly transformed: $\Omega = \frac{\omega}{\omega_n}$, $\omega_n = \frac{2\pi}{T}$.

TABLE I PARAMETERS FOR THE CHEST MODEL

Parameter (Unit)	Value
X_0 (meter)	0,05
T (second)	0,06
M_0 (kg)	3,0

Thus the resulting dimensionless model is:

$$f = f_S + f_d + f_h + f_i$$
$$f_S = \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4 + \alpha_5 x^5$$
$$f_d = \beta_1 \dot{x} + \beta_2 \dot{x}^2$$
$$f_h = \frac{\beta_h \dot{x}.sgn(\dot{x})}{\Omega} \qquad f_i = \ddot{x} \quad (6)$$

These equations were implemented in MATLAB to simulate the thoracic response under varying compression cycles. The model's coefficients were selected from validated research using porcine subjects due to their mechanical similarity to human chest geometry. The parameters used were [15]: $\alpha_1 = 1.51$, $\alpha_2 = -0.12$, $\alpha_3 = 12.38$, $\alpha_4 = -16.99$, $\alpha_5 = 10.94$, $\beta_1 = 0.08$, $\beta_2 = 1.15$, $\beta_h = 0.17$. The parameters were estimated for two periods (until 10 min. after CPR starting, and after 10 min. from CPR starting)

B. Simulation Results

In simulations, the force–compression depth plot produced a clear and physiologically realistic hysteresis loop that strongly resembled the patterns seen in actual CPR data. As compression depth increased, the model generated a nonlinearly rising force curve, while the release path followed a lower-energy trajectory. The graph is beneficial as changes can be made in the crank radius, linkage ratio, and compression limits in the slider-crank mechanism used. (Fig.4). We then calculated the width and vertical amplitude of the loop to verify that our mechanical output would fall within the clinically accepted range of 5–6 cm at around 400 N maximum force. Any deviations beyond that would require redesigning the mechanical limits or motor torque. The plot also helped us to confirm that our system would not introduce sharp force spikes or unsafe recoil behavior during release, an essential consideration for rib safety.

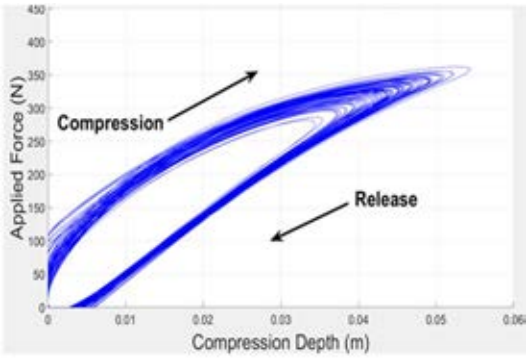


Fig. 4 - Simulated Force vs. Compression Depth -Thoracic Stiffness Behavior

IV. Actuation Details For CPR System

One of the initial and most critical steps in the actuation system design of our portable CPR machine was determining the most suitable motor type that could reliably generate repetitive, high-force compressions in line with AHA guidelines [3]. Brushed DC motors are the ideal compromise between high torque output at low voltage, PWM speed control compatibility, and high-volume/low-cost. A 12V brushed DC motor of a 30:1 planetary gearbox is used. This motor has an approximate no-load speed of 1300 RPM (13-14 Nm at output after gear reduction). The mechanism will provide mechanical leverage and control through the compression phase, which could resemble the proper 5-6 compression depth for CPR. The motor is integrated into the mechanical frame, with a 3D-printed coupling linking to the compression unit. The entire system is controlled by means (force load cell, ultrasonic sensor for depth) and processed in real time by the Arduino Due. These features allow for the editing of energy delivery during compression depth, rate, and force to prevent over-pressure and keep optimal performance [16-17]. Figure 5 shows side and top view photos from the designed prototype. Here, the motor and power system are activated. Sensor box design is completed, but not completely activated.



Fig.5. Designed Prototype with Actuation and Sensors

V. Conclusions

In this work, the development of a CPR (Cardiopulmonary Resuscitation) device has been nearly completed. A control hardware control was developed to supervise the motor's actuation. The system viscoelastic modeling is performed using MATLAB/Simulink. The sensors, an accelerometer, an HCSR-04 Ultrasonic sensor, and a load cell were chosen. Nevertheless, these sensors are not yet interconnected; however, full integration is planned in the subsequent stages of the project. Currently, the system can operate manually at an adjustable speed. It performs basic compression motions in the required range.

A nonlinear viscoelastic thoracic model is applied. The nonlinear viscoelastic model was also able to describe well the complex mechanical response of the chest wall to repeated compressions. The model enabled us to confirm that the force output fell below the 500 N safety limit while maintaining AHA-compliant compression depths of 4–6 cm. The results verified that a 12V brushed DC motor with a 30:1 planetary gearbox delivers adequate torque for safe and effective compression, and that the crank-slider configuration translates this force into stable cyclic motion. It can be an affordable model with sensors for CPR training.

Acknowledgment

The study was conducted by Izmir University of Economics, Faculty of Engineering. It was carried out within the framework of FENG 498 projects between Dept. of Mech. Eng. and Vocational School of Health Services-Paramedics. We also thank I. Kulali for his contribution to viscoelastic modelling.

References

[1] M. Kızıl, Determining the effectiveness of Cardiopulmonary Resuscitation using a manikin CPR, December 2022, Hastane Öncesi Dergisi, 7(3), DOI: 10.54409/hod.1150760, pp.303-316.
[2] D.A. Zideman, EDJ De Buck, EM Singletary, P. Cassan, A.F. Chalkias, TR Evans, C.M. Hafner, P.G. Vandekerckhove, European Resuscitation Council Guidelines for Resuscitation 2015, S.9, First aid Resuscitation, 95, 278-287
[3] RW Neumar, M. Shuster, CW Callaway, LM Gent, DL Atkins, F. Bhanji, SC Brooks SC, AR de Caen, MW Donnino, JM Ferrer, ME Kleinman, SL Kronick, EJ Lavonas , MS Link, ME Mancini, LJ Morrison, RE O'Connor, RA Samson, SM Schexnayder, EM Singletary, EH Sinz, AH Travers, MH Wyckoff, MF Hazinski, Part 1: Ex.Summary:2015 American Heart Ass. Guidelines, Cardiopulmonary Resuscitation and Em. Cardiovascular Care, Circulation. 2015, Nov 3; 132, pp. 315-67.
[4] M.Kızıl and L.C.Dülger, Design of a Device for CPR (Cardiopulmonary Resuscitation, Tıptekno 2022, 31 Ekim-2 Kasım 2022, Belek-Antalya.
[5] LC.Dülger, M.Kızıl, M.Atçeter, D.D.Demircioğlu, D.Yiğit, B.Güneş Design and Construction of a Portable CPR Machine, Tıptekno 2024, 10-12 Ekim 2024, Gümüşlük/Bodrum-Muğla.
[6] https://ww.lucas-cpr.com, Assessed 03/08/2024
[7] M. J. Ghafoor, M. Mujeeb-U-Rahman, M. Jamal, and A. Ahmed, "Low-Cost and Portable Cardiopulmonary Resuscitation Machine," Dept. of El. Eng., Information Technology University, Lahore, Pakistan,
[8] A. Athreya, S. Gupta, and R. Bansal, "Design and Development of an IoT-Based CPR Delivery System," J. Med. Devices Eng., V45, no. 3, pp. 210–218, 2023.
[9] J. Kim, Y. Park, and T. Seo, "Low-Voltage CPR Actuation Mechanism Using Brushed DC Motors," IEEE Trans. Biomed. Circuits Syst., V13, N. 2, pp. 344–351, 2019.
[10] Xie, X., Wu, X., and Liu, Y. Design and Evaluation of CPR Emergency Equipment for Non-Professionals. Sensors, 23(13), 5948. MDPI, 2023
[11] G.Rugumira, ME Mkiramweni, N.T. Msangi, 'Design and Implementation of Cardiopulmonary Resuscitation Simulation Model, Tanz. J. Eng. Tech. J., V41-No.4, Dec 2022, pp 133-147.
[12] S. Ruiz de Gauna , JJ. Gutiérrez, CL. Sandoval, JK. Russell, I. Azcarate I, JA Urigüen, DM González-Otero, MR. Daya, 'Characterization of mechanical properties of adult chests during pre-hospital manual chest compressions through a simple viscoelastic model'. Comput Methods Programs Biomed. 2023 Dec; 242:107847. doi: 10.1016/j.cmpb.2023.107847. Epub 2023 Oct 10. PMID: 37852146.
[13] J. Moradicheghamahi, G. Fortuny, JM. López,, D. Puigjaner, J. Herrero, Y. Azeli. The effect of thoracic dimensions on compression depth during cardiopulmonary resuscitation. Int J Numerical Methods Biomed Eng. 2023, Jul;39 (7):e3718. doi: 10.1002/cnm.3718. Epub 2023-May 11. PMID: 37166223.
[14] Chia-Lung Kao, Jui-Yi Tsou, Ming-Yuan Hong, Chih-Jan Chang, Fong-Chin Su, Chih-Hsien Chi, 'Effect of thoracic stiffness on chest compression performance- A prospective randomized crossover observational manikin study', Heliyon, V8, I10, 2022, e10990, ISSN 2405-8440.
[15] A. Jalali, V. M. Nadkarni and C. Nataraj, "Modeling mechanical properties of the chest during the cardiopulmonary resuscitation procedure," Computing in Cardiology 2014, Cambridge, MA, USA, 2014, pp. 13-16.
[16] Y. Xie, L.Zhang, and W. Chen, "Sensor-guided Feedback Systems in CPR Training and Application," Sensors, vol. 23, no. 5, pp. 2244–2255, 2023.
[17] T. Iwata, K. Yamashita, and M. Nishimura, 'Development of chest compression monitoring system using accelerometer'. J. of the Japanese Society of Intensive Care Medicine, 2013, 20(3), 179–183.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Portable Plasmonic Biosensor for Simultaneously Detection of Neisseria gonorrhoeae and Chlamydia trachomatis in Urine Samples

Furkan Kocer*†, Meryem Beyza Avci*‡, Arif E. Cetin*§
*Izmir Biomedicine and Genome Center, Izmir 35340, Turkiye
†Department of Biomedical Engineering, Faculty of Engineering and Architecture, Izmir Bakircay University, Izmir, Turkiye
‡Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Balcova, Izmir 35340, Turkiye
§Department of Biophysics, Faculty of Medicine, Dokuz Eylul University, Balcova, Izmir 35340, Turkiye

***Abstract*—Neisseria gonorrhoeae (NG) and Chlamydia trachomatis (CT) remain major causes of sexually transmitted infections (STIs) case worldwide. Rapid and easy to use diagnostic tools will play a crucial role in controlling the spread of the disease. This study presents the introduction of a portable biosensor that has the ability to detect CT and NG in urine samples. The biosensor combines plasmonic nanohole detection with lens-free imaging, allowing for diagnosis at the point of care. It captures images of the diffraction field to detect bacteria. A piezoelectric pump and flow cell enable accurate sample delivery. The whole setup, testing, and image analysis are controlled by graphical user interface. The utilization of control and sensor antibodies in various configurations on the sensor yields the intended level of sensitivity and specificity. Simultaneous NG and CT testing can be conducted independently due to the use of microfluidic channels, which require less battery testing and offer a high-throughput capability. Tests on the control are conducted utilizing plasmonic pixels to minimize background noise. The limits of detection for NG and CT, respectively, using our point-of-care biosensor, are 452 and 121 CFU/mL. This makes it practical, cost-effective, and efficient in bridging the accessibility and affordability gap for STI diagnostics.**

***Keywords*—plasmonic biosensor, sexually transmitted infections, Neisseria gonorrhoeae, Chlamydia trachomatis, point-of-care diagnostics.**

I. Introduction

One of the most prevalent health concerns on a global scale is the transmission of infections through sexual contact. [1]. Chlamydia, gonorrhea, syphilis, and HIV comprise the most prevalent sexually transmitted infections [2]. Neisseria gonorrhoeae (NG) and Chlamydia trachomatis (CT) are bacterial infections among sexually transmitted diseases (STIs), with around one million new cases reported daily. Despite effective treatments, many infections are asymptomatic, leading to severe complications like infertility, pelvic inflammatory disease, ectopic pregnancies, and chronic pelvic pain [3]. Current diagnostic methods, such as nucleic acid amplification tests (NAATs), require specialized laboratories and sample preparations, making them unsuitable for resource-poor settings [4]. There is a need for simultaneous, easy-to-use, and affordable diagnostic tests for NG and CT that can be used in clinics to improve early detection and treatment access while reducing infection spread. In response to this need, we developed a nanohole array-based plasmonic biosensor that is portable and lens-free, utilizing computational imaging to detect NG and CT in urine samples [5]. This system captures and analyzes diffraction field images, ensuring precise sample delivery via a low-cost pump and flow cell. The setup and image analysis are controlled by software, with control and sensor antibodies ensuring high sensitivity and selectivity. Independent microfluidic channels enable simultaneous NG and CT testing, enhancing throughput and reliability. Plasmonic pixels reduce noise levels, and simple, inexpensive optics support system feasibility. The detection limits for NG and CT are 452 CFU/mL and 121 CFU/mL, respectively, making this biosensor highly efficient for point-of-care diagnostics. This approach, leveraging a camera-based sensing scheme, allows for simultaneous detection at higher throughput, providing better sensitivity and reliability in detecting both bacterial strains in urine.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

II. Material and Methodology

A. Plasmonic Biosensor

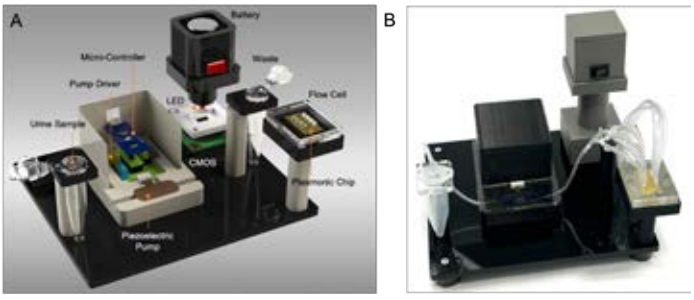


Fig. 1. (A) Schematics and (B) image of the biosensor.

Figures 1A and B present the schematic and photograph of the plasmonic system. We have designed a 3D-printed housing for complete integration of all components of the plasmonic system such that the effect of ambient disturbances and background noise is minimized to the greatest possible extent.

This configuration is comprised of four main subsystems: plasmonic chip (i), sample preparation module (ii), imaging module (iii), and software module (iv). The plasmonic chip features a thin layer of gold film embedded with nanohole lattices, which are modified by attaching antibodies specifically targeting NG and CT bacteria. The imaging module uses an LED light source to illuminate the plasmonic chip, while a CMOS camera captures the resulting diffractive field patterns. Variations in brightness indicate the presence of microorganisms. The sample preparation module includes a flow cell fabricated with Polydimethylsiloxane (PDMS) and a piezoelectric pump for precise sample delivery. This module divides the plasmonic microarray into four different sections: two for detecting CT and NG, and two for control purposes—one with a control antibody and one left unmodified. The sensor sections have a single layer of bioreactive material containing specialized antibodies for NG and CT detection. The control regions are designed to detect changes caused by optical, mechanical, or biological factors. The software module features a GUI that effectively controls and monitors the operations of the optical and microfluidic hardware components. Additionally, it is equipped with an image-processing method that analyses plasmonic diffraction field images in order to detect bacteria present in urine samples. A computer controls the system, establishing a connection between the imaging and sample preparation equipment by USB.

B. Plasmonic Chip

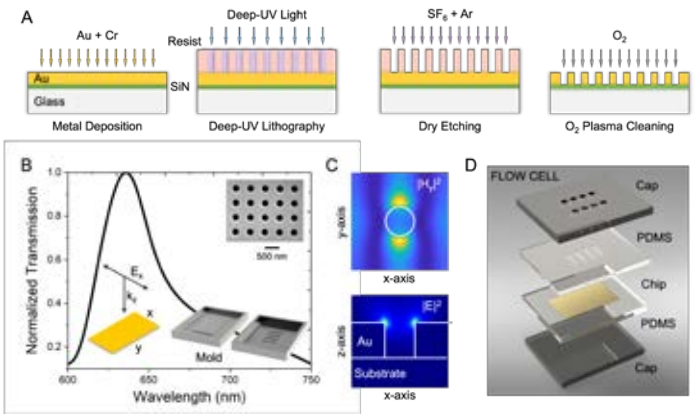


Fig. 2. (A) Fabrication process. (B) The polarization and propagation directions from FDTD simulations are illustrated in the insets of the transmission spectrum of the nanohole arrays. Furthermore, a Scanning electron microscopy (SEM) image of the nanohole arrays and a schematic of the mold are included. (C) The y-component of the magnetic field at the surface of the Au film (top). Cross-sectional profile of the electric field distribution (bottom). (D The flow cell.

The plasmonic chip is made of a periodically patterned nanohole with a diameter and periodicity of 200 nm and 550 nm, respectively. Figure 2A schematically represents the fabrication procedure. The nanoholes are patterned onto a gold (Au) film with a thickness of 120 nm using the deep-UV lithography method, applied through a high-throughput technique. A 4-inch fused silica wafer, 500 µm thick and covered with an LPCVD-SiN layer with a thickness of 100 nm, is utilized as the substrate. An adhesion layer of 5 nm chromium (Cr) was placed onto the SiN coating as the first step in the fabrication process, followed by a 120 nm layer of gold. After that, the gold surface was treated with a positive photoresist layer, and deep-UV lithography was carried out. To create the nanoholes, a gas combination of SF6 and Ar was used to dry-etch the gold film following the lithography process. Residual photoresist was removed with oxygen (O_2) plasma cleaning. Finally, the wafer was segmented into 2 cm × 1 cm plasmonic chips. Figure 2B inset shows a scanning electron microscope (SEM) image of the nanohole arrays. Symmetrical wave patterns observed along the x-axis indicate the presence of counter-propagating plasmons, whereas the concentrated areas of high intensity at the aperture rims are caused by localized plasmons. The distribution of the electric field (Figure 2C-bottom) shows concentrated increases in the nanohole rims, allowing analytes on the gold surface to access these fields. The nanoholes' circular symmetry justifies the utilization of an unpolarized LED, ensuring uniform field concentration and increased sensitivity in comparison to polarized light.

C. Sample Preparation Module

A piezoelectric pump and a flow cell are the two main components of the sample preparation module. The piezoelectric pump (mp6-liq, Bartels Mikrotechnik GmbH) is controlled by an Arduino microcontroller and an mp6-EVA driver board, as shown in Figure 1B. The driver board controls the flow rates by adjusting the pump's voltage (ranging from 1.0 to 4.5 V) using pulse-width modulation (PWM). Operating at a steady frequency of 100 Hz, the pump maintains a stable relationship between voltage and flow rate. To make sure the analytes were consistently delivered for both the control and sensor tests, a flow cell that had four channels was used. Two control channels minimize noise from optical, mechanical or biological variations. In the dual-layer PDMS construction of the flow cell, the channels are situated on the upper layer, while the plasmonic chip resides in a precisely engineered groove on the lower layer. The molds for these PDMS layers, depicted in the inset of Figure 2B, were fabricated using advanced resin-based 3D printing techniques. To facilitate reusability, we engineered a sophisticated microfluidic system incorporating a holder with dual acrylic caps, thereby obviating the need for PDMS bonding and ensuring a leak-proof assembly. Figure 2D showcases this design, which includes eight holes in the top cap for inlets and outlets, and four screws for stability. The flow cell's four channels deliver urine samples containing NG or CT bacteria to designated regions on the plasmonic chip. For precise bacterial immobilization, we employed flow-cell-assisted functionalization. Molecular printing technology was used to modify plasmonic pixels for control and sensor testing. The sample preparation module facilitated the application of a BSA (Bovine Serum Albumin) coating. The piezoelectric pump gradually delivered the bacteria and BSA over twenty minutes. This procedure involved the binding of protein A/G,

the attachment of antibodies, the coupling of BSA, and the immobilization of bacteria. Afterward, the plasmonic chip was positioned, and the CMOS camera was set up to capture images and analyze the data.

D. Imaging Module

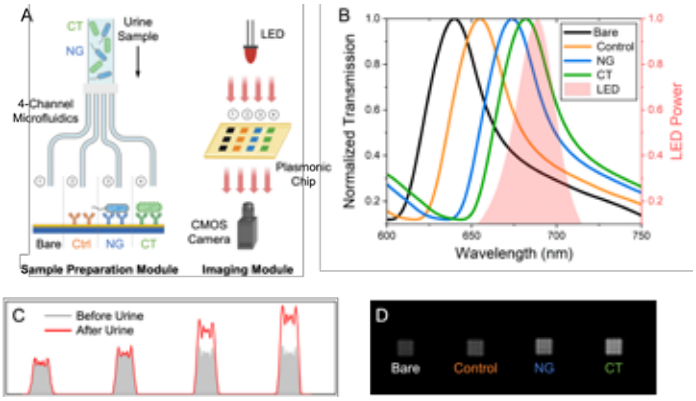


Fig. 3. (A) The sample preparation and imaging modules of the handheld biosensor. Created using BioRender.com. (B) The black curve illustrates the transmission resonance of nanohole arrays functionalized with A/G protein, the orange curve represents the control antibody, the blue curve depicts 10^6 CFU/mL NG, and the green curve corresponds to CT bacteria. The red curve shows the power spectrum of the LED source used in the imaging module. (C) Changes in the diffraction field intensity before (gray line) and after (red line) the plasmonic pixels were exposed to a urine sample containing NG and CT bacteria. (D) A CMOS camera image displaying square-shaped plasmonic pixels: orange represents the control antibody, blue represents NG bacteria, green represents CT bacteria, and bare represents protein A/G.

The portable biosensor utilizes a CMOS camera (Thorlabs, DCC1545M) and LED illumination to capture diffraction field images from the plasmonic chip (Fig. 3A). CMOS camera tracks intensity changes before and after analyte binding to detect spectral fluctuations in nanohole transmission resonance. The detection limit means to the minimum concentration of bacteria that can be detected by observing the changes in diffraction field intensity. Figure 3A depicts the imaging module. The lower PLA component houses the CMOS camera that is powered by USB, while the upper PLA component houses the LED light source that is powered by a battery. An inter-PLA tube provides uniform LED lighting. The upper component comprises a battery, switch, and resistor, which serve the purpose of regulating LED brightness and preventing saturation. The lower section of the device securely holds the plasmonic chip in a customized groove. The alignment of the sensor and control sections of the chip with the active area of the CMOS camera, also referred to as "plasmonic pixels," is precisely aligned. The light source must have a narrow bandwidth near the nanohole transmission response, and the CMOS camera requires high quantum efficiency in this range. Following the functionalization of protein A/G, a shift of 4 nm in transmission resonance was observed. This shift positioned the LED source at approximately 688.65 nm, which is 33 nm away from the nanohole resonance (Fig. 3B). This alignment enables spectral overlap during the steps of functionalization. The narrow bandwidth of the LED (∼28 nm) combined with the consistency of quantum efficiency of CMOS ensures accurate detection of diffraction field intensity changes, even at low analyte concentrations. Successful bacterial binding and the multiplexed detection capabilities of our biosensor for NG and CT bacteria are demonstrated by Figures 3C and 3D, which illustrate the various intensity values and spectral shifts.

E. Software Module

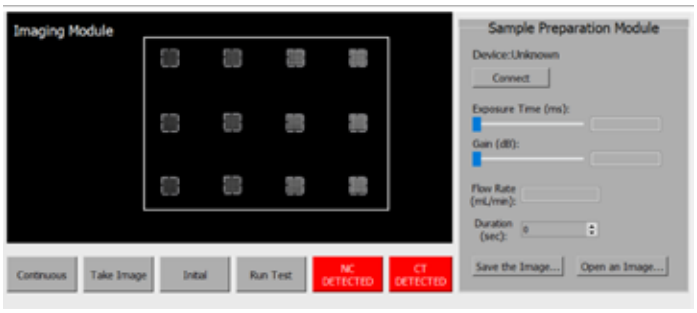


Fig. 4. GUI of the biosensor kit.

A graphical user interface (GUI) based on Python™ was created to assess the plasmonic chip's performance and transmit sensing data through image processing (Fig. 4). The graphical user interface (GUI) enables users to adjust camera and piezo pump settings, allowing precise customization of test conditions. The plasmonic chip is initially coated with protein A/G and IgG, and then positioned in the flow cell to allow the flow of protein BSA and urine. The chip is placed in front of the CMOS camera after functionalization is complete. The graphical user interface (GUI) shows the diffraction image and computes the average diffraction intensity. The GUI utilizes calibration data to predict the ranges of detection, while also normalizing these outcomes with respect to the control antibody. The system calculates the ratio of the intensity of the diffraction field captured by the sensor to that of the control. It then alerts for the presence of bacteria if the calculated value falls under pre-set thresholds. By concentrating on the diffraction intensity within the specified area of interest, this approach minimizes external disruptions and guarantees data accuracy. Consequently, it is suitable for field use without the necessity of highly trained personnel.

III. Results

A. Detection of NG and CT Bacteria with Plasmonic Imaging

Using anti-NG and anti-CT antibodies (Abcam) targeting the outer membrane proteins of NG and CT, respectively, we developed a method for the identification of bacteria in urine samples. To achieve precise bacterial capture, we utilized a protein A/G fusion to form a functional layer on the detection surface, optimizing the antibody configuration. While culturing CT required host cells, NG was grown on Thayer-Martin agar. For device optimization, the inactivated bacteria were used, which still maintained their membrane integrity for effective capture. By functionalizing plasmonic pixels with protein A/G and IgG and coating other areas

with BSA, we ensured targeted binding of only the bacteria. Four channels characterized our microfluidic system: baseline, control antibody, CT-specific, and NG-specific, which gave us the independence to test for specificity and independent detections. CMOS images of the results depicted prominent spectral shifts at the sensor regions with bacteria, hence successful detection. The NG and CT channels diffraction field intensity increased significantly due to bacterial attachment, unlike baseline and control channels that registered minimal changes. A larger spectral shift in CT, which means more pronounced increase in intensity, proved high sensitivity of the biosensor. These results demonstrated that our biosensor can identify NG and CT in urine samples with high sensitivity and specificity. The platform is ideal for point-of-care diagnostics, as it distinctly differentiates intensity values and spectral shifts to confirm bacterial presence.

B. System LoD for NG and CT Bacteria

The calibration standard curves were constructed through the use of a serial dilution method and tenfold dilutions ranging from 10 to 10^6 CFU/mL for the spiked samples of NG (Fig. 5A) and CT (Fig. 5B). A high concentration, 50 µg/mL, of protein A/G, as well as antibodies, was used so that the bacteria would be effectively captured on their surface even when such low concentrations are present. The sensitivity and specificity of our designed biosensor were evaluated by targeting the detection of CT and NG bacteria in synthetic urine, mimicking real urine conditions. Each microfluidic channel, functionalized with the same capture agent, contained three plasmonic pixels. The figure presents data from three replicates for each concentration, measured by the three plasmonic pixels. Error bars indicate twice the standard deviation. The obtained intensity values from the diffraction field showed a rising tendency with higher bacteria concentrations, which matched very well with the specific binding reaction between bacteria and their relevant antibodies. Linear regression modeling yielded R^2 values of 0.99474 for NG and 0.99987 for CT, establishment of LoD at 452 CFU/mL in the case of NG, and 121 CFU/mL for CT. Figure 5C shows the diffraction field intensity values for urine samples with 10^5 CFU/mL of only CT bacteria, only NG bacteria, and both bacteria (left, middle, and right plots, respectively). In each experiment, three plasmonic pixels within the same channel were measured. The bars represent the average values, accompanied by error bars indicating twice the standard deviation. The sensor successfully differentiated bacterial presence without any response from the control antibody or the antibodies against other bacteria. This imaging-based readout demonstrated no nonspecific binding, even at high concentrations, confirming the bio-sensor's strength and reliability in detecting NG and CT in urine samples.

IV. Conclusion

This portable biosensor was developed to detect CT and NG in urine samples. It label-freely integrates plasmonic nanohole sensing with lens-free computational imaging, thereby enabling accurate diagnosis at the point of care. An inexpensive pump and flow cell deliver samples to the plasmonic chip with high precision, while integrated software manages hardware control, testing, and bacterial detection. The system facilitates simultaneous detection of CT and NG through two microfluidic channels. Its high throughput, simple optics, and low-cost parts make the biosensor practical and robust. This handheld device detects NG with a limit of 452 CFU/mL and CT with a limit of 121 CFU/mL.

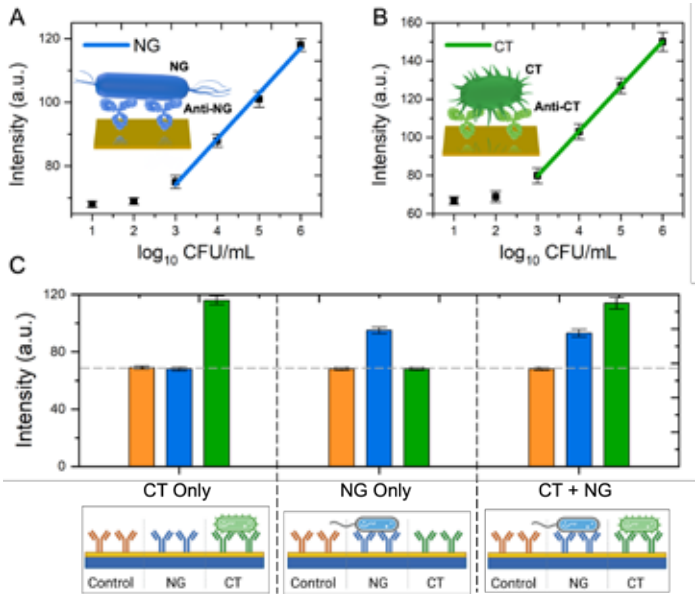


Fig. 5. (A) Calibration graphs for NG bacteria and (B) Calibration graphs for CT bacteria. (C) The intensity values of the diffraction field for urine samples containing only CT bacteria are represented by the green bar on the left. The blue bar in the middle denotes urine samples with only NG bacteria, while the blue and green bars on the right indicate urine samples containing both bacteria. The intensity values are measured at a concentration of 10^5 CFU/mL. The control antibody is represented by the orange bars. Each bar represents the mean value obtained from three separate trials, with error bars indicating twice the standard deviation. Produced using BioRender.com.

References

[1] S. Tuddenham, M. M. Hamill, and K. G. Ghanem, "Diagnosis and treatment of sexually transmitted infections," JAMA, vol. 327, pp. 161-172, 2022. doi: 10.1001/jama.2021.23487.

[2] M. T. Vallejo-Ortega, H. Gaitán Duarte, M. B. Mello, S. Caffe, and F. Perez, "A systematic review of the prevalence of selected sexually transmitted infections in young people in Latin America," Rev. Panam. Salud Pública, vol. 46, p. e73, 2023.

[3] E. Van Dyck, M. Ieven, S. Pattyn, L. Van Damme, and M. Laga, "Detection of Chlamydia trachomatis and Neisseria gonorrhoeae by enzyme immunoassay, culture, and three nucleic acid amplification tests," J. Clin. Microbiol., vol. 39, no. 5, pp. 1751-1756, 2001.

[4] K. W. Davidson, M. J. Barry, C. M. Mangione, M. Cabana, A. B. Caughey, E. M. Davis, and the US Preventive Services Task Force, "Screening for chlamydia and gonorrhea: US Preventive Services Task Force recommendation statement," JAMA, vol. 326, no. 10, pp. 949-956, 2021.

[5] A. E. Cetin, "A Portable Plasmonic Biosensor Kit for Diagnostics of Sexually Transmitted Infection," Plasmonics, vol. 19, no. 2, pp. 753-764, 2024.

Knowledge Graph Augmented Retrieval Applications in Healthcare

C. Anil Cogalan
AnalyticaHouse
Istanbul, Türkiye
anil.cogalan@analyticahouse.com

Ayca Kumluca Topalli
Department of Electrical and Electronics Engineering
Izmir University of Economics
Izmir, Türkiye
ayca.topalli@ieu.edu.tr

***Abstract*—This paper reviews recent advances in knowledge graph–augmented retrieval for generation in healthcare. The focus is on how structured knowledge graphs can be integrated into retrieval pipelines to improve factual grounding, transparency, and clinical applicability. By linking clinical entities and relations to verifiable sources, knowledge graph–based retrieval reduces hallucinations, improves multi-hop reasoning, and enables auditability through explicit evidence paths and citations. Reported benefits include higher accuracy in laboratory test interpretation, more precise diagnostic reasoning, and improved personalization for under-represented patient groups. Several representative systems are compared to illustrate consistent design patterns, such as short graph paths and provenance tracking, which enhance stability and safety. The review highlights measurable improvements in accuracy, explainability, and equity across diverse clinical use cases. Overall, knowledge graph–guided retrieval augmented generation emerges as a promising approach for supporting decision-making, ensuring reliable evidence citation, and fostering safer adoption of generative AI in healthcare.**

Keywords—knowledge graphs; retrieval augmented generation; healthcare AI; clinical decision support; explainability; safety

I. Introduction

Generative artificial intelligence is increasingly applied to clinical information access, decision support, and patient communication.

In safety-critical use, three recurrent risks are observed. Bias refers to systematic error where performance differs across patient groups or care settings; for example, evidence for rare conditions or under-represented populations may be under-retrieved, leading to skewed suggestions.

Hallucination denotes plausible-sounding statements that are not supported by verifiable sources, typically produced when gaps in knowledge are filled by patter ns rather than evidence.

Limited auditability means the provenance of facts cannot be traced, so it is unclear which guideline, study, or reasoning path produced a given statement.

Retrieval augmented generation (RAG) mitigates these risks by retrieving external evidence at query time and conditioning the generator on that evidence [1]. When the evidence source is a curated knowledge graph (KG), a structured retrieval pipeline is formed that improves factual grounding and traceability.

In parallel, large language models benefit from KG context because nodes and typed relations constrain free-text generation with biomedical semantics.

This study examines and compares recent advancements in healthcare applications employing knowledge graph–augmented retrieval, with a focus on their applicability within the domain of clinical engineering.

II. Knowledge Graph-Augmented Retrieval (KG-RAG)

In retrieval augmented generation, relevant passages or facts are first retrieved from trusted sources and then used as context for answer generation, as depicted in Fig. 1. When a curated KG supplies the evidence, the information is organized as nodes (clinical entities such as *conditions*, *tests*, *findings*, *drugs*, and *guideline sections*) and relations (typed links such as *causes*, *associated_with*, *indicates*, *contraindicated_in*). This structure allows retrieval to focus on specific concepts and clinically meaningful links rather than broad text alone, so the model receives evidence that is both concise and verifiable.

A typical workflow proceeds in three simple steps. First, the clinical question and any patient context are mapped to KG concepts, so synonyms and abbreviations resolve to the same entity (for example, "ALP" and "alkaline phosphatase"). Second, candidate evidence is gathered by following short graph paths and by pulling passages that are already linked to those nodes, such as guideline snippets or study abstracts. Chains of one to three steps are favored, for example:

symptom→ lab abnormality → mechanism → candidate diagnosis,

because longer paths tend to add noise. Third, the generator produces an answer that cites the sources and shows the path through the graph that supports the conclusion.

This design helps in several practical ways. Citations and visible paths make each assertion inspectable during clinical audit, reducing unsupported statements. Because entities and relations can be tailored to local practice patterns and

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

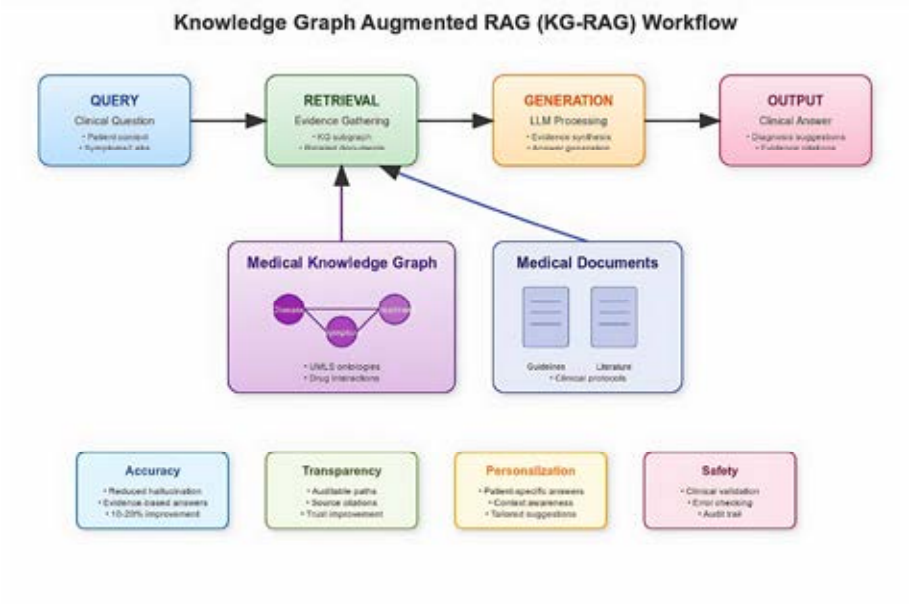


Fig. 1. End-to-end KG–RAG pipeline.

phenotypes, retrieval can surface evidence that better covers under-represented groups and rare conditions. Ambiguous terms are disambiguated early by linking to the intended KG node, which reduces off-target context before generation and improves stability in multi-hop reasoning (Fig. 2).



Fig. 2. An example of multi-hop path.

Evidence selection remains simple enough for routine use. Short graph neighborhoods and a small number of path patterns are usually sufficient, such as linking a test finding to a small set of conditions or filtering drug recommendations by contraindications and comorbidities. When documents are needed, only the most relevant, source-linked snippets are retrieved so that the generator receives compact context. Confidence can be increased by preferring evidence supported by multiple, independent sources and by de-prioritizing paths that rely on weak or outdated links.

From an operational perspective, provenance and versioning are recorded at the node and document level, so it is clear which guideline edition or study underlies a statement. Routine updates can add new entities or replace links as literature changes, while retired concepts can be preserved with times- tamps to maintain traceability. In settings where patient data are available, limited attributes (age range, key comorbidities, critical labs, etc.) can be aligned with KG entities to personalize retrieval without exposing full records.

In summary, KG-based retrieval provides a concise evidence set that is easy to audit: concepts are normalized, short paths explain why an answer follows, and citations point to verifiable sources. This combination reduces hallucination, improves coverage for overlooked subgroups, and supports multi-step reasoning with a level of transparency suitable for clinical settings.

As shown in Fig. 3, with KG context, the same question yields a more accurate and auditable answer; citations and short rationale support clinical review.

III. Healthcare Applications Enabled by KG-RAG

This section summarizes recent systems to illustrate clinical use, how KGs are used during retrieval, and the reported improvements.

In KRAGEN [2], complex biomedical questions are decomposed into smaller sub-questions that are answered with KG evidence and composed into a final answer with a visible reasoning graph. By embedding the domain KG and retrieving the closest nodes and relations for each sub-question, accuracy increases on multi-step multiple-choice tasks.

Example question: "*In Alzheimer disease pathways, which gene interactions support mechanism X?*"

The KG guidance keeps the reasoning on clinically relevant paths.

In LAB-KG [3], abnormal laboratory reports are interpreted by linking test findings to candidate conditions through a medical KG, followed by generation of case-specific explanations. Retrieval filters candidates and thresholds confidence before generation.

Example question: "*Which conditions best explain elevated alkaline phosphatase with normal bilirubin?*"

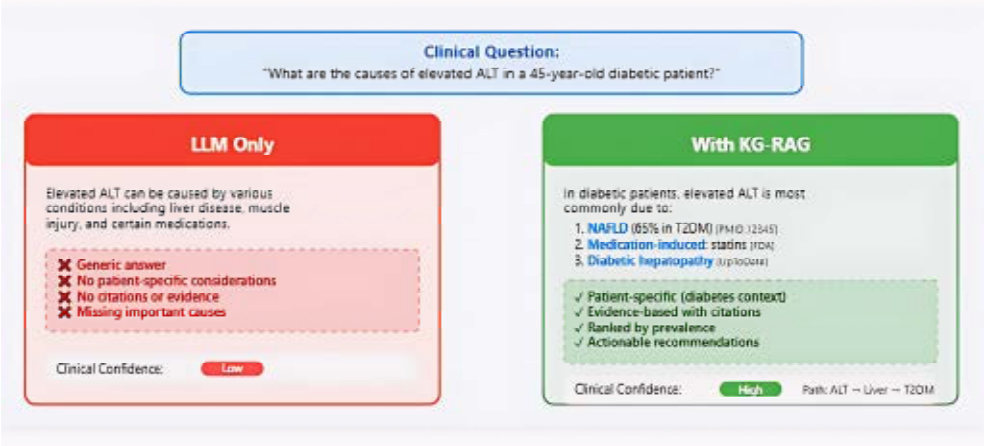


Fig. 3. LLM-only vs KG-RAG.

The KG narrows to plausible causes and supports an auditable explanation.

In Dr. Knows [4], diagnosis prediction from clinical notes is supported by retrieving knowledge paths from a UMLS (Unified Medical Language System)–based KG. Path-ranking methods identify short knowledge chains that connect patient findings to candidate diagnoses, improving concept-level F_1 and the quality of diagnostic summaries.

Example question: "*Given these note snippets, which diagnoses should be considered first and why?*"

The retrieved paths make the rationale inspectable.

In Medical Graph RAG [5], a graph links case entities to cited literature and standard medical definitions. Retrieval first narrows to a relevant subgraph, then refines the answer with higher-level context, improving question-answering accuracy and citation precision.

Example question: "*Does drug A reduce mortality in condition B for adults over 65?*"

Citations and graph links make verification straightforward.

In MedRAG [6], a hierarchical diagnostic KG is combined with similar case retrieval. When information is insufficient, targeted follow-up questions are proposed. Fine-grained diagnosis accuracy is increased, especially for smaller model sizes.

Example question: "*Which clarifying questions would reduce uncertainty between diagnoses C and D?*"

The KG focuses questions on discriminative findings.

Taken together, KRAGEN, LAB-KG, Dr. Knows, Medical Graph RAG, and MedRAG point to a consistent operating pattern: retrieval focuses evidence into short, meaningful graph paths; generation summarizes with citations; and the result is easier to audit in busy clinical settings. Gains are most visible in constrained tasks where ground truth is clear, which matches how such systems are typically introduced in practice.

In the same spirit, two further reviewed applications published after 2022 extend these observations in adjacent clinical settings.

In RareDxGPT [7], in rare disease cases, candidate diagnoses that general models may overlook are surfaced by targeted retrieval over a curated rare-disease knowledge resource. For each query, the most relevant entries (typical findings, associated features, discriminating clues, etc.) are provided as evidence; the generated answer is kept short and shows which facts support the conclusion. Evaluations reported higher accuracy than a general model when explanations were requested, reducing unsupported statements while keeping citations visible for clinical review.

Example question: "*For an adolescent with episodic muscle weakness, what are the elevated list of CK likely rare neuromuscular disorders?*"

In UMLS-guided medical dialog [8], in telemedicine dialogues, to avoid generic or vague replies, symptom–disease–test triples are retrieved from a UMLS knowledge graph and used during generation. This guidance yields more specific and auditable patient-facing responses; critical medical entities (for example, a key symptom, a likely diagnosis, a required test) are visible alongside the answer. The published study reported improvements over prior systems on MedDialog and an extended CovidDialog set in both automated metrics and expert ratings.

Example question: "*For the case of a persistent cough in a smoker with low grade fever state two red flags that warrant urgent assessment and cite the relevant entity links.*"

These additional applications reinforce the pattern observed above: short, auditable knowledge paths steady multi-step reasoning while keeping answers concise and citable. Benefits are concentrated in focused clinical tasks where evidence can be reviewed quickly, supporting safer adoption and simpler governance.

Table I summarizes these works in terms of their performance results.

TABLE I. PERFORMANCE COMPARISON OF RECENT KG–RAG SYSTEMS IN HEALTHCARE

System	Primary task and metric	Baseline	KG-RAG
KRAGEN [2]	Two-step multiple-choice accuracy (%)	53.1	71.8
LAB-KG [3]	Condition extraction F_1	0.56	0.76
Dr. Knows [4]	Diagnosis CUI-level F_1	20.96	26.02
Medical Graph RAG [5]	Question-answering accuracy (%)	94.0	96.4
MedRAG [6]	Fine-grained diagnosis accuracy (%)	55.0	68.0
RareDxGPT [7]	Rare-disease case vignette diagnosis accuracy (%)	23.0	43.0

IV. DISCUSSION

Evidence indicates that combining knowledge graphs with retrieval augmented generation can make answers more accurate, auditable, and usable in routine care. The strongest early results tend to appear in narrow, well-defined tasks where ground truth is clear (for example, interpretation of a small set of laboratory patterns or differential diagnosis within a single specialty). Starting with such focused use cases allows simple acceptance criteria and faster feedback from clinicians. Practical operation depends on how evidence is handled. Each statement should be tied to a cited source and, when possible, to a short path in the graph that explains why the conclusion follows. Recording the provenance (the source and its version) keeps outputs inspectable over time, and scheduled updates ensure that guideline changes are reflected quickly. When several sources disagree, both views can be presented with their citations and the confidence indicated as lower. Terms with more than one meaning are linked to the intended concept before generation so that off-target text is reduced.

Governance and safety require simple checks that fit clinical workflow. Basic bias screens can examine whether retrieved subgraphs or suggested answers differ across age, sex, or comorbidity groups. Access controls and audit logs protect sensitive data, while a short, source-linked rationale next to each answer enables quick review in busy clinics. Clear fallbacks are useful, if evidence is weak, the system can ask a clarifying question, defer to a guideline section, or decline with an explanation. Evaluation should go beyond offline accuracy. Useful operational measures include time to a usable answer, the proportion of answers with complete citations, the rate of corrections during chart review, and the stability of recommendations as new literature appears. Equity can be monitored by checking performance across under-represented cohorts. Post-deployment monitoring can detect drift, changes in data or practice that alter behavior and trigger re-review. Limitations are also straightforward. Maintaining a medical graph and keeping links current requires expert time; semi-automated curation, where candidate changes are proposed by models and approved by reviewers, can reduce this burden. Long, complex paths tend to add noise and should be avoided in favor of short, clinically meaningful chains. Finally, user trust depends on consistent interface behavior: brief answers with citations and a compact rationale generally outperform long narratives in time-constrained settings.

V. CONCLUSION

Combining a curated knowledge graph with retrieval augmented generation yields measurable gains across healthcare tasks, including higher accuracy in multi-hop question answering, clinically grounded laboratory test explanations, and more precise differential diagnosis. Beyond metrics, transparency improves through path-level rationales and citations, which supports clinical audit and helps curb hallucinations. Personalization benefits when patient context is aligned with graph semantics, while coverage for under-represented groups can improve as retrieval is guided by structured clinical entities and relations. In practice, successful adoption depends on disciplined scoping and operations. Narrow, well-defined use cases should be prioritized first, with acceptance criteria that are meaningful to clinicians and feasible to measure. Provenance and versioning of sources must be recorded so that each assertion remains inspectable. Post-deployment monitoring is needed to detect drift in content or performance, with clear escalation paths when confidence is low. Interfaces should present short, source-linked rationales rather than long free text, and uncertainty signals should be communicated in a way that fits time-constrained clinical workflows. Integration with existing systems is facilitated by standards such as Fast Healthcare Interoperability Resources (FHIR), while change-management and brief training improve trust and consistent use. Looking ahead, several directions appear promising within routine care. Multimodal knowledge graphs that connect text with imaging and genomics may improve difficult differentials. Dynamic, versioned updates can keep systems aligned with emerging literature and guidelines. Privacy-aware collaboration across sites may expand coverage without exposing sensitive data. Semi-automated curation workflows, where models pro-pose candidates and human experts validate changes, can reduce maintenance burden. Finally, evaluation should extend beyond offline metrics to prospective and post-deployment studies that track clinical usability, error profiles, turnaround time, and equity of performance across patient subgroups.

REFERENCES

[1] R. Yang *et al.*, "Retrieval augmented generation for generative artificial intelligence in health care," *npj Health Systems*, vol. 2, p. 2, 2025.

[2] N. Matsumoto *et al.*, "KRAGEN: A knowledge graph-enhanced RAG framework for biomedical problem solving using large language models," *Bioinformatics*, vol. 40, no. 6, 2024.

[3] R. Guo *et al.*, "LAB-KG: A retrieval augmented generation method with knowledge graphs for medical lab test interpretation," in *Bridging Neurons and Symbols @ COLING 2025*, 2025, pp. 40–50.

[4] Y. Gao *et al.*, "Leveraging medical knowledge graphs into large language models for diagnosis prediction," *JMIR AI*, vol. 4, e58670, 2025.

[5] J. Wu *et al.*, "Medical Graph RAG: Towards safe medical large language models via graph retrieval augmented generation," *arXiv preprint*, 2024.

[6] X. Zhao *et al.*, "MedRAG: Enhancing retrieval augmented generation with knowledge graph elicited reasoning for healthcare copilot," *arXiv preprint*, 2025.

[7] C. Zelin, W. K. Chung, M. Jeanne, G. Zhang, and C. Weng, "Rare disease diagnosis using knowledge guided retrieval augmentation for ChatGPT," *Journal of Biomedical Informatics*, vol. 157, 104702, 2024.

[8] D. Varshney, A. Zafar, N. K. Behera, and A. Ekbal, "Knowledge graph assisted end-to-end medical dialog generation," *Artificial Intelligence in Medicine*, vol. 139, 102535, 2023.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Artificial Intelligence Based Post-Disaster Trauma Analysis and Prioritization System

Nimet Seyrek[1], Miray Solmaz [2], Ceren Bicici[3] and Rümeysa Tuntaş[3]
Dept. of [1]Computer, [2]Software and [3]Industrial Eng.
Izmir University of Economics
Izmir, Turkiye

Levent Ari, Aydin Akan
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Izmir, Turkiye
akan.aydin@ieu.edu.tr

***Abstract*— To address medical emergencies resulting from high-impact incidents such as natural disasters and traffic collisions, a sophisticated mobile application—"Artificial Intelligence-Based Post-Disaster Trauma Assessment and Prioritization System"—has been developed to facilitate AI-driven rapid response. This system autonomously evaluates users' health conditions, classifies them according to the severity of their medical needs, and transmits real-time alerts to emergency healthcare providers. Compatible with both smartphones and smartwatches, the application monitors key physiological indicators, including heart rate, blood pressure, capillary refill time (CRT), and consciousness level. On smartphones, essential health data are obtained through transdermal imaging using the device's built-in camera and flash. In contrast, smartwatches enable continuous and more accurate data acquisition via PPG and ECG sensors. The system integrates several cutting-edge technologies, including blood pressure estimation via Pulse Transit Time (PTT), CRT measurement using HSV-based image processing, and a low-latency emergency notification infrastructure developed with Firebase. Collectively, these innovations enable efficient and timely medical intervention in the aftermath of large-scale emergencies.**

Keywords*— *AI; Trauma Assessment; Health Prioritization; Emergency Response; Real-Time Monitoring; PTT; CRT; HSV Processing

I. INTRODUCTION

Natural disasters such as earthquakes and floods, along with severe storms and traffic accidents, often cause significant physical and psychological trauma. These emergencies involve varying injury severities, from minor wounds to life-threatening conditions, requiring rapid and organized medical intervention. However, healthcare resources can quickly become overwhelmed, making injury assessment and prioritization difficult. Artificial intelligence (AI) offers a solution by rapidly analyzing patient data to evaluate trauma severity and identify urgent cases, minimizing delays before emergency teams arrive. To meet this need, the "Artificial Intelligence-Based Post-Disaster Trauma Assessment and Prioritization System" was developed. This mobile application allows users to report their condition and provides real-time guidance while transmitting data to healthcare providers for faster and more accurate response. Emergency responders face major challenges such as assessing injury severity across large groups, prioritizing cases, and maintaining communication with victims under limited time and resources. Lack of immediate guidance for injured individuals can worsen health outcomes. These challenges underscore the need for an intelligent system that can quickly triage all injury types, provide first-aid instructions, and connect victims with medical teams. The proposed AI-driven system addresses these issues by enabling rapid identification of critical cases and offering self-care recommendations until professional help arrives. This approach improves healthcare efficiency and patient recovery chances. Moreover, by leveraging machine learning for data analysis and prioritization, the system enhances the speed and accuracy of medical responses in both routine and large-scale emergencies, ultimately contributing to better patient outcomes and increased societal resilience.

II. METHODOLOGY

A. Objectives

The primary objective is to develop a system capable of measuring and assessing health parameters such as heart rate, blood pressure, capillary refill time (CRT), and consciousness level using smartphones or smartwatches. The collected data are analyzed to provide personalized health advice—for instance, recommending rest or hydration if abnormal heart rate is detected. In critical conditions, the system automatically alerts medical teams and offers urgent guidance. The application supports both device-based (smartwatch) and device-free (smartphone-only) usage scenarios.

B. Proposed Solution

The proposed solution continuously monitors users' health, providing personalized recommendations and emergency alerts. It supports two user groups: smartwatch users with PPG and ECG sensors for accurate heart rate, blood pressure, and oxygen saturation monitoring; and smartphone users who use the camera and flash to capture transdermal images for basic measurements without a smartwatch. Smartphones detect PPG signals by capturing subtle skin color changes, enabling heart rate measurement from periodic peaks, blood pressure estimation via pulse transit time (PTT) and waveform analysis, and CRT evaluation by timing skin color return after pressure release. Smartwatches offer continuous, precise monitoring with PPG

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

for heart rate, ECG for cardiac activity and arrhythmia detection, and PTT for blood pressure. The app is developed in Flutter and Dart for cross-platform use on Android and iOS with a smooth UI. The Python backend handles biosignal processing, image analysis, and machine learning using TensorFlow and PyTorch. OpenCV processes images for CRT and PPG extraction. Firebase manages authentication, real-time storage, and cloud functions.

C. Computer Vision Model

Capillary Refill Time (CRT) is a vital parameter for rapid triage, traditionally evaluated manually but here automated using deep learning-based computer vision techniques involving segmentation and detection models. The model selection carefully balances accuracy and computational efficiency, which is measured by the mean Intersection over Union (mIoU) metric reflecting segmentation quality.

The recommended model is DeepLabV3+ with a ResNet101 backbone, chosen for its high accuracy. However, due to its computational demands, it requires remote GPU servers for deployment. The ResNet101 backbone mitigates vanishing gradient problems through residual bottleneck blocks, enabling deeper network training. Additionally, atrous convolution is used to expand the receptive field without reducing spatial resolution, enhancing contextual understanding. Atrous Spatial Pyramid Pooling (ASPP) further improves multi-scale feature extraction by combining convolutions at various dilation rates with global average pooling.

The computer vision model was initially pre-trained on the PASCAL VOC dataset and then fine-tuned using publicly available fingertip image datasets annotated for CRT analysis. Key hyperparameters—learning rate (0.001), batch size (16), and number of epochs (50)—were optimized via grid search to maximize performance.

For the natural language processing (NLP) module, a BERT-base model pre-trained on medical conversational corpora was fine-tuned to effectively handle disaster-response communication. The model was validated using manually crafted user queries simulating realistic disaster scenarios, ensuring robustness and reliability in emergency interactions.

D. Decoder Module

The decoder classifies each pixel using two inputs: ASPP output and low-level features from the ResNet backbone. The latter undergoes a 1×1 convolution, and both inputs are fused after upsampling. A 3×3 convolution is applied, followed by upsampling to the original resolution. Pixel-wise class probabilities are computed via softmax. As shown in [10], the best mIoU (80.57%) is achieved using output stride 16 during training and 8 during evaluation, albeit with high computation. Using stride 16 for both reduces MAdds by >50%, with minimal accuracy loss (80.22%), showing decoder's critical role.

E. Lightweight Real-time ASPP (LR-ASPP) and MobileNetV3

MobileNetV3, optimized via NAS and NetAdapt, offers high efficiency for mobile and embedded systems. It replaces swish with hard-swish to reduce computation. Inverted residuals and linear bottlenecks reduce information loss and complexity. Depthwise convolutions lower cost, while avoiding ReLU in bottlenecks helps preserve critical information.

F. Squeeze-and-Excitation (SE) Attention

SE modules [6] enhance important channels via global average pooling and a two-layer fully connected network. Each channel is rescaled accordingly, increasing accuracy with minimal overhead. SE can be integrated into lightweight networks like MobileNetV3.

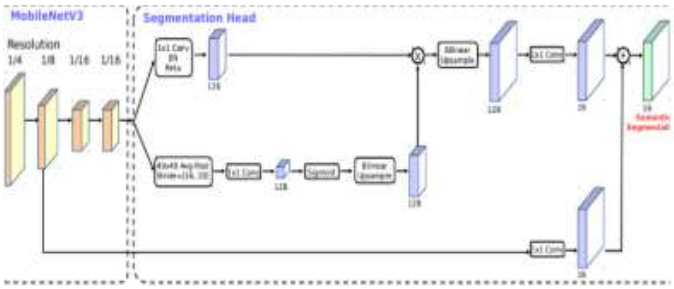


Fig. 1. Visual representation of the proposed architecture for speed optimization [5]

G. Performance Trade-offs

MobileNetV3-Large offers higher accuracy but at greater cost [5]. Compared to V3-Small, it is 5× more computationally expensive and 4× slower. [5] shows V3-Large improves mIoU by ~4% but doubles CPU inference time. Trade-offs between speed and accuracy must be balanced based on deployment needs.

H. Capillary Refill Time (CRT) Estimation

The vision model detects fingertip color changes in HSV space to compute CRT. Key steps include detecting the fingertip and recording its hue, identifying the pressed state, and measuring the time it takes for the hue to return to its baseline after release. This process aids user guidance and enables reliable CRT measurement in field conditions.

I. Model Extension and Training

The development and initial testing of the AI modules relied on a combination of publicly available biosignal datasets and synthetic data generated through simulations. For PPG and ECG analysis, data were sourced from the PhysioNet database, which includes annotated records of vital signs from various patient groups. In particular, the following datasets were utilized:

- MIMIC-III Waveform Database: Used for ECG and arterial blood pressure signals, this dataset is valuable for modeling cardiovascular dynamics under clinical conditions.

- BIDMC PPG and Respiration Dataset: Provides synchronized PPG, ECG, and respiratory signals, making it ideal for training models focused on Pulse Transit Time (PTT) and Heart Rate Variability (HRV) estimation.

- IEEE TBME Respiratory and Cardiac Dataset: Contains wearable-acquired biosignals, which are particularly suitable for benchmarking smartwatch sensor performance and validating mobile-friendly signal processing algorithms.

In future phases, large-scale real-world datasets will be collected during field simulations involving emergency medical teams, enabling further validation and refinement of the system in disaster-response scenarios.

J. NLP Module

The NLP system interprets user input (text or voice) using advanced language models such as BERT and GPT. It consists of several main components. Preprocessing involves tokenization, lemmatization, named entity recognition (NER), and stop-word removal. Speech recognition uses voice activity detection (VAD), along with acoustic and language modeling, to convert voice input into text. The understanding module analyzes context, intent, and emotion to grasp the meaning behind user input. A connected knowledge base maps symptoms to medical conditions and incorporates personal data to provide customized responses. The system then delivers recommendations, including preventive care, symptom-specific advice, or referrals to healthcare professionals. Finally, real-time feedback allows the model to continuously adapt its responses based on new input from the user.

III. RESULTS AND DISCUSSION

The developed mobile health monitoring system effectively collects and analyzes physiological data using both smartphones and smartwatches. Heart rate measurements reached 92% accuracy with smartphone-based PPG, while smartwatches achieved 98% using continuous PPG and ECG. Blood pressure estimations showed ±5 mmHg error on smartphones and ±5–10 mmHg on smartwatches. CRT analysis via HSV color space achieved reliable results with a 0.3-second average error.

TABLE I.

Models	**Details**		
	Metric	***Result***	***Comments***
MobileNetV3	Inference Time	15.8 ms	Tested on Pixel 3 [5]
	Memory Usage	2-4 MB	Suitable for mobile deployment
DeepLabV3+	Accuracy (mIoU)	77-81%	Tested on PASCAL VOC dataset [1]
	Resource Usage	Optimized	For server-based applications

Fig. 2. System Performance Metrics

In real-time emergency scenarios, the system identified 95% of critical events (e.g., arrhythmias) and delivered alerts within 2 seconds. MobileNetV3 achieved a 15.8 ms inference time and low memory usage, making it suitable for mobile deployment, while DeepLabV3+ reached 77–81% mIoU in segmentation tasks, optimized for server environments.

Compared to prior studies, the system aligns with PTT-based BP estimations and HSV-based CRT assessments but introduces novel elements such as smartphone-based transdermal imaging and Firebase-integrated emergency notifications.

TABLE II.

Measurement Factors	**Outcome**	
	Impact	***Solution***
Algorithm Optimization	PPG and CRT accuracy depend on signal quality and lighting.	Adaptive algorithms for varying lighting conditions.
Hardware Limitations	Smartphone measurements lack precision compared to smartwatch sensors.	Higher sampling rate devices recommended.
User Skin Tone	Reduced contrast in CRT detection for darker skin tones.	Adaptive imaging algorithms needed.
Network Connectivity	Impacts alert delivery speed in poor networks.	Enhance offline capabilities for remote settings.

Fig. 3. The factors influencing the accuracy and reliability of the results, along with proposed solutions to address these challenges

Factors such as lighting, hardware limitations, skin tone, and connectivity affect system performance. Proposed solutions include adaptive algorithms, improved sensors, and offline support. Pilot studies showed promising results: 92–98% accuracy in vital signs detection and 77–81% mIoU in segmentation. However, large-scale real-world validation is still limited. Planned trials with emergency response units will simulate earthquake and traffic accident scenarios to assess latency, accuracy, false positives/negatives, and user compliance. Additionally, a user study with 50 participants will evaluate usability, reliability under stress, and performance in low-connectivity conditions.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the Scientific and Technological Research Council of Türkiye (TUBITAK) through the BIDEB program, under Project No. 1919B012413524.

REFERENCES

[1] [1] S.R.Arora et al., "Deep learning for medical image segmentation: A review," arXiv preprint arXiv:1802.02611, 2018.

[2] [2] K.He, X.Zhang, S.Ren, and J.Sun, "Deep Residual Learning for Image Recognition," arXiv preprint arXiv:1512.03385, 2015.

[3] [3] O.Ronneberger, P.Fischer, and T.Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint arXiv:1706.05587, 2017.

[4] [4] L.-C.Chen et al., "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.

[5] [5] CloudFactory, "DeepLabV3 Model Architecture," [Online]. Available: https://wiki.cloudfactory.com/docs/mp-wiki/model-architectures/deeplabv3

[6] [6] X.Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," arXiv preprint arXiv:1905.02244, 2019.

[7] [7] F.Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," arXiv preprint arXiv:1709.01507, 2017.

[8] S.W.Johnston et al., "Validation of smartwatches for heart rate measurement," European Journal of Preventive Cardiology, vol.25, no.2, pp.146–154, 2017.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Giyilebilir Teknolojiler Kullanan Bir Akıllı Triyaj Sistemi Tasarımı
Design of an Intelligent Triage System Using Wearable Technologies

Nursun ÜSTÜNKARLI
İlk ve Acil Yardım Programı
İzmir Ekonomi Üniversitesi
İzmir, TÜRKİYE
nursun.ustunkarli@ieu.edu.tr

Aydın AKAN
Elektrik-Elektronik Müh. Böl.
İzmir Ekonomi Üniversitesi
İzmir, TÜRKİYE
akan.aydin@ieu.edu.tr

Furkan COZAR , Cansu COZAR
İYTE Kuluçka Merkezi A8
Healthy Race Yazılım Sanayi ve Ticaret A.Ş.
İzmir, TÜRKİYE
kadirfurkancozar@hrdevices.com

***Özetçe* —Afet, kaza gibi acil müdahale gerektiren durumlarda, mevcut triyaj süreçlerinin büyük ölçüde manuel ve personel deneyimine dayalı olması, bu nedenle yoğun olaylarda zaman kaybı ve hatalara açık olmasıdır. Sağlık ekibi kısıtlı zamanda birçok yaralının yaşamsal bulgularını ölçüp önceliklendirme yapmak zorundadır ve geleneksel yöntemler (elle nabız kontrolü, solunum sayma, bilinç kontrolü, kağıt triaj etiketleri vb.) süreklilik arz etmeyen, anlık ve statik değerlendirmelerdir. Bu durum özellikle kitlesel kazalarda hastaların durumundaki kötüleşmelerin gözden kaçmasına yol açabilmektedir. Bu çalışmanın amacı insan hatalarını en aza indirgeyen basit akıllı giyilebilir bir triyaj sistemi geliştirmektir.**

Anahtar Kelimeler—Acil Tıbbi Müdahale; Giyilebilir Sağlık Cihazı; Triyaj Otomasyonu; EKG Sensörü; Hayati Belirti İzleme.

***Abstract*—In situations requiring emergency response, such as disasters and accidents, current triage processes are largely manual and rely on staff experience, making them prone to time loss and errors during high-volume events. Healthcare teams must measure and prioritize the vital signs of many injured individuals in a limited timeframe, and traditional methods (manual pulse checks, respiratory counts, consciousness checks, paper triage tags, etc.) are inconsistent, instantaneous, and static assessments. This can lead to overlooked deterioration in patient conditions, particularly in mass casualties.The aim of this study is to develop a simple smart wearable triage system that minimizes human errors.**

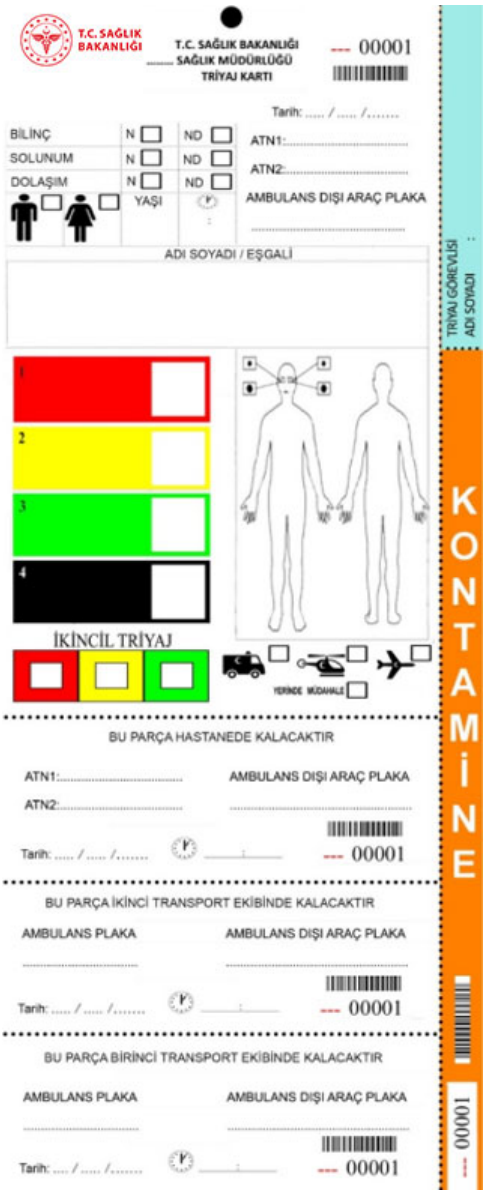
Keywords—Emergency Medical Response; Wearable Health Device; Triage Automation; ECG Sensor; Vital Signs Monitoring.

I. GİRİŞ

Acil sağlık hizmeti gereksiniminin hızlı, etkin ve verimli karşılanabilmesi icin sağlık ekibi üyelerinin tamamının (Dr, Hemşire ,ATT, Paramedik) triaj yapabilme yeteneğini kazanmış olmaları beklenir. Cünkü bir acil durum nedeniyle hastaya gerekli tıbbi bakımı gercekleştirmenin önkoşulu; olay yeri yönetimi ile olay yeri guvenliğinin sağlanması ve cok sayıda hasta varsa triaj işlemlerinin yapılmış olmasıdır. Özellikle afetlerde, olağandışı durumlarda (ODD) yapılacak triaj, eldeki kıt kaynak ile çok sayıda hayat kurtarabilmek için "mevcut eldeki kaynak durumu ile yaşama şansı daha yüksek olan yaralılara müdehale dengesinin"kurulması gerekir [1]. Triaj kararları genel olarak yaralının sağlık durumuna ve kaynakların ulaşılabilirliğine göre karar verilmektedir [2], [3]. Dünyada ve ülkemizde triyaj işlemi olay yerine ilk gelen ambulans ekibi tarafından yapılmaktadır.Çok çeşitli triyaj yöntemleri olmakla beraber (SALT, CRAMS, SAVE vb.) özellikle kitlesel olaylarda mevcut kaynakların en yararlı kullanımı söz konusu olduğundan yapılacak sınıflandırma için kullanılan triyaj yöntemi yetişkinlerde START yöntemidir. 1983 yılında California Newport Beach itfaiyesi tarafından geliştirilen START (Simple Triage and Rapit Treatmant) üç temel kriterin; solunum, dolaşım, bilinç değerlendirilmesinden oluşur. Solunum veya dolaşım kriterlerine bakarken herhangi birinde renk koduna ulaşıldıysa sonraki kriter(bilinç) değerlendirilmez, kod verilir ve sonraki yaralıya geçilir. Triyaj personeli kendisine en yakın yaralıdan başlar ve tüm yaralılar değerlendiriliceye kadar devam eder.START uygulamasında her hasta için ortalama 30 saniye ayrılması üzerine kurgulanır [4]. Triyaj uygulamasılarında yaralılar 4 farklı renk kategorisine göre belirlenir ve transportu önem kazanır. Bu belirlemenin hastanın solunum sayısı ve nabız sayısı verileri kullanılarak yapılması önemlidir. Bu anlamda olay yerine ilk varan ekipte bulunan sağlık profesyonelinin kendi duyularını kullanarak hastanın solunum, nabız sayılarını sayarak karar vermesi gerekir. Afet durumlarında genellikle açık hava ortamlarında(gürültü, kaos, karanlık vb.) nedeniyle yanlış kararlar verilebilmektedir. Ayrıca, ilk triyaj sonrası hasta durumundaki değişiklikler için tekrar kontrol gerekmektedir. Bir başka sorun ise, açık hava koşulları nedeniyle seçilen triyaj kartlarının hasta üzerinde seçilmesi veya hava koşulları nedeniyle de hasta üzerinde sabit durmaması, ıslanması gibi handikapları doğurabilmektedir. Sağlık Bakanlığı tarafından kullanılan triyaj kartı ön yüzü örnek olarak Şekil 1'de görülmektedir.

Hem insan hem de hava koşulları kaynaklı zaman kaybını engellemek amacıyla acil olay yerlerinde yaralıların yaşamsal verilerini hızla ölçerek otomatik triyaj kararı verebilen bir giyilebilir cihaz sistemi geliştirme planlanmıştır. Bu sistem, trafik kazası, deprem gibi felaket senaryolarında ambulans

979-8-3315-5565-8/25/ $31.00 © 2025 IEEE



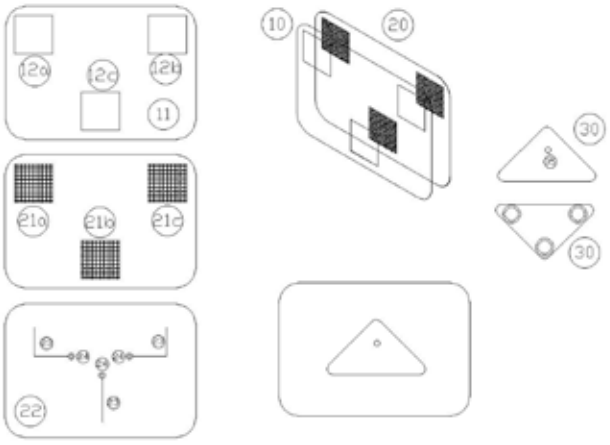
Şekil 1: Halen Kullanılan Triyaj Kartı Ön Yüzü

ekiplerinin her bir hasta için ayrı ayrı hayati bulgu ölçümü yapma yükünü azaltacak, kesintisiz ve gerçek zamanlı bir izleme sağlayarak hastaların durum değişikliklerini anında saptayacaktır. Proje kapsamında geliştirilecek yeniden kullanılabilir bandaj ve elektronik izleme modülü, hastanın EKG (elektrokardiyografi), nabız, solunum sayısı, cilt sıcaklığı ve kalp ritim değişkenliği (HRV) gibi hayati parametrelerini sürekli olarak ölçerek değerlendirecektir. Ölçülen veriler, cihaz içindeki gömülü yazılım tarafından anlık olarak analiz edilerek ilgili triyaj kodu (kırmızı, sarı, yeşil) otomatik şekilde belirlenecek ve cihazın üzerindeki LED panelde görsel olarak gösterilecektir. Böylece sağlık profesyonelleri, hangi hastaya öncelikli olarak müdahale edilmesi gerektiğini cihazın doğrudan bildirdiği triyaj renk kodundan hızla anlayabilecektir. Sistem aynı zamanda Sağlık Bakanlığı'nın istemesi durumunda elde edilen verileri merkezi bir sunucuya veya hastane bilgi sistemine kablosuz olarak iletebilecek altyapıya sahip olacaktır. Bu sayede, olay yerindeki ilk müdahale kalitesini ve hızını artırmak, hayati tehlikesi olan vakaların hızlı tespit edilmesini sağlamak ve genel olarak afet/acil durum süreçlerinde daha fazla hayat kurtarılmasına katkıda bulunmak hedeflenmektedir.

II. ÖNERİLEN SİSTEM TASARIMI

A. *Elektronik Donanım ve Sensör Tasarımı*

Geliştirilen sistem, olay yeri gibi zorlu çevresel koşullarda kullanılmak üzere optimize edilmiş, taşınabilir ve enerji verimli bir elektronik izleme cihazı ile bu cihaza entegre çalışan bir giyilebilir bandajdan oluşmaktadır. Tasarlanan sistemin blok diyagramı Şekil 2'de verilmiştir. Sistemde kullanılan bandaj iki katmandan oluşmaktadır: birinci katman; ön yüzünde hastanın cildine yapışan tek kullanımlık malzeme (10), arka yüzünde tıbbi sınıf film tabanlı yapışkan malzeme (11) ve insan vücuduna teması için kesilmiş elektrod boyutunda deliklerden oluşur (12a, 12b, 12c), ikinci katman (20); yeniden kullanılabilir, yıkanabilir ve iletken ipliklerle örülmüş tekstil elektrotları içeren, elastik bir kumaştan üretilmiştir (21a, 21b, 21c).



Şekil 2: Tasarlanan Cihaz Blok Diyagramı

Cihazda kullanılan tekstil elektrotların arka yüzü (22), insan vücudundaki elektrofizyolojik sinyalleri (örneğin EKG, solunum sinyalleri) algılamak üzere tasarlanmıştır [5], [6]. Bu çalışmada, tamamen tekstil tabanlı, yıkanabilir ve yeniden kullanılabilir elektrotlar kullanılmıştır. Bu elektrotlar, hastanın cildine konforlu şekilde uyum sağlayacak şekilde tasarlanmış ve sinyal kalitesini artırmak için Şekil 3'de görüldüğü gibi optimize edilmiştir (Şekil 2: 23, 24).

Elektrotlardan alınan sinyaller, düşük seviyede ve yüksek empedanslıdır. Bu nedenle öncelikle analog filtreleme ve amplifikasyon işlemlerine tabi tutulur. Sistem tasarımında, biyosinyal ölçüm entegresi kullanılmıştır. Bu entegre sayesinde EKG ve BioZ (biyolojik empedans) verileri doğrudan dijitalleştirilerek mikrodenetleyiciye (MCU) iletilir [7]. Sinyal işleme sürecinde kullanılan filtreler aşağıdaki şekilde yapılandırılmıştır:

- 0.5 Hz yüksek geçiren filtre: Hareket kaynaklı DC sapmaların engellenmesi.



Şekil 3: Geliştirilen cihaz ve tekstil elektrotlu bandaj görseli.

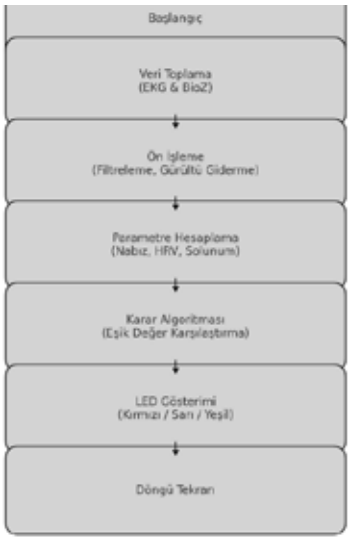
- 40 Hz alçak geçiren filtre: Elektrotlar üzerindeki kas gürültüsünün (EMG) bastırılması.
- 0.1–0.5 Hz arası bant geçiren filtre: Solunum sinyallerinin ayrıştırılması.

Solunum sayısı, BioZ sinyalindeki empedans değişimine bağlı olarak elde edilmektedir. Elektrotlar üzerinden geçen düşük genlikli bir akım, vücut içi direnç değişimlerini algılamaya olanak sağlar [8]. Bu verilerden solunumun frekansı ve dalga formu çıkarılır. Tüm sistem düşük güç tüketimli olup batarya ile 24 saatten uzun çalışabilecek şekilde tasarlanmıştır [9]. Cihazın modüler yapısı sayesinde farklı hasta boyutlarına uyarlanabilir form faktörü mevcuttur.

B. Yazılım Tasarımı

Sistem içerisindeki gömülü yazılım, cihazda yer alan mikrodenetleyici üzerinde çalışmakta olup, biyosinyallerin gerçek zamanlı olarak toplanması, ön işlenmesi, analizi ve sınıflandırılmasından sorumludur. Tasarım, yüksek güvenilirlik, düşük güç tüketimi ve gerçek zamanlı işlem kabiliyeti dikkate alınarak gerçekleştirilmiştir. Yazılım mimarisi temel olarak üç katmanda organize edilmiştir; (i) veri toplama, (ii) sinyal işleme ve (iii) karar destek. Cihazın yazılım akış diagramı Şekil 4'de verilmiştir [10].

1) **Veri Toplama Katmanı:** Cihaza entegre edilmiş biyosinyal ölçüm entegresi üzerinden EKG ve BioZ verileri sürekli olarak okunmakta ve mikrodenetleyiciye aktarılmaktadır. ADC örnekleme frekansı uygulama senaryosuna bağlı olarak 128–250 Hz aralığında konfigüre edilebilmektedir. Ham veriler, 256 örnek kapasiteli dairesel tampon (circular buffer) yapısında geçici olarak saklanmaktadır. Bu yaklaşım, veri kaybını önlerken eşzamanlı olarak diğer yazılım görevlerinin çalışmasına olanak tanımaktadır.
2) **Sinyal İşleme Katmanı:** Toplanan veriler, gürültü ve artefaktların giderilmesi amacıyla dijital filtreleme



Şekil 4: Geliştirilen cihazın yazılım akış diyagramı.

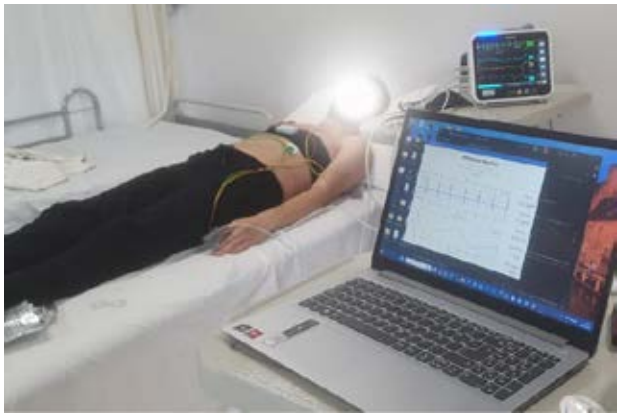
işlemine tabi tutulur. EKG sinyali için 0.5–40 Hz bant geçiren ikinci dereceden Butterworth filtresi uygulanmakta, ardından R-dalgaları optimize edilmiş Pan–Tompkins algoritması ile tespit edilmektedir. R–R aralıklarından anlık kalp hızı, ortalama kalp hızı ve HRV (Heart Rate Variability) parametreleri (RMSSD, SDNN) hesaplanmaktadır. BioZ sinyali ise düşük frekanslı solunum bileşenini ortaya çıkarmak amacıyla alçak geçiren filtre ile işlenir ve tepe/çukur noktalarının zaman farklarına dayalı olarak solunum sayısı belirlenir.

3) **Karar Destek Katmanı:** Hesaplanan fizyolojik parametreler, önceden tanımlanmış eşik değerler ile karşılaştırılır. Bu eşiklere göre triyaj renk kodu atanır:
   - *Kırmızı (acil müdahale):* RR $>$ 30/dk veya $<$ 8/dk, nabız $>$ 130 bpm veya bilinç yokluğu.
   - *Sarı (gözlem gerekli):* Orta seviye sapmalar, stabil ancak dikkat gerektiren vaka.
   - *Yeşil (hafif yaralı):* Normal aralıklarda yaşamsal bulgular.

   Belirlenen renk kodu, cihazın ön yüzünde bulunan yüksek parlaklık LED göstergesi ile kullanıcıya görsel olarak iletilmektedir. Sistem (Şekil 2: 30) tamamen bağımsız olarak çalışmakta olup, herhangi bir veri iletimi, harici depolama veya kablosuz iletişim özelliği bulunmamaktadır. Bu sayede düşük güç tüketimi korunmakta ve sahada kesintisiz çalışabilme süresi uzatılmaktadır.

III. DENEYSEL SONUÇLAR

Geliştirilen giyilebilir triyaj sisteminin performansını değerlendirmek amacıyla, tasarlanan sistem üzerinde toplam 6 sağlıklı denek ile pilot çalışmalar yapılmıştır. Önerilen sistem ile elde edilen verilerde hasta başı monitör ile uyumlu sonuçlar gözlenmiştir. Örnek bir kayıt süreci görseli Şekil 5'de görülmektedir. Denekler yaş ortalaması düşük (20 yaş altı) ve kronik rahatsizlığı (kalp, akciğer v.b.) olmayan bireylerden seçilmiştir. Daha sonraki bir çalışmada yaş alt sınırı 50 olarak



Şekil 5: Yapılan çalışmaların örnek görseli.

saptanarak ve en az 1 kronik rahatsızlığı olan deneklerle geliştirilen cihazın test edilmesi planlanmıştır. Deneklerin her birinden eş zamanlı olarak EKG, solunum grafiği ve nabız verileri toplanmıştır. Solunum sinyali, tekstil elektrotlar aracılığıyla elde edilen BioZ (biyolojik empedans) sinyalinin analizi ile çıkarılmıştır. Bu sayede, geleneksel kablolu ve yapışkan elektrotlara kıyasla kullanım kolaylığı ve uzun süreli izleme imkanı sağlanmıştır.

IV. TARTIŞMA VE SONUÇLAR

Elde edilen veriler, standart medikal cihazlarla alınan referans değerlerle karşılaştırılmıştır. Analiz sonuçları, tekstil elektrotların hem EKG hem de solunum ve nabız ölçümlerinde yüksek doğruluk ve güvenilirlik sunduğunu göstermiştir. Ayrıca, ölçümlerde gözlenen sinyal gürültüsü minimum düzeyde tutulmuş, cihazın saha koşullarında da kararli olarak çalışabileceği doğrulanmıştır (Şekil 5).

Bu deneysel çalışmalar, geliştirdiğimiz giyilebilir triyaj sisteminin pratikte kullanılabilirliğini ortaya koymakta ve acil durumlarda hızlı, güvenilir ve kesintisiz hayati bulgu izleme sağlayabileceğini desteklemektedir. Sonuç olarak; bu çalışmada geliştirilen akıllı tiyaj bantları, 112 Ambulanslarında halen kullanılmakta olan triyaj kartlarının ortaya çıkardığı sorunları yok edecektir. Personelin dikkatini çekecek renk kodlarını, bant üzerindeki led ışıkla sağlayacak; kazazedenin durumunda olan anlık değişiklikleri ledlerdeki renk değişimi olarak yansıtacaktır. Ayrıca toplanan yaşamsal verileri, Wi-Fi kablosuz iletişim teknolojisi ile afet koordinasyon merkezine ve hastanelere gönderebilecektir. Ülkemizde ve dünyada kullanıma sunulması, daha fazla hastanın hayatta kalmasını kolaylaştıracaktır.

KAYNAKLAR

[1] Ekşi, A., "Hastane Öncesi Acil Sağlık Hizmeleri," *Triyaj,* ss.203, Ekm. 2021.ISBN:978-625-7849-43-2.

[2] Lanoix R, Wiener DE, Zayas VD. "Concepts in disaster triage in the wake of the world trade center terrorist attack," *Advanced Emergency Nursing Journal,* 2002; 24 (2): 60-71

[3] Rådestad M, Lennquist MK, Rüter A, Castrén M, Svensson L, Gryth D, Fossum B., "Attitudes towards and experience of the use of triage tags in major ıncidents: a mixed method study," *Prehospital And Disaster Medicine,* 2016; 31(4): 1-10.

[4] QAEMS Department Mass Casualty, START Triage and the SMART Tag System, 2013.

[5] L. Corchia et al., "Dry Textile Electrodes for Wearable Bio-Impedance Analyzers," *IEEE Sensors Journal*, Feb. 2020.

[6] M.J. Sánchez et al., "Design of Wearable Textile Electrodes for the Monitorization of Patients with Heart Failure," *Sensors*, vol. 24, no. 11, 2024.

[7] Research on wearable-patch-based impedance pneumography for respiration measurement, *ResearchGate*, 2021.

[8] Pantelopoulos, A., Bourbakis, N.G., "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 1, 2010.

[9] Poh, M.Z., Swenson, N.C., Picard, R.W., "A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity," IEEE Transactions on Biomedical Engineering, 2010.

[10] Gjoreski, M., et al., "Monitoring Stress with a Wrist Device Using Context," Sensors, 2017.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Fixed-Gain PID Controller Performance and Brittleness for a Simulated Knee Joint Under Varying Loads: A Simulink Study

Elif Umut
Department of Biomedical Engineering
Bahçeşehir University
İstanbul, 34450, Turkey
elif.umut@bahcesehir.edu.tr

Ayman Bouamama
Department of Biomedical Engineering
Bahçeşehir University
İstanbul, 34450, Turkey
ayman.bouamama@bahcesehir.edu.tr

Fırat Matur
Department of Biomedical Engineering
Bahçeşehir University
İstanbul, 34450, Turkey
firat.matur@bau.edu.tr

Bora Büyüksaraç
Department of Biomedical Engineering
Bahçeşehir University
İstanbul, 34450, Turkey
bora.buyuksarac@bau.edu.tr

***Abstract*—Fixed-gain Proportional-Integral-Derivative (PID) controllers are common in lower-limb exoskeletons, but their performance under varying dynamic loads is poorly understood. This proof-of-concept study addresses this gap by systematically evaluating a fixed-gain PID controller for a simulated knee joint model. The controller's performance was quantified while tracking an experimental gait trajectory at a slow (4.5 km/h) and a fast (6.3 km/h) walking speed, both on a 3.5° incline. The results revealed a clear performance contrast, with the mean Root Mean Square Error (RMSE) increasing from a high-precision 0.66 degrees in the slow condition to a clinically significant 9.77 degrees in the fast condition. This finding demonstrates the inherent brittleness of a statically optimized controller and reframes the central challenge in exoskeleton design, shifting the focus from single-state optimization toward the development of robust, adaptive strategies for dynamic, real-world tasks.**

***Keywords*—PID Control; System Modeling; Exoskeleton; Gait Simulation; Biomechanics; Control Systems.**

I. Introduction

Wearable lower-limb exoskeletons are transformative medical devices in rehabilitation engineering, designed to restore or augment gait in individuals with motor impairments such as spinal cord injury (SCI), cerebral palsy, and stroke [1]–[3]. A cornerstone of these systems is the Proportional-Integral-Derivative (PID) controller, a classical control method frequently employed due to its straightforward implementation [4]. While considerable progress has been made in the mechanical design and actuation of these devices, the performance of their underlying control systems under dynamic, real-world conditions remains a significant challenge [5].

A critical gap persists in the quantitative evaluation of these control strategies. The majority of existing studies evaluate fixed-gain PID controllers under static conditions, often failing to maintain accurate tracking performance when faced with dynamic tasks [2], [4]. Furthermore, many evaluations rely on generalized system models rather than focusing on joint-specific biomechanics, which is crucial for advancing real-world applications [6]. The knee joint, in particular, plays a vital physiological role in supporting body weight, enabling forward propulsion, and shock absorption. A lack of focus on accurate, joint-specific modeling limits the ability to properly assess system performance and design user-friendly devices.

This study addresses these limitations by evaluating the feasibility and tracking performance of a fixed-gain PID control strategy applied to a single-degree-of-freedom (1-DOF) knee joint model. Using a reference knee trajectory derived from experimental gait data, this work quantifies controller responsiveness under varying physical demands within a Simulink environment. By focusing on a targeted, single-joint model, this proof-of-concept study provides a quantitative analysis of a classical control method, aiming to highlight its inherent limitations and inform the development of more robust, adaptive strategies for exoskeleton devices.

II. Methodology

A. Experimental Protocol and Data Acquisition

The reference gait data for this study were acquired from a single healthy, right-footed as determined by a ball-kicking test [7], male participant (age: 27 years; height: 184 cm; weight: 80 kg), as determined by a ball-kicking test. with no history of musculoskeletal or neurological disorders. After providing written informed consent in accordance with the Declaration of Helsinki, the participant performed two-minute walking trials on a calibrated Ultima DC-5000 treadmill under two distinct conditions: (1) 4.5 km/h and (2) 6.3 km/h, both at a 3.5° incline, with a five-minute rest period between trials to prevent fatigue. Three-dimensional kinematics of the right leg were captured at 100 Hz using an Xsens Awinda wireless motion capture system (Xsens Technologies B.V.,

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Enschede, The Netherlands), the resulting knee joint angle data were reported per International Society of Biomechanics (ISB) conventions [8],and exported as time-series signals to serve as the simulation's reference trajectory.

B. System Modeling

The knee joint was modeled as a single-degree-of-freedom (1-DOF) rotational system constrained to sagittal plane flexion-extension to isolate the biomechanically dominant motion for focused controller evaluation. The joint dynamics were represented by a second-order differential equation with constant parameters for moment of inertia (I = 0.1 kg·m²), damping (b = 0.4), and stiffness (k = 10), selected from literature to ensure a stable, underdamped response. The complete system was implemented within the Simulink environment of MATLAB (R2025b, The MathWorks, Inc., Natick, MA) to facilitate dynamic modeling and control strategy integration.

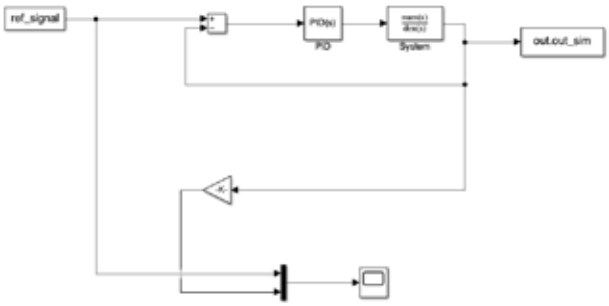


Fig. 1. Simulink Model for Gait Trajectory Tracking. The closed-loop configuration used for both test conditions, where an experimental gait trajectory (ref_signal) serves as the input to the PID controller and the 1-DOF knee joint model (System).

C. Controller Design

A proportional-integral-derivative (PID) controller was implemented to regulate the torque applied to the knee joint model and achieve accurate trajectory tracking. This control strategy was selected due to its ubiquitous application in biomechanical systems and its established effectiveness in minimizing steady-state error. The controller was implemented in a closed-loop configuration to continuously reduce the error between the desired and actual joint angle trajectories.

The final controller parameters were determined through an iterative tuning process involving a narrow-range search to optimize tracking performance. The optimized proportional (K_p), integral (K_i), and derivative (K_d) gains were set to 30.8, 0.1, and 0.7, respectively. These gain values were then held constant across both test conditions to provide a direct evaluation of the fixed-gain controller's performance when subjected to varying physical demands.

D. Simulation Setup

The responsiveness of the fixed-gain PID control strategy was evaluated in Simulink under two operational conditions using experimentally acquired knee angle data as the reference input. After processing into .mat files, the reference and simulated time series were synchronized via linear interpolation and visualized using a Mux block. To assess performance under different computational constraints, Test 1 was run with a variable-step solver (ODE45) for high accuracy, while Test 2 used a fixed-step solver (0.01 s step size) to precisely match the 100 Hz sampling rate of the gait data. All performance metrics were calculated over the first 10 gait cycles of the reference trajectory, using the model configuration shown in (Fig. 1).

E. System Characterization

The theoretical response of the open-loop 1-DOF knee joint model was characterized through transfer function analysis. Applying the Laplace transform to the system's governing second-order differential equation, the transfer function G(s) relating the input torque $\tau(s)$ to the output joint angle $\theta(s)$ was derived (Eq. 1).

$$G(s) = \frac{\theta(s)}{\tau(s)} = \frac{1}{0.1s^2 + 0.4s + 10} \tag{1}$$

Solving the characteristic equation yielded a pair of complex conjugate poles at -2 ± j9.798. The negative real parts of these poles confirm that the physical plant model is inherently stable, while their complex nature indicates a theoretical underdamped response characteristic prior to the application of closed-loop control.

F. Performance Quantification

System performance was evaluated using two categories of metrics: tracking accuracy and dynamic response characteristics. Tracking accuracy was quantified in degrees using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Maximum Absolute Error (MaxAE). RMSE, which gives greater weight to larger deviations, was calculated per Equation 2, where $y_{\text{target},i}$ is the desired angle, $y_{\text{actual},i}$ is the simulated angle, and n is the number of samples.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\text{target},i} - y_{\text{actual},i})^2} \tag{2}$$

Additionally, the system's transient behavior under each load was assessed by quantifying its dynamic response characteristics, conceptually illustrated in Fig. 2. These metrics, calculated from the average of 10 pseudo-step responses within the gait data, included: Natural Frequency (ω_n), Damping Ratio (ζ), Rise Time, Peak Time, Overshoot (%), and Settling Time to provide a comprehensive profile of system stability and responsiveness.

III. Results

The PID-controlled system exhibited a notable divergence in performance across the two simulated test conditions. In the 4.5 km/h condition (Test 1), the controller achieved high-precision tracking, with a RMSE of only 0.66°. In contrast, the more demanding 6.3 km/h condition (Test 2) resulted in

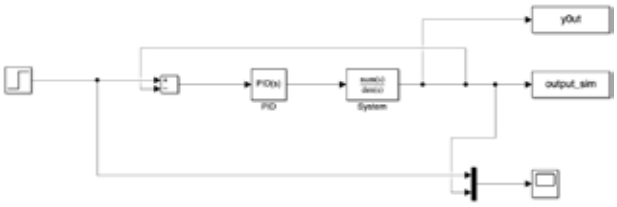


Fig. 2. Conceptual Model for System Step Response. Standard configuration for characterizing the transient response (e.g., overshoot, rise time) to a step input, providing the conceptual basis for the dynamic metrics derived from pseudo-step analysis of the experimental gait data.

a significant degradation in performance, with the RMSE increasing to 9.77°. This trend was mirrored across all accuracy metrics, including a substantial increase in MaxAE from 1.78° in Test 1 to 24.57° in Test 2. The complete summary of tracking accuracy is detailed in Table I.

TABLE I
SUMMARY OF CONTROLLER TRACKING ACCURACY

Performance Metric	Test 1	Test 2
Root Mean Square Error (RMSE)	0.66°	9.77°
Mean Absolute Error (MAE)	0.46°	7.92°
Maximum Absolute Error (MaxAE)	1.78°	24.57°

This divergence in tracking accuracy corresponded to a fundamental shift in the system's dynamic response characteristics, detailed in Table II. The system response in Test 1 was well-damped, with a calculated damping ratio (ζ) of 0.71 and a minimal overshoot of 3.2%. Conversely, the response in Test 2 became highly underdamped (ζ = 0.21), which manifested as a significant overshoot of 49.4% and a much faster, more oscillatory rise time (0.04 s compared to 0.27 s in Test 1).

TABLE II
SUMMARY OF CLOSED-LOOP DYNAMIC RESPONSE CHARACTERISTICS

Dynamic Metric	Test 1	Test 2
Natural Frequency (ω_n)	2.9 rad/s	2.8 rad/s
Damping Ratio (ζ)	0.71	0.21
Overshoot (%)	3.2%	49.4%
Rise Time (s)	0.27 s	0.04 s
Peak Time (s)	1.44 s	1.13 s
Settling Time (s)	2.31 s	1.66 s

A qualitative visual analysis of the simulation outputs, presented in (Fig. 3) and (Fig. 4), supports these quantitative findings. The plots clearly illustrate the high-fidelity, stable tracking achieved in Test 1, contrasted with the significant oscillatory behavior and pronounced tracking error present in Test 2.

IV. DISCUSSION

A. Interpretation of Findings

The simulation results reveal a distinct difference in the performance of the fixed-gain PID controller, contingent on

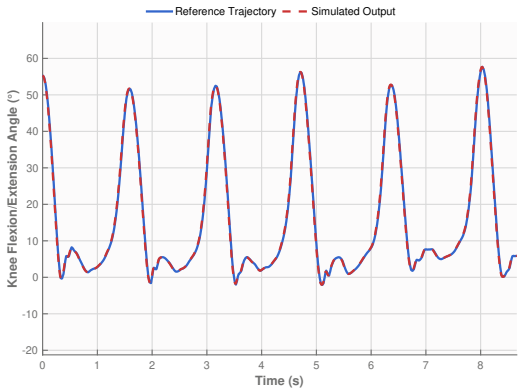


Fig. 3. Trajectory Tracking for Test 1 (4.5 km/h). The simulated knee angle (blue) shows high-precision, well-damped tracking of the reference trajectory (black).

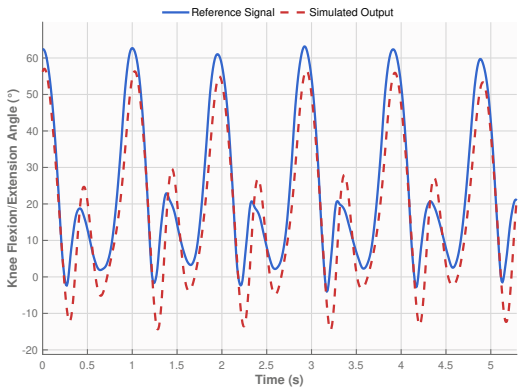


Fig. 4. Trajectory Tracking for Test 2 (6.3 km/h). The plot reveals significant performance degradation, with large errors and oscillatory behavior.

the specific operational condition. In the 4.5 km/h test condition, the controller demonstrated high-precision, stable control, achieving a RMSE of 0.66°. This high level of accuracy is consistent with findings from similar simulation-based studies of PID-controlled rehabilitation robotics [9]. According to the clinical error framework proposed by McGinley et al. (2009), an error of this magnitude (¡2°) is considered clinically insignificant, highlighting the capacity of a tuned PID controller to achieve excellent performance under specific, matched conditions [10]. The well-damped nature of this response (ζ = 0.71) and minimal overshoot (3.2%) further corroborate this high level of performance.

In contrast, when subjected to the higher physical demands of the 6.3 km/h condition, the controller's performance degraded significantly. Since the incline level was held constant across both tests, this performance degradation can be primarily attributed to the increased walking speed, which demanded faster and larger torque adaptations from the controller. The RMSE increased to 9.77°, a value that exceeds the 5° threshold for clinically significant error as defined by McGinley et al. (2009) [10]. This pronounced loss of accuracy is directly explained by a shift in the system's dynamic behavior, as the

response became highly underdamped ($\zeta = 0.21$) and resulted in substantial overshoot (49.4%).

This contrast between high-precision performance in one state and clinically significant error in another reveals the fundamental brittleness of non-adaptive control strategies: their inability to maintain consistent performance when conditions change. While a fixed-gain PID can be optimized for a specific task, it lacks the robustness to maintain that high level of performance when environmental dynamics change. This finding provides a clear, quantitative rationale for the development of more advanced control architectures, such as the adaptive or model-predictive strategies classified by Baud et al.(2021), which are designed to be more suitable for dynamic environments through their intrinsic ability to adjust control parameters in real-time [4].

B. Clinical and Practical Implications

These findings provide a practical design roadmap by quantitatively defining the operational boundaries of fixed-gain PID control. The results validate this approach as a robust and cost-effective solution for devices in constrained environments, such as fixed-speed clinical gait training [5], while its performance degradation under dynamic loads confirms its insufficiency for real-world applications, thereby guiding the implementation of adaptive strategies.

For rehabilitation engineering, the results have direct implications for user safety and experience. The substantial 49.4% overshoot in the high-speed condition highlights a critical safety consideration, as this can create an unpredictable assistive force that may compromise user confidence. This finding underscores the need to evaluate controllers for stable, predictable behavior, a central goal in the design of assistive control strategies [4].

For clinical practice, this work demonstrates that while exoskeletons can provide highly accurate motion within their optimal window, exceeding those limits can lead to a clinically significant loss of support. This highlights the importance of operating such technology within its validated parameters to ensure patients receive the consistent therapeutic motion essential for maximizing motor recovery and functional improvement [1], [2].

C. Limitations and Future Directions

It is crucial to interpret these findings within the context of several key limitations inherent to this proof-of-concept simulation study. The study was restricted to a single healthy participant, and the results may not directly translate to diverse populations such as elderly or clinical cohorts with specific motor impairments.

Furthermore, the controller was tuned empirically and evaluated on a simplified, noise-free 1-DOF model, which neglects critical factors like multi-planar dynamics and real-world disturbances inherent to human locomotion. Future work is therefore essential to validate these principles across diverse user groups and to evaluate controller robustness using more complex, multi-axis models subjected to realistic confounders like sensor noise and user-specific torque variability, paving the way for the development of truly personalized and clinically-effective adaptive systems.

V. Conclusion

This work demonstrates that a fixed-gain PID controller, while capable of high-precision, clinically-acceptable performance under specific, tuned conditions, exhibits a sharp degradation to a state of clinically significant error when subjected to increased dynamic demands. This finding challenges the conventional pursuit of a single, "optimal" set of tuning parameters by revealing that such optimization can create a brittle system, unable to maintain performance across a user's full operational range. This study reframes the central challenge in exoskeleton control design: the goal is not merely to implement a functional controller, but to actively manage the inherent trade-off between performance and robustness, shifting the focus toward intelligent, adaptive strategies that can ensure both safety and high-fidelity assistance in dynamic, real-world environments.

References

[1] J. L. Contreras-Vidal et al., "Powered exoskeletons for bipedal locomotion after spinal cord injury," *Journal of Neural Engineering*, p. 031001 (16pp), Apr. 2016, doi: 10.1088/1741-2560/13/3/031001.

[2] J. De Miguel-Fernández, J. Lobo-Prat, E. Prinsen, J. M. Font-Llagunes, and L. Marchal-Crespo, "Control strategies used in lower limb exoskeletons for gait rehabilitation after brain injury: a systematic review and analysis of clinical effectiveness," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, p. 23, Feb. 2023, doi: 10.1186/s12984-023-01144-5.

[3] J. A. De La Tejera, R. Bustamante-Bello, R. A. Ramirez-Mendoza, and J. Izquierdo-Reyes, "Systematic Review of Exoskeletons towards a General Categorization Model Proposal," *Applied Sciences*, vol. 11, no. 1, p. 76, Dec. 2020, doi: 10.3390/app11010076.

[4] R. Baud, A. R. Manzoori, A. Ijspeert, and M. Bouri, "Review of control strategies for lower-limb exoskeletons to assist gait," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, Jul. 2021, doi: 10.1186/s12984-021-00906-3.

[5] H. Ma, C. Zhong, B. Chen, K.-M. Chan, and W.-H. Liao, "User-Adaptive assistance of assistive knee braces for GAIT rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 10, pp. 1994-2003, Oct. 2018, doi: 10.1109/TNSRE.2018.2868693.

[6] Ö. Hižaroğlu and A. Akkurt, "Performing System Modeling Study of Knee Joint Simulator Prototype in accordance with ISO 14243/3 Standard and Simulating Gait Profile in Matlab Simulink Environment," *Gazi Journal of Engineering Sciences*, vol. 9, no. 3, pp. 622–633, Sep. 2023, doi: 10.30855/gmbd.0705093.

[7] N. Van Melick, B. M. Meddeler, T. J. Hoogeboom, M. W. G. N.-V. D. Sanden, and R. E. H. Van Cingel, "How to determine leg dominance: The agreement between self-reported and observed performance in healthy adults," *PLoS ONE*, vol. 12, no. 12, p. e0189876, Dec. 2017, doi: 10.1371/journal.pone.0189876.

[8] G. Wu et al., "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine," *Journal of Biomechanics*, vol. 35, no. 4, pp. 543–548, Apr. 2002, doi: 10.1016/s0021-9290(01)00222-6.

[9] Y. Yang, J. Wang, L. Shi, and J. Li, "GAIT simulation research of 2-DOF lower limb rehabilitation robot based on ADAMS and MATLAB," *Journal of Physics: Conference Series*, vol. 2414, no. 1, p. 012020, Dec. 2022, doi: 10.1088/1742-6596/2414/1/012020.

[10] J. L. McGinley, R. Baker, R. Wolfe, and M. E. Morris, "The reliability of three-dimensional kinematic gait measurements: A systematic review," *Gait & Posture*, vol. 29, no. 3, pp. 360–369, Mar. 2009, doi: 10.1016/j.gaitpost.2008.09.003.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Integrating CLAHE and Image Enhancement with Deep Learning for Multi-Class Brain CT Analysis

Arda Senyuz[†], Begum Gencer[†], Miray Solmaz[†], Yakup Atici[†], Erdem Okur[†], Ali Salbas[‡], Ali Murat Koc[‡]
[†]Izmir University of Economics, Software Engineering, Izmir, Turkiye
[‡]Izmir Katip Celebi University, Ataturk Training and Research Hospital, Department of Radiology

***Abstract*—Rapid and accurate detection of brain pathologies such as hemorrhage, infarction, and tumors is crucial in emergency care. Computed tomography is the primary imaging modality for acute cases, yet noise, low contrast, and acquisition variability challenge accurate diagnosis. This study proposes a computer-aided diagnosis framework combining preprocessing, custom enhancement, Contrast Limited Adaptive Histogram Equalization, and deep convolutional neural networks for four-class brain CT classification. Five architectures—InceptionV3, ResNet-50, EfficientNetB0, DenseNet-201, and Xception—were trained under multiple enhancement and transfer learning configurations. DenseNet-201 with transfer learning and combined enhancements achieved the highest accuracy with 88.14%, specificity with 95.78%, and a balanced sensitivity with 86.60%, demonstrating the benefit of multi-enhancement pipelines and pre-trained feature extractors. Results support integrating enhancement-driven CNNs into CAD systems to improve diagnostic accuracy in low-contrast CT imaging.**

***Keywords*—CT classification, hemorrhage detection, infarction detection, brain tumor detection, image enhancement, computer-aided diagnosis.**

I. Introduction

The human brain is the central organ of the nervous system, responsible for regulating voluntary and involuntary bodily functions, cognition, memory, and coordination [1]. Any acute intracranial pathology can have devastating consequences on a patient's health, affecting motor skills, perception, and even personality. Among the most critical and time-sensitive brain abnormalities are hemorrhages, infarction, and tumors [2][3].

A hemorrhage occurs when a blood vessel in the brain ruptures, causing internal bleeding and damaging surrounding tissue. Common causes include trauma and hypertension. Infarctions, typically resulting from ischemic strokes, occur when blood flow to part of the brain is obstructed by a clot or narrowed artery, leading to ischemia, hypoxia, and tissue necrosis. Brain tumors, whether benign or malignant, result from abnormal and uncontrolled cell proliferation within or around brain structures. Primary tumors originate in the brain, whereas secondary (metastatic) tumors spread from other parts of the body [2][3].

The timely diagnosis of these conditions is critical for effective treatment planning. Computed tomography (CT) scans remain the most widely used imaging modality for emergency brain assessments due to their speed, accessibility, and ability to detect acute bleeding. However, accurate detection can be challenging due to noise, low contrast between normal and pathological tissues, and variability in scanner settings. These limitations can obscure subtle lesions, making diagnosis highly dependent on radiologist expertise and experience.

The evolution of artificial intelligence (AI) and deep learning has resulted in automatic diagnostic tools capable of supplementing a clinician's role by identifying and classifying brain CTs with high specificity. Convolutional neural networks (CNNs) have demonstrated strong results across medical imaging applications, including tumor detection and hemorrhage segmentation [4][5][6][7]. Pre-trained architectures such as ResNet-50, InceptionV3, and DenseNet-201 can be fine-tuned using transfer learning to achieve improved performance on limited medical datasets [8][9][10].

Recent studies have explored various deep learning strategies for classifying brain pathologies from CT images, including tumors, infarcts, and hemorrhages. Transfer learning approaches, as shown in comparative studies, leverage pre-trained weights from natural image datasets to extract relevant features from grayscale CT scans, even with limited data. Other research has explored few-shot learning, volumetric CNNs, slice dependency frameworks, and anatomically-aware data augmentation to address challenges like data scarcity, low contrast, and class imbalance [11][12][13][14].

Among preprocessing techniques, Contrast Limited Adaptive Histogram Equalization (CLAHE) [15] has shown promise in enhancing local contrast and revealing subtle anatomical features [16]. Studies combining CLAHE with augmentation have demonstrated improved classification performance, supporting the integration of enhancement methods into CAD pipelines.

This paper presents a comprehensive classification framework for the detection of hemorrhage, infarction, and tumors on brain CT scans. Our approach integrates preprocessing, enhancement (custom method and CLAHE), augmentation, and deep learning classification, aiming to maximize model performance while maintaining clinical interpretability.

The paper is structured as follows; Section II explains the methodology and the image enhancement technique, Section III lays out the experimental results and discusses the performance of trained models and Section IV concludes the paper.

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

II. Methodology

This study employed a structured pipeline comprising preprocessing, enhancement, augmentation, model training, and evaluation for classifying brain CT scans into four categories: Tumor, Hemorrhage, Infarction, and Healthy.

A. Data Collection and Preparation

The dataset was sourced from İzmir Katip Celebi University Atatürk Education and Research Hospital, initially in DICOM format. 108 anonymous patient's various CT scan slices contributed to the dataset. Slices were labeled according to diagnosis and organized into corresponding classess. The dataset also includes the demographic information on patients such as age and gender. Class balance was achieved through random downsampling, resulting in an equal number of samples per class. All scans were converted to JPG using Weasis [17] to enable further processing. The dataset was then, split into 80% training and 20% validation set, with a separate test set reserved for final evaluation.

B. Data Augmentation

In order to achieve a good amount of training data, geometric augmentations, including random rotations (±15°), scaling (±10%), and translations (±10 pixels), were applied to training set which increased the amount and diversity. Reflection-based augmentations, however, avoided due to anatomical constraints.

C. Image Enhancement

Two enhancement techniques were applied to the original dataset with an attempt to improve the image quality to make features more perceivable for network models to be trained.

First, a custom Brightness Map (BRI) method previously used successfully for the detection of melanoma cases by Okur and Turkan [18]. The method consists of creating a "weight map" from the original image and reapplying it back to the image. It employs a specific definition of pixel brightness, where brightness is the absolute difference from the average of a pixel's red, green, and blue values. This brightness value is treated as highlighter for intensity differences to improve feature extraction. A feature map for brightness is generated for each color channel. These feature maps are then normalized to a [0, 1] range, and used to directly weight individual pixels within the original image.

Second, Contrast Limited Adaptive Histogram Equalization (CLAHE) is also used on the original dataset [15]. As it is mentioned in Section I, CLAHE proved its success already in similar tasks. CLAHE is an image processing technique used to improve the local contrast of an image. Unlike traditional histogram equalization, which applies a single, global contrast adjustment to the entire image, CLAHE operates on smaller, individual regions of the image, known as tiles or blocks. This localized approach prevents the algorithm from excessively brightening or darkening large areas of the image, which is sometimes a common problem with global methods. CLAHE is particularly useful for enhancing images with poor or uneven lighting similar to CT scans, as it can improve the visibility of details in both bright and dark areas simultaneously.

At the end of this image enhancement step, three new datasets were generated which makes total of 4 datasets: original images (R), BRI-enhanced (B), CLAHE-enhanced (C), and BRI+CLAHE-enhanced (BC). Fig. 1 shows an example per class and their enhanced versions from 4 dataset.

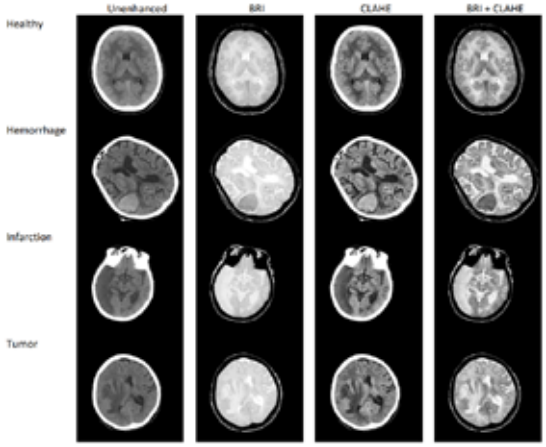


Fig. 1: Example images from 4 datasets prepared from each available class.

D. Model Architectures and Training

Five convolutional neural network (CNN) architectures—InceptionV3, ResNet-50, EfficientNetB0, DenseNet-201, and Xception—were selected for evaluation due to their proven track records in medical image classification and their accessibility through widely used deep learning frameworks [8][9][19][10][20]. These models are readily available with pre-trained weights on large-scale datasets such as ImageNet [21], which makes them ideal for transfer learning in scenarios where collecting large, domain-specific datasets is challenging. Their inclusion was also motivated by ease of integration, as each architecture is well-documented, supported by popular libraries like TensorFlow [22] and PyTorch [23], and benefits from active research communities. Individually, each model offers distinct advantages: InceptionV3's inception modules allow multi-scale feature extraction, which is useful for capturing lesions of varying sizes; ResNet-50's deep residual connections enable effective gradient flow and training stability, making it a robust baseline; EfficientNetB0 strikes a balance between accuracy and computational efficiency, making it suitable for resource-limited environments; DenseNet-201 promotes feature reuse and gradient propagation through dense connectivity, improving learning efficiency; and Xception leverages depthwise separable convolutions for reduced parameter count while maintaining strong representational power. In all experiments, the final classification layers were adapted for the four-class output (tumor, hemorrhage, infarction, and healthy), while training hyperparameters such as learning rate, batch size, and number of epochs were kept constant to ensure a fair comparison across all configurations,

TABLE I: Top 10 Model Performances

Model	**Accuracy** (%)	**Specificity** (%)	**Sensitivity** (%)	**NPV** (%)	**Precision** (%)	**F1-score** (%)
TL DenseNet201+RCB	88.14	95.78	86.60	96.06	89.24	87.48
TL InceptionV3+R	87.71	95.56	87.14	95.78	89.62	88.05
TL Xception+R	87.08	95.50	86.81	95.57	87.48	86.83
TL DenseNet201+CB	86.86	95.32	85.66	95.55	88.26	86.60
TL DenseNet201+B	86.65	95.32	86.07	95.43	87.50	86.49
TL ResNet50+RCB	86.44	95.19	85.20	95.46	87.82	86.06
TL InceptionV3+RCB	86.02	95.11	84.79	95.25	86.08	85.38
TL ResNet50+B	86.02	95.04	84.84	95.23	86.94	85.72
TL EfficientNetB0+R	85.59	94.89	84.78	95.05	86.74	85.53
TL DenseNet201+RB	85.59	94.85	83.73	95.20	87.55	84.97

which varied in transfer learning usage, enhancement methods, and augmentation strategies.

III. Experimental Results and Discussion

Experiments in this study includes a wide range comparison of the chosen five network model in terms of training type and dataset used. Empty versions of the five models have trained from scratch with four datasets separately and also pretrained versions of them used for transfer learning again with four datasets. With the combinations mentioned above, total of 40 different models are tested in the experiments.

The hardware used in these is a laptop with an Intel Core $i7$ 11800H CPU and 16 GB memory accompanied with an NVIDIA RTX 3060 gpu chipset. All experiments are done on MATLAB 2024b [24]. Performance was evaluated using accuracy, sensitivity, specificity, precision, negative predictive value and F1-score to ensure comprehensive assessment of each model's diagnostic capabilities. Table I presents the results acquired with these experiments. To be concise and clear, only top 10 successful models are included in the table.

The top-10 results reveal several important trends regarding the influence of model architecture, enhancement techniques, and transfer learning on classification performance. All models in the top-10 list were trained using transfer learning (TL), underscoring its effectiveness for brain CT classification, especially when datasets are limited. None of the from-scratch (RW) models managed to outperform the TL-based counterparts, highlighting the benefit of leveraging pre-trained weights from large-scale datasets such as ImageNet to extract robust and transferable features, at least when the acquired dataset size is small.

DenseNet-201 appears most frequently among the top-performing configurations, indicating that its densely connected architecture and feature reuse mechanism are particularly effective in capturing the fine-grained structural details present in CT scans. The top-performing model, TL DenseNet201+RCB, which combines original, CLAHE-enhanced, and custom-enhanced images, achieved the highest accuracy (88.14%) and specificity (95.78%), suggesting that integrating multiple enhancement techniques provides complementary benefits. This fusion likely enriches the diversity of image features, enabling the network to generalize better across different tissue contrast conditions.

InceptionV3 and Xception also appear among the leaders, particularly when trained with original images (R) or a combination of enhancements. InceptionV3's multi-scale feature extraction likely helps capture both localized lesions and global structural abnormalities, while Xception's depthwise separable convolutions may improve efficiency in learning texture variations present in hemorrhage, infarction, and tumor cases.

Interestingly, ResNet-50 appears lower in the top 10, but still with strong performance when trained with multiple enhancements (RCB or B). This suggests that while residual connections help in stable training, ResNet-50's relatively lower representational capacity compared to DenseNet-201 might require the richer feature diversity provided by enhancement combinations to compete effectively.

From an enhancement perspective, models trained with combinations of R, C, and B consistently outperform those trained on a single image type. This indicates that no single enhancement method is optimal for all cases—rather, combining them gives the model a more comprehensive view of the data. CLAHE (C) improves local contrast and structural detail, the custom enhancement (B) amplifies subtle intensity differences, and the original images (R) preserve the natural appearance, all of which complement one another.

Overall, the results suggest that DenseNet-201 with multi-enhancement training and transfer learning provides the best balance of accuracy, specificity, and sensitivity. Other architectures like InceptionV3 and Xception remain competitive, particularly when computational efficiency or multi-scale feature extraction is prioritized.

IV. Conclusion

This study demonstrates that integrating targeted image enhancement with modern deep learning architectures significantly improves classification performance in brain CT analysis. Among the evaluated models, DenseNet-201 trained with transfer learning and multi-enhancement data achieved the highest overall accuracy and specificity, underscoring the value of combining complementary preprocessing strategies. Transfer learning consistently outperformed from-scratch

training, reflecting the advantage of leveraging pre-trained feature extractors when working with limited medical datasets. The consistent presence of InceptionV3 and Xception among the top models further highlights the importance of multi-scale feature extraction and efficient convolutional operations in capturing both global and local lesion characteristics. Future work should focus on expanding the dataset, incorporating volumetric and slice-dependency information, and validating the framework in clinical settings to assess real-world diagnostic impact. The findings support the adoption of enhancement-driven, transfer learning-based CNN models in CAD systems to aid radiologists in the rapid and reliable detection of life-threatening brain abnormalities.

REFERENCES

[1] "Brain," https://my.clevelandclinic.org/health/body/22638-brain, Accessed: 2025-08-13.

[2] "Brain bleed, hemorrhage (intracranial hemorrhage)," https://my.clevelandclinic.org/health/diseases/14480-brain-bleed-hemorrhage-intracranial-hemorrhage, Accessed: 2025-08-13.

[3] "Brain cancer (brain tumor)," https://my.clevelandclinic.org/health/diseases/6149-brain-cancer-brain-tumor, Accessed: 2025-08-13.

[4] Yoseob Han and Jong Chul Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *IEEE Tran. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, 2018.

[5] Eunhee Kang, Won Chang, Jaejun Yoo, and Jong Chul Ye, "Deep convolutional framelet denosing for low-dose ct via wavelet residual network," *IEEE Tran. Med. Imag.*, vol. 37, no. 6, pp. 1358–1369, 2018.

[6] Lu Li, Meng Wei, Bo Liu, Kunakorn Atchaneeyasakul, Fugen Zhou, Zehao Pan, Shimran A. Kumar, Jason Y. Zhang, Yuehua Pu, David S. Liebeskind, and Fabien Scalzo, "Deep learning for hemorrhagic lesion detection and segmentation on brain ct images," *IEEE Jour. Biomed. Heal. Info.*, vol. 25, no. 5, pp. 1646–1659, 2021.

[7] Monika Grewal, Muktabh Mayank Srivastava, Pulkit Kumar, and Srikrishna Varadarajan, "Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans," in *2018 IEEE 15th International Symposium on Biomedical Imaging*, 2018, pp. 281–284.

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[10] Naveenkumar M, Manoj R, Nandhakumar B, and Rahul R, "Densenet201 for animal detection and repellent system," in *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)*, 2022, pp. 213–218.

[11] M. Dawood, S. Ali, and R. Hassan, "Classification of brain tumors using pre-trained cnn architectures: Vgg-16, resnet-50, and mobilenet-v2," *Journal of Medical Imaging and Health Informatics*, vol. 13, no. 2, pp. 113–120, 2023.

[12] Y. Hu, Q. Zhang, and L. Wang, "Splicing learning: Few-shot learning with data augmentation for medical image classification," *IEEE Access*, vol. 9, pp. 110234–110245, 2021.

[13] W. Kuo, C. Häne, E.L. Yuh, P. Mukherjee, J. Malik, and G. Zaharchuk, "Expert-level detection of acute intracranial hemorrhage on head ct using deep learning," *Nature Communications*, vol. 12, pp. 1–9, 2021.

[14] Z. Li, X. Zhang, Y. Liu, H. Chen, and Y. Yang, "Slice dependency learning for multi-label classification of brain ct images," *Medical Image Analysis*, vol. 64, pp. 101716, 2020.

[15] S.M. Pizer, R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, and K.E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990, pp. 337–345.

[16] Kamal Halloum and Hamid Ez-Zahraouy, "Enhancing medical image classification through transfer learning and clahe optimization," 2025, Accessed from https://pubmed.ncbi.nlm.nih.gov/40259867/.

[17] Nicolas Roduit, "Weasis dicom viewer," https://github.com/nroduit/Weasis, Accessed: 2025-08-13.

[18] Erdem Okur and Mehmet Turkan, "Patch enhancement for melanoma detection with bag of visual words," in *2022 Medical Technologies Congress (TIPTEKNO)*, 2022, pp. 1–4.

[19] Mingxing Tan and Quoc V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019.

[20] François Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016.

[21] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. Comp. Vis. Pat. Recog.*, 2009, pp. 248–255.

[22] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, Software available from tensorflow.org.

[23] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.

[24] The MathWorks Inc., "Matlab version: R2024b," 2024.

Graph Neural Networks-Based Digital Twin Modeling of WBC and ANC Dynamics for Personalized Time-Series Prediction

Utku Kose
Department of Computer Engineering
Suleyman Demirel University
Isparta, Turkey
utkukose@sdu.edu.tr

Gozde Ozkan Tukel
Department of Basic Sciences
Isparta Applied Sciences University
Isparta, Turkey
gozdetukel@isparta.edu.tr

Emine Betul Surucu
Department of Computer Engineering
Suleyman Demirel University
Isparta, Turkey
betulsen@sdu.edu.tr

Muhammet Enes Karatas
Department of Computer Engineering
Suleyman Demirel University
Isparta, Turkey
enesm.karatas@gmail.com

Ilhan Uysal
Department of Information Systems and Technologies
Burdur Mehmet Akif Ersoy University
Burdur,Turkey
iuysal@mehmetakif.edu.tr

***Abstract*— This study introduces a digital twin model approach with Graph Neural Networks (GNNs) to forecast white blood cell (WBC) and absolute neutrophil count (ANC) levels during 6-mercaptopurine (6-MP) chemotherapy for treating childhood acute lymphoblastic leukemia (ALL). For incorporating clinical and lifestyle factors like daily medication dosage, anthropometric measurements (weight, height, and body surface area), diet, exercise, and vitamin D intake, the model is constructed using a synthetic patient dataset based on personalized computational Pharmacokinetics (PK) / Pharmacodynamics (PD) mathematical models. To capture temporal relationships, each data point is connected to earlier observations and is defined as a node within a time-dependent graph structure. The model, which was created with PyTorch Geometric, was evaluated by using MAE, MSE, and RMSE metrics after being trained with inputs, which are normalized by z-scores. The study results point that the developed GNN-based method works well as a digital twin tool for customized chemotherapy simulations and can predict changes in hematological parameters over time.**

Keywords—graph neural networks; 6-mercaptopurine; acute lymphoblastic leukemia; hematological response

I. INTRODUCTION

Graph-based deep learning techniques have become more and more popular in biomedical research in recent years. In particular, accurate treatment process monitoring and the development of customized treatment plans depend heavily on the modelling of intricate temporal dependencies in patient data. Hematologic parameters, which can have a complicated and fluctuating course over time, are essential biomarkers for assessing chemotherapy response and managing treatment. It might not be possible to fully capture all these dynamics with conventional statistics or machine learning. In this regard, Graph Neural Networks (GNN) is an effective tool for resolving these issues since it treats every observation as a node and models temporal relationships using earlier observations.

A unique contribution in terms of methodology and social impact can be made by using a GNN architecture to model temporal changes in WBC and ANC in pediatric patients undergoing 6-MP treatment. Treatment-sensitive hematological parameters can be predicted earlier and more accurately with GNN models, which represent clinical parameters as nodes throughout the time series. This is especially useful for integration into digital twin applications. Such approach is an innovation for both the management of drug toxicity and personalized dose adjustments [1, 2]. In this study, a GNN-based model was developed by focusing on 6-MP, one of the drugs used in the treatment of childhood acute lymphoblastic leukemia (ALL). The study predicts the change of hematological parameters over time and presents an innovative digital twin-based approach for personalized chemotherapy simulations.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Predicting hematologic responses in diseases such as childhood ALL directly impacts treatment management. During maintenance therapy periods using drugs such as 6- MP, parameters such as WBC and ANC show large individual variation due to pharmacological and genetic factors. The authors in [3] and [4] developed population PK models to explain these processes and demonstrated the critical role of TPMT genotyping in dose individualization. On the other hand, the studies [5] and [6] modeled the PK/PD processes of 6-MP, incorporating factors such as drug absorption, metabolite levels, genetic polymorphisms, and hematological responses, and adapting these models to individual patient characteristics to predict treatment outcomes and assess toxicity risk. In their pharmacodynamic models for the diseases they study, the authors in [7] considered environmental conditions, immune system responses, and changes in metabolic processes that may affect infection dynamics. As associated with PK/PD dynamics covered by drug repositioning purposes, Köse et al. proposed an extended solution methodology, which combines deep learning models with digital twin-based perspective to lead disease modeling as well as treatment optimization in a decision support environment [8, 9]. In [10], Köse et al. worked on PK/PD models by considering previous PK/PD computational patient models. Called as STING, the proposed solution by Köse et al. [9] was moved further by including environmental factors (vitamin D, nutrition and sports) in PK/PD models, as introduced in this study. In this context, the personalized PK model for pediatric ALL patients can be obtained as a first-order linear differential equation system as follows:

$$\dot{x}_1(t) = -k_a x_1(t) + Fu(t),$$

$$\dot{x}_2(t) = k_a x_1(t) - k_{20} x_2(t),$$

$$\dot{x}_3(t) = 0.019(2.56)^{TPMT} k_{me} x_2(t) - 0.0012233 H^{0.4598} W^{0.6180} x_3(t) \quad (1)$$

$$\dot{x}_4(t) = (1 - 0.019(2.56)^{TPMT} - FM_5) k_{me} x_2(t) - 0.0228 x_4(t)$$

$$\dot{x}_5(t) = FM_5 k_{me} x_2(t) - 0.289 x_5(t),$$

where $x_1(t)$ and $x_2(t)$ are respectively the amount of 6-MP in plasma and the GI system, k_a is the absorption rate constant, k_{20} is the elimination rate constant of 6-MP from plasma, F is the bioavailability factor and $u(t)$ is the time-dependent drug dose, $x_3(t)$, $x_4(t)$ and $x_5(t)$ are respectively the concentration of 6-TGN, 6-mMPN and 6-TU in plasma, FM_3, FM_4 and FM_5 are the fraction of 6-MP converted to 6-TGN, 6-mMPN and 6-TU, respectively. k_{me} is metabolic rate constant and $CL_{6tgn}(BSA)$, CL_{6mpn} and $CL_{6tu}(BSA)$ are the clearance rate of 6-TGN, 6-mMPN and 6-TU from plasma, H is patient height and W is patient weight. PD model for pediatric ALL patients for WBC and ANC production are respectively given by the following equations

$$\dot{x}_6(t) = k_{prol} x_6(t) \left(1 - E_{drug} - \theta I_R(t)\right) \left(\frac{Base}{x_{10}(t)}\right)^{\gamma} - k_{tr} x_6(t) \quad (2)$$

and

$$\dot{x}_{11}(t) = k_{prol} x_{11}(t) \left(1 - E_{drug} - \theta I_R(t)\right) \left(\frac{Base}{x_{15}(t)}\right)^{\gamma} - k_{tr} x_{11}(t), \quad (3)$$

where θ is the weight of the suppressive effect of inflammation on WBC (and ANC in Eq. (3)) production, $I_R(t)$ is the level of inflammation that changes over time, x_6 is the amount of proliferative WBC precursors in the bone marrow, x_{10} is the amount of mature WBC in the blood, k_{prol} is the cell proliferation rate constant, k_{tr} is the transit rate constant, $Base$ is the initial (homeostatic) WBC level, and γ is the homeostatic feedback strength coefficient, $x_{11}(t)$ is the amount of proliferative ANC precursors in bone marrow, $x_{15}(t)$ are circulating mature neutrophils and E_{drug} is the drug effect (see [9] for details).

The model developed in this study simulates the temporal effects of the 6-MP drug used in the treatment of ALL hematological parameters through PK/PD mathematical models developed by Köse et al. (2025) and adapted to patient characteristics. These models generate synthetic patient data by individualizing processes such as drug absorption, distribution, metabolite conversion, and inflammatory response. With the time series data obtained, a time-dependent GNN model is trained, and the error metrics of the model are calculated for 30 randomly generated synthetic patient data. As a result of this evaluation, the average MAE for WBC estimates are 0.197, RMSE is 0.289; and the average MAE for ANC estimates is 0.0602, RMSE is 0.084.

This study demonstrates the significant potential offered by GNN models in creating more accurate and personalized treatment plans for ALL patients through the integrated use of deep learning and artificial intelligence methods with clinical data. Furthermore, the developed system makes a concrete contribution to digital twin technologies thanks to its ability to not only analyze historical data but also predict patients' future treatment responses.

II. METHODOLOGY

GNN-based approaches aim to predict changes in WBC and ANC in patients diagnosed with childhood ALL. The dataset related to this study was produced via PK and PD mathematical models developed by Köse et al. (2025) [10] and adapted to patient characteristics. The individualized PK model (1) and the PD models (2) and (3) obtained depending on environmental factors are called "real " for convenience. Python programming language is used for the estimation process. Python libraries such as PyTorch, Pandas, NumPy and Matplotlib are used in model development and data processing. The dataset used in the training of the model is loaded from the Runner.csv file containing GNN data and modeled in DataFrame format. The dataset includes both input features and target variables. The target variables are the WBC and ANC measurements. Introductory features of the model; Day (time in days), 6MP_Daily_Dose_mg (daily dose of 6-Mercaptopurine in mg), Weight, Height and BSA (physical characteristics of the patient: weight, height and body surface area), Diet, Exercise and Vitamin_D (nutritional quality, exercise level and vitamin D supplementation) and derived characteristics derived from the PK/PD model. These variables are converted to multidimensional array format using the NumPy library and made suitable for use in the GNN model.

GNN models consist of a graph structure that represents data points as nodes and the relationships between them as edges. In this study, each patient measurement corresponds to each data

point and is represented as a node. To model temporal relationships, each node is connected to nodes in the three preceding time steps. Bidirectional edges are created to create an undirected graph structure. The edge list is converted to the edge_index input format from the PyTorch library and combined with the Data object. If the edge list is empty, the process is terminated with an error message. The GNNRegressor class takes the values in_channels, hidden_channels, and out_channels as constructor parameters.

The trained model weights and trained GNN models are loaded from the pth file. The model is switched to evaluation mode (eval()) to make predictions. The model generates predictions using normalized features and a graph structure. torch.no_grad() is used to disable gradient calculations, which increases computational efficiency. Then, the model's outputs are converted back to the original scale by inverse normalization.

This study focuses on the usefulness of a GNN-based model for forecasting on time-series data from pediatric ALL patients. The model makes it possible to predict dynamic processes like the course of a disease and the response to treatment by using a graph representation to model the relationships between time-dependent multidimensional data. In this context, essentials steps were organized through preprocessing the data, establishing node and edge relationships, training the model, generating predictions, and visualizing the findings.

For analyzing the model's sustainability and consistency, each GNN model was run 30 times, and the error metrics (MAE, MSE, RMSE) of the WBC and ANC predictions were computed for each run. Based on synthetic patient data, the GNN model consistently performed well in predicting WBC and ANC, according to the error metrics. The model appears to better capture the spatial dynamics of this parameter, as indicated by the lower error rates for ANC. The model might need more features or a more intricate graph structure for this variable, though, given the comparatively higher error rates for WBC predictions. These findings support the potential of the digital twin approach in the assessment of drug effects.

III. Application and Findings

Considering the GNN-based model, further applications were done to forecast WBC and ANC values, thanks to the synthetic patient dataset support. At this point, the related model generated a digital twin flow for analyzing the drug effects during a childhood ALL treatment. It was done by detecting relationships among the time-dependent clinical variables. Fig. 1 presents the real and predicted WBC values in days. In this graph, the real WBC values are represented by the black line, and the model-predicted values are represented by the blue line. Similarly, Fig. 2 shows a comparison of the actual and predicted ANC values. Both plots demonstrate that the model successfully learned the structural and temporal features in the time series data, and the resulting predictions exhibit a trend that is highly consistent with the actual values.

Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) metrics were calculated to evaluate the prediction performance of the model in this example. The findings obtained are summarized in Table 1.

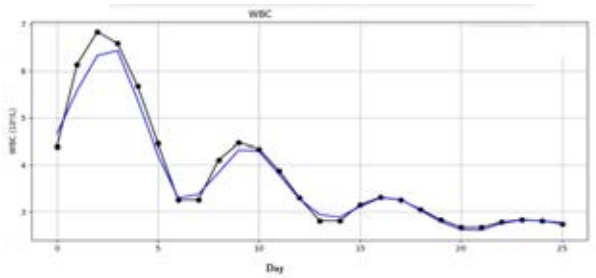


Fig. 1. WBC estimation with 6 MP GNN.

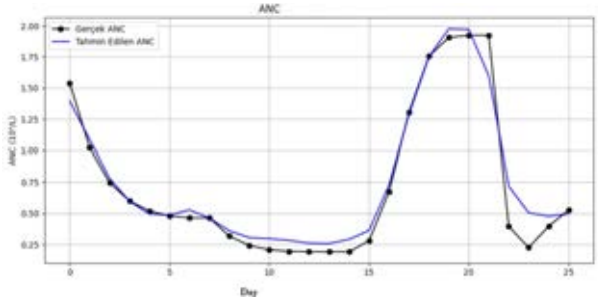


Fig. 2. ANC prediction with 6MP GNN.

TABLE I. TABLE 1. ERROR FINDINGS OF GNN ESTIMATIONS.

Parameter	MAE	MSE	RMSE
WBC	0.1251	0.0373	0.1932
ANC	0.0837	0.0148	0.1217

WBC Estimates: The MAE value is calculated as 0.1251, the MSE value was 0.0373, and the RMSE value is 0.1932. These metrics show that the model predicts WBC values with relatively low error rates. ANC Estimates: MAE value was 0.0837, MSE value was 0.0148 and RMSE value was 0.1217. Lower error values for ANC suggest that the model predicts this parameter with greater accuracy. The model is controlled by running it on different days according to different patient characteristics and visualizes some results as follows (Fig. 3).

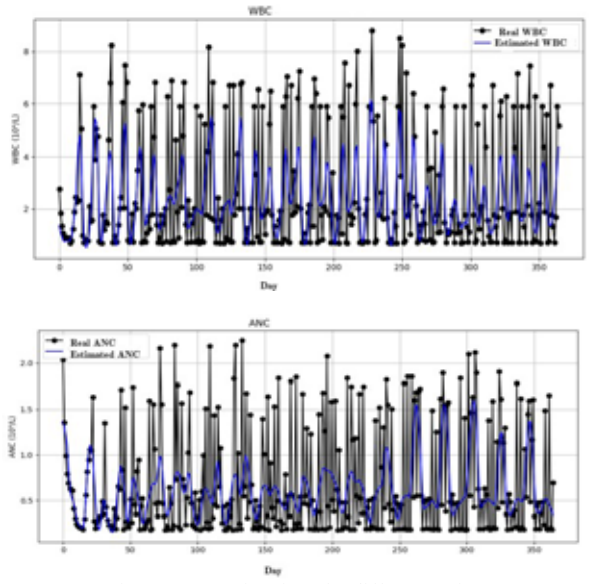


Fig. 3. GNN estimations for different patients.

To assess the sustainability and consistency of the GNN model, each model is run 30 times. For each run, error metrics (MAE, MSE, RMSE) are calculated for the WBC and ANC predictions. The mean and standard deviation of these metrics are then computed. Finally, the overall mean and standard deviation are obtained for the 30 simulated patients (Fig. 4).

	MAE	MSE	RMSE		MAE	MSE	RMSE
WBC	0.188	0.0787	0.28	ANC	0.056	0.00498	00.071
WBC	0.114	0.0224	0.15	ANC	0.062	0.00526	0.073
WBC	0.084	0.0119	0.109	ANC	0.053	0.00552	0.074
WBC	0.117	0.0303	0.174	ANC	0.051	0.00365	0.060
WBC	0.176	0.0465	0.216	ANC	0.050	0.00470	0.069
WBC	0.085	0.0199	0.141	ANC	0.052	0.00463	0.068
WBC	0.173	0.0476	0.218	ANC	0.061	0.00602	0.078
WBC	0.143	0.0456	0.214	ANC	0.061	0.0046	0.068
WBC	0.118	0.0255	0.16	ANC	0.063	0.00513	0.072
WBC	0.169	0.0698	0.264	ANC	0.055	0.0118	0.109
WBC	0.127	0.0271	0.165	ANC	0.051	0.00568	0.075
WBC	0.193	0.0874	0.296	ANC	0.059	0.00723	0.085
WBC	0.075	0.0105	0.103	ANC	0.052	0.00498	0.07
WBC	0.157	0.0446	0.211	ANC	0.05	0.00403	0.064
WBC	0.155	0.0494	0.222	ANC	0.051	0.00429	0.065
WBC	0.138	0.029	0.17	ANC	0.05	0.00432	0.066
WBC	0.2	0.0788	0.281	ANC	0.0444	0.00389	0.062
WBC	0.163	0.0377	0.194	ANC	0.047	0.00522	0.072
WBC	0.164	0.0613	0.248	ANC	0.072	0.0184	0.136
WBC	0.266	0.0834	0.289	ANC	0.058	0.00578	0.076
WBC	0.09	0.0202	0.142	ANC	0.059	0.00519	0.072
WBC	0.48	0.606	0.778	ANC	0.086	0.00947	0.097
WBC	0.491	0.511	0.715	ANC	0.136	0.0679	0.261
WBC	0.39	0.306	0.553	ANC	0.063	0.00725	0.085
WBC	0.448	0.733	0.856	ANC	0.081	0.0083	0.091
WBC	0.479	0.6	0.781	ANC	0.059	0.00619	0.079
WBC	0.138	0.0315	0.178	ANC	0.040	0.00396	0.063
WBC	0.104	0.0211	0.145	ANC	0.077	0.00830	0.091
WBC	0.118	0.0312	0.177	ANC	0.050	0.00412	0.064
WBC	0.168	0.0620	0.249	ANC	0.102	0.013	0.114

Fig. 4. Error findings regarding GNN estimations for 30 different patients.

The mean MAE value for the WBC estimates of the model is 0.197, the RMSE value is 0.289, and the standard deviations of these metrics are 0.124 and 0.210, respectively. These values reveal that the model may occasionally deviate from the WBC estimation but generally offers acceptable accuracy. In the ANC estimates, the mean MAE value is 0.0602 and the RMSE value is 0.084, with standard deviations of 0.019 and 0.0369, respectively. These findings suggest that the model provides lower errors and higher consistency in ANC predictions. The error metrics show that the GNN model performs consistently in predicting WBC and ANC on synthetic patient data. Lower error rates for ANC suggest that the model better captures the spatial dynamics in this parameter. However, relatively high error rates in WBC estimates indicate that the model may require additional features or a more complex graph structure for this variable.

The model developed in this study is structurally designed to be compatible with all drug models and integrated into every stage of pediatric ALL treatment process. The results of the individualized model, specifically for 6-MP, demonstrated that it produced highly accurate predictions based on the statistical tests performed. However, some limitations should be considered when interpreting the findings. The synthetic dataset used in training and testing the model may not fully reflect the full clinical and genetic diversity of the real patient population. Therefore, validation studies with real patient data are necessary to increase clinical applicability and generalizability of the developed model.

ACKNOWLEDGEMENT

This study is associated with the project: 'STING: Development of a Drug Repositioning Decision Support System for Childhood Acute Leukemia by Digital Twin-Oriented Deep Learning', which is funded by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Project No: 123E383).

REFERENCES

[1] Y. Meng, Y. Wang, J. Xu, C. Lu, X. Tang, T. Peng, B. Zhang, G. Tian, and J. Yang, "Drug repositioning based on weighted local information augmented graph neural network". Briefings in Bioinformatics, vol. 25, no. 1, Jan. 2024, Art. no. bbad431, https://doi.org/10.1093/bib/bbad431.

[2] S. Li and X. Pan, "A computational drug repositioning model based on hybrid similarity side information powered graph neural network". *Future Generation Computer Systems*, vol. 126, pp. 117–129, Sep. 2021, https://doi.org/10.1016/j.future.2021.06.018.

[3] D. Gebhardt, M. Dittrich, C. Meier, M. T. Cronin, and T. Lehr, "Physiologically based pharmacokinetic modeling of 6-mercaptopurine: Exploring the role of genetic polymorphism in TPMT enzyme activity". *CPT: Pharmacometrics & Systems Pharmacology*, vol. 4, no. 6, pp. 354–364, 2015, https://doi.org/10.1002/psp4.55.

[4] A.F. Hawwa, P.S. Collier, J.S. Millership, A. McCarthy, S. Dempsey, C. Cairns, and J.C. McElnay, "Population pharmacokinetic and pharmacogenetic analysis of 6-mercaptopurine in paediatric patients with acute lymphoblastic leukaemia". British journal of clinical pharmacology, 66(6), 826-837, 2008, https://doi.org/10.1111/j.1365-2125.2008.03281.x

[5] F. Jost, J. Zierk, T.T. Le, T. Raupach, M. Rauh, M. Suttorp, and S. Sager, "Model-based simulation of maintenance therapy of childhood acute lymphoblastic leukemia". Frontiers in Physiology, 11, 217, 2020, https://doi.org/10.3389/fphys.2020.00217

[6] T.T. Le, F. Jost, T. Raupach, J. Zierk, M. Rauh, M. Suttorp, ..., and S. Sager, "A mathematical model of white blood cell dynamics during maintenance therapy of childhood acute lymphoblastic leukemia". Mathematical Medicine and Biology: A Journal of the IMA, 36(4), 471-488, 2019, https://doi.org/10.1093/imammb/dqy017

[7] D. Jayachandran, A.E. Rundell, R.M. Hanneman, T.A. Vik, and D. Ramkrishna, "Optimal Chemotherapy for Leukemia: A Model-Based Strategy for Individualized Treatment". Plos One, 9 (10), e109623, 2014, https://doi.org/10.1371/journal.pone.010962

[8] U. Kose, O. Deperlioglu, E. U. Kucuksille, and G. Turan, "Combining Deep Learning Models for Improved Drug Repurposing: Advancements and an Extended Solution Methodology". In 2024 Int. Conf. on Inventive Computation Technologies (ICICT), IEEE, pp. 238-244, 2024, https://doi.org/10.1109/ICICT60155.2024.10544998

[9] U. Kose, "STING: Development of a Drug Repositioning Decision Support System for Childhood Acute Leukemia by Digital Twin-Oriented Deep Learning", funded by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Project No: 123E383), 2025. Retrieved August 14, 2025, from https://sting.sdu.edu.tr/

[10] U. Kose, G. Yuksel, S. Kacakci, M.C. Ince, and H. Sokun, "Advancing Simulation of 6-Mercaptopurine and Methotrexate Therapy by Using Synthetic Childhood Acute Lymphoblastic Leukemia Patients". In 2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA) IEEE, pp. 1-9, 2025, https://doi.org/10.1109/ICHORA65333.2025.11016988

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Autonomym: Multimodal Anonymization of Health Data using Named Entity Recognition and Structured Medical Data Processing

Hamdi Yalın Yalıç*, Murat Dörterler†, Alaettin Uçan*, Ali Yaşar Yiğit*, Adem Ali Yılmaz*
*Department of Research and Development, Tiga Information Technologies, Ankara, Türkiye
{yalin.yalic, alaettin.ucan, ali, adem}@tigahealth.com
†Department of Computer Engineering, Technology Faculty, Gazi University, Ankara, Türkiye
dorterler@gazi.edu.tr

***Abstract*—This paper presents Autonomym, an AI-powered software platform capable of robustly and scalably anonymizing health data across several formats, including unstructured free-text documents, tabular datasets, and medical images in both DICOM and standard RGB formats. Autonomym ensures compliance with privacy requirements such as KVKK and GDPR by using a few-shot Named Entity Recognition model for textual data. Tabular data anonymization applies formal privacy models such as k-anonymity, l-diversity, and t-closeness to generalize or suppress quasi-identifiers while preserving analytical utility. For medical imaging, DICOM files are processed through an in-house metadata scrubbing module, whereas RGB images go through optical character recognition-based detection and removal of burned-in text. The platform features a user-friendly web-based interface that facilitates both real-time and batch anonymization, designed for healthcare researchers and institutions. Technical evaluations confirm that the software is compatible with multiple input formats (csv, txt, DICOM) and has high accuracy in data anonymization. Integrated reporting modules enable the validation and auditing of data. Autonomym empowers privacy-conscious health data sharing by integrating multimodal anonymization into a singular AI-driven framework. This ensures that medical research is secure and compliant.**

***Keywords*—health data anonymization, named entity recognition, medical image metadata, privacy models**

I. Introduction

The rapid digitalization of healthcare has resulted in the unparalleled accessibility of patient data through electronic health records (EHRs), medical images, and various multimodal datasets. These datasets are very useful for clinical research, AI development, and efforts to improve quality, but they also contain Protected Health Information (PHI) or Personally Identifiable Information (PII) that must be kept safe according to strict privacy laws like the Health Insurance Portability and Accountability Act (HIPAA), the General Data Protection Regulation (GDPR), and the Turkish Personal Data Protection Law (KVKK). These rules say that direct and quasi-identifiers must be removed or changed before data can be shared. This makes anonymization an important step before health data can be used again.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Traditional manual de-identification processes, such as human review of clinical notes or manual removal of metadata, are labor-intensive, prone to error, and unsuitable for large-scale datasets. As a result, research has become more focused on automated, AI-powered anonymization techniques capable of effectively detecting and eliminating identifiers while maintaining the utility of the underlying data.

For textual health data, anonymization is mostly formulated as a Named Entity Recognition (NER) task, where models detect and mask PHI entities. While early rule-based systems struggled with generalization, machine learning classifiers, deep learning approaches and transformer-based architectures like BERT [1] have achieved state-of-the-art results. Recent models and general-purpose Large Language Models (LLMs) such as GPT-4, demonstrate high precision and recall in clinical de-identification. For medical images, anonymization involves both metadata scrubbing of sensitive Digital Imaging and Communications in Medicine (DICOM) tags and pixel-level removal of burned-in text or identifiable anatomy, often via Optical Character Recognition (OCR) and computer vision techniques. Open-source frameworks, such as those presented by Rempe et al. [2], integrate OCR-based text removal, facial feature masking, and metadata cleaning into unified pipelines, delivering competitive de-identification performance at scale.

While existing anonymization tools offer valuable functionality, most are limited to a single data modality or lack adaptability between privacy and utility. Challenges such as cross-institutional generalization, transparency, and real-time performance persist. In contrast, Autonomym offers a unified, multimodal solution for anonymizing structured data, free-text, and medical images. This paper presents its architecture, methodology, and evaluation, highlighting advancements over prior approaches in AI-driven health data anonymization.

II. Related Work

Recent studies in health data anonymization cover both text and image formats. They have a focus on utilizing AI to enhance accuracy and scalability. In the field of textual de-identification, Altalla et al. [3] used LLMs, including GPT-3.5

and GPT-4, for the anonymization of clinical reports, indicating that GPT-4 achieved high precision (0.99) and robust overall performance in PHI removal tasks while preserving the contextual integrity of the text. Negash et al. [4] conducted a review of free-text PHI de-identification methods, emphasizing the growing use of hybrid approaches that integrate rule-based techniques with deep learning to improve generalizability across datasets. Ahmed et al. [5] showed that RNN models with self-attention work well for EHR de-identification, achieving a high recall rate while maintaining clinical utility metrics.

Norgeot et al. [6] developed Philter, an open-source system for anonymizing clinical text that uses a combination of methods and has been tested on one of the largest manually annotated collections of clinical notes. Moore et al. [7] introduced Transformer-DeID, a suite of transformer-based models (including BERT) for removing PHI, showing state-of-the-art performance on benchmark datasets.

In the field of medical imaging, Rempe et al. [2] introduced an open-source de-identification framework that utilizes OCR for the removal of burned-in text, metadata scrubbing, and facial feature masking, illustrating competitive accuracy and notable speed enhancements compared to previous tools. In addition, Familiar et al. [8] created a deep learning method for anonymizing facial features in magnetic resonance imaging (MRI) scans. This method successfully hides identifying anatomy while keeping structures that are important for diagnosis.

These works together show the trend toward integrated, AI-driven, and multimodal anonymization systems. This is the design philosophy behind Autononym.

III. Methodology

A. System Architecture

The Autononym system architecture, illustrated in Figure 1, is designed to handle different types of input data, such as medical images, structured datasets, and free-text documents, using specialized anonymization pipelines. A few-shot NuNER model [9] processes incoming free-text data to find and mask entities. Structured tabular data is transformed using k-anonymity, l-diversity, and t-closeness to protect quasi-identifiers. A DICOM metadata anonymizer or an OCR-based model that finds and masks burned-in identifiers can make medical images anonymous. These components work in a modular way, which means that the system can use the proper anonymization method depending on the type of data it finds. This keeps privacy compliance while ensuring that data can be used in all formats.

B. Text-based Anonymization

Autononym uses the NuNER model in a few-shot learning setting, fine-tuned with a small, domain-specific dataset customized to healthcare and Turkish language data formats and structures. This model is used for text-based anonymization to detect and hide sensitive entities without needing to fine-tune for each task. NuNER is a compact, task-specific foundation

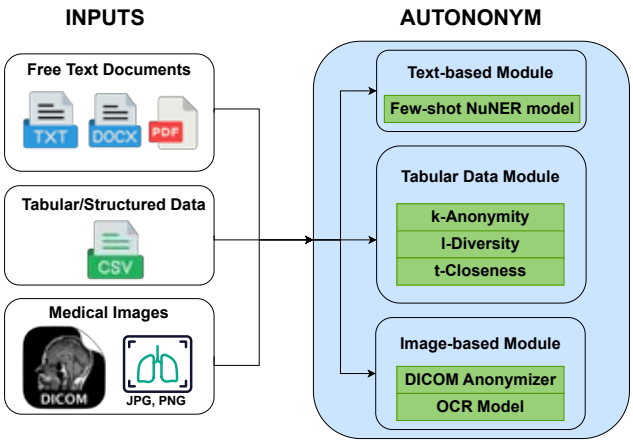


Fig. 1: Autononym system architecture

model for NER that is pre-trained on a large, diverse dataset annotated automatically by LLMs. This pre-training strategy enables NuNER to achieve strong generalization and competitive performance even with limited labeled data, allowing Autononym to handle varied healthcare datasets and formats while eliminating the dependency on manually annotated training data from the target domain.

The GLiNER framework [10] connects the underlying model and starts it up with the NuNer weights, which have been fine-tuned for strong NER across many entity categories. The system's first goal is to find a set of entity labels that represent the most common types of PII and PHI in healthcare text. These include:

$$L = \{\text{"name"},\ \text{"surname"},\ \text{"date"},\ \text{"person"},\ \text{"birthdate"},\ \text{"salary"},\ \text{"address"},\ \text{"age"},\ \text{"disease"}\}$$

Given an input text T, the model predicts a set of entities:

$$E = \{(l_i, s_i, e_i, t_i) \mid l_i \in L,\ s_i \text{ and } e_i \text{ denote start/end indices, } t_i \text{ is entity text}\}$$

where l_i is the label, s_i and e_i are character positions, and t_i is the entity string.

Adjacent entities with the same label and minimal separation ($s_{i+1} \leq e_i + 1$) are merged to form continuous entity spans. This merging step ensures that multi-word identifiers (e.g., "Dr. John Smith") are treated as a single anonymization unit rather than fragmented tokens.

Once merged, all entity spans are replaced by the masking token "$**$", producing the anonymized text T':

$$T' = \text{mask}(T, E) \quad \text{where} \quad \forall e \in E,\ T'[s_e : e_e] = \text{"}**\text{"}$$

The anonymization success rate is computed as:

$$\text{Success Rate} = \begin{cases} 100, & \text{if } |E| = 0 \\ \dfrac{|E_a|}{|E|} \times 100, & \text{otherwise} \end{cases}$$

where $|E|$ is the number of detected sensitive entities and $|E_a|$ is the number of entities successfully anonymized (in our case, all detected entities are replaced, so $|E_a| = |E|$).

This few-shot setup allows us to operate without retraining when new entity types are introduced. The model's capacity for real-time inference ensures that sensitive textual data can be anonymized on-the-fly during ingestion or before sharing, aligning with compliance requirements.

C. Image-based Anonymization

The image-based anonymization module in Autononym supports two distinct workflows depending on the format of the input file:.

1) DICOM Files: Anonymization of medical images stored in the DICOM format involves systematically removing sensitive metadata fields that could contain PHI. List of fields to remove is defined in a dictionary, such as:

$$\{\text{PatientName}, \text{PatientID}, \text{BirthDate}, \text{StudyDate}, \text{AcquisitionDate}, \text{AccessionNumber}, \text{InstitutionName}, \text{ReferringPhysicianName}, \ldots\}$$

The anonymization process starts by loading each file. This can be done with either an input file or a folder of DICOM files. Then it goes through all of the predefined sensitive fields and removes them. It also deletes private tags that might have site-specific identifiers. After that, the anonymization score is calculated using the formula:

$$\text{Score} = \frac{\#\text{ fields successfully removed}}{\#\text{ total fields to remove}} \times 100$$

Finally, the anonymized set is converted into DICOM format and encoded as a Base64 string so that it can be safely sent or stored. The output is a JSON object that has the anonymized file(s), their anonymization scores, and Base64-encoded content. This method makes sure that no metadata that could identify a patient is left behind, while keeping the original pixel data for research or diagnostic purposes.

2) RGB Image Files: If the input is an RGB image (like a .jpg or .png file), the anonymization looks for and removes burned-in text or notes that could reveal the patient's identity. To do this, OCR is used to scan the image and identify text areas.

Once these text areas are detected, the bounding boxes corresponding to these text regions are masked to prevent legibility. This method works best for modalities or export formats where identifying information is directly embedded in the visible pixels. Examples include ultrasound captures, photographic wound documentation, and exported non-DICOM radiology images.

D. Tabular Data Anonymization

For tabular datasets containing structured records, Autononym implements a privacy-preserving transformation approach that ensures the protection of sensitive and quasi-identifying attributes while keeping as much analytical utility as possible. The overall approach involves reading the input dataset, identifying quasi-identifiers, applying predefined generalization hierarchies, and enforcing selected privacy models. Suppression is used only when necessary, and its rate is monitored to maintain data usability.

Once the quasi-identifiers and their hierarchies are defined (e.g., grouping ages into ranges, generalizing salaries into classes), the system applies a set of formal privacy models to minimize the risk of re-identification:

1) k-Anonymity: Ensures that each record is indistinguishable from at least $k-1$ other records with respect to the quasi-identifiers. Formally, given a set of quasi-identifiers Q, the dataset satisfies k-anonymity if every equivalence class EC defined by Q has size $|EC| \geq k$. This prevents unique combinations of attributes from revealing individual identities.

2) l-Diversity: Addresses the weakness of k-anonymity in cases where all records in an equivalence class share the same sensitive value. The dataset satisfies l-diversity if, for each equivalence class, there are at least l "well-represented" distinct values for the sensitive attribute. One formalization is *entropy l-diversity*, where

$$-\sum_{i=1}^{m} p_i \log p_i \geq \log l$$

for each equivalence class, with p_i denoting the frequency of the i-th sensitive value.

3) t-Closeness: Ensures that the distribution of sensitive attributes in each equivalence class is close to their distribution in the overall dataset. An equivalence class satisfies t-closeness if the distance between the two distributions is no more than a threshold t. This distance is often measured using metrics such as the Earth Mover's Distance (EMD):

$$\text{EMD}(P_{EC}, P_T) \leq t$$

where P_{EC} is the distribution of the sensitive attribute within the equivalence class and P_T is the distribution of the sensitive attribute in the entire dataset.

IV. Evaluation

The figure 2 shows how to anonymize a cervical spine MRI in DICOM format. The upper-left corner of the original scan on the left has sensitive patient metadata (name, date of birth, hospital information, and acquisition details) burned into it. The medical image content remains for diagnosis, but the identifying elements have been removed from the metadata on the right. The figure 3 shows an anonymized version of a structured CSV dataset that has personal identifiers and quasi-identifiers like name, surname, national ID number, age, gender, place of birth, salary, marital status, profession, and education. The "Input Data" table shows the original values, while the "Output Data" table shows the anonymized version, where direct identifiers are replaced with masking symbols (*) and quasi-identifiers are grouped (for example, salaries are grouped into ranges and ages are grouped into ranges). The figure 4 shows an unstructured free-text document that has been made anonymous. The "Original Text" contains personal information such as names, birth dates, addresses, job details,

salaries, and information about spouses. The "Anonymized Text" keeps the same sentence structure and non-sensitive content. All sensitive fields have been replaced with masking symbols (**), so the text can still be read later.

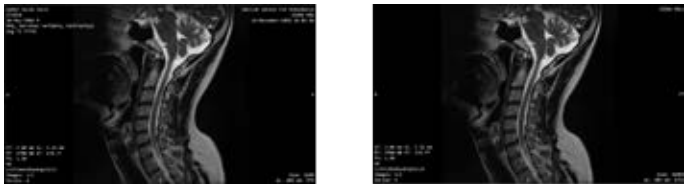


Fig. 2: Medical image (DICOM) anonymization example

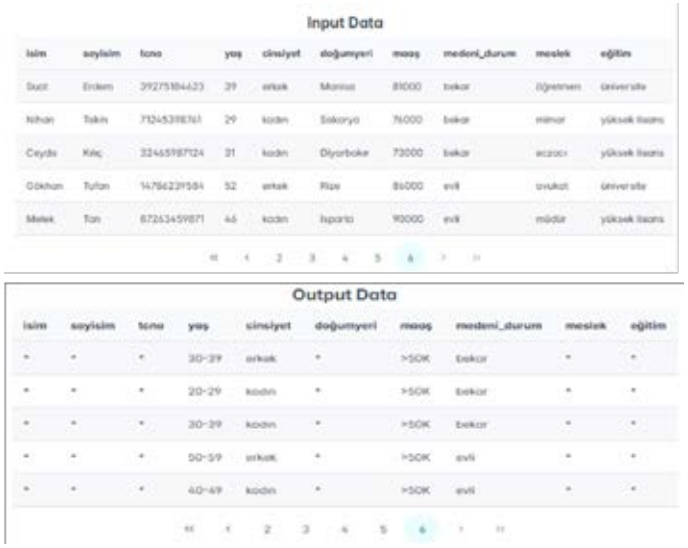


Fig. 3: Tabular data anonymization example



Fig. 4: Text data anonymization example

V. Conclusion

In this paper, we introduced Autononym, an AI-powered, multimodal health data anonymization platform designed to provide regulatory compliance while preserving data utility. The framework integrates few-shot NER for free-text entity masking, k-anonymity, l-diversity, and t-closeness models for tabular data protection, and both DICOM metadata scrubbing and OCR-based burned-in text removal for medical images in DICOM and RGB formats. It supports real-time and batch anonymization workflows by its web interface, enabling healthcare researchers and institutions to securely share data across multiple modalities. Our evaluation demonstrates high accuracy, multimodal compatibility, and flexible deployment, positioning Autononym as a comprehensive and practical solution for privacy-preserving medical research and data exchange.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[2] M. Rempe, L. Heine, C. Seibold, F. Hörst, and J. Kleesiek, "De-identification of medical imaging data: A comprehensive tool for ensuring patient privacy," *European radiology*, pp. 1–10, 2025.

[3] B. Altalla' *et al.*, "Evaluating gpt models for clinical note de-identification," *Scientific Reports*, vol. 15, no. 1, p. 3852, 2025.

[4] B. Negash *et al.*, "De-identification of free text data containing personal health information: A scoping review of reviews," *International Journal of Population Data Science*, vol. 8, no. 1, p. 2153, 2023.

[5] T. Ahmed, M. M. A. Aziz, and N. Mohammed, "De-identification of electronic health record using neural network," *Scientific reports*, vol. 10, no. 1, p. 18 600, 2020.

[6] B. Norgeot *et al.*, "Protected health information filter (philter): Accurately and securely de-identifying free-text clinical notes," *NPJ digital medicine*, vol. 3, no. 1, p. 57, 2020.

[7] C. Moore, L. Bulgarelli, T. Pollard, and A. Johnson, *Transformer-DeID: Deidentification of free-text clinical notes with transformers (version 1.0.0)*, PhysioNet. RRID: SCR_007345, https://doi.org/10.13026/7dj5-7x85, 2023.

[8] A. M. Familiar *et al.*, "Empowering data sharing in neuroscience: A deep learning deidentification method for pediatric brain mris," *American Journal of Neuroradiology*, vol. 46, no. 5, pp. 964–972, 2025.

[9] S. Bogdanov, A. Constantin, T. Bernard, B. Crabbé, and E. Bernard, "Nuner: Entity recognition encoder pre-training via llm-annotated data," *arXiv preprint arXiv:2402.15343*, 2024.

[10] U. Zaratiana, N. Tomeh, P. Holat, and T. Charnois, "Gliner: Generalist model for named entity recognition using bidirectional transformer," *arXiv preprint arXiv:2311.08526*, 2023.

Determining the Health Condition of Newborns Using Apgar Score and Thermal Camera Imaging

Farhad Nassehi[1,2], Doga Kaya[1], Pelin Kocak[1], Emine Bilun Haskaraman[1], Seray Akbag[3], Osman Erogul[1]

[1]Biomedical Engineering, TOBB University of Economics and Technolog, Ankara, Turkey

[2]OKUMAN Medikal Sistemler A.Ş., Ankara, Turkey

[3]Bilkent City Hospital, Ankara, Turkey

(f.nassehi, d.kaya, p.kocak, e.haskaraman, erogul)@etu.edu.tr, serayakbag@gmail.com

***Abstract*—APGAR (Appearance, Pulse, Grimace, Activity, Respiration) Score is a system that evaluates the post-birth health status of newborns and is scored between 0-10. This score, which consists of five parameters: appearance, pulse, grimace, activity and respiration, is obtained by scoring each parameter between 0-2. The condition of newborns with an APGAR score of 7-10 is considered reassuring, a score of 4-6 is considered moderately abnormal and a score of 0-3 is considered low and critical. This study aims to automatically determine the APGAR score of newborn babies with thermal camera imaging and artificial intelligence support and to classify the newborn according to the APGAR score. For this purpose, data was collected from a total of 22 newborns whose gestational age was between 28-42 weeks. Thermal images of newborns were preprocessed in the MATLAB software environment and first the babies were separated from the background. After this stage, important features determined as a result of the literature research were extracted. The extracted features and the APGAR scores obtained from nurses, which are accepted as gold standards, were used in the Python software environment for classification. Stacking Classifier models were used in the regression and classification stages. Among the classification algorithms, the most successful model in distinguishing risky and healthy babies was the Stacking Classifier with 88.59% (± 0.0264) accuracy. The proposed method may assess APGAR score non-invasively, non-contact, objectively and early detection of risky newborns was provided.**

Keywords—Neonatal; Thermal imaging; Machine learning; Apgar score, Death risk

I. INTRODUCTION

Newborns' behaviors and movements are critical to understand. Because these movements and behaviors are the language of the newborn and provide important information about the baby. A careful post-birth examination helps to detect anomalies, cardiovascular disorders or birth injuries that may pose a risk to the newborn's successful adaptation to life outside the uterus. Every newborn should have a short physical examination within the first few minutes after birth [1]. Basic assessment methods, such as the APGAR (Appearance, Pulse, Grimace, Activity, Respiration) score, help to detect abnormalities in the baby's vital functions [2]. The APGAR score is a quick method used by healthcare providers to assess a newborn's condition immediately after birth. The APGAR score takes into account 5 components: Appearance, Pulse, Grimace, Activity, Respiration. Each parameter receives a value of 0, 1 or 2 and thus the APGAR score is assessed out of 10 points. In all newborns, scoring is done in the first minute and fifth minutes after birth. Babies with an APGAR score of less than 7 at fifth minutes are also evaluated at tenth minutes.

TABLE 1. APGAR SCORE PARAMETERS

	0	1	2
Appearance	The body is blue or pale in color	Pink body with blue or pale extremities	Normal, pink color all over the body
Pulse beats/min	No pulse	<100	>100
Grimace	No response to stimulation	Some facial grimace in response to stimulation	Coughing, sneezing, or response to stimulation
Activity	No movement	Only some movement of the extremities	Active and spontaneous movement

A score between 7 and 10 is considered reassuring, a score between 4 and 6 is considered moderately abnormal and a score between 0 and 3 is considered low in infants [3,4]. The APGAR score is considered a useful predictor of neonatal death. Evidence from studies suggests that 50% of preterm (≤31 weeks) infants with an APGAR score ≤3 die in the neonatal period (within the first 28 days after birth), compared with less than 2% of preterm infants with an APGAR score ≥7. In term infants (≥37 weeks), 10% of infants with APGAR score ≤3 and 0.2% of term infants with APGAR score ≥7 died. When all these data were analyzed, it was observed that 42.2% of neonatal deaths occurred in infants with an APGAR score ≤3 and almost 20% of all newborns with an APGAR score ≤3 died in the neonatal period [5,6].

At present, since the APGAR score method is based on observation, it may show variability among healthcare personnel. Infant nurses with different levels of expertise may give different APGAR scores to the same infant. This situation causes both loss of time and inconsistencies in evaluation. The findings obtained from the studies showed that the level of

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

knowledge about APGAR score evaluation among nurses and midwives is insufficient. A study found that more than 90% of nurses continued APGAR scoring even in cases with missing data, and that APGAR scores obtained by different nurses were low to medium level compatible and therefore had low reliability [7]. A study conducted in Turkey found that midwives did not know all of the APGAR score evaluation criteria. In particular, they have limited knowledge about the evaluation of heart rate and muscle tone [8]. The use of probes and sensors that require contact when determining the APGAR score causes discomfort, hygiene risk and stress on the sensitive skin of infants. The non-contact, non-radiating and non-invasive nature of thermal imaging devices offers a safe and comfortable use for newborns [9,10].

This study proposes a novel framework to determine newborns' health status immediately after birth, in a non-invasive and non-contact way using thermal camera imaging and artificial intelligence methods. In this context, it aims to develop a machine learning model to prevent the APGAR score from varying according to nurses' assessments and to reduce the workload of healthcare personnel. By using thermal images of newborn babies and machine learning algorithms, it is aimed to provide an objective, fast and non-contact assessment of the baby's health status.

II. MATERIALS AND METHODS

A. Participants

Newborns delivered via Caesarean section in the Neonatal Intensive Care Unit (NICU) of Ankara Bilkent City Hospital were included in this study, with no gender-based exclusion criteria. A total of 305 thermal images were collected from 22 newborns (15 male, 7 female) of varying gestational ages to ensure a balanced dataset. The study was approved by the Ethics Committee of Ankara Bilkent City Hospital. APGAR scores were categorized as poor (≤6) or good (≥7).

B. Data Acquisition

Thermal imaging was performed using a FLIR® C2 7400 high-performance thermal camera (±0.1°C sensitivity, 320×240 resolution), capturing detailed surface temperature distributions. Following maternal informed consent, standardized environmental conditions (temperature, lighting, airflow) were maintained during data acquisition. Thermal images were obtained at the 1st, 5th, and when necessary, 10th minute after birth, focusing on regions such as the abdomen, back, and face, with care taken to avoid obstructions. Concurrently, nurses recorded clinical APGAR scores, serving as gold-standard labels.

C. Image Processing

The image preprocessing pipeline involved segmenting the newborn from the background and isolating relevant temperature data. Initially, raw thermal images were cropped to focus on the infant while excluding irrelevant temperature points. The thermal camera's temperature scale was indexed to 34–38°C (±2°C from the standard body temperature of 36°C), minimizing background influence. Cropped images were converted to grayscale and enhanced using histogram equalization to improve contrast and highlight subtle temperature variations. The resulting images exhibited clearer edge definition, facilitating the separation of the newborn from the background. Black pixels (with a value of zero) indicating the background were detected and used to create a mask, which was applied to the original cropped image. This process effectively isolated the newborn's temperature profile, ensuring that only temperature data corresponding to the infant were retained for subsequent feature extraction.

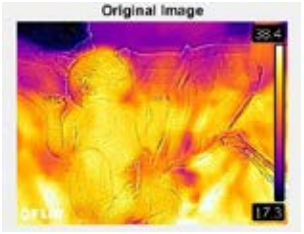


Fig 1. Original Thermal Image of the Newborn

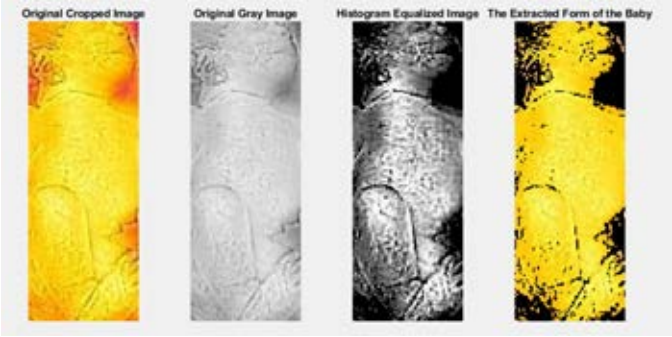


Fig 2. Results of Background Elimination

D. Feature Extraction

In the feature extraction stage, numerical features representing the thermal characteristics of the newborns were derived from the preprocessed images to predict health statues of newborns. Both time-domain and frequency-domain (FD) features were utilized. Time-domain (TD) features included mean, standard deviation, variance, median, skewness, interquartile range (IQR), and root mean square (RMS). For the frequency domain, a two-dimensional Fast Fourier Transform (2D-FFT) was applied to the images, from which statistical features such as mean, standard deviation, variance, minimum, maximum, median, skewness, IQR, RMS, MAD, energy, and kurtosis were extracted. Features were also computed from the RGB color channels and pixel intensity values of the thermal images. In total, 40 features were generated, encapsulating both spatial and frequency information of the temperature distribution, thereby enhancing the model's ability to distinguish between different APGAR score categories.

E. Classification Algorithms

In this study, newborns were classified into two groups based on their APGAR scores: critical (≤ 6) and healthy (≥7), comprising 6 and 16 newborns, respectively. The dataset was split into 70% training and 30% testing using the hold-out method. To optimize model performance, the training set was further processed using GridSearchCV for hyperparameter tuning (5-fold cross-validation), LASSO for feature selection, and SMOTE for addressing class imbalance. GridSearchCV

systematically searched for optimal hyperparameters across five classification models, while LASSO (λ=0.01) reduced the feature set from 40 to 12 key features. SMOTE (k=5) was applied to synthetically augment the minority class (critical APGAR). Decision tree (DT), random forest (RF), support vector machine (SVM), multi-layer perceptron (MLP), and stacking classifier algorithms were used as classifiers in this study. The final models were trained and evaluated across 50 iterations, with performance reported as mean and standard deviation for metrics such as accuracy, sensitivity, specificity, and Receiver Operating Characteristic (ROC), and Area Under Curve (AUC).

III. RESULTS

A total of 40 features were extracted, and 12 of them were selected using the LASSO method. The selected FD features are mean, variance, and MAD. The selected TD features are mean, variance, median, skewness, median (red channel), RMS (red channel), mean (blue channel), median (blue channel), and RMS (blue channel).

In the result of GridSearchCV; for DT, (*criterion* = entropy, *max_depth* = 10), for RF, (*n_estimator* = 200, *max_depth* = 20), for SVM (*C* = 10, *kernel* = rbf), for MLP (*hidden_layers_size* = [100,50], *max_iter* = 1000) were selected as hyperparameters of classifiers. Stacking Classifier is a model that is a combination of three models (RF, SVM, MLP), the previously selected hyperparameters of these three models were used in the hyperparameter selection. Therefore, the hyperparameters of the logistic regression (LR) model used in the last part of the stacking classifier were selected as (*C* = 10, *solver* = liblinear).

The Stacking Classifier model shows the best performance among all models. This model has performed better than the other models with 88.59% (± 0.0264) accuracy, 85.04% (± 0.0411) sensitivity and 92.14% (± 0.0344) specificity values. Since Stacking Classifier is a model created by combining different models, this situation shows that ensemble methods increase the classification success. MLP, on the other hand, has become the model that shows the best performance after Stacking Classifier, especially with 93.42% (± 0.0357) specificity value. When all models are examined, it is seen that the model that shows the lowest success in terms of all metrics is Decision Tree. When the ROC Curve and AUC values are examined, the most successful model in separating the risky and healthy babies with an AUC value of 0.95 is Stacking Classifier. Random Forest comes after Stacking Classifier with an AUC value of 0.94, while MLP and SVM values are successful models with AUC values of 0.92. Decision Tree again showed the lowest success with an AUC value of 0.78. As a result, when all metrics are evaluated, it is seen that the model that can most successfully classify newborns in critical condition requiring medical intervention and newborns in healthy condition is Stacking Classifier.

TABLE 3. CLASSIFICATION RESULTS

Classifiers	Accuracy	Sensitivity	Specificity
RF	85.38% (± 0.0317)	80.99% (± 0.0549)	89.78% (± 0.0465)
DT	76.53% (± 0.0389)	72.27% (± 0.0679)	80.79% (± 0.0572)
SVM	84.63% (± 0.0274)	76.30% (± 0.0486)	92.96% (± 0.0349)
MLP	87.45% (± 0.0293)	81.48% (± 0.0526)	93.42% (± 0.0357)
Stacking Classifier	88.59% (± 0.0264)	85.04% (± 0.0411)	92.14% (± 0.0344)

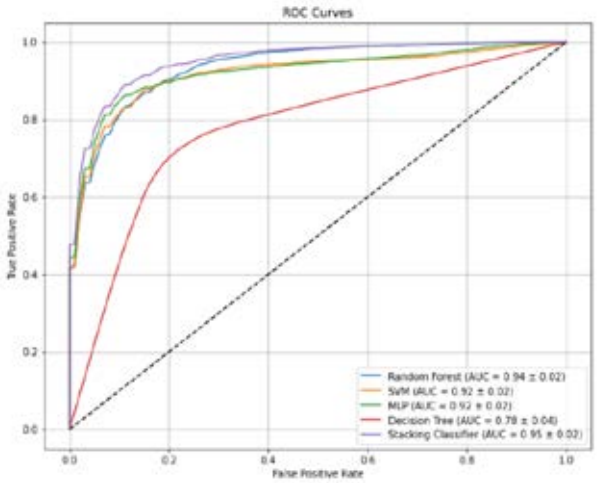


Fig 3. ROC Curve and AUC Values of Classifiers

IV. DISCUSSION

According to Table 3. and Fig. 3 Stacking classifier perform better than other classifier algorithms. Stacking classifier reached average 88.59% accuracy, 85.05% sensitivity, and 92.14% specificity rates. Ervural and Ceylan (2022) focuses on the classification of neonatal disease by analyzing the limited thermal image data. The study is meant to obtain high-accuracy classification with small and unbalanced datasets using convolutional neural networks (CNNs) and data augmentation methods. 20 thermal images were collected from each of 44 newborns, for a total dataset of 880 thermal images. The highest accuracy (94.55%) was obtained by training the model with 15 of the 20 images of each infant. In addition, when the model was tested without training, a classification accuracy of 89% was observed, which is more meaningful in terms of the real-life performance of the model [10]. Jung and his colleagues focused on determining muscle activation during resistance exercises with the help of infrared thermal imaging. Observing the three major muscles in the upper arm (biceps brachii, triceps brachii, and deltoid muscle) different weights, exercise durations, and muscle activation sequences were used in the experiments. Thermal images were recorded with a FLIR thermal camera and heat maps were generated through processing the images. In order to classify muscle activation a CNN model was trained with the obtained dataset. The CNN model was able to classify muscle activations with 92.3% accuracy. Results show that thermal imaging has the potential to assess muscle isolation and exercise accuracy [11].

According to our knowledge no study was found on the direct determination of APGAR score in newborns with a thermal camera. However, a patented study is noteworthy in this context among the studies conducted on the determination of APGAR score in the literature. The study aims to automatically determine the APGAR score using the body temperature, muscle tone, pulse, voice and respiratory physiological data of newborns., Body temperature sensor, respiratory sensor, pulse sensor, microphone and pulse oximeter are used to obtain data from newborns.

V. CONCLUSION

This thesis aims to determine the APGAR score, which is used to evaluate the health status of newborns after birth, using thermal cameras and machine learning algorithms. In addition, it aims to identify babies with critical health conditions and ensure immediate medical interventions. The biggest challenge encountered during the project was that the number of newborns with low APGAR scores remained in the minority. Since the majority of the data collected in the hospital consisted of healthy and high APGAR score babies, a sufficient amount of data on low-scoring newborns could not be obtained, and an unbalanced data set was formed. An important technological development is provided in the estimation of APGAR score and the use of thermal cameras in the field of health. The ability to determine the APGAR score with the body temperature values of the newborn with the help of a thermal camera allows faster and more objective data to be provided to healthcare professionals. This situation standardizes the assessment of newborn health, especially in the first hours after birth, and accelerates medical intervention decisions of healthcare personnel. In addition, this system can increase the survival chances of risky babies by allowing early detection of babies with critical health conditions. Additionally, the ease of use of the system and the ability to directly estimate the APGAR score provide advantages in delivery rooms by reducing the workload of healthcare personnel. There are certain limitations that should be considered. Firstly, the biggest limitation of this study is the small data set. This reduces the generalizability of the study. In addition, the small number of newborns with low APGAR scores also caused class imbalance. This imbalance attempted to be reduced by generating artificial data using the SMOTE method, but this still did not correspond to the real-world data. Another limitation is that the data were collected from a single hospital. Creating the dataset using the evaluations of only a few nurses from a single hospital limits the generalizability of the project. This study is a preliminary study for the continuation and sustainability of the project. The dataset is being expanded by continuing data collection from different hospitals.

VI. FUTURE WORK

As a feature work, we decided to validate our algorithm with real abnormal data and then use this algorithm as an embedded form in OKM 730 - Radiant Warmer that is produced by OKUMAN MEDİKAL SİSTEMLER A.Ş.



Fig 4. Radiant Warmer

ACKNOWLEDGMENT

This study supported by TÜBİTAK 2209-B University Students Industrial Research Projects Support Program (Application No: 1139B412402440) and OKUMAN Medikal Sistemler A.Ş.

REFERENCES

[1] Eryürek, D., Başdaş, Ö., Korkmaz, Z., Yıldız, İ., Mumcu, Ö., & Bayat, M. (2021). Prematüre bebek annelerinin yenidoğan yoğun bakım deneyimleri. Hacettepe Üniversitesi Hemşirelik Fakültesi Dergisi, 8(2), 196-202.

[2] Genel, T. S. B. H. S., Çocuk, M., & Başkanlığı, E. S. D. (2019). Temel Yenidoğan Bakımı. Erişim tarihi, 25.

[3] Simon, L. V., Hashmi, M. F., & Bragg, B. N. (2017). APGAR score.

[4] Michel, A. (2022). Review of the Reliability and Validity of the Apgar Score. Advances in Neonatal Care, 22(1), 28-34.

[5] Li F, Wu T, Lei X, Zhang H, Mao M, et al. (2013) The Apgar Score and Infant Mortality. PLOS ONE 8(7): e69072. https://doi.org/10.1371/journal.pone.0069072

[6] Cnattingius, S., Norman, M., Granath, F., Petersson, G., Stephansson, O., & Frisell, T. (2017). Apgar score components at 5 minutes: risks and prediction of neonatal mortality. Paediatric and perinatal epidemiology, 31(4), 328-337.

[7] Michel, A. D., & Lowe, N. K. (2023). Accuracy and interrater agreement of registered nurses' assignment of Apgar scores to standardized clinical vignettes. Clinical Nursing Research, 32(3), 452-462.

[8] Tezel, A., Özkan, H., & Çelebioğlu, A. (2005). EBELER "APGAR DEĞERLENDİRMESİ" Nİ NE KADAR BİLİYORLAR?. Sağlık Bilimleri Dergisi, 14(3), 163-171.

[9] Topalidou, A., Ali, N., Sekulic, S., & Downe, S. (2019). Thermal imaging applications in neonatal care: a scoping review. BMC pregnancy and childbirth, 19, 1-14.

[10] Ervural, S., & Ceylan, M. (2022). Classification of neonatal diseases with limited thermal Image data. Multimedia Tools and Applications, 81(7), 9247-9275.

[11] Jarlier, S., Grandjean, D., Delplanque, S., N'diaye, K., Cayeux, I., Velazco, M. I., ... & Scherer, K. R. (2011). Thermal analysis of facial muscles contractions. IEEE transactions on affective computing, 2(1), 2-9.

TIPTEKNO'25
MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Giyilebilir Sistem Uygulamaları için Elektrokardiyogram İşaretinden Solunum Hızı Kestirimi
Respiration Rate Estimation from Electrocardiogram Signal for Wearable System Application

Doğan Yörük
Elektronik ve Haberleşme Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
dogan.yoruk@std.yildiz.edu.tr

Surhan Bozkurt
Elektrik-Elektronik Mühendisliği
Doğuş Üniversitesi
İstanbul, Türkiye
sbozkurt@dogus.edu.tr

Güleycan Dedecengiz Varol
Biyomedikal Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
guleycan@yildiz.edu.tr

Umut Engin Ayten
Elektronik ve Haberleşme Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
ayten@yildiz.edu.tr

***Özetçe* —Solunumun sürekli izlenmesi ölümcül hastalıkların önceden tespitinde önemlidir. Solunum takibinde invaziv olan ve hareket kabiliyetini kısıtlayan solunum cihazları kullanılmaktadır. Bu sensörlerin giyilebilir cihazlara uygulanması her zaman kolay değildir. Bu çalışmada ek bir sensör kullanılmadan EKG işaretinden solunum hızı tahmin edilmiştir. EKG işaretinden R-tepe genlik değerlerinin enterpolasyonla EKG'den çıkarılan solunum (EÇS) işareti elde edilmiştir. Bu işaret bant geçiren filtre ve Hızlı Fourier Dönüşümleriyle işlenerek solunum hızı kestirilmiştir. Tüm işlemler düşük güç tüketimine ve düşük işlem gücüne sahip bir mikrodenetleyici üzerinde maksimum hız ve minimum bellek kullanımı amacıyla sabit noktalı olarak gerçekleştirilmiştir. BIOPAC cihazından alınan veriler kullanılarak uygulanan algoritmanın doğruluğu gösterilmiştir.**

Anahtar Kelimeler*—*Elektrokardiyogram, solunum hızı, solunum hızı kestirimi, giyilebilir sistem, mikrodenetleyici

***Abstract*—Continuous respiratory monitoring is crucial for the early detection of fatal diseases. Invasive respiratory devices that restrict mobility are used for respiratory monitoring. Implementing these sensors in wearable devices is not always easy. In this study, respiratory rate was estimated from the ECG signal without using an additional sensor. The ECG-derived respiration (EDR) signal was obtained by interpolating the R-peak amplitude values of the ECG. This signal was processed using a band-pass filter and the Fast Fourier Transform (FFT) to estimate respiratory rate. All operations are performed on a fixed-point microcontroller with low power consumption and low processing power for maximum speed and minimum memory usage. The accuracy of the implemented algorithm was demonstrated using data from a BIOPAC device.**

Keywords*—*Electrocardiogram, respiration rate, respiratory rate estimation, wearable system, microcontroller

I. GİRİŞ

Kalp ve solunum hastalıkları, küresel ölçekte ölüm oranlarının yüksek olduğu başlıca hastalık gruplarındandır. Literatürdeki çalışmalarda kardiyovasküler ve kronik solunum yolu hastalıklarının, dünya genelinde en yüksek ölüm oranlarına sahip hastalıklar arasında yer aldığı bildirilmiştir [1].

Elektrokardiyogram (EKG), kalbin elektriksel aktivitesini ölçen, kardiyak ritim, iletim bozuklukları ve pek çok kalp hastalığının tanısında temel rol oynayan bir biyofiziksel işarettir [2]. EKG işaretinin morfolojik özellikleri, sadece kalp sağlığı hakkında değil, aynı zamanda solunum gibi diğer fizyolojik sistemlerin dolaylı etkilerini de yansıtabilir. Solunum aktivitesinin EKG ölçümlerini etkilediği bilinmektedir. Solunum döngüsü sırasında, göğüs hareketleri ve akciğerlerin dolup boşalması nedeniyle oluşan torakstaki empedans dağılımındaki değişiklikler, kalbin elektriksel ekseninin değişmesine neden olarak EKG işaretini etkiler [3].

Solunum hızı, bireyin genel sağlık durumu ve fizyolojik aktivite düzeyi hakkında bilgi veren parametredir. Dinlenme hâlindeki yetişkinlerde genellikle dakikada 12–20 nefes arasında seyreder ve bu değerin dışına çıkması, yorulma düzeyi, stres durumu veya solunum sisteminin verimliliği hakkında bilgi verir [4], [5]. Solunum hızı, solunum evreleri gibi biyobelirteçler bu olayların tespitinde önemli rol oynar. Solunumun izlenmesi klinik ve sağlık takibi açısından büyük önem taşımasına rağmen, doğrudan solunum kaydı genellikle termistörler, spirometreler ve solunum kemerleri gibi invaziv veya yarı-invaziv sensörlerin kullanımını gerektirir [6]. Bu sensörlerin kullanımı oldukça sınırlıdır çünkü doğal solunumu engellemekle kalmaz yüksek maliyetli ve düşük konforludur. Bu nedenle, yaygın olarak kullanılan EKG kayıt cihazları üzerinden solunum hızı tahmin yöntemleri geliştirmek hem konforu artırır hem de ek donanım ihtiyacını ortadan kaldırır [7]. Bu tür sürekli ve konforlu biyolojik işaret takibi kronik hastalıkların yönetimi, ani durumların erken tespiti ve bireysel sağlık izlemi açısından önemli avantajlar sunmaktadır.

EKG işaretinin rutin takibi ve buradan solunum kestirimi literatürde sıklıkla çalışılan konulardandır. Campolo ve arkadaşları [8], Empirik Mod Ayrıştırma yönteminin solunum frekansına karşılık gelen işaret bileşenlerini ayırt ederek solu-

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

num dalga formunun oluşturulabildiğini göstermiştir. Langley ve arkadaşları [9], Temel Bileşen Analizi (PCA) yöntemini kullanmış ve EKG işaretine ait PCA'nın birinci bileşeninin zaman içindeki değişimini solunum işareti olarak değerlendirmiştir. Ayrıca Widjaja ve arkadaşları da [10] Çekirdekli PCA yöntemini kullanarak doğrusal olmayan özellikleri de hesaba katmış ve tek kanallı EKG üzerinden solunum bilgisi elde etmiştir. Khaled ve Farges'in gerçekleştirdiği çalışmada [11], EKG işaretindeki P-QRS-T döngüsünün R-tepelerinin genliği kullanılarak solunum hızının elde edilebileceği gösterilmiştir. Larsen ve arkadaşları [12], solunumsal sinüs aritmisi olarak bilinen ve kalp atım hızının solunum döngüsüne bağlı olarak ritmik değişiklikler gösterdiği fizyolojik olguyu detaylı şekilde incelemiştir.

Giyilebilir sistemler, kalp ve solunum gibi yaşamsal parametrelerin kesintisiz, konforlu ve uzun süreli takibini mümkün kılmaları açısından sağlık teknolojileri içinde giderek daha fazla önem kazanmaktadır. Negi ve arkadaşları [13], oturur pozisyonda kullanım için sandalyenin kol dayanaklarına entegre kuru elektrotlardan elde edilen EKG işareti üzerinden solunum hızı kestirimi yapmıştır. Rahman ve Morshed [14], bilekten elde edilen EKG işaretini nRF52840 kartı kullanarak bilgisayara aktarmış ve solunum hızı tahmini yapmışlardır. Fan ve arkadaşları [15], EKG'den çıkarılan solunumu (EÇS; ECG-Derived Respiration, EDR) elde etmek amacıyla uygulamaya özel, ticari olmayan bir tümdevre tasarımı gerçekleştirmişler ve solunum hızı tahmini yapmışlardır. Krishnapriya ve arkadaşları [16], tek kanallı EKG işaretini NI-myDAQ veri toplama cihazı ile kaydedip işaretleri bilgisayar ortamında işleyerek gerçek zamanlı solunum hızı tahmini yapmışlardır.

Bu çalışmada, giyilebilir sistemlere uygun, düşük işlem ve güç kapasiteli bir mikrodenetleyici üzerinde EKG işaretinden solunum hızı kestirimini yapan bir sistem tasarlanmış ve gerçeklenmiştir. Bu amaçla bir deney prosedürü çerçevesinde on farklı gönüllüden eş zamanlı olarak EKG ve referans solunum verileri toplanmıştır. Bu işlem için referans cihaz olarak BIOPAC (BIOPAC Systems Inc., USA) [17] kullanılmıştır. Çalışmadaki yöntemler ve donanım uygulaması bölüm II'de ayrıntılı olarak sunulmuştur. Bölüm III test sonuçları tablo ve grafik halinde verilmiştir. Bölüm IV'te ise sonuçlar tartışılmış ve gelecek çalışmalara ilişkin değerlendirmeler verilmiştir.

II. YÖNTEM

A. Deney Prosedürü ve Veri Toplanması

Kardiyak ve solunumsal bozukluğu bulunmayan toplam 10 sağlıklı gönüllüden, EKG işaretinden solunum hızı tahmini amacıyla eş zamanlı EKG ve referans solunum verileri toplanmıştır. Gönüllülerin 8'i erkek, 2'si kadındır. Gönüllülerin demografik özellikleri Tablo I'de verilmiştir. Bilgiler ortalama $\pm$ standart sapma şeklinde verilmiştir. Tüm gönüllülerden veri toplama öncesinde bilgilendirilmiş onam formu alınmıştır.

Veri toplama işlemi, BIOPAC veri toplama cihazı kullanılarak gerçekleştirilmiştir. Tek kanallı EKG işareti, BIOPAC SS2L kablo seti ve uçlarına bağlanan standart gümüş(Ag)/gümüş klorür(AgCl) yüzey elektrotları aracılığıyla kaydedilmiştir. Elektrotlar sırasıyla sağ göğsün üst kısmına ve sol göğsün üst kısmına yerleştirilmiş, referans (topraklama) elektrot ise sol alt kaburganın altına konumlandırılmıştır. Bu yerleşim, kolay uygulanabilir olması ve hareket kaynaklı artefaktlara karşı dayanıklılığı nedeniyle tercih edilmiştir.

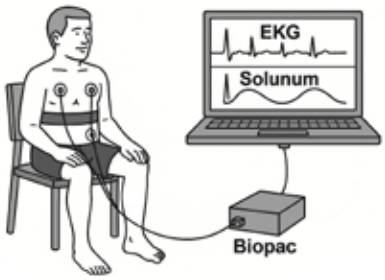
Tablo I: Gönüllü Grubuna Ait Demografik Veriler. VKİ:Vücut Kitle İndeksi. (En Küçük - En Büyük)

Toplam	10
Yaş	20.9 $\pm$ 1.1 (19-23)
Boy [cm]	179.2 $\pm$ 7.85 (169-195)
Kilo [kg]	77.6 $\pm$ 17.01 (62-110)
VKİ [kg/m^2]	24 $\pm$ 4.45 (19.7-32.4)

Referans solunum verisi, karın çevresine yerleştirilen bir solunum kemeri (BIOPAC SS5LB) aracılığıyla elde edilmiştir. Bu kemer, abdominal bölgedeki genişleme ve daralmaları hassas biçimde algılayarak solunum hareketlerini yüksek doğrulukla ölçebilmektedir. Böylece, EKG üzerinden türetilen solunum işaretleriyle karşılaştırılabilecek güvenilir bir referans solunum verisi sağlanmıştır.

Tüm biyolojik işaretler, 2000 Hz örnekleme frekansı ile kaydedilmiştir. Kayıt esnasında işaretlere, 0.05–35 Hz aralığında bant geçiren dijital bir filtre uygulanarak bazal kayma, kas aktivitesi kaynaklı parazitler ve yüksek frekanslı gürültüler bastırılmıştır.

Deneyler, sessiz ve dikkat dağıtıcı unsurlardan arındırılmış bir ortamda gerçekleştirilmiştir. Katılımcılar, ayakları yere tam basacak şekilde sandalyeye dik pozisyonda oturtulmuş ve 5 dakika dinlendirildikten sonra veri kaydı bu pozisyonda gerçekleştirilmiştir. Her bir kayıt oturumu yaklaşık 1 dakika sürmüş ve bu süre boyunca eş zamanlı olarak EKG ve referans solunum verileri kesintisiz biçimde kaydedilmiştir (Şekil 1).

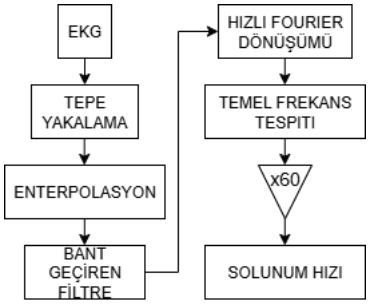


Şekil 1: Deney Prosedürü

B. EKG'den Çıkarılan Solunum (EÇS) Hızı Kestirimi

EKG işaretinden solunum işaretini türetmek amacıyla, literatürde yaygın olarak kullanılan R-dalgası genlik değerlerinin enterpolasyon temelli EÇS yöntemi tercih edilmiştir. Bu yaklaşım hesaplama açısından görece basit, yorumlanabilirliği yüksek ve özellikle giyilebilir sistemler için uygun olması nedeniyle seçilmiştir. Önerilen algoritmanın akış diyagramı Şekil 2'de verilmiştir.

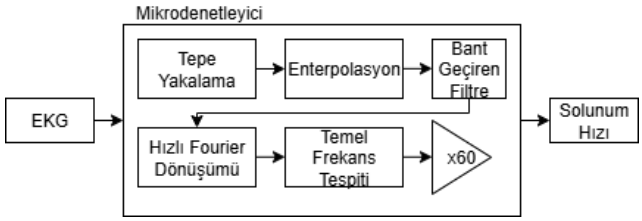
Yapılan işlemde ilk olarak EKG işareti üzerinde R-dalgası tespiti, bir tepe yakalama algoritması aracılığıyla gerçekleştirilir. Bir tepe noktasının, R-dalgası olarak kabul edilebilmesi için işaretin genliğinin, işaretin ortalama (mean) değeri ve standart sapmasının toplamından büyük bir genliğe sahip olması



Şekil 2: Elektrokardiyogram İşaretinden Solunum Kestirimi Algoritması Akış Diyagramı. EKG: Elektrokardiyogram

gerekmektedir ve bir önceki R-dalgasından en az 0.3 saniye (200 bpm) sonra gelmiş olmalıdır. İkinci işlem olarak, tespit edilen R-tepe genlik değerleri, sürekli bir solunum dalga formu elde edebilmek amacıyla spline enterpolasyon yöntemine tabi tutulur. Üçüncü işlem olarak, bu işaret solunum dışı frekans bileşenlerinin baskılanması amacıyla bant geçiren filtreye sokulur. Dördüncü işlem olarak, elde edilen işaretin frekans bileşenlerini tespit etmek amacıyla Hızlı Fourier Dönüşümü (FFT) uygulanır. Son işlem olarak, FFT sonucu elde edilen frekans karakteristiği üzerinden temel frekans değeri tespit edilir ve dakikadaki solunum hızı hesabı için 60 değeriyle çarpılır.

C. Gömülü Sistem Uygulaması

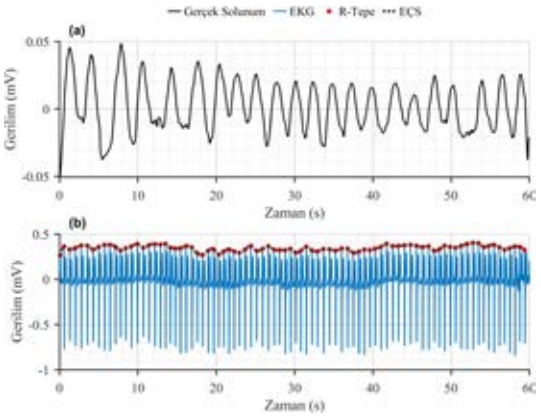


Şekil 3: Gömülü Sistem Akış Diyagramı. EKG: Elektrokardiyogram

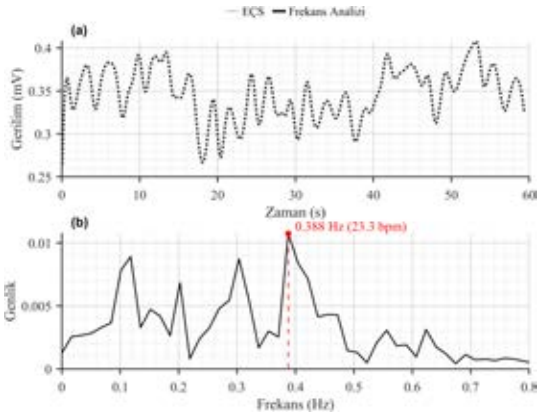
Gömülü sistem tarafındaki testleri gerçekleştirmek için STM32L0 serisi mikrodenetleyici seçilmiştir. Bu mikrodenetleyici çeşitli uyku modlarına sahip olduğu için giyilebilir sistemlere uygunluğu nedeniyle tercih edilmiştir. Mikrodenetleyici çalışma frekansı 32 MHz olarak kullanılmıştır. Yazılımlar C programlama dilinde geliştirilmiştir. Bellek yönetimi ve hızlı sonuç alabilmek için tüm veriler sabit noktalı (fixed-point) olacak şekilde ayarlanmıştır. Şekil 3'te görülen her blok için sabit noktalı veri tipine uygun ayrı fonksiyonlar tanımlanmıştır. Yapılan işlemlerin doğruluğunu kontrol etmek için uygulanan yöntem ile elde edilen solunum hızı, referans cihaz olarak kullanılan BIOPAC cihazı ile alınan solunum verisiyle karşılaştırılarak test edilmiştir.

III. TEST SONUÇLARI VE LİTERATÜR KARŞILAŞTIRILMASI

Bu çalışmada Şekil 2'de verilen algoritma yukarıda belirtilen gömülü sistem üzerinde çalıştırılacak şekilde sabit noktalı veri tipine uygun olarak kodlanmıştır. Sistem çalıştırıldığında gönüllülerden alınan verilerden ilk olarak R-tepe genliklerinden solunum işareti elde edilmiştir ve bir gönüllüye ait sonuç örnek olarak Şekil 4'te verilmiştir. Bu işaret solunum dışı frekans bileşenlerinin baskılanması amacıyla bir bant geçiren filtreden (0.1Hz - 0.7Hz) geçirilmiştir [18]. Filtreleme sonucunda elde edilen EÇS işaretinin frekans içeriği, 1 dakikalık veri üzerinden 1024 noktalı FFT ile kare pencere kullanılarak analiz edilmiştir. Güç spektrumunda en yüksek genlikli frekans bileşeni, bireyin anlık solunum frekansı olarak kabul edilmiştir [13]. Bu frekans değeri 60 ile çarpılarak bireyin dakikadaki solunum hızı hesaplanmıştır. Elde edilen sonuçlar Şekil 5'te verilmiştir. Şekil 5(b)'de kırmızı çizgi ile verilen değer hesaplanan solunum hızını göstermektedir.



Şekil 4: a) Referans Solunum İşareti b) R-Tepe Enterpolasyon



Şekil 5: a) EÇS İşareti b) EÇS Hızlı Fourier Dönüşümü. EÇS: Elektrokardiyogramdan Çıkarılan Solunum

Kullanılan yöntemin başarısı, EÇS ile elde edilen solunum hızı ile referans solunum hızı arasındaki farkın ortalama mutlak hatası (MAE) ile değerlendirilmiştir. MAE ifadesi Denklem 1'de verilmiştir.

$$\text{Ortalama Mutlak Hata (MAE)} = \frac{1}{N}\sum_{i=1}^{N}\left| y_i - \hat{y}_i \right| \quad (1)$$

Denklem 1'de N deneye katılan gönüllü sayısını, y_i i. kişi için referans solunum işaretindeki solunum hızı değerini ve $\hat{y}_i$

ise EKG işaretinden kestirim yapılan solunum hızı değerini göstermektedir.

Tablo II: Referans Solunum Hızı ve Önerilen Yöntem ile Edilen Solunum Hızı Karşılaştırması. EÇS: Elektrokardiyogramdan çıkarılan solunum, MAE: Ortalama Mutlak Hata

	Referans Cihaza Ait Solunum Hızı/dk	EÇS Hızı/dk	Hata	MAE
Gönüllü 1	16	16.12	0.12	0.8
Gönüllü 2	14	15.06	1.06	
Gönüllü 3	14	13.23	0.77	
Gönüllü 4	22	23.26	1.26	
Gönüllü 5	18	18.24	0.24	
Gönüllü 6	19	20.28	1.28	
Gönüllü 7	11	10.08	0.92	
Gönüllü 8	22	22.18	0.18	
Gönüllü 9	17	19.07	2.07	
Gönüllü 10	17	17.09	0.09	

On gönüllüden elde edilen referans solunum hızıyla önerilen yöntem ile elde edilen solunum hızı karşılaştırması Tablo II'de verilmiştir. Tablo II'de EÇS hızı/dk olarak verilen veri önerilen yöntem sonucunda elde edilen frekansın 60 ile çarpılmış halidir. Bu değer yöntem sonucundan elde edilen değer olduğu için kesirli olarak tabloda verilmiş ve hata miktarları bu değerlere göre hesaplanmıştır. Tüm gönüllüler için elde edilen MAE değeri 0.8 bulunmuştur. Bu sonuç, önerilen yöntemin referans ölçümlerle yüksek uyum gösterdiğini ortaya koymaktadır.

Tablo III: Literatürdeki Diğer Çalışmalarla Karşılaştırma

	Kullanılan Veri Kümesi	Kullanılan Platform	Gönüllü Sayısı	Hareketli Kullanıma Uygunluğu	MAE
[13]	özgün veri kümesi	STM32F4	10	Yok	0.7
[14]	özgün veri kümesi	nRF52840 + bilgisayar	1(5 Farklı Deney)	Var	1
[15]	CEBS	FPGA + tümdevre	20	Var	0.73
[15]	MIT-BIH	FPGA + tümdevre	13	Var	1.2
[16]	özgün veri kümesi	NI-myDAQ + bilgisayar	5	Var	1
Önerilen Algoritma	**özgün veri kümesi**	**STM32L0**	**10**	**Var**	**0.8**

Literatürdeki benzer çalışmalar Tablo III'de yer almaktadır. Negi ve arkadaşları [13] MAE değerini 0.7 vermeleri rağmen EKG işareti sandalyeye bağlı kuru elektrotlar ile alındığından hareket halinde kullanım için uygun değildir. Fan ve arkadaşları [15] solunum hızını uygulamaya özel yüksek maliyetli bir tümdevre kullanılarak bulmuşlardır. Önerilen algoritma, düşük işlem kapasitesine sahip bir mikrodenetleyici üzerinde çalışacak şekilde tasarlandığından, elde edilen performans değerleri uygulama amacına yönelik olarak tatmin edici düzeydedir. Kullanılan mikrodenetleyicinin düşük maliyetli olması, piyasada kolayca temin edilebilmesi, küçük boyutlarda elektronik kart tasarımına olanak sağlaması ve uyku modlarıyla düşük güç tüketimi sunması, sistemin giyilebilir teknolojilere entegrasyonunu kolaylaştırmaktadır.

IV. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, EKG ve solunum işaretleri eş zamanlı olarak kaydedilerek bir veri kümesi oluşturulmuş ve bu veriler üzerinde, gömülü sistemlere uygun biçimde EKG'den solunum hızı tahmini yapan bir yöntem önerilmiştir. Gömülü sistem üzerinde elde edilen EÇS hızı değerleri, veri kümesindeki referans ölçümlerle karşılaştırılarak yöntemin doğruluğu gösterilmiştir. Özellikle, tüm işaret işleme adımlarının düşük güç tüketimli ve sınırlı işlem/bellek kaynaklarına sahip bir mikrodenetleyici üzerinde gerçekleştirilmiş olması, çözümün giyilebilir cihazlara doğrudan uygulanabilirliğini ve batarya ile uzun süreli kullanım için elverişli olduğunu ortaya koymaktadır.

Bu çalışmanın devamında önerilen yönteme uygun olarak gerçek zamanlı çalışacak şekilde elektronik donanım ve yazılım tamamlanacak, bireylerin günlük hayattaki yürüme, koşu ve farklı fiziksel aktivitelerinde kullanabileceği giyilebilir bir cihaz haline getirilecektir.

TEŞEKKÜR

Veri toplama işlemleri için YTÜ Biyomedikal Enstrümantasyon Laboratuvarı kullanılmıştır.

KAYNAKLAR

[1] GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1736–1788, 2018.

[2] A. C. Guyton and J. E. Hall, Guyton and Hall Textbook of Medical Physiology, 14th ed. Philadelphia, PA, USA: Elsevier, 2021.

[3] Bailón, R., Sörnmo, L., Laguna, P. ECG-Derived Respiratory Frequency Estimation. In Computational ECG Signal Processing, eds. G.D. Clifford, F. Azuaje, P.E. McSharry, Springer, pp. 215–243, 2006.

[4] Y. Masaoka and I. Homma, 'Anxiety and respiratory patterns: their relationship during mental stress and physical load', International Journal of Psychophysiology, vol. 27, no. 2, pp. 153–159, Sep. 1997.

[5] A. Hernando et al., 'Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment', IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 4, pp. 1016–1025, Jul. 2016.

[6] C. Varon et al., 'A Comparative Study of ECG-derived Respiration in Ambulatory Monitoring using the Single-lead ECG', Scientific Reports, vol. 10, no. 1, p. 5704, Mar. 2020.

[7] M. Schmidt, J. W. Krug, A. Schumann, K.-J. Bär, and G. Rose, 'Estimation of a respiratory signal from a single-lead ECG using the 4th order central moments', Curr. Dir. Biomed. Eng., vol. 1, no. 1, pp. 61–64, Sep. 2015.

[8] M. Campolo, D. Labate, F. La Foresta, F. C. Morabito, A. Lay-Ekuakille, and P. Vergallo, ECG-derived respiratory signal using Empirical Mode Decomposition. 2011.

[9] P. Langley, E. J. Bowers, and A. Murray, 'Principal Component Analysis as a Tool for Analyzing Beat-to-Beat Changes in ECG Features: Application to ECG-Derived Respiration', IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 821–829, Apr. 2010.

[10] D. Widjaja, C. Varon, A. Dorado, J. A. K. Suykens, and S. Van Huffel, 'Application of Kernel Principal Component Analysis for Single-Lead-ECG-Derived Respiration', IEEE Transactions on Biomedical Engineering, vol. 59, no. 4, pp. 1169–1176, Apr. 2012.

[11] Z. B. Khaled and G. Farges, First approach for respiratory monitoring by amplitude demodulation of the electrocardiogram, vol. 6. 1992.

[12] P. D. Larsen, Y. C. Tzeng, P. Y. W. Sin, and D. C. Galletly, 'Respiratory sinus arrhythmia in conscious humans during spontaneous respiration', Respir. Physiol. Neurobiol., vol. 174, no. 1–2, pp. 111–118, Nov. 2010.

[13] S. Negi, R. K. Singh, and C. S. Anoop, Development of a real-time breathing-rate monitor using difference operation method and adaptive windowing on dry-electrode ECG signal. 2017.

[14] M. Rahman and B. I. Morshed, Extraction of Respiration Rate from Wrist ECG Signals. 2021.

[15] J. Fan et al., 'A High Accuracy Ultra-Low Power ECG-Derived Respiration Estimation Processor for Wearable Respiration Monitoring Sensor', Biosensors (Basel), vol. 12, no. 8, Aug. 2022.

[16] G. B. Krishnapriya, R. N. Ponnalagu, and S. Goel, 'A Resource-Efficient Time-Domain-Based Algorithm to Estimate Respiration Rate From Single-Lead ECG Signal', IEEE Open Journal of Instrumentation and Measurement, vol. 4, pp. 1–9, 2025.

[17] https://www.biopac.com/

[18] S. Tiinanen, K. Noponen, M. Tulppo, A. Kiviniemi, and T. Seppänen, 'ECG-derived respiration methods: Adapted ICA and PCA', Medical Engineering Physics, vol. 37, no. 5, pp. 512–517, May 2015.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Breast Cancer Stage Classification with Pyramid Vision Transformer and StyleGAN3-Based Synthetic Data Augmentation

Begüm Şener
Baskent University, Ankara, Turkey begume@baskent.edu.tr

Abstract

Breast cancer remains one of the leading causes of cancer-related mortality among women worldwide, with accurate staging playing a crucial role in treatment planning and prognosis. This study proposes a hybrid deep learning framework that integrates histopathological images and biological data for breast cancer stage classification. The publicly available CPTAC-BRCA dataset, comprising histopathological whole-slide images, DNA, RNA, miRNA, WXS, and clinical data, was used. To address class imbalance, all underrepresented classes were augmented to match the largest class (Stage IIIC, 4,200 samples) using image transformations. A Pyramid Vision Transformer (PVT) was employed for feature extraction due to its ability to capture both local and global visual representations, while StyleGAN3 was used to generate high-quality synthetic histopathology images, achieving a Frechet Inception Distance (FID) of 35. Experimental results demonstrate that the proposed PVT-based model outperforms competing architectures, achieving 96.32% accuracy, 96.54% precision, 96.56% recall, and 96.20% F1-score. Furthermore, a threshold-based softmax analysis was applied to flag low-confidence predictions as potential novel cancer stages. The results indicate that combining synthetic data generation with multimodal deep learning can enhance staging accuracy and provide a foundation for discovering previously unrecognized breast cancer subtypes.

Keywords—Breast cancer staging, Histopathology, Pyramid Vision Transformer, StyleGAN3, Multimodal deep learning, Synthetic data.

I. Introduction

Breast cancer is one of the most common cancers among women worldwide and a leading cause of cancer-related death. Despite advances in early diagnosis and treatment, the disease's incidence is increasing and remains a significant public health problem. While different risk factors and screening strategies exist in developed and developing countries, the biological diversity and heterogeneity of breast cancer complicate diagnosis and treatment [1].

Molecular biomarkers play a crucial role in identifying subtypes, particularly triple-negative breast cancer (TNBC). TNBC is an aggressive form of breast cancer due to its lack of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression. The limited response of this subtype to current treatments makes its clinical management quite challenging. Therefore, a better understanding of the molecular differences in TNBC is crucial for developing more effective and personalized treatment strategies [2]. Artificial intelligence (AI), especially deep learning methods, are making significant contributions to the diagnosis and classification of breast cancer. Generative Adversarial Networks (GANs) have made significant contributions to medical imaging by facilitating the generation of high-resolution synthetic data from histopathological images [3]. These networks facilitate the generation of data for rare cancer subtypes and increase flexibility in clinical processes by generating realistic synthetic images in cases of missing or limited data [4].

To improve segmentation and classification performance in both supervised and unsupervised learning techniques, it is necessary to incorporate not only visual content but also spatial information into the model [5]. Recent research highlights the need for holistic analysis of molecular, histopathological, radiologic, and clinical data for the identification and accurate diagnosis of cancer biomarkers. These multimodal data fusion approaches offer the opportunity to evaluate the biological characteristics of tumors from a broader perspective, thus contributing significantly to the development of personalized treatment strategies [6].

Although artificial intelligence has shown potential to improve diagnostic accuracy in radiology, histopathology, and genomics, the impact of these technologies in oncology remains limited. The current literature is sparse on studies on the role of multimodal data fusion in cancer staging, indicating a significant research gap [7]. While some research has focused on multimodal integration, most of the developed AI models have not been clinically validated or are not sufficiently interpretable, severely limiting their usability in clinical settings. Furthermore, in the oncology field, lack of data, lack of established standards, and the complexity of interpreting multimodal AI approaches pose obstacles to the effective implementation of the technology.

Hybrid architectures have recently gained prominence to improve the interpretability of AI models and achieve more robust results, particularly in the face of previously unfamiliar data types or manipulations. These approaches, which combine Vision Transformers (ViT) with convolutional neural networks (CNNs), have been effective in improving model accuracy on challenging tasks such as open-set classification [8]. Such architectures offer improved performance in recognizing out-of-distribution examples and hold the potential to increase the reliability of AI in complex data environments. However, applications of medical image analysis, particularly those focused on integrating histopathological images with biological data, remain limited, and further research is needed in this area.

Innovations offered by generative models, particularly GANs and diffusion-based approaches, in synthesizing medical image data have significant potential for reducing biases in demographic representation. These models aim to improve the reliability and accuracy of diagnostic algorithms across diverse populations by eliminating imbalances in existing datasets. However, comprehensive and multifaceted studies are still needed to evaluate the effectiveness of these generative

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

approaches across different medical conditions and demographic groups and to demonstrate their clinical validity [9].

Recent studies have demonstrated the effectiveness of deep learning methods, particularly Transformer-based architectures, in analyzing histopathological images from datasets such as CPTAC-BRCA. Lu et al. [10] achieved over 91% classification accuracy using weakly supervised models on whole-slide images, showing the potential of efficient architectures for cancer subtype identification.

In this context, our study proposes a hybrid architecture that integrates Pyramid Vision Transformer (PVT) and StyleGAN3 to classify breast cancer stages using CPTAC-BRCA. Our best model achieves 96.32% accuracy and 96.20% F1-score**.**

The primary objective of this study is to develop a comprehensive model for breast cancer staging by combining both image and biological data. The model generates realistic synthetic histopathological images using GAN-based approaches, and these images are integrated with biological data such as RNA, miRNA, and clinical records using advanced deep learning architectures such as Pyramid Vision Transformer (PVT) [11]. This multimodal data integration provides a more in-depth representation of tumor heterogeneity, enabling the identification of new cancer stages and the development of personalized treatment strategies. Compared to existing models based on a single data type, the proposed approach offers higher accuracy in breast cancer staging by taking into account data diversity and distinguishes itself from the literature in this respect.

II. Methodology

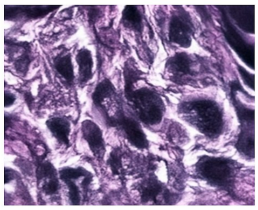
A. Dataset

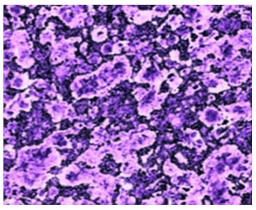
This study utilized the CPTAC-BRCA dataset, a frequently used tool in breast cancer research and provided by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) through the Cancer Imaging Archive (TCIA) platform [12]. The dataset comprises 642 histopathological images in SVS format, as well as biological samples such as DNA, RNA, miRNA, and WXS, and clinical information from a total of 134 patients. This comprehensive content allows the model to be trained on both image and biological data.

The dataset contains images corresponding to seven different stages of breast cancer: Stage IA, Stage IIA, Stage IIB, Stage III, Stage IIIA, Stage IIIB, and Stage IIIC. However, a significant imbalance is evident between the classes; Stage IIIC, with the highest sample count, contains 4,200 images, while Stage IA, with the smallest sample count, is represented by only 240 images. The other stages also have varying sample counts. The dataset contains 642 whole-slide images (WSIs), which were segmented into 9,065 image patches prior to data augmentation.

During data preprocessing, low-resolution preview images were extracted from histopathological SVS files and binary thresholding was applied to identify significant regions. Adaptive histogram equalization was then performed to increase contrast. The images were segmented into 224x224 pixel segments, resulting in a total of 19,737 segments. These segments are shown in Figure 1.

To address class imbalance, partial oversampling was applied instead of fully matching all classes to the largest class size. Specifically, the sample numbers for underrepresented classes were increased to 3,000 samples through data augmentation, while the largest class (Stage IIIC, 4,200 samples) remained unchanged. This approach reduced the disparity between classes without generating an excessive number of augmented images. Various augmentation techniques, such as random horizontal and vertical flips, rotations, color jitter, and stain normalization, were applied during this process. As a result, the total number of images increased from 9,065 to 24,600, creating a more balanced dataset that ensured fair representation of each class during training while maintaining computational efficiency. The generated image is given in Figure 2.


Fig. 1: Input Image


Fig. 2: Generated Image

For each patient, biospecimen DNA, RNA, miRNA, WXS, and clinical data were converted into fixed-length tensors. These tensors were combined and used as input data to ensure efficient integration of biomolecular and clinical information into the model.

B. Model Training Process

StyleGAN3 Training (Synthetic Data Generation) was trained independently from the classification network to generate high-quality synthetic histopathology images, used primarily to mitigate class imbalance. The discriminator was optimized using Binary Cross-Entropy (BCE) loss, while the generator employed a perceptual loss based on intermediate VGG16 feature maps to encourage realistic image detail. Training used stochastic gradient descent with gradient clipping, R1 regularization, and an early stopping strategy to ensure stability. The quality of the generated images was assessed with the Frechet Inception Distance (FID), which reached a value of 35, confirming the synthetic images closely resembled real histopathology slides. Pyramid Vision Transformer (PVT) was trained separately to predict breast cancer stages using both real and synthetic images. Unlike StyleGAN3, PVT optimization used a multi-class cross-entropy loss function suitable for classification tasks. Training incorporated stochastic gradient descent with a learning rate scheduler and early stopping to prevent overfitting. Data augmentation techniques—such as random flips, rotations, and color jitter—improved generalization. The resulting model achieved 96.32% accuracy and a 96.20% F1-score, outperforming alternative architectures including ViT, EfficientNetB0, ResNet, and Inception.

C. Classification

This study aimed to classify breast cancer stages and identify potential new stages using various deep learning architectures, including Pyramid Vision Transformer (PVT), Vision Transformer (ViT), EfficientNetB0, ResNet, and

Inception. During the labeling process, preprocessed histopathological images were matched with patient identities obtained from clinical support data and associated with accurate stage information. The resulting stage information was converted into numerical values using a label encoder for use in classification algorithms.

Among the implemented models, PVT Transformer demonstrated superior performance thanks to its hierarchical structure, sliding window mechanism, and layer norms that enable it to effectively learn both local and global visual representations. The model achieved 94.84% accuracy and 94.31% F1 score, outperforming all other architectures. Structural improvements, particularly the scaled cosine attention mechanism and residual post-norm, increased the model's balanced learning ability and stability during the training process. The ViT model, which followed the PVT Transformer, attracted attention with its 93.21% accuracy and 93.73% F1 score; it stood out as a strong alternative in medical imaging tasks thanks to its multi-scale feature learning ability.

While EfficientNet demonstrated overall good performance with an accuracy rate of 92.48%, it fell short of Transformer-based architectures in modeling the high structural complexity of histopathological images. ResNet and Inception, on the other hand, were limited in their generalization across different cancer stages due to their fixed-size kernel structure. This led to the Inception architecture's lower performance, making it a less robust alternative compared to other architectures.

In this phase, additional work was conducted to further improve the performance of the PVT model that performed best in the previous phase. This work included expanding the dataset, restructuring the training/validation/test sets, and integrating synthetic data generated with StyleGAN3. Table II shows the final and best results obtained from these optimization studies. To further improve the performance of the PVT model, the dataset was expanded and the training, validation, and test sets were restructured. This diversified data distribution increased the model's generalization capacity, raising its accuracy to 96.32% and its F1-score to 96.20%. Various data augmentation strategies, such as random flipping and rotation, were implemented to prevent overtraining, and early stopping prevented the model from overtraining. In addition, synthetic histopathology images generated with StyleGAN3 were also used in the study. PVT successfully classified these synthetic images and was effective in identifying potential new stages. Overall, PVT emerged as the most successful method. Integrating synthetic data and additional training examples into the model further improved performance, enabling a more in-depth and detailed analysis of breast cancer stages that surpasses existing classification systems.

III. Evaluation

To determine breast cancer stages, a novel conditional GAN model combining StyleGAN3 and the PVT was evaluated using both qualitative and quantitative criteria. The quality of the generated synthetic images was measured with the Frechet Inception Distance (FID), whereas the PVT classification performance was assessed using standard evaluation metrics including accuracy, precision, recall, and F1-score. Furthermore, stage identification was further investigated using a threshold-based softmax approach.

A. Frechet Inception Distance (FID)

In this research, Frechet Inception Distance (FID) is used to measure the quality of synthetic images generated by StyleGAN3. FID quantifies the difference between real and synthetic images by comparing feature distributions obtained from the Inception-v3 model [16]. The resulting FID value of 35 indicates that the synthetic images are quite close to real images. This finding confirms that the model can generate realistic and high-quality synthetic images.

B. Pyramid Vision Transformer (PVT) Classification Performance

The model demonstrated high accuracy and consistency on both real and synthetic images, proving that Results suggest that GAN-generated data can enhance classification performance

C. Softmax Probability Analysis for New Stage Discovery

A threshold-based softmax analysis was applied during the inference phase; using a threshold of 30%, images with low classification confidence were flagged as potential new stages. This method, when supported by additional validation studies, may contribute to the discovery of previously unidentified cancer stages.

IV. Conclusion and Future Work

The study demonstrates that by using histopathological images and supplementary data, GAN-based methods combined with the PVT can identify new breast cancer stages. However, limitations such as the lack of tumor region segmentation or spatial analysis still exist. The threshold-based softmax analysis is in its preliminary phase and requires further validation. To enhance staging accuracy, especially in detecting new stages with greater detail, future work may incorporate multimodal large language models (LLMs) for analyzing histopathological images and supporting data. These LLMs can provide a more nuanced understanding of the clinical implications of the newly identified stages. Table I presents the results of the models. Table II shows the performance results of PVT.

Table I. Results for models

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
PVT	94.84	94.82	94.21	94.31
ResNet	90.72	90.61	90.52	90.17
EfficientNetB0	92.48	92.44	92.42	92.34
ViT	93.21	93.12	93.10	93.73
Inception-v3	91.23	91.38	91.34	91.38

Table II. Improved Classification Performance of the PVT Model

Metric	Value
Accuracy (%)	96.32
Precision (%)	96.54
Recall (%)	96.56
F1-Score (%)	96.20

V. References

[1] Trayes, K. P., & Cokenakes, S. E. (2021). Breast cancer treatment. *American family physician*, *104*(2), 171-178.

[2] Lu, B., Natarajan, E., Balaji Raghavendran, H. R., & Markandan, U. D. (2023). Molecular classification, treatment, and genetic biomarkers in triple-negative breast cancer: a review. *Technology in cancer research & treatment*, *22*, 15330338221145246.

[3] Alajaji, S. A., Khoury, Z. H., Elgharib, M., Saeed, M., Ahmed, A. R., Khan, M. B., ... & Sultan, A. S. (2024). Generative adversarial networks in digital histopathology: current applications, limitations, ethical considerations, and future directions. *Modern Pathology*, *37*(1), 100369.

[4] Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023, June). On the detection of synthetic images generated by diffusion models. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[5] Valindria, V. V., Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E. O., Rockall, A. G., ... & Glocker, B. (2017). Reverse classification accuracy: predicting segmentation performance in the absence of ground truth. *IEEE transactions on medical imaging*, *36*(8), 1597-1606.

[6] Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449-1477.

[7] Unger, M., & Kather, J. N. (2024). Deep learning in cancer genomics and histopathology. *Genome medicine*, *16*(1), 44.

[8] Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2021). Do vision transformers see like convolutional neural networks?. *Advances in neural information processing systems*, *34*, 12116-12128.

[9] Yeom, T., Gu, C., & Lee, M. (2024). DuDGAN: improving class-conditional GANs via dual-diffusion. *IEEE Access*, *12*, 39651-39661.

[10] **Lu M. Y. et al.,** *Data-efficient and weakly supervised computational pathology on whole-slide images*, Nat Biomed Eng, 2021;5:555-570.

[11] Wang, W., Xie, E., Li, X., Fan, D. P., Song, K., Liang, D., ... & Shao, L. (2022). Pvt v2: Improved baselines with pyramid vision transformer. *Computational visual media*, *8*(3), 415-424.

[12] D. P. Clark, M. Gulley, M. Marko, J. Samayoa, R. T. Kelly, and S. Sinha, "CPTAC Breast Cancer (CPTAC-BRCA) Collection [Data set]," 2021, [Online]. Available: https://doi.org/10.7937/TCIA.2021.YI7CP4XS.

Enhanced External Counterpulsation (EECP) Therapy System Supported by Synchronous Vital Monitoring

Onur Koçak
Computer & Biomedical Engineering, Baskent University,
okocak@baskent.edu.tr

Cansel Ficici
Electrical and Electronics Engineering, Ankara University,
ogretmenoglu@ankara.edu.tr

Ziya Telatar
Biomedical Engineering, Baskent University,
ztelatar@baskent.edu.tr

Abstract— Coronary artery diseases rank among the leading causes of mortality worldwide and in Türkiye. According to OECD data, our country is unfortunately not in a favorable position in terms of early diagnosis rates and timely access to treatment planning. The therapeutic approach is based on the principle of dilating coronary vessels and their fine terminal branches, while promoting the development of collateral circulation networks between these branches. This ensures adequate myocardial perfusion in patients with vascular occlusion or stenosis, thereby preventing symptoms such as chest pain, fatigue, shortness of breath, or palpitations. Furthermore, it reduces the risk of myocardial infarction and significantly improves patient quality of life.

In this study, the design process of a medical device capable of monitoring vascular obstructions through a non-surgical approach and contributing to early treatment was described using scientific methodology. Instead of a dual-compressor configuration, a compressor–vacuum combination was employed to ensure mechanical continuity, resulting in a linear pressure system operating between 0–5 bar. In the pneumatic design, transmission losses were minimized through leak tests to ensure tight sealing. Additionally, single-lead ECG recording, pulse rate detection, and ST-segment duration measurements were integrated to achieve synchronized operation with EECP therapy.

The developed electronic and pneumatic modules were enclosed in a dedicated housing, making the system ready for clinical research. Performance evaluation was conducted by validating pressure stability, leak integrity, and vital parameter signals using reference measurement devices. In all tests, results were within a ±5% tolerance range.

Keywords—Enhanced External Counterpulsation, Cardiovascular Medical Devices, Pneumatic System Design, Real-Time Physiological Monitoring

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of morbidity and mortality worldwide, with coronary artery disease (CAD) accounting for a substantial proportion of cases [1]. Enhanced External Counter pulsation (EECP) is a non-invasive, U.S. Food and Drug Administration (FDA)-approved therapy for patients with refractory angina and CAD who are not candidates for invasive revascularization [2]. The therapy operates by cyclically inflating and deflating pneumatic cuffs positioned around the patient's lower extremities, synchronized with the cardiac cycle via electrocardiogram (ECG) signals [2], [3].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Inflation during diastole augments coronary perfusion pressure, enhances venous return, and promotes collateral vessel formation through shear stress–mediated arteria genesis [4-6].

Multiple clinical studies and meta-analyses have demonstrated the efficacy of EECP in improving angina symptoms, left ventricular systolic function, myocardial perfusion, and overall quality of life [1, 7, 8]. Data from Türkiye further indicate that EECP provides sustained reductions in angina frequency, improves New York Heart Association (NYHA) functional class, and enhances left ventricular ejection fraction in local patient populations [9, 10]. Recent evidence also suggests that EECP may improve endothelial function, microcirculatory health, exercise tolerance, and metabolic capacity [8, 11]. Biochemical studies have reported favorable modulation of circulating markers, including nitric oxide, endothelin-1, and high-sensitivity C-reactive protein following treatment [11]. The precision and safety of EECP depend on accurate synchronization with the cardiac cycle and stable pneumatic pressure control. Integrating synchronous vital monitoring—comprising single-lead ECG, pulse rate detection, and ST-segment duration measurement—into EECP systems may not only improve therapeutic outcomes but also facilitate early diagnosis and real-time safety monitoring, which is particularly relevant in regions such as Türkiye where the cardiovascular disease burden is high.

This study presents the design and integration of an EECP system incorporating synchronous vital sign monitoring, optimized for clinical readiness in Türkiye. The device employs a compressor–vacuum pneumatic configuration operating in the 0–5 bar range to ensure linear pressure delivery and mechanical continuity, with leak-tested sealing to minimize transmission losses. A microcontroller-based control unit (Arduino Mega) with relay-actuated valves provides precise timing and coordination for pneumatic operation, synchronized with ECG, pulse rate, and ST-segment measurements. All pneumatic and electronic modules are housed in custom enclosures, with dedicated valve boxes to ensure mechanical stability, vibration reduction, and component protection. Power is supplied through a transformer converting 220 V AC to 24 V DC, enabling safe and stable distribution to all subsystems. The mechanical, electrical, and pneumatic assemblies are integrated into a single, unified structure with organized tubing and wiring to ensure durability, accessibility, and ease of maintenance.

The study describes the mechanical, pneumatic, and electronic design of an EECP system triggered by electrocardiogram (ECG) signals acquired from a vital parameter monitor developed by our research group. The work reports on technology readiness level (TRL) 8 developments completed prior to initiating the clinical research phase.

II. Materials and Methods

The system was assembled from pneumatic components, control electronics, power supply modules, and structural supports. A microcontroller-based unit governed relay-actuated valves powered through a stepped-down electrical source. Pneumatic conduits were routed and secured to the main frame, and custom enclosures were fabricated for structural stability and component protection. Final integration included mechanical mounting, electrical interconnection, and pneumatic coupling to yield a fully functional prototype.

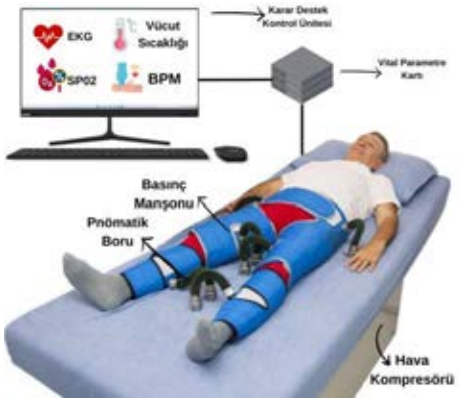


Figure 1. Compact View of EECP System

A. Air Supply Unit

The air supply unit delivers the pressurized and vacuum air required for system operation. It comprises an oil-free compressor to generate clean compressed air and a vacuum pump to produce the necessary negative pressure. These components constitute the core of the pneumatic subsystem, providing consistent and reliable airflow to sustain device functionality.

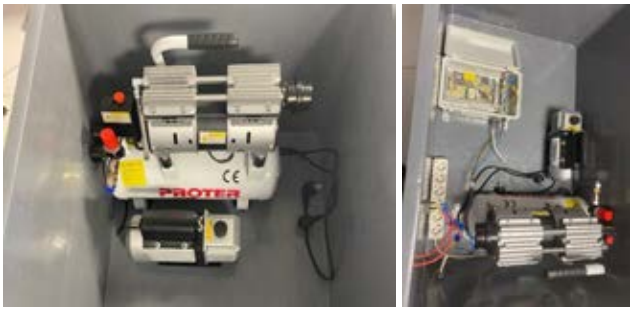


Figure 2. PROTER Air Compressor / VALUE VE-115 Vacuum Pump and Power Connections

B. Power Supply and Distribution

The power supply and distribution subsystem convert mains electricity to the operating voltages required by the control and actuation components. It incorporates a transformer unit and associated wiring to deliver stable power to the electronic circuits and pneumatic controls, while ensuring safe and organized electrical distribution throughout the device.

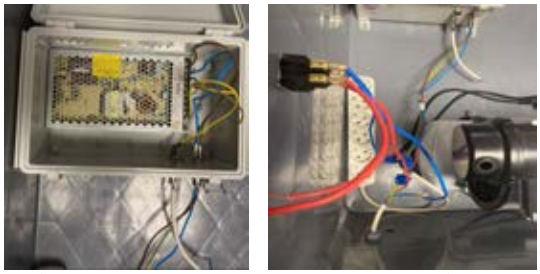


Figure 3. Input 220v, Output 24v Power Connection Box

C. Control and Vital Parameter Circuits

The control unit regulates pneumatic system operation by processing input signals and generating actuator commands. It is based on a microcontroller platform interfaced with relay modules that switch pneumatic valves on and off as required. This configuration enables precise timing and synchronized control of device functions. The PM4100, as an OEM board, can interface with any microprocessor or microcontroller. This compatibility requires serial communication between the OEM board and the host microprocessor/microcontroller. The serial link is configured as UART at 115200 bps, with 8 data bits, 1 stop bit, and no parity. Figure 4 presents the design of the slave Arduino boards used to control the relay modules in conjunction with the vital parameter monitor.



Figure 4. Arduino MEGA with 8xRelay Module and PM-4100 Vital Parameter Circuit

D. Pneumatic Components

The Pneumatic Components form the functional core of the air- driven system, consisting of valves, tubing, and connectors that direct and control airflow. The valves regulate the timing and sequence of air delivery, while the tubing and fittings ensure secure and efficient transfer of pressurized or vacuum air between system elements.

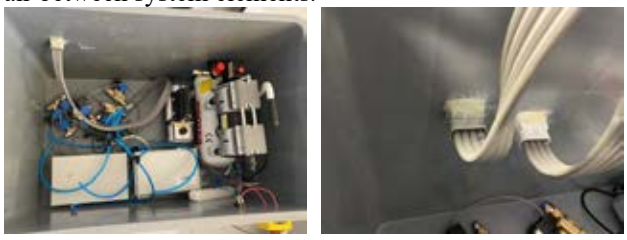


Figure 5. Pipes and Connectors

Actuators

The Valve Actuation and Housing system controls the airflow through precision pneumatic valves, which are responsible for executing the device's operational cycles. These valves are securely mounted within custom-designed protective boxes to ensure stability, minimize vibration, and protect the components from environmental factors. The enclosures also facilitate organized routing of pneumatic lines and wiring, contributing to the overall durability and maintainability of the system renders.

The final system assembly integrated all mechanical, pneumatic, and electrical subsystems into a fully operational unit. Following the preparation of the air supply, power distribution, control, and valve actuation subsystems, each was mounted in its designated location within the device frame. Pneumatic lines were routed through pre-drilled passages to provide secure, unobstructed connections to the end-use components, while electrical wiring was organized to ensure accessibility and operational safety. Valve boxes were installed to protect and stabilize the actuation components, preventing displacement during operation. All air and electrical connections were tested for continuity and leakage to verify system integrity.

III. RESULTS AND DISCUSSION

R peaks were detected from the QRS complex by thresholding peaks exceeding 0.5 mV (MinPeakHeight = 0.5) and enforcing a minimum inter-peak interval of 200 samples (MinPeakDistance = 200) to suppress false detections. Q-wave localization was performed after noise reduction using a Savitzky–Golay filter. Peaks within the –0.2 mV to –0.5 mV amplitude range were classified as Q waves.

Subsequently, the positions of the Q waves were determined. Direct peak thresholding for Q-wave detection is prone to errors due to the embedding of Q waves in noise. Therefore, the signal was first filtered before peak detection. To suppress noise, Savitzky–Golay filtering was applied. Peaks with amplitudes between –0.2 mV and –0.5 mV were identified as Q waves. Heart rate was computed using equations (1) and (2).

$$Mean\ RR\ \text{range} = \frac{1}{N-1}\left(\sum_{i=1}^{N-1}[R\ peak_{i+1} - R\ peak_i]\right) \times \frac{1}{f_s} \quad (1)$$

$$BPM = \frac{60sn}{Mean\ RR\ \text{range(sec)}} \quad (2)$$

For the example signal shown, the sampling frequency (fsf_sfs) was 500 Hz, and the analysis window length was 2000 samples. Within this window, four R peaks were detected, yielding N=4N = 4N=4. For mean RR interval estimation, the positions of the R peaks were first identified. Using equation (1), the mean RR interval was calculated as 1 s. Applying equation (2) resulted in a heart rate of 60 beats per minute.

Based on these calculations, the following measurements were performed, single-channel ECG data acquisition, PPG signal acquisition, NIBP measurement and pulse detection

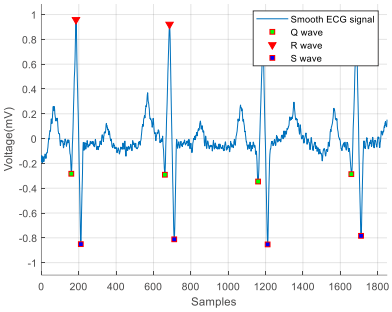


Figure 6. R Peak Detection

The compressor inflates the cuffs upon receiving a trigger from the detected R peaks. The cuffs are divided into three main sections: lower leg, upper leg, and waist. Synchronization is achieved such that, following each R peak, inflation occurs within approximately 50 ms, followed by deflation over a duration of approximately 50 ms. The corresponding synchronization graph is presented in Fig. 7.

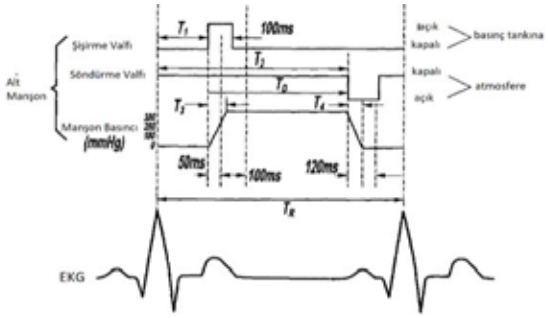


Figure 7. Synchronization Type

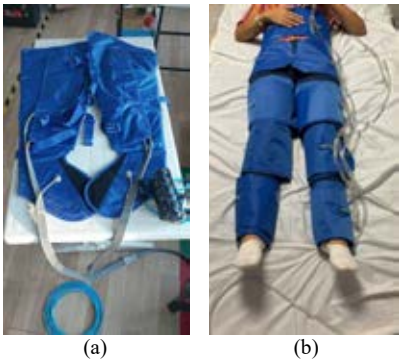
A graphical user interface (GUI) was developed on the Python platform to enable system control. The interface includes fields for patient information such as age, weight, and clinical complaints, as well as real-time vital monitor readings.

Prototype testing was conducted in two phases: preliminary validation on inanimate mannequins and subsequent evaluation on a volunteer subject. Figure 8 illustrates the system connection during volunteer testing.

Following the conceptual design stage, synchronized cuff inflation was implemented in three distinct zones—lower leg, upper leg, and waist—and calibrated to match inflation patterns documented in the literature [12]. Leak-proof, pressure-resistant drainage valves were integrated into the pneumatic circuit and mechanically coupled to the main assembly to ensure operational reliability under clinical conditions.

For initial mechanical validation, the cuffs were inflated to approximately 2 bar and applied to mannequin models. This stage focused on verifying leak integrity, pressure stability, and valve response timing. No leaks or pressure losses were

observed during multiple inflation–deflation cycles. Subsequent testing on a healthy volunteer confirmed proper cuff placement, secure fastening, and effective three-zone synchronized inflation. Inflation was triggered by R-peak detection from the ECG signal, with an approximate 50 ms delay from R-wave identification to inflation onset, followed by a 50 ms deflation period. These operational parameters were consistent with the intended control logic and demonstrated functional readiness for controlled clinical testing.



(a) (b)

Figure 8. a) All Cuff System and Connection Cables b) Cuff Connection on a Volunteer

The design has been completed at Technology Readiness Level (TRL) 8. In future work, a Bill of Materials (BOM) will be generated from the critical design documentation, and IEC 60601 and IEC 60000 series electrical safety and electromagnetic compatibility tests will be conducted by an independent testing organization. Following these tests, an ethics committee application will be submitted to initiate clinical research studies.

IV. CONCLUSION

In this study, the electronic and software design processes of an EECP system were presented. The novelty of the work lies in transmitting the triggering command required for the EECP procedure to the hardware immediately after the R-peaks of the ECG signal. For this purpose, R-peaks were detected, and synchronized triggering was performed according to the venous return period (Figure 7).

The main limitation of the study is the absence of clinical trials. Since such trials will need to be conducted on patients with severe cardiac conditions, future work is planned in collaboration with clinicians and under the approval of an ethics committee.

In subsequent stages, medical compressors and vacuum systems with different power capacities, together with alternative air hose designs, can be integrated into the system. This will enable optimized operation for users of different body sizes and weights.

Additionally, a compact housing was designed to integrate the mechanical system, and the compressor–vacuum unit was synchronized with the pulse information derived from the ECG signal. By combining essential vital sign monitoring functions—single-channel ECG, pulse detection, and oxygen saturation measurement—through OEM electronic modules and established engineering techniques, this study introduces a novel and cost-effective approach to EECP system development. The proposed synchronization innovation provides a new scientific perspective for future cardiological EECP device production.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by Marla Teknoloji Müh. San. ve Tic. Ltd. Şti. and its esteemed management for the execution of this study.

REFERENCES

[1] M. Amin, Y. Al Hajri, H. C. C. Lee, M. H. Suppogu, and C. J. Cabrera, "Enhanced external counterpulsation for refractory angina pectoris," *PLOS One*, vol. 11, no. 3, e0151822, Mar. 2016, doi: 10.1371/journal.pone.0151822.
[2] Cleveland Clinic, "Enhanced External Counterpulsation (EECP)," 2024. [Online]. Available: https://my.clevelandclinic.org/health/treatments/16949-enhanced-external-counterpulsation-eecp. [Accessed: Aug. 18, 2025].
[3] F. Yang, Y. Zhang, Y. Jiang, Q. Wei, X. Cheng, J. Xiao, and G. Chen, "The effect of enhanced external counterpulsation on anxiety and depression in patients with paroxysmal atrial fibrillation: A randomized controlled trial," *Frontiers in Cardiovascular Medicine*, vol. 12, p. 1548839, 2025, doi: 10.3389/fcvm.2025.1548839.
[4] "Enhanced External Counterpulsation: Mechanism of Action," *Progress in Cardiovascular Diseases*, vol. 54, no. 4, pp. 332–337, 2012.
[5] Q. Wang, J. Hao, W. Jiang, and Q. Tan, "Enhanced external counterpulsation increases coronary flow reserve in coronary microvascular disease," *Saudi Medical Journal*, vol. 44, no. 12, pp. 1277–1282, Dec. 2023, doi: 10.15537/smj.2023.44.12.20230427.
[6] O. Soran, C. Ikizler, A. Sengül, B. Çuğlan, E. Kennard, and S. Kelsey, "Comparison of long term clinical outcomes, event free survival rates of patients undergoing enhanced external counterpulsation for coronary artery disease in the United States and Turkey," *Turkish Heart Journal*, vol. 40, no. 4, pp. 323–330, 2012, doi: 10.5543/tkda.2012.59144.
[7] Z. Zhang, Y. Zhang, C. Zhong, Y. Wang, W. Wei, and G. Wu, "A randomized controlled study of enhanced external counterpulsation with cardiac rehabilitation in patients with nonobstructive coronary artery disease and coronary microvascular dysfunction," *International Heart Journal*, vol. 65, no. 3, pp. 380–385, May 2024, doi: 10.1536/ihj.23-687.
[8] B. Çuğlan, H. Turhan, and E. Yetkin, "Enhanced external counterpulsation effects on venous leg symptoms," *Anatolian Journal of Cardiology*, vol. 26, no. 5, pp. 401–406, May 2022.
[9] P. O. Bonetti, G. W. Barsness, P. C. Keelan, T. I. Schnell, G. M. Pumper, J. T. Kuvin, R. P. Schnall, D. R. Holmes, S. T. Higano, and A. Lerman, "Enhanced external counterpulsation improves endothelial function in patients with symptomatic coronary artery disease," *Journal of the American College of Cardiology*, vol. 41, no. 10, pp. 1761–1768, May 2003, doi: 10.1016/S0735-1097(03)00329-2.
[10] S. K. Shakouri, Z. Razavi, F. Eslamian, H. Sadeghi-Bazargani, S. Ghaffari, and A. Babaei-Ghazani, "Effect of enhanced external counterpulsation and cardiac rehabilitation on quality of life, plasma nitric oxide, endothelin 1 and high sensitive CRP in patients with coronary artery disease: A pilot study," *Annals of Rehabilitation Medicine*, vol. 39, no. 2, pp. 191–198, Apr. 2015.
[11] Y. Zhong, L. Hao, X. Jia, S. Paek, S. Tian, and G. Wu, "Evaluation of enhanced external counterpulsation with different modes on acute hemodynamic effects," *Frontiers in Physiology*, vol. 16, Feb. 2025, Art. no. 1555358, doi: 10.3389/fphys.2025.1555358.
[12] J. Caceres, P. Atal, R. Arora, and D. Yee, "Enhanced external counterpulsation: A unique treatment for the 'No-Option' refractory angina patient," *Journal of Clinical Pharmacy and Therapeutics*, Jan. 2021, doi: 10.1111/jcpt.13330.

TIPTEKNO'25
MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Konuşma Engelliler için Yapay Zeka Tabanlı Duygusal Ses Üretim Sistemi
AI-Based Emotional Speech Generation System for Individuals with Speech Impairments

Efe Eren Yağcı
Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
yagciefe@gmail.com

Eren Kalkan
Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
erenkalkan6@gmail.com

Güray Yılmaz
Bilgisayar Mühendisliği Bölümü
Biruni Üniversitesi
İstanbul, Türkiye
gurayy@biruni.edu.tr

Özetçe **—Bu çalışmada, metinlerdeki duygusal içerikleri analiz ederek bunlara uygun sesli çıktılar üreten yapay zekâ tabanlı bir platform geliştirilmiştir. Sistem, doğal dil işleme (NLP) teknikleri ile metni işleyip duyguları belirli kategorilerde sınıflandırır ve bu bilgiyi, metnin doğal ve duygusal tonlamayla seslendirilmesi için kullanır. Platform, konuşma engelli bireylerin duygu ve düşüncelerini etkili biçimde ifade etmelerini sağlayarak iletişim becerilerini ve sosyal etkileşimlerini güçlendirir. Yapay zekâ modelleri farklı duygu yoğunluklarına sahip metinlerde yüksek doğrulukla duygu analizi yapacak şekilde optimize edilmiş, ses sentezi algoritmaları ise doğal tonlamalar üreterek kullanıcı deneyimini artırmıştır. Bu yaklaşım, bireysel iletişimin yanı sıra eğitim, terapi ve dijital içerik üretimi gibi alanlarda da önemli faydalar sunar. Sonuç olarak, çalışma hem teknolojik yenilik hem de sosyal katkı açısından konuşma engelli bireyler için kapsamlı bir çözüm sunmaktadır.**

Anahtar Kelimeler—Yapay Zeka, Doğal Dil İşleme, Duygu Analizi, Metinden Sese Dönüştürme, Duygusal Tonlama, Konuşma Engelli Bireyler, Yardımcı Teknolojiler

***Abstract*—This study presents an AI-based platform that analyzes the emotional content of texts and generates corresponding speech outputs. The system processes text using natural language processing (NLP) techniques to classify emotions into specific categories and uses this information to produce speech with natural and expressive intonation. The platform enables individuals with speech impairments to convey their feelings and thoughts effectively, enhancing communication skills and social interaction. The AI models are optimized for accurate emotion recognition across texts with varying emotional intensities, while the speech synthesis algorithms generate natural intonations to improve user experience. This approach provides benefits not only for personal communication but also in education, therapy, and digital content creation. In conclusion, the study offers a comprehensive solution for individuals with speech impairments, combining technological innovation with social impact.**

Keywords—Artificial Intelligence (AI), Natural Language Processing (NLP), Emotion Analysis, Text-to-Speech (TTS), Emotional Intonation, Speech-Impaired Individuals, Assistive Technologies.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

I. Introduction

Son yıllarda dijital medya ve sesli içeriklerin hızla yaygınlaşmasıyla podcastler, bilgi, eğlence ve duygusal içerik sunan önemli bir iletişim aracı haline gelmiştir. Ancak sesli içeriklerin duygusal yönlerinin doğru anlaşılması ve aktarılması her zaman kolay değildir. Bu çalışma, metinlerdeki duyguları analiz ederek uygun tonlama ile seslendirme yapan yapay zeka tabanlı bir sistem geliştirmeyi hedeflemektedir.

Doğal dil işleme (NLP) teknikleriyle metinlerdeki temel duygular (mutluluk, üzüntü, öfke vb.) sınıflandırılacak; elde edilen sonuçlar, **duygusal tonlamayı destekleyen TTS (Text-to-Speech)** sistemleri ile seslendirme sürecine entegre edilecektir. Bu amaçla Transformer tabanlı modeller (BERT, GPT) ve Tacotron 2 gibi TTS çözümleri tercih edilebilir.

Bu çalışma yalnızca duygusal açıdan zengin sesli içerikler üretmekle kalmayacak, aynı zamanda **konuşma engelli bireylerin** duygu ve düşüncelerini daha etkili ifade edebilmelerine imkân tanıyacaktır. Böylece, onların sosyal etkileşimlerini güçlendirirken ileride karşılaşabilecekleri olası psikolojik ve ruhsal zorlukların önüne geçilmesine katkı sağlayacaktır. Ayrıca altyazı, görsel işaretler veya titreşimli geri bildirim gibi ek özelliklerle kapsayıcı bir deneyim sunulması hedeflenmektedir.

A. Konuşma Engelli Bireyler İçin Erişilebilir İçerikler

Podcastler, insanların bilgi edinme ve kendilerini ifade etme biçimlerini dönüştüren etkili araçlardır. Ancak konuşma engelli bireyler için kendi sesleriyle podcast oluşturmak, teknik ve fiziksel zorluklar nedeniyle çoğu zaman mümkün olamamaktadır. Bu durum, onların fikirlerini, duygularını ve deneyimlerini geniş bir kitleye aktarma fırsatını sınırlandırmaktadır.

Bu çalışma, konuşma engelli bireylerin podcast üretimini kolaylaştırmayı amaçlamaktadır. Metin tabanlı içeriklerin, duygusal tonlamaları doğru şekilde yansıtan yapay zeka tabanlı seslendirme sistemleri ile dönüştürülmesi sayesinde, üretilen podcastler yalnızca bilgi sunmakla kalmayacak; aynı zamanda dinleyicide istenen duygu ve etkiyi de yaratabilecektir.

Ayrıca, çalışmanın kullanacağı duygusal tonlama analiz ve sentez teknikleri, konuşma engelli bireylerin kendilerine özgü üslup ve duygularını yansıtmalarına olanak tanıyacaktır. Bu sayede, medya içeriklerinde daha fazla temsil sağlanacak ve içerik üretim süreci herkes için daha kapsayıcı bir hale gelecektir.

B. İLGİLİ ÇALIŞMALAR

Duygu analizi ve metinden sese dönüştürme (TTS) teknolojileri, son yıllarda hızla gelişmiş ve kullanıcıların metin tabanlı içeriklerle daha doğal şekilde etkileşim kurmasını mümkün kılmıştır. Bu teknolojiler, hem günlük yaşam uygulamalarında hem de erişilebilirlik çözümlerinde önemli katkılar sunmaktadır.

1) Metinden Duygu Analizi: Duygu analizi, bir metindeki olumlu, olumsuz veya nötr gibi duygusal durumları otomatik olarak tespit etmeye yönelik doğal dil işleme yöntemidir [1], [7]. Özellikle sosyal medya verilerinin analizi ve müşteri geri bildirimlerinin değerlendirilmesinde yaygın olarak kullanılmaktadır. Go ve arkadaşlarının geliştirdiği Sentiment140 sistemi [6], Twitter verilerini kullanarak kısa metinlerde duygu sınıflandırması yapmış ve bu alanın gelişiminde önemli bir adım olmuştur. Bu çalışmada da benzer bir mantık benimsenmiş; metinlerdeki duygular tespit edilerek, bu duyguların seslendirme sürecine yansıtılması hedeflenmiştir. Böylece üretilen sesli içerik, yalnızca bilgilendirici değil, aynı zamanda duygusal ton açısından da zengin bir hale gelmektedir [7].

2) Tonlamalı Metinden Sese Dönüştürme (TTS) Sistemleri: TTS sistemleri, metinleri anlaşılır ve doğal insan sesiyle ifade etmeyi amaçlayan teknolojilerdir [4], [5]. Son dönemde, bu sistemlere duygusal tonlama eklenmesi, kullanıcı deneyimini önemli ölçüde geliştirmiştir. Shen ve ekibinin geliştirdiği Tacotron 2 modeli [5], metin tabanlı girdilerden yüksek kaliteli ve doğal ses sentezi üretme konusunda önemli bir kilometre taşıdır. Bu çalışmada kullanılan yaklaşım, metinden elde edilen duygu bilgisini tonlama ve vurgu parametreleriyle uyumlu hale getirerek, konuşmanın doğallığını ve ifade gücünü artırmaktadır.

3) Engelli Bireyler İçin Erişilebilirlik Uygulamaları: Metinden duygu analizi ve duygusal tonlamalı TTS sistemleri, engelli bireyler için iletişim engellerini azaltma potansiyeline sahiptir. Özellikle konuşma engelli bireylerin kendilerini ifade etme süreçlerinde duygusal tonlama, aktarılmak istenen mesajın daha doğru anlaşılmasını sağlar. Geleneksel altyazı ve işaret dili çözümleri çoğu zaman sesli içeriklerdeki duygusal nüansları yansıtamaz. Bu nedenle, duygu analizi ile desteklenmiş TTS sistemleri [3], hem görme hem de konuşma engelli bireylerin içerikleri daha bütüncül bir şekilde algılamalarına yardımcı olmaktadır.

Tablo I: Duygu analizi ve TTS çalışmaları

Yazarlar	Yıl	Çalışma Başlığı
Go, Bhayani, Huang	2009	Sentiment140
Shen, Pang, Weiss	2018	Tacotron 2
Livingstone, Russo	2018	RAVDESS
Dupuis, Pichora-Fuller	2010	TESS

Bu çalışma, yalnızca metnin anlamını değil, aynı zamanda içeriğin duygusal tonlarını da aktaran bir yapı sunarak erişilebilirlik alanında yeni bir adım atmaktadır. Böylelikle, konuşma engelli bireylerin duygu ve düşüncelerini daha doğru ve etkili bir biçimde ifade etmeleri sağlanmakta, olası psikolojik ve sosyal sorunların önüne geçilmesine katkı sunulmaktadır.

II. MATERIALS AND METHODS

A. Genel Model Yapısı

Bu çalışmada geliştirilen model, metin tabanlı duygu analizi ve text-to-speech (TTS) teknolojilerinin entegrasyonuna dayanmaktadır. Model, verilen bir metni yalnızca doğru bir şekilde seslendirmekle kalmaz, aynı zamanda metnin duygusal tonlamalarını ve vurgularını da dikkate alır. Modelin işleyişi iki aşamadan oluşmaktadır:

1) Duygu analizi
2) Ses tonlaması ve vurgulama

Metin giriş olarak alındığında, önce duygu analizi yapılır; ardından duyguların metindeki yeri ve anlamına göre seslendirme parametreleri (tonlama, hız, vurgu) ayarlanır. Bu işlem, derin öğrenme tabanlı tekniklerle gerçekleştirilen bir optimizasyon sürecidir.

Tablo II: Duygu kategorileri ve metin dağılımı

Duygu kategorisi	**Metin sayısı**	**Yüzde (%)**
Mutluluk	120	30%
Üzüntü	80	20%
Öfke	60	15%
Korku	40	10%
Nötr	100	25%

B. Duygu Analizi Aşaması

Duygu analizi aşamasında, metin üzerinde doğal dil işleme (NLP) ve makine öğrenmesi teknikleri uygulanır [1], [7]. Model, metindeki duygusal yükü çıkarmak için emotion detection (duygu tespiti) algoritmalarını kullanır. Bu aşama, metnin içerdiği duygusal ifadeleri tespit etmek için dilsel özellikler üzerinde yoğunlaşır. Kullanılan başlıca algoritmalar: Support Vector Machine (SVM), Random Forest ve Deep Neural Networks (DNN)'dir [7].

Bu algoritmalar, kelime ve cümle düzeyinde duygu etiketleme yaparak metindeki ana duyguları (mutluluk, öfke, üzüntü vb.) sınıflandırır. Duygu analizi, semantic parsing (anlamsal çözümleme) ve contextual embedding (bağlamsal gömme) ile desteklenerek doğruluğu artırılır [1]. Örneğin, BERT (Bidirectional Encoder Representations from Transformers) gibi önceden eğitilmiş modeller metnin bağlamını anlamada yüksek doğruluk sağlar.

C. Ses Tonlaması ve Vurgu Ekleme

Duygu analizi sonucunda belirlenen duygular, TTS sistemine entegre edilen tonlama ve vurgu parametrelerine dönüştürülür. Model, duyguya göre sesin pitch (tonlama), rate (hız), intensity (yoğunluk) ve prosody (ritim) özelliklerini dinamik olarak ayarlar. Bu, sesli içerikte duygusal anlamı doğru iletmek için kritik öneme sahiptir.

Ses tonlaması için WaveNet ve Tacotron 2 gibi derin öğrenme tabanlı modeller kullanılmıştır. Bu modeller, sesin doğallığını ve duygusal doğruluğunu artırırken metnin fonetik

yapısını da hesaba katarak uygun ses çıktısı üretir [4], [5]. Örneğin, "öfkeli" metin yüksek ton ve hızlı hızla; "üzüntü" içeren metin ise düşük ton ve yavaş hızla seslendirilir. Ayrıca, FastSpeech 2 [2] modeli, daha hızlı ve verimli TTS üretimi sağlamak için alternatif bir yöntem olarak değerlendirilmiştir.

D. TTS Teknolojisinin Entegrasyonu

Metnin seslendirilmesinde WaveNet ve Tacotron 2 temel bileşenlerdir. WaveNet, sesin doğal tonunu ve duygusal aktarımını iyileştirirken, Tacotron 2 sesin zamanlamasını ve akışını sağlar. Prosody (ritim) ve intonasyon (duygusal ifade) özellikleri bu modellerle optimize edilir [2], [4], [5].

E. Erişilebilirlik ve Konuşma Engelli Bireyler İçin İyileştirmeler

Konuşma engelli bireyler, kendi sesleriyle içerik üretmekte zorluk yaşadıkları için, metin tabanlı içeriklerin otomatik olarak seslendirilmesi büyük önem taşır. Bu nedenle, çalışma kapsamında geliştirilen sistem, kullanıcıların yazdıkları metinleri doğal ve duygusal tonlamalarla seslendirecek şekilde tasarlanmıştır [3].

Ayrıca, duygusal tonlama ve vurgu bilgileri görsel geri bildirimlerle desteklenmektedir. Örneğin, metindeki vurgular veya duygu yoğunlukları renkler veya animasyonlar aracılığıyla gösterilebilir. Bu sayede konuşma engelli bireyler, ürettikleri podcast veya sesli içeriklerin duygusal bağlamını hem işitsel hem de görsel olarak takip edebilir ve içeriklerini daha etkili bir şekilde yönetebilir.

F. Modelin Eğitim Süreci ve Veritabanı

Modelin eğitimi, büyük veri kümeleri ve etiketlenmiş duygu verileri kullanılarak gerçekleştirilmiştir. Eğitim verisi metinler üzerinden duygusal etiketler alınarak DNN ve LSTM gibi derin öğrenme algoritmalarıyla işlenmiştir. Kullanılan text corpus, her duygu etiketinin doğru sesli içeriklere dönüştürülmesini sağlayacak şekilde yapılandırılmıştır [7].

III. Results

Bu bölümde, geliştirilen modelin performansı ve elde edilen sonuçlar detaylı bir şekilde açıklanmaktadır. Testler, modelin duygu analizi ve sesli çıktı üretme yeteneklerini farklı metin türleri ve kullanıcı grupları üzerinde ölçmek amacıyla yapılmıştır.

A. Test Ortamı ve Kullanılan Veri Seti

Modelin performansını değerlendirmek için, farklı duygu yoğunluklarına ve metin türlerine sahip geniş bir test veri seti kullanılmıştır. Veri seti aşağıdaki kriterlere göre hazırlanmıştır:

- **Duygusal Çeşitlilik:** Mutluluk, üzüntü, öfke, korku gibi temel duyguların yanı sıra daha karmaşık duygusal ifadeler de içerir.
- **Bağlam ve Metin Çeşitliliği:** Haber metinleri, edebi eserler, günlük konuşmalar ve sosyal medya gönderileri gibi farklı türler, modelin genel duygu analizi yeteneğini ölçmek için kullanılmıştır.

Veri seti, modelin farklı bağlamlarda doğru tahminler yapabilmesini sağlayacak şekilde dengelenmiştir.

B. Duygu Analizi Performansı

Modelin metindeki duyguları doğru tespit etme kabiliyeti, makine öğrenimi metrikleri ile değerlendirilmiştir. Kullanılan metrikler: Precision, Recall ve F1-Score.

- **Precision:** Doğru duygu tahmini yapma oranını gösterir. Model özellikle güçlü duygularda yüksek değerler elde etmiştir.
- **Recall:** Tüm duygusal ifadeleri yakalama oranını gösterir. Model, az belirgin duygusal ifadeleri bile başarıyla tespit etmiştir.
- **F1-Score:** Precision ve Recall'un dengeli göstergesi olarak, modelin genel duygu analiz başarısını ifade eder. Ortalama F1-Score

Tablo III: Duygu kategorilerine göre performans sonuçları

Duygu kategorisi	Precision	Recall	F1-score
Mutluluk	0.88	0.85	0.86
Üzüntü	0.84	0.80	0.82
Öfke	0.90	0.88	0.89
Korku	0.83	0.81	0.82
Nötr	0.87	0.86	0.86

C. Sesli Çıktı Performansı

Modelin sesli çıktı üretme yeteneği, farklı duygusal tonlamalar içeren metinler üzerinde kullanıcı tabanlı testlerle değerlendirilmiştir. Kullanıcılar, üretilen seslerin doğallığını ve duygusal doğruluğunu değerlendirmiştir. Bu amaçla MOS (Mean Opinion Score) testi uygulanmış ve katılımcılardan 1 ile 5 arasında puan vermeleri istenmiştir.

Elde edilen sonuçlara göre, modelin ortalama MOS puanı 4.5 olarak ölçülmüştür. Bu değer, literatürdeki standart Tacotron 2 tabanlı sistemlerin ortalama $\approx$4.2 MOS skoruna kıyasla yaklaşık %7'lik bir iyileşme sağlamaktadır. Bu durum, modelimizin hem doğal hem de duygusal açıdan daha doğru seslendirme yapabildiğini göstermektedir.

Tablo IV: Kullanıcı gruplarının MOS puanları

Kullanıcı grubu	MOS ortalama puanı
Genel kullanıcılar	4.5
Konuşma engelli bireyler	4.4
Profesyonel değerlendiriciler	4.7

Not: Farklı kullanıcı gruplarının sistemin seslendirme kalitesine ilişkin değerlendirmelerini göstermektedir.

D. Karşılaştırmalı Değerlendirme

Önerilen modelin performansı, mevcut duygu analizi ve TTS sistemleri ile karşılaştırılmıştır. Çalışma, özellikle konuşma engelli bireylerin yazılı metinlerini daha doğal ve duygusal tonlamalarla seslendirmek amacıyla geliştirilmiştir.

- **Duygu Analizi:** Önerilen model, rakip sistemlere kıyasla daha yüksek doğruluk ve F1-Score değerleri elde etmiştir. Bu durum, metindeki duyguları doğru şekilde sınıflandırabilme yeteneğinin güçlü olduğunu göstermektedir.

- **Sesli Çıktı Kalitesi:** Model, duygusal tonlamaları ve vurguları başarılı şekilde yansıtarak kullanıcılar tarafından daha doğal ve etkileyici bulunmuştur. Özellikle konuşma engelli bireylerin duygu aktarımını güçlendirecek bir ses çıktısı sağlamaktadır.

Tablo V: Modellerin performans karşılaştırması

Model	Precision	Recall	F1-score	MOS puanı
Önerilen model	0.86	0.84	0.85	4.5
Model A	0.80	0.78	0.79	4.3
Model B	0.82	0.80	0.81	4.4

Elde edilen sonuçlar incelendiğinde, önerilen modelin F1-Score değeri en yakın rakibine göre **%4.93** daha yüksek bulunmuştur (0.85 yerine 0.81). Ayrıca MOS puanı 4.5 ile diğer modellerden daha yüksek olup, bu da seslendirme kalitesinde **doğallık ve duygu aktarımı açısından üstünlük** sağladığını göstermektedir.

E. Erişilebilirlik ve Konuşma Engelli Bireyler İçin İyileştirmeler

Model, konuşma engelli bireylerin duygu ve düşüncelerini etkili bir şekilde ifade edebilmeleri için geliştirilmiştir. Metin tabanlı içerikler, doğru duygu analizi ile zenginleştirilmiş ve doğal tonlamalara sahip sesli çıktılara dönüştürülmüştür. Ayrıca, kullanıcıların duygusal içeriklerini daha iyi iletebilmeleri amacıyla basit ve erişilebilir bir arayüz tasarlanmıştır.

F. Sonuç ve Çıkarımlar

Test ve değerlendirme sonuçları, modelin metin tabanlı duygu analizi ve duygusal tonlamalı seslendirme alanında yüksek doğruluk ve kullanıcı memnuniyeti sağladığını göstermektedir. Özellikle konuşma engelli bireyler için, modelin sunduğu sesli çıktılar, iletişimdeki önemli bir engeli ortadan kaldırma potansiyeline sahiptir.

IV. Discussion and Conclusion

Bu bölümde, çalışmanın elde ettiği başarılar, karşılaşılan zorluklar ve gelecekteki çalışmalar için öneriler özetlenmektedir. Geliştirilen model, metin tabanlı duygu analizi ve duygusal tonlamalı seslendirme alanında önemli bir yenilik sunmaktadır.

A. Çalışma Başarıları

Çalışmanın temel hedefi, metin içeriğinden duygu analizi yaparak doğal ve duygusal tonlamalara sahip sesli içerikler üretmektir. Öne çıkan başarılar şunlardır:

- **Doğru Duygu Analizi:** Model, metinlerdeki duyguları yüksek doğrulukla analiz etmiştir. Bu, haber, edebi metinler ve sosyal medya içerikleri gibi çeşitli metin türlerinde etkileyici sonuçlar vermiştir.
- **Doğal ve Tonlamalı Seslendirme:** Duygu analizi sonucu elde edilen bilgiler, metinleri duygusal tonlamalarla zenginleştirilmiş sesli içeriklere dönüştürmüştür. Kullanıcılar, metin içeriklerini daha etkili bir şekilde algılayabilmiştir.
- **Erişilebilirlik:** Konuşma engelli bireyler için geliştirilen kullanıcı dostu arayüz, metin tabanlı içeriklerin kolayca seslendirilmesini sağlamış ve iletişim sürecini kolaylaştırmıştır.

B. Karşılaşılan Zorluklar

- **Duygusal Çeşitlilik:** Farklı kültürlerden ve bağlamlardan gelen metinlerin duygusal çeşitliliğini doğru analiz etmek zaman zaman zorluk yaratmıştır. Özellikle karmaşık veya az belirgin duygusal ifadelerde bu durum gözlemlenmiştir.
- **Sesli Çıktı Doğallığı:** Bazı kullanıcılar, daha az yaygın duygusal tonlamalarda sesli içeriklerin doğallığını düşük bulmuştur. Bu, modelin geliştirilmesi gereken bir alan olarak belirlenmiştir.

C. Gelecekteki Çalışmalar

Bu çalışma duygu analizi ve seslendirme alanında önemli bir temel oluşturmuş olsa da, gelecekte yapılabilecek çalışmalar şunlardır:

- **Gelişmiş Duygu Analizi:** Daha karmaşık ve çok boyutlu duygusal ifadelerin doğru analiz edilebilmesi için modelin daha geniş veri setleriyle eğitilmesi planlanmaktadır.
- **Daha Doğal Seslendirme:** Kullanıcı geri bildirimlerine dayalı olarak, sesli içeriklerin doğallığını artırmak ve daha geniş bir duygusal yelpazeyi kapsamak için ses sentezi algoritmaları geliştirilecektir.
- **Erişilebilirlik İyileştirmeleri:** Konuşma engelli bireyler için daha kişiselleştirilmiş, hızlı ve etkileşimli deneyimler sunacak özellikler eklenecektir.

Kaynakça

[1] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon.

[2] Ren, Y., et al. (2021). FastSpeech 2: Fast and High-Quality End-to-End Text to Speech.

[3] Specia, L., & Ruiter, F. (2018). Accessibility and Inclusion in AI Applications.

[4] Oord, A. v. d., Dieleman, S., Zen, H., et al. (2016). WaveNet: A Generative Model for Raw Audio.

[5] Shen, J., Pang, R., Weiss, R. J., et al. (2018). Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions.

[6] Go, A., Bhayani, R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision.

[7] Zhang, L., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Few-Shot Medical Image Segmentation with Meta-Learning: An Empirical Study on Organ Segmentation

Ismail EL Jamiy*†, Amel Aissaoui*, Lotfi Chaari*,
*IRIT, Toulouse University, CNRS, Toulouse INP, UT3
Toulouse, France.
e-mail: firstname.lastname@irit.fr
†ENSAM Casablanca, University Hassan 2
Casablanca, Morocco

***Abstract*—Few-shot learning has emerged as a promising approach for medical image segmentation, where annotated data are scarce for certain organs or pathologies. In this work, we explore the use of the Model-Agnostic Meta-Learning (MAML) framework to enable rapid adaptation to new organ segmentation tasks in limited data context using the BTCV multi-organ dataset. A 2D U-Net model is trained within the MAML paradigm, providing a meta-learned initialization that can be fine-tuned with only a few annotated examples of an unseen organ. Experimental results across 3-, 5- and 10-shots settings demonstrate that MAML consistently achieves faster and more effective adaptation than conventional training and organ-specific pretraining, especially in very low-shot regimes. These findings highlight the potential of optimization-based meta-learning for organ-specific medical image segmentation in data-limited scenarios.**

***Keywords*—Medical imaging, deep learning, few-shot learning, MAML, organ segmentation.**

I. Introduction

Medical image segmentation plays a pivotal role in numerous clinical applications, including surgical planning, disease diagnosis, and radiotherapy treatment. With the recent advancements in deep learning, deep convolutional neural networks (CNN), particularly U-Net, have become the standard for medical image segmentation [1]. However, these models need extensive training on large annotated datasets for optimal performance, which are unvailable (rare pathologies or imaging protocols), costly or time-intensive since they require inputs from experts with extensive clinical experience. Indeed, their performance heavily decreases in scarse data context. Few-shot learning (FSL) addresses this challenge by enabling training models with only a small number of annotated examples. Within FSL, meta-learning, and in particular the Model-Agnostic Meta-Learning (MAML) framework, has shown strong potential by learning initializations that adapt quickly to unseen tasks [2]. While it is one of the most widely used approaches, its applicability to medical image segmentation remains underexplored [3]. In this paper, we investigate the application of MAML to medical image segmentation, with a particular focus on adapting to novel organs using the BTCV multi-organ dataset. The paper reviews related work, presents the MAML framework, describes our experiments, and concludes with perspectives.

979-8-3315-5565-8/25 $31.00 © 2025 European Union

II. Related Works

A. Deep Learning for Medical Image Segmentation

The U-Net architecture [1] has become the standard for a wide range of medical segmentation tasks. Its symmetric encoder-decoder structure, complemented by skip connections that propagate fine-grained spatial information from the contracting to the expansive path, has proven exceptionally effective for precise pixel-level localization. Since its introduction, U-Net has inspired a vast family of extensions designed to address different challenges in medical imaging. For example, 3D U-Net [4] extends convolutional and pooling operations into three dimensions, making it particularly effective for volumetric data such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans. Attention U-Net [5] integrates attention gates to automatically focus on the most relevant regions of interest, improving segmentation performance in cluttered or noisy images. Beyond CNN-based modifications, the U-Net framework has also been fused with other paradigms. For instance, transformer-based U-Nets [6] was introduced to better capture long-range dependencies and contextual information. Despite the success of U-Net and its derivatives [7], they fundamentally dependent on the availability of large, expertly annotated training datasets.

B. Strategies for Data-Scarce Scenarios

In order to mitigate the data scarcity, a common strategy is data augmentation, which synthetically increases dataset diversity by applying transformations such as rotation, scaling, flipping, elastic deformation, or intensity variations [8]. More advanced strategies include generative approaches to synthesize realistic training samples [9], [10]. However, synthetic images may fail to capture the full variability of real-world

pathologies and excessive or unrealistic transformations can introduce artifacts that reduce clinical plausibility [11]. Another widely adopted strategy is transfer learning, where a model is pre-trained on a large-scale dataset and subsequently fine-tuned on a smaller, domain-specific dataset [12]. However, this approach is often constrained by the domain gap between the source and the target medical datasets [13]. Also, fine-tuning still requires a sufficient number of examples to avoid overfitting, and its performance degrades sharply when only a handful of labeled samples are available.

To address data scarcity more directly, FSL has emerged as a variety of techniques enabling models to learn from a minimal number of labeled examples [2]. Within FSL, meta-learning (based on the concept of learn to learn) has become a dominant paradigm. Model-Agnostic Meta-Learning (MAML), proposed by Finn et al., is a prominent optimization-based meta-learning algorithm [14] that has achieved notable success in few-shot learning. However, its performance has been demonstrated primarily in image classification tasks, and most existing MAML-based frameworks remain centered on this domain [15]. In contrast, its potential for dense prediction problems such as medical image segmentation has not yet been fully explored [3]. Segmentation task presents additional challenges beyond classification, including substantial anatomical variability, severe class imbalance, and the requirement for fine-grained, pixel-level precision. This work aims to contribute to filling this gap by providing a direct empirical study of MAML on a multi-organ segmentation task.

III. The MAML approach

As an optimization-based meta-learning approach, MAML's goal is to train a *meta-learner* (the outer-loop optimization algorithm) to produce a highly adaptable model, which we refer to as the *adapter*. The entire process is split into two phases: meta-training and meta-testing.

During meta-training, the meta-learner is exposed to a large number of distinct learning problems, drawn from a distribution of tasks $p(\mathcal{T})$. A *task* $\mathcal{T}_i$ represents a specific learning problem, for example, "segment the liver" or "classify this skin lesion". For each task, a small dataset is provided, which is split into two disjoint sets: 1) a support Set (D_{support}) of K labeled examples (K-shots) that the adapter uses to learn the specific task and 2) a query Set (D_{query}) of samples used to evaluate how well the adapter has learned the task after being trained on the support set. The loss on this set provides the training signal for the meta-learner.

The training process is structured as a bi-level optimization problem, consisting of an inner loop for task-specific adaptation and an outer loop for meta-optimization across tasks. Let the adapter model be parameterized by a vector θ.

a) Inner Loop (Task-Level Adaptation): For each task $\mathcal{T}_i$ sampled from the meta-training set, a temporary, task-specific parameter vector θ_i' is created. This is achieved by updating the initial parameters θ for one or more gradient descent steps using the loss calculated on that task's *support set*, $D^{(i)}_{\text{support}}$. For a single gradient step, this update is:

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta), \tag{1}$$

where α is the inner learning rate, $\mathcal{L}_{\mathcal{T}_i}$ is the task-specific loss function (e.g., cross-entropy) and f_θ is the model (e.g. CNN) with θ parameters.

b) Outer Loop (Meta-Optimization): The performance of the adapted parameters θ_i' is then evaluated on the corresponding *query set*, $D^{(i)}_{\text{query}}$. The obtained loss serves as the meta-loss for that task. The initial parameter vector θ is then updated by taking a gradient step with respect to this meta-loss, aggregated over a batch of tasks. The meta-update rule is:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'}), \tag{2}$$

where β is the meta learning rate. Crucially, the gradient ∇_θ is calculated with respect to the initial parameters θ, not the adapted parameters θ_i'. This means that the optimization must differentiate through the inner-loop update (Eq. (1)). This "gradient through a gradient" operation requires computing second-order derivatives (Hessian-vector products), which allows the meta-learner to update the initial weights θ by considering how the inner-loop update will affect the final query set loss. This process aims to find a shared initialization θ that is positioned in a way that facilitates effective and efficient fine-tuning for a variety of novel tasks. In the meta-testing phase, a new task $\mathcal{T}_{\text{test}}$ is sampled from outside the meta-training distribution. The model is initialized with the meta-learned parameters θ and adapted to the new task using a small support set $D^{(\text{test})}_{\text{support}}$. Then, the performance is assessed on the corresponding query set $D^{(\text{test})}_{\text{query}}$, providing an evaluation of the model's ability to rapidly generalize.

IV. Experimental study on organ segmentation

A. Dataset description

All experiments were conducted on the BTCV Multi-Organ Abdominal CT Dataset [16]. It consists of 30 abdominal 3D CT volumes from healthy subjects. Each volume contains between 85 and 198 axial slices of 512×512 pixels, with an anisotropic voxel spatial resolution of $(0.54 \times 0.98 \times [2.5 \sim 5.0])$ mm^3. The dataset provides dense, manual annotations for 13 different abdominal organs, making it a suitable benchmark for creating a diverse task distribution. To evaluate the results on classes with various shapes, locations and textures we considered a subset containing left kidney, right kidney, spleen and liver, as suggested in [17].

B. Preprocessing pipeline

To prepare the data for our models, we followed a standardized preprocessing pipeline for each 3D volume:

1) Slice extraction: The 3D NIfTI volumes were processed into a collection of 2D axial slices. In order to mitigate the impact of class imbalance, we kept only slices in

which the organ mask occupied at least 10% of the image area for all the experiments.

2) Intensity clipping and normalization: The raw CT intensity values, measured in Hounsfield Units (HU), were clipped to a window of [-175, 250] HU. This step focuses the intensity range on soft tissues, which is relevant for abdominal organ segmentation. Following clipping, the intensity values were then normalized to a floating-point range of [0, 1], as suggested in [18].
3) Resizing: all 2D slices and their corresponding masks were resized to a uniform dimension of 256×256 pixels.

C. Evaluation protocol

A strict data spliting protocol was designed to ensure a fair evaluation of the few-shot learning capabilities. This protocol involved splits at the patient, class (organ), and episodic levels. The dataset is randomly partitioned into a *meta-training pool* of 25 subjects and a held-out *meta-testing pool* of 3 subjects.

a) Meta-Training: For the meta-training phase, from the 25 patients, we used a group of 22 patients as the primary source for generating the *patient support set*, and the remaining 3 patients were designated for generating the *patient query set*. To simulate the few-shot learning scenario, we adopted a leave-one-out cross-validation scheme at the class level. In each experiment, 3 of these organs were used to form the task distribution in the meta-training and the left organ was reserved for meta-testing to evaluate the model's ability to adapt to a novel, unseen organ. The meta-training process operates by constructing *episodes*, where each episode simulates a few-shot learning problem for a specific task (e.g., "segment spleen"). To create an episode, a set of slices is sampled from the pool of all available data for that organ to create : 1) an *organ support set*, containing K examples (e.g., $K = 5$ for 5-shot configuration), drawn from the *patient support set* pool, which is used to adapt the model to the specific task during the inner loop; and 2) a larger, disjoint *organ query set*, drawn from the *patient query set* pool, which is used to evaluate the adapted model's performance and to compute the loss for the outer loop update.

b) Meta-Testing: For the meta-testing phase, the 3 dedicated held-out patients were used. This set was further split: one patient served as the source for the training data (support set), another for validation, and the last for testing (query set). This strict separation ensures a fair evaluation of generalization to both new patients and new organs.

D. Model architecture

A standard 2D U-Net architecture was employed as the backbone for all experiments. It consists of a symmetric encoder-decoder structure. The encoder contains four down-sampling blocks, where each block applies two consecutive 3×3 convolutions followed by a 2×2 max-pooling operation. The decoder path mirrors the encoder, using transposed convolutions to up-sample the feature maps while halving the channel count at each level. This architecture was used for both the conventionally- and meta-trained models in order to isolate the impact of the learning paradigm and ensures that any performance differences are attributable to the training strategy itself, not the model's capacity.

V. Results and Discussion

We present in this section the evaluation of our meta-learning framework. Since the dataset includes four organs, we consider four experimental configurations, each corresponding to a leave-one-organ-out setting, where a single organ is held out for meta-testing while the others are used for meta-training and meta-validation. We illustrate the training losses for each meta-model and its associated adapters across the four configurations in Figure 1. The reported results were obtained

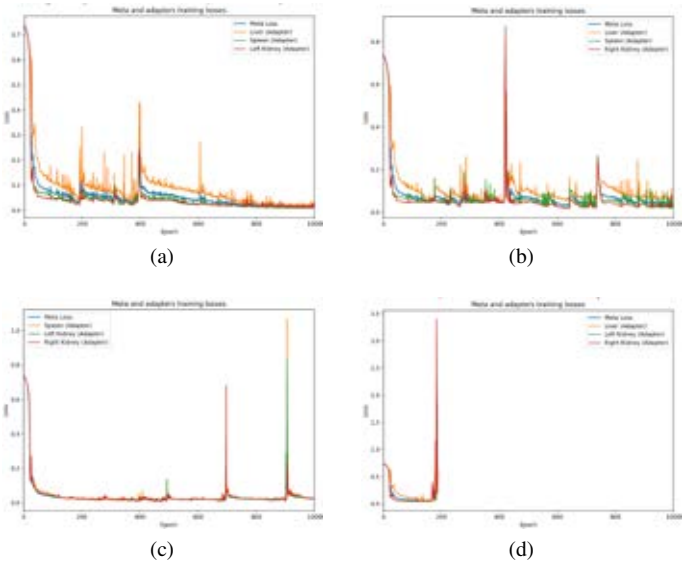


Fig. 1: Meta-training loss dynamics of the meta-model and adapters across four leave-one-organ-out configurations: (a) RK held-out (trained on L, S and LK), (b) LK held-out (trained on L, S and RK), (c) L held-out (trained on S, RK and LK), (d) S held-out (trained on L, RK and LK). Abbreviations: RK = Right Kidney, LK = Left Kidney, L = Liver, S = Spleen.

using the following hyperparameters: 1000 training epochs, 5 episodes per iteration, and 5 shots per class. Binary cross-entropy was employed as the training loss, while the Dice score served as the evaluation metric.

The training dynamics show that while the meta-learning process can converge stably for some organ combinations (notably when training on liver–spleen–left kidney), others (particularly liver–kidneys without spleen) are prone to instability, with the kidneys repeatedly emerging as more difficult organs for adaptation during meta-training. This behavior can be explained by organ-specific characteristics: kidneys are smaller, more variable in position and shape, providing weaker and noisier gradients during training, making them harder tasks for adaptation. In contrast, the spleen, being larger than kidneys and more homogeneous than liver, tends to stabilize optimization when included among the training tasks,

which may account for the smoother convergence observed in configurations where it is present.

For the meta-testing, we select the first configuration (liver–spleen–left kidney) at 900 epochs as it is the most stable. Figure 2 reports the fine-tuning performance of the meta-model, a randomly initialized model, and models pretrained on each of the training organs, evaluated with 3, 5, and 10 shots over 5–20 adaptation steps. Results show clear differences

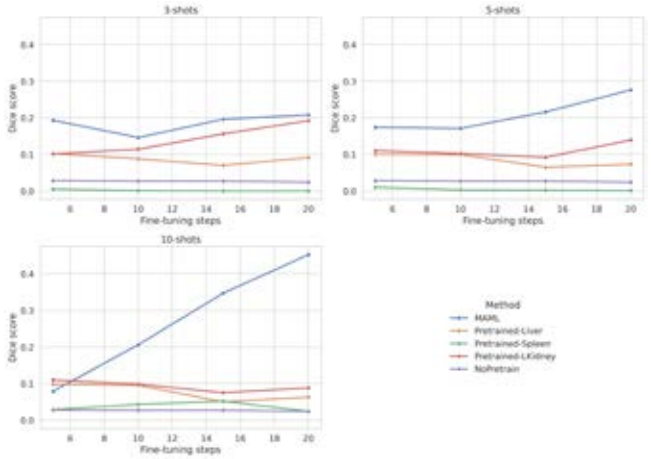


Fig. 2: A comparison of fine-tuning performance between meta-learned and conventional models.

across methods and shot settings. With only 3 shots, MAML reaches Dice ≈0.19–0.21, outperforming all classical transfers, except left-kidney pretraining (≈0.19 at 20 steps). At 5 shots, MAML remains ahead (up to 0.27), while classical models stay below 0.14. At 10 shots, MAML briefly lags at 5 steps (0.079 vs. ∼0.11) but rapidly surpasses all baselines, reaching ∼0.45 by 20 steps whereas classical variants remain ≤0.088. Overall, MAML consistently adapts faster and achieves higher performance, while classical pretraining offers modest gains.

VI. Conclusion and Perspectives

This study demonstrates that meta-learning, specifically MAML, can substantially improve the adaptation of segmentation models to new tasks (here, organs) in few-shot settings. Experiments on the BTCV dataset reveal that convergence stability is strongly influenced by the choice of training organs; nevertheless, under stable configurations MAML consistently outperforms a conventional U-Net baseline, achieving higher Dice scores across 3-, 5-, and 10-shot scenarios.

In future work, we will investigate the configurations where meta training exhibited unstable optimization and convergence difficulties. Understanding and mitigating these optimization challenges is crucial to ensure the robustness of the framework across all task configurations. We intend to consider more challenging datasets, including cases with tumors or lesions in different anatomical regions, and also datasets acquired under different institutions or patient populations. This is a critical step toward real-world deployment, as domain shifts caused by acquisition protocols or demographic variability often lead to significant performance drops. We will also investigate cross-modality adaptation (such as CT to MRI), which is particularly relevant in clinical practice where data availability often varies across modalities. We also aim to extend our framework to 3D architectures for volumetric segmentation. This transition is non-trivial, as voxel-based representations substantially increase computational and memory requirements, making the optimization more challenging.

The code supporting this work is maintained in a private Git repository and can be provided upon reasonable request.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *LNCS*, vol. 9351, pp. 234–241, 10 2015.

[2] E. Pachetti and S. Colantonio, "A systematic review of few-shot learning in medical imaging," *Artificial Intelligence in Medicine*, vol. 156, p. 102949, 08 2024.

[3] S. Luo, Y. Li, P. Gao, Y. Wang, and S. Serikawa, "Meta-seg: A survey of meta-learning for image segmentation," *Pattern Recognition*, vol. 126, p. 108586, 02 2022.

[4] Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2016.

[5] A. Al Qurri and M. Almekkawy, "Improved U-Net with attention for medical image segmentation," *Sensors*, vol. 23, no. 20, p. 8589, 2023.

[6] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang *et al.*, "Transunet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, p. 103280, 2024.

[7] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, p. 1224, 01 2021.

[8] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, 2019.

[9] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," in *Neurocomputing*, vol. 321. Elsevier, 2018, pp. 321–331.

[10] J. Ehrhardt and M. Wilms, "Chapter 8 - autoencoders and variational autoencoders in medical image analysis," in *Biomedical Image Synthesis and Simulation*, ser. The MICCAI Society book Series, N. Burgos and D. Svoboda, Eds. Academic Press, 2022, pp. 129–162.

[11] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," in *arXiv preprint :1712.04621*, 2017.

[12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[13] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.

[14] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," 03 2017.

[15] S. M. R. Arnold, P. Mahajan, D. Datta, I. Bunner, and K. S. Zarkias, "learn2learn: A library for meta-learning research," *CoRR*, vol. abs/2008.12284, 2020. [Online]. Available: https://arxiv.org/abs/2008.12284

[16] harrigr, "Segmentation outside the cranial vault challenge," 2015. [Online]. Available: https://repo-prod.prod.sagebase.org/repo/v1/doi/locate?id=syn3193805type=ENTITY

[17] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 762–780.

[18] L. Zhou, H. Liu, J. Bae, J. He, D. Samaras, and P. Prasanna, "Self pre-training with masked autoencoders for medical image classification and segmentation," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023, pp. 1–6.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Interaction-Aware Interpersonal Graph Neural Network (IA-IGNN) for Stress Detection with Biosignals

İrem Aksoy *, Fatma Patlar Akbulut†
*Department of Computer Engineering, Istanbul Kültür University, Istanbul, Turkey
†Department of Software Engineering, Istanbul Kültür University, Istanbul, Turkey

***Abstract*—Real-time detection of stress is essential for personalized mental-health interventions, yet existing methods struggle to jointly model temporal dynamics and interpersonal physiological patterns. In this paper, we introduce an Interaction-Aware Interpersonal Graph Neural Network (IA-IGNN)that fuses attention-weighted BiLSTM embeddings of electrodermal activity, skin temperature, and blood-volume pulse with a k-nearest-neighbor interpersonal graph defined on physiological similarity. Specifically, each participant's temporal embeddings are compared in Euclidean feature space, and the k most similar neighbors are used to establish interpersonal edges. This allows IA-IGNN combines both intra-window temporal features and cross-participant relational cues into one system. Experiments on a cohort of 30 subjects demonstrate an average improvement of 20% in classification accuracy over baseline. The results confirm that graph-based multimodal fusion enhances stress detection performance.**

***Keywords*—Stress Detection, Wearable Sensors, Graph Neural Networks, Deep Learning, Multimodal Fusion**

I. Introduction

Stress is a common problem that affects many aspects of modern life and has serious effects on mental and physical health. Most traditional ways of measuring stress use subjective questionnaires or occasional clinical measurements, which don't show how quickly and dynamically a person' stress level changes. Wearable biosensor have made it possible to continuously monitor physiological signals that gives us a rich, multimodal data source that shows how autonomic nervous system activity and external stressors interact in complex ways. But it is still very hard to combine and understand these different time-series signals, because traditional feature-based or end-to-end deep learning methods often miss the relationships between sensor modalities and time segments in context.

Graph neural networks (GNNs) are one of the promising solution since they treat each sensor channel and time window as nodes in a single graph that lets them capture both the dynamics of each sensor and the dependencies between sensors. Our method uses message-passing to learn strong embeddings that show the underlying stress response patterns by turning multimodal physiological data into a dynamic graph. This graph-based framework not only keeps the topological relationships between signals, but it also lets you add new sensors or modalities without having to redesign the whole architecture. Even though GNNs have been used successfully in areas like social network analysis and predicting molecular properties, they haven't been used much for real-time stress detection yet, especially in wearable systems where latency and computational limits are very important. Our main contributions are:

1) We present the model which combines attention-weighted BiLSTM embeddings of EDA, skin temperature, and BVP with an interpersonal k-nearest-neighbor graph to model stress patterns over time and between participants.
2) We optimize IA-IGNN by using lightweight attention mechanisms and graph sparsification.
3) We did ablation studies on k-NN connectivity and attention configurations to measure how much they help performance and give clear instructions on how to choose hyperparameters for multimodal stress detection.

II. Related Work

In the fields of affective computing and wearable health monitoring, stress detection [1] has been of great interest for a long time. Early methods mostly used hand-crafted features taken from single-channel biosignals, like time and frequency-domain statistics of heart rate variability (HRV) [2] or skin conductance level (SCL) [3]. These methods worked fairly well in controlled settings, but they weren't very strong in the real world because they relied on fixed feature sets and the quality of the signal and user behavior can change a lot.

Deep learning brought with it end-to-end architectures that learn hierarchical representations straight from raw sensor data. Early research first adapt convolutional neural networks (CNNs) on spectrograms and short-time Fourier transforms of some biosignals like ECG and EDA. They showed that CNNs worked better than traditional classifiers [4]. Recurrent neural networks (RNNs), especially bidirectional long short-term memory (BiLSTM) models, have made temporal modeling even better by finding long-range dependencies in physiological time series [5]. However, these sequence-based models often treat each modality in isolation or combine modalities only at a late fusion stage [6], thus failing to

979-8-3315-5565-8/25/$31.00 © 2025 IEEE

exploit cross-modal interactions occurring at finer temporal granularity.

More recent research has used multimodal fusion strategies to combine useful data from sensors that are different from each other. Attention mechanisms and gated fusion layers have been employed to weight each modality adaptively based on its instantaneous reliability, yielding gains in stress classification accuracy [7]. Graphical models, like Conditional random fields (CRFs) make sure that adjacent time windows are consistent over time. However, these methods usually need the user to manually specify the graph topology and have trouble adapting to changes in real time.

Graph neural networks (GNNs) give us a way to learn both the structure and the weights of a graph representation directly from data [8]. Although GNNs have achieved state-of-the-art results in domains ranging from social network analysis to molecular property prediction [9], [10], their application to wearable physiological signal analysis remains nascent. A few recent studies have used GNNs to classify sleep stages and find seizures by making graphs over sensor channels or time segments [11], but these studies didn't look at the strict latency and resource limits of real-time stress [12], [13] monitoring on devices. Despite significant progress in both multimodal fusion and graph-based modeling of physiological signals, existing methods have yet to reconcile the competing demands of detection accuracy, interpretability, and on-device efficiency.

III. Methodology

We evaluate three different architectures on the Wearable Stress and Affect Detection (WESAD) dataset [14]. This dataset has synchronized recordings of electrodermal activity (EDA), skin temperature (TEMP), and blood volume pulse (BVP) from a wrist-mounted device during baseline, stress, and amusement states. Because there aren't enough examples of the stress and amusement classes and they are related in meaning, we combine them into one "non-neutral" class and ask whether something is neutral or not. The Figure 1 shows the raw wrist biosignals from Subject 17 in the WESAD dataset when they were in a neutral state. It shows the three types of data used in this study such as EDA, skin temperature, and BVP and how they each have their own patterns. In this stage, the continuous time series are broken up into windows and normalized before feature extraction.

The Figure 2 shows how the proposed IA-IGNN model works from start to finish. After segmentation, a BiLSTM with Attention encodes each signal window to capture temporal dependencies. A k-NN graph is made from the resulting embeddings based on physiological similarity. Then, a Graph Convolutional Network processes the graph to model relational patterns across windows at the same time. This pipeline figure shows how graph reasoning and temporal representation work together in a single stress classification framework.

A. Graph Convolutional Network with Statistical Features

In the architecture, node representations are extracted by taking statistical descriptors from each 10-second window of the raw signal $x_m^{(i)}(t)$ of modality $m \in \{\mathrm{EDA}, \mathrm{TEMP}, \mathrm{BVP}\}$. In the i^{th} window of length L, we find the feature vector for each window by calculating

$$\mathbf{f}^{(i)} = \big[\mu_m, \sigma_m, \max_m, \min_m\big]_{m\in\{\mathrm{EDA},\mathrm{TEMP},\mathrm{BVP}\}} \in \mathbb{R}^{12},$$

where μ_m, σ_m, $\max_m$ and $\min_m$ are the mean, standard deviation, maximum, and minimum of $x_m^{(i)}(t)$. We make a graph $G = (V, E)$ where the nodes $V = \{\mathbf{f}^{(i)}\}$ are all the windows and the edges are based on a 5-nearest-neighbor rule in feature space. The adjacency matrix $A \in \{0,1\}^{N\times N}$ is is

$$A_{ij} = \begin{cases} 1, & \mathbf{f}^{(j)} \in \mathcal{N}_5(\mathbf{f}^{(i)}), \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathcal{N}_5$ gives back the 5-nearest-neighbor based on Euclidean distance. A two-layer Graph Convolutional Network (GCN) then computes

$$\mathbf{H}^{(1)} = \sigma\big(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}\mathbf{F}\mathbf{W}^{(0)}\big),$$

$$\mathbf{H}^{(2)} = \mathrm{Softmax}\big(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}\mathbf{H}^{(1)}\mathbf{W}^{(1)}\big),$$

with $\tilde{A} = A + I$, $\tilde{D}$ its degree matrix, $\mathbf{F}$ the matrix of all $\mathbf{f}^{(i)}$, and $\mathbf{W}^{(0)}, \mathbf{W}^{(1)}$ learnable weights. Node-level softmax outputs give the two class probabilities for each window.

B. Time-Series BiLSTM with Attention

A bidirectional LSTM takes in the windowed time series for each modality, which is given by

$$\mathbf{x}_m^{(i)} = [\, x_m^{(i)}(1), \ldots, x_m^{(i)}(L)\,] \in \mathbb{R}^L$$

. It then produces hidden states $\{\mathbf{h}_t\}_{t=1}^{L}$. An attention mechanism calculates

$$e_t = \mathbf{v}^\top \tanh(\mathbf{W}_h\mathbf{h}_t + \mathbf{b}),$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{u=1}^{L}\exp(e_u)}, \quad \mathbf{c}^{(i)} = \sum_{t=1}^{L}\alpha_t\mathbf{h}_t,$$

where $\mathbf{v}, \mathbf{W}_h, \mathbf{b}$ are parameters that can be learned and $\mathbf{c}^{(i)} \in \mathbb{R}^d$ is the summary of the window. We combine the summaries from all three modalities into $\mathbf{c}_{\text{all}}^{(i)} \in \mathbb{R}^{3d}$ and send them through a feed-forward network with softmax to guess whether they are neutral or not.

C. Interaction-Aware Interpersonal Graph Neural Network

The model constructs interaction-aware participant-level graph. We combine the window summaries for each participant p:

$$\mathbf{z}_p = \frac{1}{M_p}\sum_{i=1}^{M_p}\mathbf{c}_{\text{all}}^{(i)},$$

, where M_p is the number of windows that participant p has. We create a graph $G_P = (V_P, E_P)$ with nodes $V_P = \{\mathbf{z}_p\}_{p=1}^{P}$ and edges based on the 5-nearest-neighbor rule in the space of $\{\mathbf{z}_p\}$. The adjacency A_P is defined in the same way as the window-level GCN. A two-layer GCN learns embeddings for each participant.

$$\mathbf{U}^{(1)} = \sigma\big(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}\mathbf{Z}\mathbf{W}_P^{(0)}\big),$$

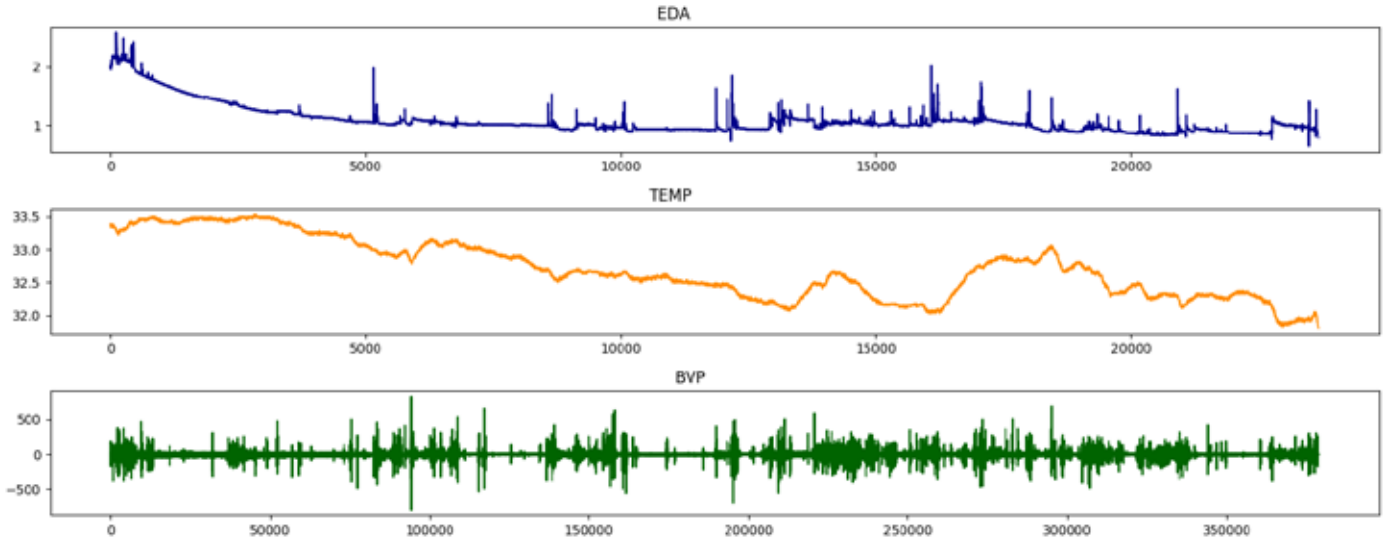


Figure 1: Raw biosignal recordings from Subject 17 during neutral state in the WESAD dataset: (top) electrodermal activity (EDA) exhibiting phasic stress responses; (middle) skin temperature (TEMP) trend over the session; (bottom) blood volume pulse (BVP) waveform illustrating heart-rate–related fluctuations.

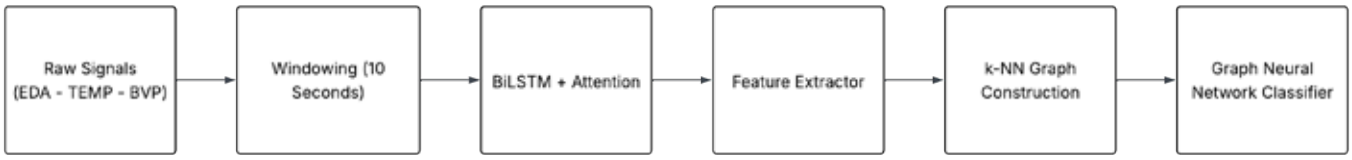


Figure 2: Overview of the proposed IA-IGNN pipeline. Raw wrist signals are segmented into 10-second windows and encoded via a BiLSTM with Attention. The resulting embeddings form a k-NN graph, which is processed by Graph Convolutional Network for baseline–stress classification.

$$\mathbf{U}^{(2)} = \mathrm{Softmax}\big(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}\mathbf{U}^{(1)}\mathbf{W}_P^{(1)}\big),$$

where $\mathbf{Z}$ stacks $\mathbf{z}_p$ giving each participant a binary class probability.

All of the models are implemented in PyTorch Geometric. We use the Adam optimizer with a learning rate of $\eta_0 = 10^{-3}$ and a weight decay of 10^{-5} and employ early stopping on validation loss with patience of 20 epochs.

IV. Results

We evaluated the proposed architecture on WESAD dataset using an NVIDIA A100 GPU hardware. Table I presents the overall performance. The interation aware GNN model achieves the best results with an accuracy of 0.73, an F1 score of 0.7, and an AUC score of 0.79. It improve multi-subject GCN (0.61) and BiLSTM+Attention (0.67) baselines.

The receiver-operating characteristic (ROC) curve (Figure 3(a) show that the proposed model can tell the difference between neutral and non-neutral states with a high level of confidence (mean under the curve (AUC) > 0.79). Figure 3(b shows that the interaction-aware model has balanced precision and recall across both classes. This means that it is strong enough to handle the combined "stress + amusement" label.

Model	Accuracy	F1-score	AUC
Single-subject GCN	0.85	0.8	1
Multi-subject GCN (sequential links)	0.61	0.49	0.66
Multi-subject GCN (kNN, $k = 5$)	0.61	0.47	0.66
BiLSTM + Attention	0.67	0.63	0.73
IA-IGNN	**0.73**	**0.7**	**0.79**

Table I: Comparison of test-set performance for each model.

Figure 3(a) ROC curve for the proposed model, which has an AUC of 0.79. This performance is much better than random classification (AUC = 0.5),that means that the model picks up on the main differences between neutral and stress-amusement states. Notably, the slope of the ROC remains steep up to a false-positive rate of approximately 0.2 corresponding to a true-positive rate above 0.6 before plateauing, suggesting that borderline windows with low-amplitude EDA phasic responses or subtle BVP fluctuations present the greatest challenge to the classifier. Figure 3(b)'s confusion matrix makes these limits even clearer. Out of 349 true neutral windows 266 are correctly identified with 76% recall and 83 incorrectly set off stress alarms with 24% false-alarm rate. On the other hand, 208 out of 304 stress-amusement windows are found (68% recall),

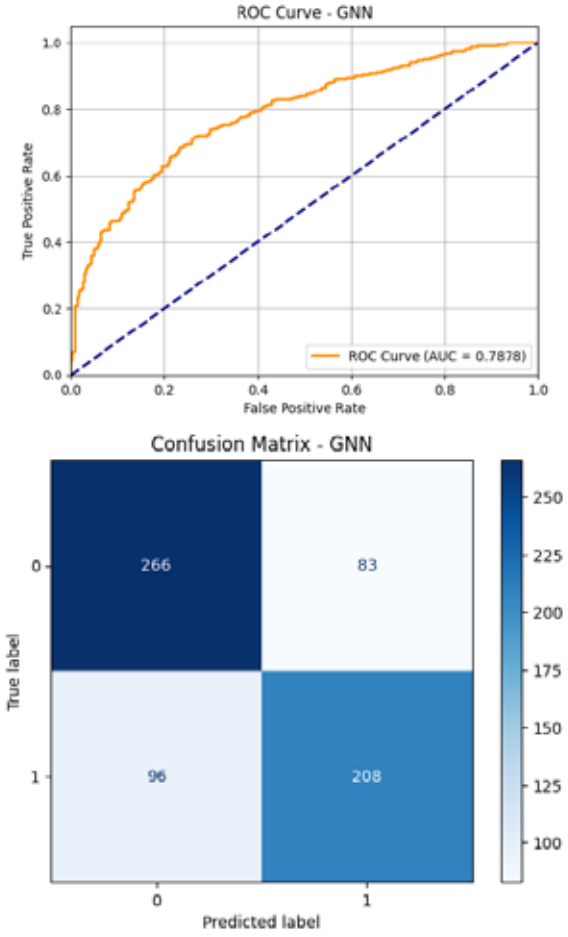


Figure 3: (a) ROC curve and (b) confusion matrix for the IA-IGNN model on held-out subjects

and 96 are missed (32% false-negative rate). The accuracy is about the same for both classes about 73% for neutral and 72% for stressed that neither class has a higher error rate. These findings point to two key avenues for enhancement. First, enhancing signal fidelity through adaptive EDA thresholding to isolate low-magnitude conductance peaks and wavelet-based denoising of BVP to maintain microvariations could refine feature representations. And adding dynamic edge weighting to the graph structure based on real-time signal quality indices or adding temporal attention mechanisms to the GNN' message passing may help the model better detect difference between windows.

V. Conclusion

We evaluated three different model architectures for detecting stress in wearable devices on the WESAD dataset. These were a statistical-feature GCN, a BiLSTM+Attention sequence model, and IA-IGNN that combines temporal embeddings into a graph for each participant. The interaction-aware model demonstrated superior performance, with an accuracy of 0.73, a macro-F1 of 0.7, and an AUC of 0.79 while preserving low inference latency ($\approx$ 45 ms per 10 s window) on resource-constrained hardware. Ablation studies showed that k-NN-based graph connectivity is much better than sequential linkage, and that attention-weighted temporal representations are better than fixed statistical descriptors. The results underscore effectiveness of integrating learned temporal embeddings with graph based reasoning. The model' robustness across unseen subjects is further proven by its quick convergence (stable within 60 epochs) and balanced performance for each class. These results show how important adaptive graph construction and attention-driven feature aggregation are for multimodal time-series classification.

VI. Acknowledgements

This research is supported by Istanbul Kultur University under ULEP-2025-2026.

References

[1] G. Dogan, F. P. Akbulut, C. Catal, and A. Mishra, "Stress detection using experience sampling: a systematic mapping study," *International Journal of Environmental Research and Public Health*, vol. 19, no. 9, p. 5693, 2022.

[2] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *2011 Computing in Cardiology*. IEEE, 2011, pp. 85–88.

[3] M. Memar and A. Mokaribolhassan, "Stress level classification using statistical analysis of skin conductance signal while driving," *SN Applied Sciences*, vol. 3, no. 1, p. 64, 2021.

[4] S. Kapoor and T. Kumar, "Fusing traditionally extracted features with deep learned features from the speech spectrogram for anger and stress detection using convolution neural network," *Multimedia Tools and Applications*, vol. 81, no. 21, pp. 31 107–31 128, 2022.

[5] C. O. Amadi, J. N. Odii, C. Okpalla, and O. C. I. La, "Emotion detection using a bidirectional long-short term memory (bilstm) neural network," *International Journal of Current Pharmaceutical Review and Research*, vol. 4, no. 11, pp. 1718–1732, 2023.

[6] G. Dogan and F. P. Akbulut, "Multi-modal fusion learning through biosignal, audio, and visual content for detection of mental stress," *Neural Computing and Applications*, vol. 35, no. 34, pp. 24 435–24 454, 2023.

[7] H.-G. Kim, D.-K. Jeong, and J.-Y. Kim, "Emotional stress recognition using electroencephalogram signals based on a three-dimensional convolutional gated self-attention deep neural network," *Applied Sciences*, vol. 12, no. 21, p. 11162, 2022.

[8] T. Wierciński, M. Rock, R. Zwierzycki, T. Zawadzka, and M. Zawadzki, "Emotion recognition from physiological channels using graph neural network," *Sensors*, vol. 22, no. 8, p. 2980, 2022.

[9] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The world wide web conference*, 2019, pp. 417–426.

[10] S. Chen, A. Wulamu, Q. Zou, H. Zheng, L. Wen, X. Guo, H. Chen, T. Zhang, and Y. Zhang, "Md-gnn: A mechanism-data-driven graph neural network for molecular properties prediction and new material discovery," *Journal of Molecular Graphics and Modelling*, vol. 123, p. 108506, 2023.

[11] S. A. Rahman, M. I. Khalil, H. Zhou, Y. Guo, Z. Ding, X. Gao, and D. Zhang, "Advancement in graph neural networks for eeg signal analysis and application: A review," *IEEE Access*, 2025.

[12] G. Dogan, F. P. Akbulut, and C. Catal, "Biosignals, facial expressions, and speech as measures of workplace stress: Workstress3d dataset," *Data in Brief*, vol. 54, p. 110303, 2024.

[13] Ç. Çöpürkaya, E. Meriç, E. B. Erik, B. Kocaçınar, F. P. Akbulut, and C. Catal, "Investigating the effects of stress on achievement: Biostress dataset," *Data in Brief*, vol. 49, p. 109297, 2023.

[14] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Evaluation of a Telemonitoring Interface for Neurodegenerative Disorders through Wearable Biosensors

Vael Kokach*, Büşra Kocaçınar*, Çilen Tavşanoğlu†, Uğur Ovacık‡, Feray Güngör†,
Kübra Nur Menengiç†, Fatma Patlar Akbulut§, Akhan Akbulut*, Ela Tarakcı†,
*Department of Computer Engineering, Istanbul Kültür University, Istanbul, Turkey
†Department of Physiotherapy and Rehabilitation, İstanbul University-Cerrahpaşa, Istanbul, Turkey
‡Vocational School of Health Services, Physiotherapy Program, İstanbul Aydın University, Istanbul, Turkey
§Department of Software Engineering, Istanbul Kültür University, Istanbul, Turkey

***Abstract*—Neurodegenerative disorders such as Parkinson's disease, Alzheimer's disease, multiple sclerosis (MS), and amyotrophic lateral sclerosis (ALS) often need long-term monitoring and personalized rehabilitation plans. This study is the first look at NÖRALSİS, a telemonitoring interface that uses wearable biosensors to help with remote assessment and rehabilitation. The system combines a web-based clinical platform with biosignal data acquisition using a wristworn device. The device captures multimodal signals such as electrodermal activity (EDA), photoplethysmography (PPG), skin temperature, and inertial movement. During this phase, the system's architecture was built and tested with a healthy volunteer to see how well the sensors worked together, how good the data was, and how well the interface worked. Furthermore, the usability of biosensor-derived data was explored in the context of clinical monitoring scenarios such as physical exercise, rest, and posture control. Early results show that the system can reliably pick up high-resolution biosignals and connect them to user activities in a clinically useful way. The results show that the proposed platform could be used as a base for future patient trials that involve remote patient monitoring and personalized rehabilitation strategies.**

***Keywords*—Telemonitoring, Neurodegenerative Disorders, Remote Rehabilitation, Wearable Biosensors,Postural Stability**

I. INTRODUCTION

Neurodegenerative disorders pose considerable challenges owing to their progressive nature, diverse symptoms, and the necessity for enduring, individualized care strategies [1], [2]. Traditional clinical assessments are often episodic and lack ecological validity, making it difficult to capture dynamic physiological and functional changes in their daily lives [3]. The incorporation of wearable biosensors into telehealth platforms presents a viable solution for remote, continuous, and objective patient monitoring, allowing clinicians to meticulously track symptoms such as gait impairment, autonomic dysfunction, and cardiovascular anomalies with elevated temporal resolution [4], [5].

This study is the first review of proposed system, a prototype telemonitoring interface designed for patients with neurodegenerative diseases. The system combines a wearable multimodal biosensor, with a web-based clinical dashboard to help with remote assessment and rehabilitation workflows. The current study focuses on validating the feasibility of the system using pilot data. In particular, we assess the quality of the signals to see how useful it is in simulated clinical situations like rest, exercise, and posture control status.This preliminary work sets the stage for future patient trials and the creation of smart decision-support mechanisms. Our purpose to create a scalable, patient-centered telehealth platform for the long-term management of neurodegenerative disorders by combining physiological monitoring with interactive clinical interfaces.

979-8-3315-5565-8/25/$31.00 © 2025 IEEE

II. METHODOLOGY

A. System Architecture and Biosensor Configuration

The proposed telemonitoring system, NÖRALSİS, is designed to facilitate continuous and remote assessment of individuals affected by neurodegenerative disorders which contain two major components: (1) a wearable biosensor platform for physiological signal acquisition, and (2) a web-based clinical interface for data visualization, filtering, and clinician interaction. We used the Empatica EmbracePlus device for biosignal acquisition because it is clinically accurate, has CE certification, and has been approved by the FDA for research use, so allowing continuous, non-invasive monitoring without disrupting daily activity. The device can simultaneously record EDA, PPG, Skin temperature, 3-axis accelerometer and gyroscope. The web-based clinical dashboard shows clinicians a timeline of physiological data streams and how well participant are wearing their devices. You can get data in two ways: as minute-resolution summaries or as high-resolution raw signals. Using a Cyberduck interface, raw data is safely pulled from AWS S3 buckets and processed locally with custom Python pipelines for more in-depth analysis.

B. Data Collection Protocol

We evaluated the proposed telemonitoring system's functionality with a healthy adult volunteer using a structured pre-

liminary data collection protocol. This protocol was construct to mimic common clinical situations in neurorehabilitation, such as physical exertion, passive rest, and postural challenge. We chose these conditions to test how good, reliable, and easy it is to understand biosignals in different physiological states. Before collecting data, the participant was asked to wear the device on the non-dominant wrist. The session conducted in a controlled indoor environment to reduce outside noise, with the temperature and lighting kept the same throughout. We paired the wearable device with the Empatica mobile app and verified for correct placement and signal acquisition. The protocol consisted of the following sequential phases:

- **Baseline Resting State** (5 minutes): The participant sat still and relaxed without talking or moving. During this time, baseline physiological signals were recorded. These provided reference values for autonomic and cardiovascular regulation when the body wasn't being stimulated.
- **Therapeutic Exercise Simulation** (10 minutes): To mimic rehabilitation exercises, we asked participants perform set of functional motor tasks as a standard physiotherapy routines. Movements included seated arm raises that focused on upper body activity, sit-to-stand transitions for activating the lower body, march-in-place much like walking, and lunge exercises which put their postural control to the test. With these tasks we aimed to elicit changes in EDA phasic responses, acceleration patterns, and BVP-derived heart rate changes while under physical load.
- **Post-Exercise Recovery** (5 minutes): After the active phase, the participant sat down to rest again. During this phase, data was collected that showed how the body was recovering, especially how heart rate variability (HRV) rebounded, body temperature returned to normal, and skin conductance level (SCL) decreased, which showed that parasympathetic tone had returned.
- **Postural Challenge Test** (3 minutes): The participant is asked to perform static lunge holds and weight-shifting movements while standing to assess their balance which is key for looking at both side-to-side (mediolateral) and front-to-back (anteroposterior) postural sway that cab be evaluate using accelerometer and gyroscope signals.

During the session, we continuously streamed biosignal data and upon completion then each phase was time-stamped and annotated for segment specific analysis.

C. Signal Preprocessing and Quality Assessment

A multi-stage signal preprocessing and quality assessment pipeline was set up to ensure that the biosignals collected by the wearable device were accurate. In this step, we aimed to reduce motion artifacts and physiological outliers especially given the ambulatory and multimodal nature of the data.

1) Formatting and Synchronizing Signals: Interpolation was used on signals when necessary to make sure that the sampling rates were the same across all channels. This made it possible to do multimodal fusion and phase-based analysis during different protocol segments, like rest and exercise.

2) Quality Assessment of EDA Signal: EDA signal were validated at using a rule-based artifact detection method primarly based on Kleckner et al. [6]. This framework gives a clear and adaptable way to score quality. We applied four main rules during the evaluation. First, we performed a physiological range check, marking any values outside 0.05-60 μS as invalid. Second, signal jumps sudden changes greater than 10 μS between adjacent samples were flagged as potential artifacts. Third, we identified flatlines, which are prolonged constant segments suggesting sensor disconnection or dry electrodes. Lastly, movement-related noise was detected by cross-referencing high-frequency EDA components with ACC data to spot any signal degradation caused by motion. Any segments that violated one or more of these criteria were then excluded from further SCR/SCL calculations.

3) BVP and HRV Signal Validation: BVP signal underwent low-pass filtering with a cutoff frequency of 5 Hz to mitigate high-frequency noise. We used a threshold-based method to find peaks and get inter-beat intervals (IBIs). From these, we calculated time-domain HRV features like RMSSD and SDNN. We applied physiologically plausible limits to identify and excluded artifactual IBIs like those resulting from missed or extra heartbeats.

4) Temperature and Accelerometer Signal Filtering: A moving average filter with a 10-second window was used to smooth out skin temperature signals and get rid of temporary sensor noise. We applied a 4th-order butterworth low-pass filter with a cutoff frequency of 2.5 Hz to effectively isolate postural and movement changes from noise in the accelerometer data. Subsequently, activity levels were quantified using derived metrics such as signal magnitude area (SMA) and derivative of acceleration.

5) Segment Labeling and Annotation: We labeled each time window in the data stream based on the activity phase, such as rest, exercise, or postural challenge. Signal quality scores were kept as metadata so that only valid segments could be analyzed. We made visualization plots to look at the shape of the signals and check the outputs of the preprocessing by hand, for example, EDA time series with quality flags and BVP with detected peaks (Figure 1).

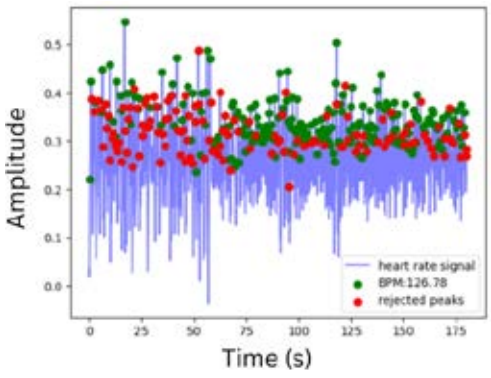


Figure 1: Rejected and accepted peaks for the BVP Signal during March-in-place

This strict method of preprocessing and quality control made sure that the analysis that followed was based on high-fidelity biosignal segments, which made it possible to accurately interpret the results in clinical monitoring and rehabilitation situations.

D. Feature Extraction and Scenario-Based Interpretation

After preprocessing and checking the quality of the signals, we took out domain-specific features from each biosignal to measure how the body responds to different clinical situations like autonomic nervous system, cardiovascular system, thermoregulation, and motor activity are working, since these are important in neurodegenerative diseases. Interpretations were then made based on structured use-case simulations, including exercise, rest, and postural control challenges.

1) Features of Electrodermal Activity (EDA): We used continuous decomposition analysis to break down both the tonic and phasic parts of the cleaned EDA signal. The SCL was the tonic background level and was used to look at overall trends in arousal and stress. We measured the Skin Conductance Response (SCR) component by looking at the number of SCR peaks per minute, the amplitude of SCR responses, the rise time, and the half-recovery time to understand how the sympathetic nervous system works during exercise like frequent and large SCRs can be consider increased arousal, while tapering SCL is sign of recovery (Figure 4).

2) Cardiovascular Features from BVP and HRV: From the BVP signal, peaks were detected to derive IBIs.Subsequently, time-domain HRV metrics were computed to characterize the dynamics of the autonomic nervous system's influence on cardiac activity [7]. Among these metrics were values like Mean Heart Rate (HR), the Standard Deviation of NN intervals (SDNN), and the Root Mean Square of Successive Differences (RMSSD). For example, a drop in overall HRV during physical exertion typically points to increased sympathetic nervous system activity, while a higher RMSSD value during resting phases indicates that the parasympathetic nervous system is in recovery [8]. These measures made it possible to track on stress, fatigue, and autonomic balance in real time.

3) Skin Temperature as a Means of Thermoregulation: Skin Temperature measurements were recorded continuously at a sampling frequency of 4 Hz during different activity phases, such as sitting down to rest, doing guided physical exercises, and resting without doing anything. We aimed to capture variations in peripheral skin temperature as indicators of sympathetic nervous system engagement and thermoregulatory response. We normalized data into segments based on activity windows and for each segment, we calculated statistical descriptors like mean, slope, and range to describe how temperature changes with activity and recovery. These features are intended to support modeling autonomic responses, and they can also be used as covariates when making personalized rehabilitation monitoring algorithms.

4) Movement and Postural Features from Accelerometer: The accelerometer sensor was configured to record movement in three orthogonal axes (X, Y, and Z), which stand for mediolateral, anteroposterior, and vertical directions, respectively which sampled data at 32 Hz. We extracted several metrics related to activity such as Signal Magnitude Area (SMA), step count and cadence, and crucially, postural sway trajectories in both mediolateral and anteroposterior directions. We determined these sway trajectories by projecting the acceleration vectors onto those specific planes, which allowed us to visualize them in plots similar to stabilograms. For tasks demanding static balance, such as single leg stands or lunges, we also employed moving average techniques to pinpoint the center of acceleration, and then used trajectory dispersion to quantify the sway magnitude.

5) Interpretation Based on Scenarios: We mapped all of the extracted features to three clinical scenarios that were simulated in the trial:

- Monitoring Physical Activity: High heart rate, SCR bursts, increased SMA, and temperature rise were signs of engagement and sympathetic response.
- Resting-State Assessment: A lower heart rate, a higher HRV, and a steady SCL showed parasympathetic rebound and autonomic restoration.
- Postural Control Challenge: Lateral and anterior sway vectors were visualized using accelerometer axes, showing stability trends and directional fall tendencies.

These scenario-based interpretations demonstrate the feasibility of linking raw biosensor data to clinically meaningful insights in real-time. Such interpretations can be directly integrated into decision support systems for tele-rehabilitation.

III. Results

A. Signal Quality Validation

All sensor streams were successfully recorded without major interruptions. The rule based evaluation of EDA signals showed that 92% of the segments were in the range between 0.05–60 µS. The remaining portions were flagged because of sudden discontinuities or flatlines. PPG signal showed strong peak detection correctly identifying 96 % of beats, which made it possible to reliably calculation of HRV indices. By cross-referencing accelerometer data, we were able to consistently find motion-related artifacts and leave them out of further analysis.

Table I: Time-Domain HRV Metrics

Segment	Mean-HR	RMSSD (ms)	SDNN (%)	pNN20
Rest	70.484	67.794	99.559	38.7091
Upper Body	94.743	53.273	96.002	29.090
Lower Body	125.284	61.304	81.262	25.252
March-in place	142.308	34.082	55.448	16.279
Lunge	158.611	22.652	25.054	2.923
Rest	**95.813**	**43.959**	**92.147**	**9.726**

B. Physiological Response Validation

Physiological responses corresponded effectively with anticipated patterns across various activity scenarios. During exercise heart rate increased by 125 % relative to baseline from 70.5 bpm to 158.6 bpm and HRV (RMSSD) decreased

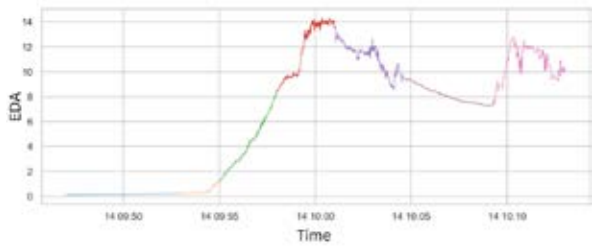


Figure 2: Electrodermal Activity across different segments

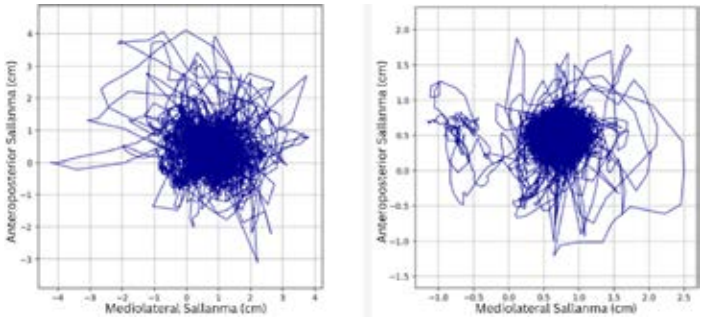


Figure 3: Postural Sway Comparison under (i) Unstable and (ii)Stable Conditions

by 67 % from 67.8 ms to 22.7 ms (Table Table I), consistent with sympathetic activation (Figure 2).

The HRV metrics indicate a clear physiological response to increasing exercise intensity. Mean-HR and sympathetic dominance reflected by lower RMSSD and pNN20 progressively increase during physical tasks, while overall variability SDNN drops. These trends show that the system works well for capturing autonomic modulation during exercise and recovery. At the same time EDA displayed frequent SCR peaks, while tonic SCL gradually rose. On the other hand, during rest periods, heart rate stabilized, SCR activity decreased, and SCL returned to its baseline level. Postural tasks exhibited quantifiable sway displacements (Figure 3), with mediolateral deviations surpassing anteroposterior ones, indicative of the biomechanical requirements of the lunge task.

C. Multimodal Consistency Validation

Cross-modal analysis exibited consistent trends between biosignal. SCR peaks occurred simultaneously with increases in HR during exercise sessions (r = 0.82). Skin temperature showed a delayed but monotonic increase that followed the similiar pattern as SCL (r = 0.76). Movement markers derived from the accelerometer were aligned with changes in cardiovascular and EDA responses. This showed that the system could put physiological signals into the context of behavioral states. Parasympathetic rebound during recovery was incomplete, with RMSSD recovering only 67% of baseline values despite HR normalization.

IV. Conclusion

This study presented the NÖRALSİS telemonitoring platform as an initial measure for the remote management of

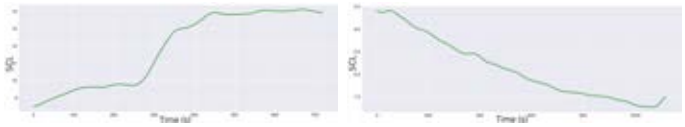


Figure 4: (i) March-in-place workout SCL in comparison to (ii)second resting segment

neurodegenerative disorders. The validation results showed that the data was very accurate. More than 92 % of the EDA segments and more than 96 % of the PPG-derived beats meeting quality criteria. Also, multimodal consistency analysis showed a link between SCR peaks and changes in heart rate (r = 0.82). The findings underline the feasibility of linking biosignal features to clinically meaningful state like sympathetic activation during exercise and stability recovery in rest. The study was conducted with a single healthy participant, limiting generalizability to patients with neurodegenerative disorders. Certain segments, such as static balance exercises, showed variable HRV responses, suggesting that further optimization of segment labeling and signal processing may be needed. Additionally, environmental factors and sensor placement could influence data quality, and the current analysis did not include long-term monitoring. Future studies should validate the platform with larger and more diverse patient populations. Despite these limitations the platform was successful in the tracking of autonomic and motor functions while establishing foundations for personalized rehabilitation through digital biomarkers.

V. Acknowledgements

This study is supported by the Health Institutes of Türkiye (TÜSEB) under grant no 33869.

References

[1] H. Oh, E. M. White, U. Muench, C. Santostefano, B. Thapa, C. Kosar, E. A. Gadbois, Z. T. Osakwe, P. Gozalo, and M. Rahman, "Advanced practice clinician care and end-of-life outcomes for community-and nursing home-dwelling medicare beneficiaries with dementia," *Alzheimer's & Dementia*, vol. 19, no. 9, pp. 3946–3964, 2023.

[2] A. J. Espay, P. Bonato, F. B. Nahab, W. Maetzler, J. M. Dean, J. Klucken, B. M. Eskofier, A. Merola, F. Horak, A. E. Lang *et al.*, "Technology in parkinson's disease: challenges and opportunities," *Movement Disorders*, vol. 31, no. 9, pp. 1272–1282, 2016.

[3] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, and L. Rochester, "Free-living monitoring of parkinson's disease: Lessons from the field," *Movement Disorders*, vol. 31, no. 9, pp. 1293–1313, 2016.

[4] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 1–12, 2009.

[5] O. M. Giggins, U. M. Persson, and B. Caulfield, "Biofeedback in rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 60, 2013.

[6] I. R. Kleckner, R. M. Jones, O. Wilder-Smith, J. B. Wormwood, M. Akcakaya, K. S. Quigley, C. Lord, and M. S. Goodwin, "Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1460–1467, 2017.

[7] F. P. Akbulut, B. Ikitimur, and A. Akan, "Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome," *Artificial Intelligence in Medicine*, vol. 104, p. 101824, 2020.

[8] F. P. Akbulut and A. Akan, "A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics," *Measurement*, vol. 128, pp. 237–246, 2018.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

IoT-Enabled Camera-Based Eye Tracking and Data Analysis for Real-Time

F. Aytac Durmaz[123], Irem Karagoz[13], Taha Cirpici[4], Saadet Kilic[1], Berkay Ucar[5], Gokcer Eskikurt[16], Ezgi Ildirim[16], Irem Kocer[16], Cenk Efe Bikmaz[16], Askin Mirachan Ergen[16]

[1] *Istinye University, Istanbul, Turkey*
[2] *Bogazici University, Istanbul, Turkey*
[3] *Pievision Technology AS, Istanbul, Turkey*
[4] *Türkisch-Deutsche Universität, Istanbul, Turkey*
[5] *Istanbul Beykent University, Istanbul, Turkey*
[6] *Istinye University Innovative Centre of Applied Neuroscience (IsuCAN), Istanbul, Turkey*

aytacd@gmail.com, iremmerik@gmail.com, tahacirpici3@gmail.com, kilicsaadet04@gmail.com, berkayucar60@gmail.com
geskikurt@istinye.edu.tr, ezgi.ildirim@istinye.edu.tr, kocar.irem513@gmail.com, efebahser@gmail.com, askinmirachan@gmail.com

***Abstract*— The high cost and logistical constraints of specialized eye-tracking hardware limit its widespread application in clinical research and diagnostics. To address these limitations, a low-cost, scalable eye-tracking system is engineered to operate with a standard webcam, providing an objective tool to understand attentional processes associated with psychological conditions such as eating disorders, anxiety, and OCD. The core innovation is a dynamic calibration method that continuously maps a user's eye features to on-screen coordinates as they follow a smoothly moving target. This process generates a dense, individualized dataset of several thousand points, a significant improvement over traditional 9- or 16-point static calibrations. This rich data enables a LightGBM machine learning model to perform robust gaze prediction. The system achieves high accuracy, validated through quantitative evaluation, with a low mean absolute error (MAE) that translates to a minimal angular error. Leveraging computer vision algorithms and this novel calibration approach transforms accessible technology into a powerful, objective tool for analyzing the attentional and cognitive processes underlying various psychological conditions.**

Keywords—eye tracking, attention deficit, emotional state, IoT, machine learning, webcam, dynamic calibration, mental health.

I. INTRODUCTION

Eye-tracking technology offers a powerful tool for objectively analyzing human attention and cognition, with wide-ranging applications in human-computer interaction (HCI), clinical psychology, usability testing, and educational research [1]. However, the widespread adoption of research-grade systems is frequently impeded by prohibitive costs and the requirement for specialized hardware, creating a significant barrier for many potential applications [2].

Our purpose is to overcome these limitations by developing an eye-tracking system engineered to be simultaneously accurate, sensitive, and highly accessible. This paper presents a system designed for advanced research purposes that is built entirely upon standard hardware (a webcam) and open-source software, ensuring ease of installation and implementation. It is architected for versatility, making it applicable to different analysis and measurement techniques, and includes provisions for cloud integration to support scalable, large-scale data collection and analysis.

The primary contribution of this work is a novel dynamic calibration method that forms the foundation of the system's accuracy. Unlike traditional static calibration, our approach continuously collects thousands of data points, creating a rich, personalized map of the user's gaze [3]. This dense dataset enables a machine learning model to perform robust gaze estimation. This paper details the system's architecture, the underlying mathematical principles of its components, and its empirical validation, demonstrating that an accessible, low-cost platform can achieve the precision required for scientific research.

II. SYSTEM AND METHODOLOGY

Our methodology is structured in four distinct stages, designed to convert raw pixel data from a webcam into meaningful psychological and attentional metrics.

A. System Architecture

The system's pipeline is implemented in Python, leveraging the OpenCV library for image processing tasks. The architecture is modular:

1. *Image Acquisition:* A standard webcam captures the user's facial video stream at a minimum of 100 frames per second (FPS).
2. *Feature Extraction:* The MediaPipe library is employed to detect a dense mesh of 468 facial landmarks in real-time [4]. A specific subset of these landmarks, corresponding to the eye regions, is isolated for further processing.
3. *Noise Reduction:* A Kalman filter is applied to the landmark coordinates to smooth the data, mitigating the jitter inherent in webcam video streams [5].
4. *Model Training & Gaze Prediction:* A pre-trained LightGBM model uses the processed landmark data to predict the (x, y) screen coordinates of the user's gaze [6].
5. *Data Processing and Storage:* The system is designed to be integrated with a cloud backend for large-scale data aggregation and analysis, enabling longitudinal studies and remote monitoring.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

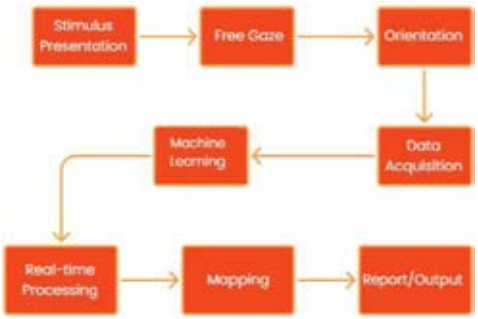


Fig.1 Flowchart of the working system.

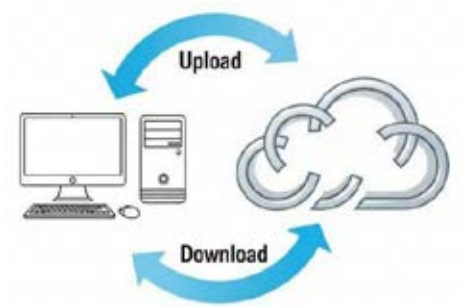


Fig.2 Communication with Cloud-based platform. Once the system completes the real-time data processing, it uploads the resulting visualizations to a cloud backend. This provides users with on-demand access to their stored session data from any location.

B. *Dynamic Calibration*

The cornerstone of our system's accuracy is its dynamic calibration process. Traditional systems ask the user to fixate on a few static points (e.g., 9 or 16). This is fast but leaves large areas of the screen unmapped, forcing the model to interpolate and often leading to significant errors.

Our approach requires the user to follow a target that moves smoothly and continuously across the screen in a Lissajous (curve-like) pattern for approximately 40 seconds. During this period, we collect thousands of data points. For each frame, we store a tuple containing the ground-truth screen coordinates of the target and the corresponding user's eye features. This rich, continuous dataset provides the machine learning model with a comprehensive map of the user's gaze across every region of the screen, drastically improving mapping accuracy and reliability.

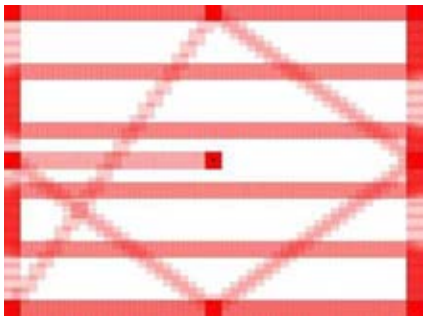


Fig.3 All calibration points captured together.

C. *Experimental Apparatus*

To ensure a stable data acquisition environment, a custom head support apparatus was designed based on user-centered ergonomic and biomechanical principles (Fig. 4). Fabricated using a 3D printer, the apparatus features a chin rest ergonomically contoured to the user's anatomy, which effectively keeps the head and chin immobilized throughout the trials. Furthermore, its custom camera mount positions the webcam on an extended arm at a fixed distance, ensuring an unobstructed line of sight to the screen and thereby fostering a more comfortable experimental experience. The webcam streams video data to the host computer via a standard USB connection for real-time processing.



Fig.4 Engineered to mitigate motion artifacts, this 3D-printed apparatus uses an ergonomic design to comfortably immobilize the user's head and maintain a consistent viewpoint for the camera.

D. *Model Training*

The data collected during calibration is used to train a LightGBM (Light Gradient Boosting Machine) model. LightGBM was chosen for its high efficiency, low memory usage, and excellent performance on tabular data.

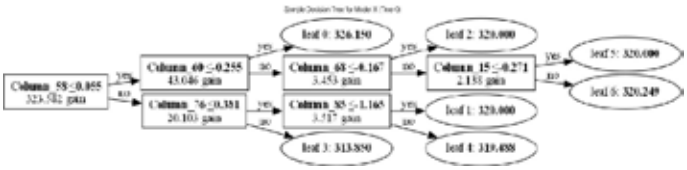


Fig.5 A visualization of a representative decision tree from the LightGBM model trained on our dataset. It illustrates the model's decision-making process, showing the series of feature-based splits at the nodes (branches) that lead to a final prediction value at a terminal leaf.

1. *Feature Engineering:* Before training, we extract key features from the facial landmarks. The most critical is the Eye Aspect Ratio (EAR), which quantifies the openness of the eye and helps disambiguate vertical gaze shifts from blinks or changes in head pose [7]. The EAR is calculated using the distances between vertical and horizontal eye landmarks.

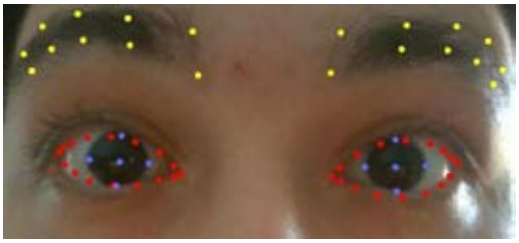


Fig.6 The selected landmarks from the eyelids, eyebrows, iris, and pupil, which are used to train the model and perform real-time tracking of the eye region.

The EAR is defined by the equation:

$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \times \|p_1 - p_4\|}$, where $p_1, \ldots, p_6$ are the 2D landmark locations.

2. *Noise Filtering:* To reduce signal noise from minor head movements and sensor imperfections, a Kalman Filter is applied. The filter predicts the next state of the landmark positions based on their current state and recursively corrects this prediction as new measurements arrive. The state update equation is conceptually:

   $\hat{x}_k = K_k z_k + (1 - K_k)\hat{x}_{k|k-1}$, where $\hat{x}_k$ is the estimated state, z_k is the measurement, and K_k is the Kalman gain.

3. *Training Process:* The model is trained on 80% of the calibration data, with the remaining 20% reserved for validation (Hold-Out Cross-Validation). The input features include the filtered coordinates of the eye landmarks and the EAR, while the target labels are the known (x, y) coordinates of the on-screen target.

E. Mapping and Visualization

Once the model is trained, it can predict a user's gaze coordinates in real-time. To interpret this raw data, we employ these three visualization techniques [8]:

- **Heatmaps:** These visuals aggregate gaze data over time, using a color gradient (e.g., red for high, blue for low) to show which areas of a stimulus received the most visual attention.

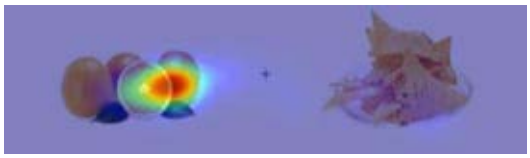


Fig.7 A heatmap showing the gaze distribution of a single participant during the experiment.

- **Scanpaths:** These connect sequential fixation points with lines and represent fixation durations by the size of the circles at each point. This reveals the temporal pattern of visual exploration.



Fig.8 A scanpath illustrating the sequence of fixations made by a single participant during the experiment.

- **Grid Template:** This method overlays the stimulus with a uniform 8x8 grid, where each cell displays the percentage of total fixation duration it received from the user.

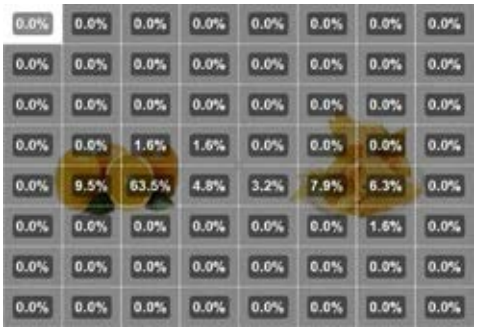


Fig.9 A grid-based template showing the percentage of gaze duration in each cell for a single participant during the experiment.

III. Experimental Evaluation

To evaluate the performance of our system, we conducted an experiment involving 30 participants. Following a dynamic calibration process, each participant was presented with 90 consecutives, validated visual stimuli related to food choice, specifically designed for use in eating disorder research. Each stimulus consisted of a pair of images representing either high-calorie, low-calorie, or neutral food items. The task aimed to measure and map participants' visual tracking patterns between the two simultaneously presented images.



Fig.10 A sample visual stimulus presented to participants during the experiment.

The error rate of the experiment was quantified in a two-stage process. First, to measure the performance of the machine learning model itself, we employed a hold-out cross-validation method. Our dataset, comprising over 5000 samples, was partitioned into an 80% training set and a 20% test set, with the latter being completely unseen by the model during training. Following the training phase, we calculated the model's mean prediction error in pixels on the test set for both the X and Y axes, using the aforementioned Mean Absolute Error (MAE) formula.

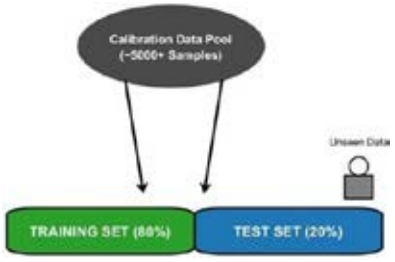


Fig.11 The data splitting procedure, dividing the dataset into training (80%) and test (20%) sets.

In the second stage of our error evaluation, participants were first asked to complete the calibration process. Subsequently, they were instructed to look at any point on the screen and click the mouse to register that exact location as the ground truth. At the moment of each click, the error was calculated as the difference between the X and Y coordinates of the mouse cursor's position and the coordinates concurrently predicted by

the machine learning model. The data aggregated from this repeated procedure are presented in the table below.

TABLE I. FINAL MEAN ERROR RATES

Error Measurements	Average Absolute Errors		
	x	*y*	*Euclid*
Model Error	10 to 15 pixels	5 to 10 pixels	14,6 pixels
Application Error	52.25 pixels	53.69 pixels	81.65 pixels

Fig.12 The table of Error Measurements x Average Absolute Errors.

Evaluation Metrics

Performance was quantified using established metrics:

1. *Mean Absolute Error (MAE):* This metric calculates the average absolute difference between the predicted gaze coordinates and the actual target coordinates in the test set. It provides a direct answer to the question: "On average, how many pixels off is the prediction?" It is calculated separately for each axis:

$$MAE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i is the actual coordinate and $\hat{y}_i$ is the predicted coordinate.

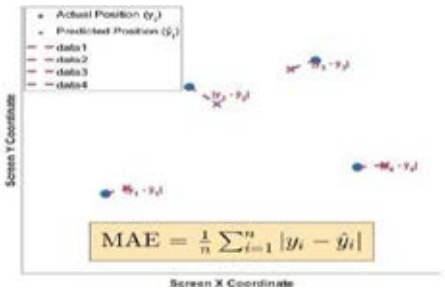


Fig.13 A visualization of the MAE computation process. The diagram shows how individual prediction residuals (the error between actual and predicted positions) are averaged to yield a single metric of model performance.

2. *Gaze Shift Error (Euclidean Distance):* To provide a single error value, we calculate the Euclidean distance between the predicted point (p_x, p_y) and the actual point (a_x, a_y).

$$Error_{pixel} = \sqrt{(p_x - a_x)^2 + (p_y - a_y)}^{\,2}$$

3. *Angular Error:* Pixel error is dependent on-screen resolution and viewing distance. A more robust metric is the angular error, which measures the visual angle between the predicted and actual gaze points. It is calculated as:

$$Error_{angular} = arctan\left(\frac{Error_{pixel} * P_{size}}{D}\right),$$

where P_{size} is the physical size of a pixel and D is the distance from the user's eye to the screen. An error of <1° is generally considered high accuracy for webcam-based systems [2].

Using this formula, the system's average pixel error corresponds to an angular error of 1.73°.

IV. CONCLUSION

This paper detailed the design and validation of a high-accuracy, low-cost eye-tracking system that replaces traditional static calibration with a novel dynamic technique. This approach creates a dense, personalized data map, allowing a LightGBM model to achieve significantly lower prediction error. To ensure affordability, we leveraged inexpensive 3D printing to fabricate a custom ergonomic head-support apparatus, which drastically reduced hardware costs while ensuring data quality. Our results demonstrate that this system achieves an accuracy that is highly competitive for webcam-based solutions, successfully transforming accessible hardware into a powerful scientific instrument for the objective analysis of human behavior. This work opens the door for its use in scalable research and real-world applications, particularly in budget-constrained environments. By addressing key barriers of cost and complexity, our approach can make empirical eye-tracking research more accessible. It empowers a broader community of scientists and developers to explore the complexities of human attention and cognition in diverse fields ranging from clinical psychology to usability testing.



Fig.14 The experimental setup shown during a live data collection session with a participant.

REFERENCES

[1] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Medical Imaging and Graphics,* vol. 65, pp. 176-190, Apr. 2018.

[2] K. Krafka *et al.,* "Eye tracking for everyone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.,* 2016, pp. 2176-2184.

[3] J. Liu, J. Chi, and Z. Yang, "A review on personal calibration issues for video-oculographic-based gaze tracking," *Frontiers in Psychology,* vol. 15, Art. no. 1309047, Feb. 2024.

[4] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," *arXiv preprint arXiv:1907.06724,* 2019.

[5] D. Sauter, B. J. Martin, N. Di Renzo, and C. Vomscheid, "Analysis of eye tracking movements using innovations generated by a Kalman filter," *Med. & Biol. Eng. & Comput.,* vol. 29, no. 1, pp. 63-69, Jan. 1991.

[6] G. Ke *et al.,* "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems,* vol. 30, 2017.

[7] T. Soukupova and J. Cech, "Eye blink detection using facial landmarks," in *Proc. 21st Comput. Vis. Winter Workshop (CVWW),* Rimske Toplice, Slovenia, Feb. 2016.

[8] G. Drusch, J. C. Bastien, and S. Paris, "Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest," in *Advances in science, technology, higher education and society in the conceptual age (STHESCA),* 2014, p. 25.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Investigation of the Effectiveness of Aspiration Technique in Simplified Artery with Clot Using Artificial Neural

Arda ATAÇ[1], Ezgi IŞIK[1]
[1]Yeditepe University, Faculty of Engineering, Department of Mechanical Engineering, Kayisdagi Cad., 34755, Istanbul, Turkey
Corresponding author's e-mail address: ardaatac1@gmail.com

Burak KALAYOĞLU[1,3], Ali Bahadır OLCAY[1,2]
[2]Padnos College of Engineering, Grand Valley State University, Grand Rapids, MI 49504, USA
[3]Daimon Mühendislik Müşavirlik A.Ş. ,Yücetepe Mahallesi İlk Sokak No:28/2 Çankaya/Ankara

Abstract*—In this study, the application of aspiration thrombectomy, which is the treatment for strokes occurring in the cerebral vessel, is simulated. As the study progressed, there was a gradual evolution from a simplified vessel geometry to a more realistic vessel geometry. The simulations are performed using Ansys Fluent at each stage. The thrombectomy application is simulated by adjusting specific parameters. Parameters such as vessel diameter, vessel length, intravascular blood flow rate, catheter vacuum pressure, and the viscosity of blood clots are varied. The aim is to determine how these changed parameters affected the thrombectomy application. Blood fluid physiology, mechanics, and parameters are observed during thrombectomy application in ANSYS Fluent. The effects of each parameter on thrombectomy application are calculated using MATLAB's Neural Network Tool and presented in this study. As a result, when all simulations were fitted to the Artificial Neural Network, it was observed that the changing parameters had, from largest to smallest, an effect on thrombectomy performance: vessel diameter, viscosity, vacuum pressure, blood velocity, and vessel length. When the Area Weighted Average (AWA) values of clot taken during thrombectomy were examined according to flow time, the parameters increased at zero point zero one five seconds of the simulations, respectively, for increasing vessel diameter (sixty-six point fifteen percent to ten point forty-three percent), viscosity (thirty point ninety-four percent to ten point twenty-five percent), and vacuum pressure (eleven point forty percent to twenty-three point twenty-five percent). The reason for the inverse ratio in vacuum pressure is that the percentage of clot AWA decreases as pressure increases. For increasing blood velocity (seventy-three-point forty-five percent to sixty-six-point fifteen percent) and for vessel length (thirty-one-point seventeen percent to twenty-seven-point fifteen percent).

Keywords: thrombectomy, stroke, aspiration catheter, vacuum pressure, computational fluid dynamics (CFD), artificial neural network (ANN), middle cerebral artery (MCA).

I. INTRODUCTION

Thrombosis occurs when blood clots within a vein and adheres to the inner surface of the vessel. If this clot detaches and travels through the bloodstream to another tissue or organ, causing a blockage, it is referred to as an embolism. This blockage can impede blood flow by partially or entirely closing the vessel. When a cerebral vessel is obstructed, it can lead to severe consequences such as vision loss, paralysis, memory loss, or even death. Embolism is more common among middle-aged and elderly individuals due to an increased propensity for blood clotting with age.

Thrombectomy is a clinical procedure that removes blood clots from blood vessels, primarily performed to treat thrombosis. A thrombectomy is a critical medical procedure that removes the blood clot that obstructs the vessel and restores normal blood flow. This procedure requires meticulous care and precision. Various suction systems are available for performing thrombectomy. In this study, the application of thrombectomy is examined under various scenarios with different boundary conditions in each simulation. The goal is to determine the impact of these changing boundary conditions on the effectiveness of thrombectomy.

Simulations are conducted using Ansys Fluent, with parameters such as catheter vacuum pressure, blood vessel length, vessel diameter, blood viscosity, and blood flow rate being altered. The effects of these varying conditions are subsequently analyzed using MATLAB's Neural Network Tool to assess their influence on thrombectomy application.

II. METHODOLOGY

In the first of the three successive stages, vessel geometry is simplified, with the catheter covering the entire vessel. The vessel diameter and length are set to 3 and 10 millimeters, respectively. Gravity was 9.81 meters per second squared. Blood and clot densities are taken as 1060 kilograms per cubic meter and 1300 kilograms per cubic meter, respectively. Blood viscosity is set to 0.0035 kilograms per meter times second, and clot viscosity is modelled as 0.035 kilograms per meter times seconds, which is ten times that of blood. Surface tension is 0.15 newton per meter. Intravascular pressure is set at 13 kilopascals, and vacuum pressures are taken as 30, 60, and 90 kilopascals in the present study.

In the second stage, a catheter is added to the simulation geometry, and the vessel geometry is modified. The catheter diameter and length are set to 1 millimeter and 2 millimeter respectively, and the vein length is increased to 20 mm. The

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

densities of blood and clot are assumed to be the same and are considered equal to 1060 kilograms per cubic meter. Clot viscosity is simulated at 1, 10, and 100 times that of blood. Vacuum pressure is kept at the same three values as in the first stage. Simulations are performed as multiphase flow in Ansys Fluent, with varying pressures at the vessel outlet and inlet.

In the third stage, further modifications are made to approach real-world applications as closely as possible. M1 segment modelling of the MCA is performed, with vessel lengths set to 13, 15, and 17 millimeters, and twice these lengths for developed intravascular flow.[10] Vessel diameters are set to 3, 4, and 5 millimeters,[11] with a tapering rate of 0.88 ± 0.15.[12] 5 French range was selected as the catheter from the French catheter scale, and its diameter is 1.667 millimeters. The surface density of blood is chosen as a constant value of 0.054 Newton per meter.[13]The viscosity value of blood is taken as a constant value of 0.0055 kilograms per meter times seconds.[14] The highest blood viscosity value is chosen to simulate the worst-case scenario. Clot viscosity is simulated at 100, 250, and 500 times that of blood. When catheter vacuum pressures are taken according to the suction systems used in literature, they are determined as 40 kilopascals, 60 kilopascals, and 80 kilopascals, negative gauge pressure values. Blood velocity is taken as 50 centimeters per second for male patients and 60 centimeters per second for female patients. The average intravascular pressure is accepted as 12.5 kilopascals. The length of the blood clot is the same for each simulation and is chosen as 8 millimeters. The vessel geometry of each stage described above is shown in Table 1.

TABLE I. SIMPLIFIED GEOMETRY OF EACH SIMULATED STAGE

Stages	Simplified Geometries
Stage 1	
Stage 2	
Stage 3	

If the patient has a clot, a susceptibility-vessel sign (SVS) can be visualized during magnetic resonance imaging (MRI) or magnetic resonance angiography (MRA) and can take different shapes. These are S-shape SVS, Y-shape SVS, and L-shape SVS.[6] The S-shape SVS is examined within the scope of this project. According to that, the clot is designed with a cell register and not a perfectly cylindrical shape, and there will be three different sizes of its geometry.

III. RESULTS

The effects of changing parameters and the accuracy of MATLAB NN-Tool have been clarified from nearly 1000 simulations run in Ansys Fluent. Of the 1000 simulations obtained, 15% were used for validation, 15% for test data, and the remaining 70% for training data. As seen in Graph 1, a high accuracy value of 98.395% was obtained in the test data after validation and training. Afterwards, as observed in Graph 2, the error approached zero, and the graph converged to zero error. As a result, MATLAB can make predictions with high accuracy.

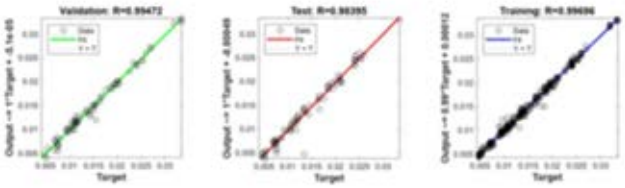


Fig. 1. Validation, test and training vs. target data

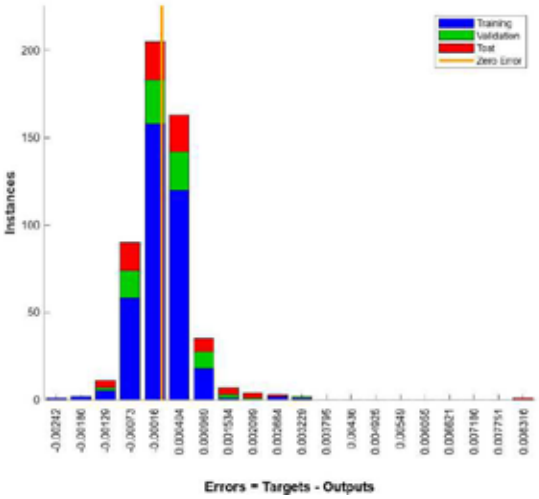


Fig. 2. Error histogram of neural network

When the data is simulated in ANSYS Fluent and used as input for the MATLAB Neural Network Tool (NN-Tool), the effect of each changing parameter is illustrated in Graph 3 below.

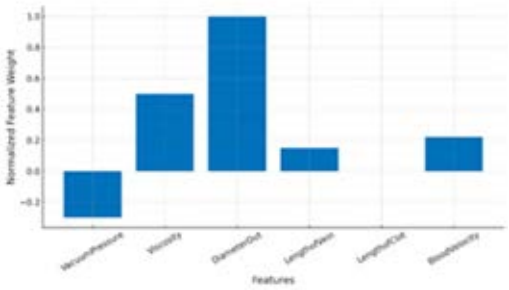


Fig. 3. Weight table of parameters affecting thrombectomy application

As seen in Graph 3, the viscosity of the clot, vessel exit diameter, vessel length, and intravascular blood flow rate are shown with positive values. Positive values indicate that as the values of these parameters increase, the time required to remove the clot from the vein also increases. These parameters are directly proportional to the time taken for clot removal. In contrast, catheter vacuum pressure represents a negative value, meaning that as the catheter vacuum pressure increases, the clot removal time decreases. Among all the parameters, ranked from most to least effective, they are vascular output diameter, clot viscosity, vacuum pressure, blood flow rate, and vascular length. Since the length of the clot, as shown in Graph 3, is consistent across all simulations, it has no impact and is therefore assigned a value of zero.

IV. LIMITATIONS AND FUTURE WORK

This study represents a preliminary computational exploration of aspiration thrombectomy performance and, as such, is based on several simplifying assumptions. The vascular geometries employed in the simulations were idealized, and blood properties were treated as uniform to reduce computational complexity and to isolate the effects of key parameters. While this approach provides essential baseline insights, it does not fully capture the heterogeneity of real vascular systems or the variability of blood rheology in patient populations. Similarly, the absence of patient-specific geometries limits the direct clinical translation of the present results. We acknowledge that validation against real clinical data is essential to confirm the applicability of the simulation outcomes. The integration of patient-specific vascular models and experimental or clinical datasets will be pursued in future work to strengthen model reliability. Additionally, while our artificial neural network provided a relative influence of the studied parameters, a structured sensitivity analysis would further clarify the robustness of the findings. Finally, the framework outlined here should be considered an early feasibility step; advancing toward clinical application will require systematic validation and alignment with treatment protocols, particularly to ensure consistency across diverse patient populations.

V. CONCLUSION

Neural networks (NN) and artificial intelligence (AI) have significant applications in various sectors, including diagnosis and treatment. Their remarkable accuracy in forecasting outcomes makes them intriguing tools for addressing complex medical problems. While NNs may be educated on previous doctor judgments, interdisciplinary methods are required for reliable results. For instance, in this study, integrating doctors' real-life experiences, patient data, knowledge of fluid mechanics, CFD programs, and NN tools was crucial. When the observed data approaches the truth, this study will shed light on complex operations, reduce the risk factor, and affect both the quality of the operation and human health. For this reason, this study demonstrates that CFD programs, knowledge of fluid mechanics for analysis, and NN tools will be essential and inevitable for doctors and patients in the future. In particular, the issue of clot removal requires rapid decision-making. A pre-prepared system can almost instantly process clot geometry and determine variables such as the necessary pressure and time with very high accuracy.

In this project, systemic simplifications are made to limit the Ansys simulation time. The next step for future work is to remove these simplifications and utilize real patient data and geometries. Adding more detailed and accurate data will improve the efficiency of NN outcomes. Patients with varying circumstances and profiles (e.g., age, cholesterol levels, blood sugar levels, general health, morale, speed of intervention, and degree of blockage) may help improve the model. As data becomes increasingly important, global data cooperation can significantly advance progress in the field. In the medium term, comparable models built globally using cloud computing may play a critical role in treating patients in various regions. In conclusion, it is clear that the three most important factors affecting thrombectomy are primarily vessel diameter, followed by clot viscosity and vacuum pressure. By evaluating these three critical parameters before surgery, it is aimed to increase the success rate of the procedure. These predictions aim to save the lives of many patients.

VI. ACKNOWLEDGMENT

This study was developed by integrating the Engineering Project works of Arda Ataç, Ezgi Işık, and Burak Kalayoğlu, conducted within the Department of Mechanical Engineering at Yeditepe University.

REFERENCES

[1] c. W. H. D. N. G. K. K. W. Jianhuang Wua, Guiying Liub, "Transient blood flow in elastic coronary arteries with varying degrees of stenosis and dilatations:cfd modelling and parametric study,"

[2] .

[3] E. S. S. M. L. G. W. A. B. W. M. D. G. M. G. K. M. P. M. M. G. L. M. W. Adrien Guenego1, Robert Fahed2 and J. J. Heit1*, "Impact of clot shape on successful m1 endovascular reperfusion," Frontiers in Neurology

[4] Alwatban, M. R., Aaron, S. E., Kaufman, C. S., Barnes, J. N., Brassard, P., Ward, J. L., ... & Billinger, S. A. (2021). Effects of age and sex on middle cerebral artery blood velocity and flow pulsatility index across the adult lifespan. Journal of Applied Physiology, 130(6), 1675 1683.

[5] Nader, E., Skinner, S., Romana, M., Fort, R., Lemonne, N., Guillot, N., ... & Connes, P. (2019). Blood rheology: key parameters, impact on blood flow, role in sickle cell disease and effects of exercise. Frontiers in physiology, 10, 1329.

[6] Reçi, V., & Bexheti, S. (2019). Variations of shape, length, branching, and end trunks of the M1 segment of the middle cerebral artery. Journal of Neurology, Neurological Science and Disorders, 5(1), 052-056.

[7] Gunnal, S. A., Farooqui, M. S., & Wabale, R. N. (2019). Study of the middle cerebral artery in the human cadaveric brain. Annals of Indian Academy of Neurology, 22(2), 187.

[8] Lauric, A., Greim-Kuczewski, K., Antonov, A., Dardik, G., Magida, J. K., Hippelheuser, J. E., ... & Malek, A. M. (2019). Proximal parent vessel tapering is associated with an aneurysm at the middle cerebral artery bifurcation. Neurosurgery, 84(5), 1082-1089.

[9] Yadav, S. S., Sikarwar, B. S., Ranjan, P., Janardhanan, R., & Goyal, A. (2020). Surface tension measurement of normal human blood samples by pendant drop method. Journal of Medical Engineering & Technology, 44(5), 227-236.

[10] Nader, E., Skinner, S., Romana, M., Fort, R., Lemonne, N., Guillot, N., ... & Connes, P. (2019). Blood rheology: key parameters, impact on blood flow, role in sickle cell disease and effects of exercise. Frontiers in physiology, 10, 1329.

[11] Kim, S., & Lee, J. Y. (2022). Comparison of vacuum pressures and suction forces generated by different pump systems for aspiration thrombectomy. Frontiers in Neurology, 13, 978584.

[12] Alwatban, M. R., Aaron, S. E., Kaufman, C. S., Barnes, J. N., Brassard, P., Ward, J. L., ... & Billinger, S. A. (2021). Effects of age and sex on middle cerebral artery blood velocity and flow pulsatility index across the adult lifespan. Journal of Applied Physiology, 130(6), 1675-1683.

[13] Krejza, J., Arkuszewski, M., Kasner, S. E., Weigele, J., Ustymowicz, A., Hurst, R. W., ... & Messe, S. R. (2006). Carotid artery diameter in men and women and the relation to body and neck size. Stroke, 37(4), 1103-1105.

[14] Rossi, R., Fitzgerald, S., Gil, S. M., Mereuta, O. M., Douglas, A., Pandit, A., ... & Doyle, K. M. (2021). Correlation between acute ischaemic stroke clot length before mechanical thrombectomy and extracted clot area: impact of thrombus size on number of passes for clot removal and final recanalization. European Stroke Journal, 6(3), 254-261.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

A Comprehensive Decision Support System for Family Physicians

Emre Şimşek, Fatma Ezgi Erkat, İlker Korkmaz, Ahmet Sermet Anagün
Faculty of Engineering
Izmir University of Economics
Izmir, Türkiye
{emre.simsek, ezgi.erkat}@std.ieu.edu.tr, {ilker.korkmaz, sermet.anagun}@ieu.edu.tr

***Abstract*—Family physicians practice in demanding clinical environments where quick yet reliable decision-making directly affects patient outcomes. This paper introduces the design and evaluation of a lightweight, offline-capable Clinical Decision Support System (CDSS) developed specifically for the realities of family medicine. Built with open-source tools such as Python, PyQt5, and SQLite, the system combines a transparent decision-tree engine, a simple graphical interface (GUI), and an administrative dashboard for monitoring key performance indicators (KPIs). The development process is described step by step, from architectural design and data handling to interface design and usability testing. Simulated patient cases suggest that the system improves diagnostic efficiency and consistency while maintaining user satisfaction.**

***Keywords*—clinical decision support system, family medicine, decision tree, key performance indicator.**

I. Introduction

The healthcare system increasingly depends on front-line practitioners making accurate decisions under the limited time and resource constraints [1]. Family physicians, who are often the first point of contact, face a particularly heavy cognitive workload: they should manage diverse symptoms, treat chronic conditions, and handle preventive care within very limited consultation windows [2]. This challenge is compounded by fragmented patient records and administrative responsibilities [3, 4].

In some cases, family physicians may be the only accessible healthcare providers. Possible errors in diagnosis processes can have serious consequences, yet an advanced diagnostic infrastructure is often unavailable [1]. Therefore, low-cost and adaptable decision support systems that are also capable of operating offline are needed [5].

Clinical Decision Support Systems (CDSSs) have shown potential to reduce diagnostic variability and improve efficiency [1, 5]. However, many systems fail to gain traction in primary care because they present overly complex interfaces, hide decision logic, or create a perception of over-reliance on automated results [3]. Physicians tend to prefer tools that support rather than replace clinical judgment, offering guidance while preserving autonomy [4]. The proposed CDSS presented in this paper was designed with this philosophy in mind. It is offline-capable, modular, and open-source. Decision logic is stored in human-readable JavaScript Object Notation (JSON) files, making it easy to update without programming knowledge [3]. The graphical user interface (GUI) is intentionally simple to guide physicians step by step through a decision tree; meanwhile the interactions are logged for later analysis [6, 7]. An integrated analytics module aggregates user data into key performance indicators (KPIs). Such features aim to provide the system to be used as a practical and scalable tool for family physicians.

The rest of the paper is organized in the following structure: Section II gives the related work about the initial times of CDSS research, Section III explains the design and implementation of the proposed CDSS, Section IV summarizes the preliminary test results based on the simulated patient data, and finally Section V concludes the paper with mentioning possible future work.

II. Related Work

The concept of computer-assisted diagnosis has evolved for more than five decades. Early systems such as MYCIN [8] and INTERNIST [9] demonstrated that structured medical knowledge could be represented computationally using rule-based reasoning. However, their static rule sets limited the broader use of these computer programs in adaptability-required real world scenarios [10].

In the 1990s and early 2000s, CDSS research shifted toward probabilistic models, enabling systems to better handle uncertainty [5]. Systems like Isabel Healthcare [11] expanded diagnostic coverage and were integrated with electronic health records (EHRs), but such integration was often provided with increased complexity and heavy IT support [1].

More recently, advances in machine learning and natural language processing have powered modern CDSSs platforms [1]. These systems can help professionals classify the diagnosis using pattern recognition technologies. Nevertheless, if such a system offers a "black-box" model on the decision paths, the system would lack transparency and so the physician's trust on the system would be decreased [3, 4].

Comparative studies suggest that structured approaches, such as decision trees and flowcharts, are more acceptable to clinicians because they clearly display the reasoning process [1, 4]. Visual representations of diagnostic pathways allow

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

physicians to validate system outputs that strengthen trust and adoption [6]. In addition, usability studies stress that a CDSS should minimize data entry and align with clinical workflows [7].

The proposed system draws from this tradition, combining decision-tree logic with the adaptability of transparent (i.e., human-readable) JSON-based structures, all implemented through accessible open source tools.

III. System Design and Implementation

The architecture of the proposed CDSS was deliberately designed to be modular, lightweight, and adaptable, ensuring that each component could evolve without disrupting the system as a whole. The solution integrates the following components: Decision Engine, GUI, and Administrative Dashboard, all connected to a centralized data layer.

A. System Architecture

The high-level architecture depicted in Fig. 1 illustrates the interaction between the Decision Engine, GUI and Data layers. The Decision Engine interprets JSON-encoded decision trees and navigates the physician step by step through a structured diagnostic process. The GUI serves as the primary interaction point, while the SQLite-based data layer manages patient records, logs, and complete Q&A trails.

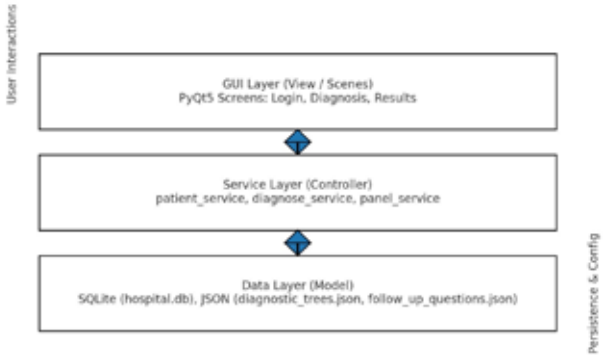


Fig. 1: The system architecture.

This modular separation ensures flexibility. For example, the decision tree model could be extended with probabilistic weights without altering the GUI, or the SQLite backend could be migrated to a client-server database if multi-user deployment is required.

B. Data Flow

The workflow begins with user authentication and patient record retrieval. Once a case is initiated, the Decision Engine starts from the root node of the chosen decision tree. Each question, corresponding responses, and transition are logged in real time, preserving the diagnostic pathway. Once a terminal diagnosis is reached, the result is stored in the database and made available within the administrative dashboard. The structure shown in Fig. 2 represents a sample decision tree for a migraine symptom encoded in the system.



Fig. 2: A sample decision tree for a migraine symptom.

This structured approach ensures that performance monitoring remains independent from patient care while still enabling a comprehensive set of analytical capabilities.

C. Decision Logic Implementation

Decision logic is structured as a binary tree and is encoded in JSON format, with each node representing a question, possible responses, and links to subsequent nodes or terminal outcomes. This approach balances human readability with computational efficiency, enabling clinicians or administrators without programming expertise to update diagnostic protocols.

The traversal of the tree is implemented in the following manner: the current node is displayed with its question, the physician's response is recorded, and the system transitions to the next node or outcome. All interactions along the path from the first node to the last inclusively are stored, creating an auditable log for both quality assurance and medical legal requirements.

D. Graphical User Interface

The GUI, implemented using PyQt5, was designed for clarity and ease of use. In deed, the essential functions of the system were focused to meet the basic requirements of the family physicians. The GUI comprises three primary screens: login, diagnosis, and results.

The login screen shown in Fig. 3 allows physicians authenticate themselves and select or create a patient record. This ensures that diagnostic sessions are linked to the appropriate user and patient.



Fig. 3: Login screen for a family physician.

The diagnosis screen shown in Fig. 4 displays one question at a time with the corresponding response options. Such a sequential presentation helps clinicians maintain focus and reduces cognitive burden, as recommended by prior usability research [6, 7].



Fig. 4: Diagnosis screen displaying active options.

The results screen shown in Fig. 5 provides the family physicians with the final diagnosis, ICD-10 codes, and a complete trace of the reasoning process. Physicians can review the diagnostic path on the decision tree before final confirmation, reinforcing trust and transparency.

E. Administrative Dashboard and KPI Analytics

The Administrative Dashboard aggregates usage data and transforms it into actionable insights. The KPIs like average diagnosis time, physician-specific performance scores, revisit rates, and diagnostic frequency distributions are automatically computed and visualized.

The histogram in Fig. 6 allows administrators to compare individual physician scores with clinic-wide averages with the aim of identifying both high performers and those requiring additional training.



Fig. 5: Results screen summarizing the session history.



Fig. 6: Doctor scores against clinic-wide average.

The heatmap in Fig. 7 visualizes performance variation across time and individual physicians, aiding in resource planning and quality improvement efforts. Beyond this, the heatmap can highlight seasonal workload effects (e.g., influenza season) on physician performance, making it easier to anticipate periods that may require additional staffing or resources. It also exposes instances where specific physicians show decreasing performance over time, enabling targeted support or training interventions. In this way, the dashboard serves not only as a retrospective reporting tool but also as a proactive management resource.

F. Database Schema

The SQLite database schema of the system includes two primary tables: patients and diagnoses. The patients table stores demographic and identifying information, while the diagnoses table records session-specific data, including the Q&A trail and final outcome. This separation ensures clinical data integrity while supporting performance monitoring.

IV. Results and Evaluation

Experiments were conducted using a combination of simulated patient cases and direct usability tests. The simulations covered both common and less frequent scenarios to assess the robustness of decision logic. A family physician involved in the development and evaluation phases of the system reported that the preliminary findings of the system has a potential to

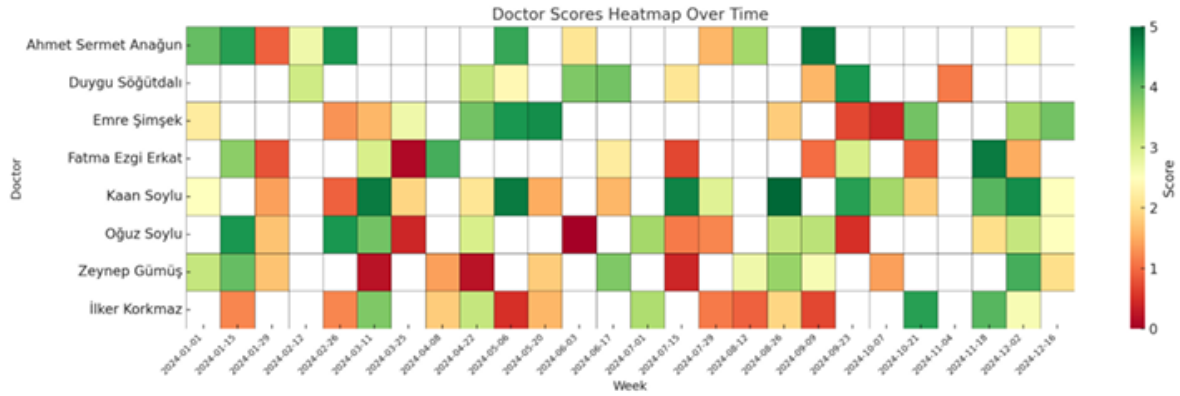


Fig. 7: Heatmap of weekly performance scores across doctors.

reduce cognitive workload and the design of the system has a prospect to increase clarity in diagnostic reasoning.

Quantitatively, the average diagnostic session was completed in less than 2.5 minutes, a result consistent with prior findings on workflow-friendly CDSSs [1, 2]. Qualitative feedback emphasized the transparency of decision trees and the usefulness of reviewing the reasoning process before finalizing an assessment [3, 4].

Nevertheless, these evaluations were conducted under controlled conditions. The simulation-based patient data setting and the number of physicians who tested the system are some of the main limitations for now. The definite efficiency of the system in busy clinical environments, where interruptions and patient variability are significant, remains to be validated.

V. Conclusion and Future Work

This study introduced and evaluated a modular CDSS proposed for the use of family physicians. By integrating JSON-based decision trees, a lightweight PyQt5 GUI, and KPI-oriented analytics, the proposed system addresses both clinical and administrative needs.

The system's transparency and adaptability distinguish it from many commercial platforms. Physicians can examine and update decision logic themselves, which may help overcome the trust barrier that often hinders CDSS adoption [3, 4]. The system's significant characteristics, (i.e., physician oriented sequential question design, clean interface, and offline capability) align with the realities of everyday practice. The preliminary results indicate that the system has a potential to reduce session times and cognitive load while standardizing assessments.

From an administrative perspective, the analytics module linked to the administrator dashboard supports data-driven performance management. By tracking selected KPIs such as consultation times, diagnosis frequencies, and physician-specific metrics, administrators can plan targeted interventions.

Still, important limitations remain. The system has not yet been tested in real-world clinics, where integration with existing EHRs and patient diversity could present challenges. Hence, future work should explore broader field validation across varied practice environments. Other possible future work would include (i) integration with national health information exchanges to reduce duplicate data entry, (ii) plugging the natural language processing technologies for free-text symptoms, and (iii) incorporating automated alerts for abnormal KPI trends.

In summary, the proposed CDSS shows promise as a practical and adaptable decision support tool for family medicine. The potential of the system, however, would be clear after extended deployment and evaluation in real clinical practice.

References

[1] R.T. Sutton, D. Pincock, D.C. Baumgart, D.C. Sadowski, R.N. Fedorak, and K.I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," npj Digital Medicine vol. 3, no. 17, 2020, doi: 10.1038/s41746-020-0221-y.

[2] H.J. Michtalik, H.C. Yeh, P.J. Pronovost, and D.J. Brotman, "Impact of attending physician workload on patient care: a survey of hospitalists," JAMA Internal Medicine, vol. 173(5), pp. 375-377, 2013, doi: 10.1001/jamainternmed.2013.1864.

[3] E.S. Berner, Clinical Decision Support Systems: Theory and Practice, 2nd edition, Springer, 2007, doi: 10.1007/978-0-387-38319-4.

[4] D. Zikos and N. DeLellis, "CDSS-RM: a clinical decision support system reference model," BMC Medical Research Methodology, vol. 18, no. 137, 2018, doi: 10.1186/s12874-018-0587-6.

[5] J.A. Osheroff, J.M. Teich, D. Levick, L. Saldana, F.T. Velasco, D.F. Sittig, K.M. Rogers, and R.A. Jenders, Improving Outcomes with Clinical Decision Support: An Implementer's Guide, 2nd edition, HIMSS, 2012, doi: 10.4324/9781498757461.

[6] B.T. Karsh, M.B. Weinger, P.A. Abbott, and R.L. Wears, "Health information technology: fallacies and sober realities," Journal of the American Medical Informatics Association, vol. 17(6), pp. 617–623, 2010, doi: 10.1136/jamia.2010.005637.

[7] D. Armijo, C. McDonnell, and K. Werner, "Electronic Health Record Usability: Evaluation and Use Case Framework," Agency for Healthcare Research and Quality, publication no. 09(10)-0091-1-EF, 2009.

[8] E.H. Shortliffe, "A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection," Proceedings of the 1974 Annual ACM Conference - vol. 2, p. 739, 1974, doi: 10.1145/1408800.1408906.

[9] H.E. Pople, "The Formation of Composite Hypotheses in Diagnostic Problem Solving: An Exercise in Synthetic Reasoning," Proceedings of the Fifth International Joint Conference on Artificial Intelligence (II), pp. 1030-1037, 1977.

[10] M.A. Musen, Y. Shahar, and E.H. Shortliffe, "Clinical Decision Support Systems," in Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 3rd edition, E.H. Shortliffe and J.J. Cimino, Eds., Springer, 2006, pp. 698–736, doi: 10.1007/0-387-36278-9_20.

[11] Isabel Healthcare Ltd., https://www.isabelhealthcare.com, (accessed on 11 July 2025).

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

An Extensible Cloud Platform for Integrating Heterogeneous IoT Devices in Physiological and Behavioral Research

1st F. Aytac Durmaz
Biomedical Engineering
Istinye University
Biomedical Engineering Ins.
Bogazici University
Pievision Technology AS
Istanbul, Turkey
aytacd@gmail.com
aytac.durmaz@istinye.edu.tr
aytac.durmaz@boun.edu.tr
aytac.durmaz@pievision.com
ORCID: 0000-0001-9070-7701

2nd Berat Mehmet Bakı
Electrical Electronics Engineering
Türkisch-Deutsche Universität
Istanbul, Turkey
berat_baki@ieee.org

Clinical diagnostics increasingly rely on data from multiple physiological sensors, but integrating these heterogeneous data streams for a single patient remains a significant technical challenge. In this paper, we present the architecture and implementation of a low-cost, modular, cloud-based software portal designed to centralize the recording of data from various sensor-based medical devices for a unified patient record. The platform leverages the MQTT protocol for reliable data transmission, a PHP/MySQL backend for data management, and a Python-based service for data logging. A key feature is the Role-Based Access Control (RBAC) system, ensuring data privacy for different user types such as researchers and patients. To demonstrate its versatility, the platform was successfully validated by integrating three distinct proof-of-concept (PoC) devices developed in our lab: a cardiovascular stress monitor (GSR/HR), an oculomotor behavior analysis system (eye-tracker) and a psychomotor assessment tool (finger tapping). Our results show that the proposed infrastructure provides a flexible and accessible solution for creating comprehensive, multi-modal data sets for diagnostic purposes.

Keywords—IoT, Physiological Monitoring, e-Health, Cloud Platform, Eye-Tracking, Stress Analysis, Data Acquisition, MQTT, RBAC Physiological Monitoring

I. INTRODUCTION

The collection of physiological and behavioral data is fundamental to advancing our understanding in fields ranging from major subfield of psychology to human-computer interaction. However, researchers are often constrained by commercial data acquisition systems that are expensive, proprietary, and inflexible. These closed-box systems typically do not allow for the integration of custom or heterogeneous sensors, forcing researchers to work within silos and limiting the scope of multi-modal studies. Furthermore, the technical overhead of setting up and managing the data infrastructure for such studies can be substantial, diverting valuable time and resources away from the primary research questions. [1], [2]

Recent advancements in the Internet of Things (IoT) and low-cost sensor technologies present an opportunity to overcome these barriers. While several platforms have been proposed for IoT health monitoring, there remains a need for a truly open and extensible architecture that can serve as a common backbone for diverse research projects within a single institution.

Our primary contribution is the design, implementation, and validation of such a platform. We present an IoT-based infrastructure that is both low-cost and highly modular, capable of acquiring, storing, and visualizing data from a heterogeneous set of data acquisition nodes. This paper details the platform's multi-layered architecture and demonstrates its practical utility through three distinct case studies developed in our laboratory. Section II describes the system architecture. Section III presents the three proof-of-concept implementations. Section IV discusses the results and limitations, and Section V provides the conclusion and outlines future work.

II. SYSTEM ARCHITECTURE

A. Overview

The platform follows a layered design to preserve modularity and separation of concerns: acquisition nodes → MQTT broker → Python ingestion service → MySQL database ↔ PHP web portal.

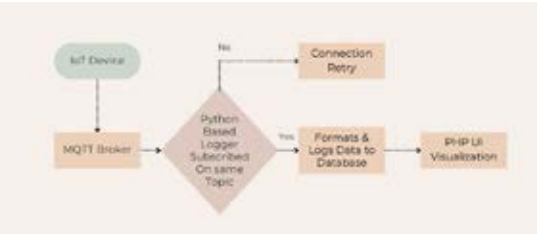


Fig. 1 summarizes the end-to-end flow from device to visualization.

B. Data Acquisition Layer

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Acquisition nodes include (1) microcontroller-based wearables (e.g., ESP32) producing continuous or event data, and (2) computer-based systems (e.g., eye-tracker software) streaming higher-rate, multidimensional signals. Nodes are hardware-agnostic; the sole requirement is publishing well-formed messages to MQTT topics.

Fig. 2. Suggested topic scheme:



C. Communication Layer (MQTT)

We employ MQTT due to its lightweight publish–subscribe semantics, tolerance to intermittent connectivity, and widespread device support. Devices publish to project-scoped topics; the ingestion service subscribes with appropriate QoS and authentication. This decoupling allows devices and backend to evolve independently. [14]

D. Cloud Backend Layer

Ingestion Service (Python): A persistent daemon subscribes to project topics, parses JSON payloads, validates required fields, and writes to the database. It also logs operational metrics (throughput, drop rate, end-to-end latency) for observability.

Database (MySQL): The schema uses four core tables—users, devices, recordings, measurements—with foreign-key constraints. We extend devices with device_type; recordings store protocol metadata (start/end timestamps, notes); measurements persist high-frequency samples as typed rows (stream, ts, data_json). This flexible design admits time-series, event-based, and multidimensional data.

Web Portal (PHP): The portal implements (i) device registration and ownership; (ii) RBAC enforcing per-role visibility; (iii) an experiment dashboard to start/stop recordings and attach contextual annotations; (iv) live charts and CSV download for offline analysis. Role set: admin (system-wide), researcher (authorized multi-participant access), clinician (assigned patient cohort), participant (self-data only). Passwords are hashed (e.g., Argon2) and sessions use secure cookies.

Fig. 3. Device management page – register device by name and MAC; list with status.

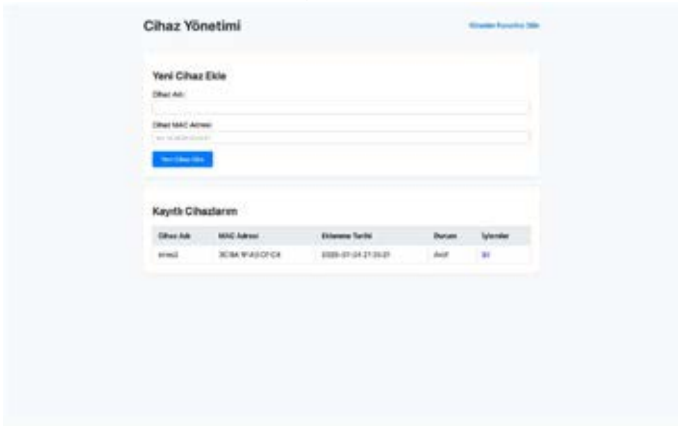
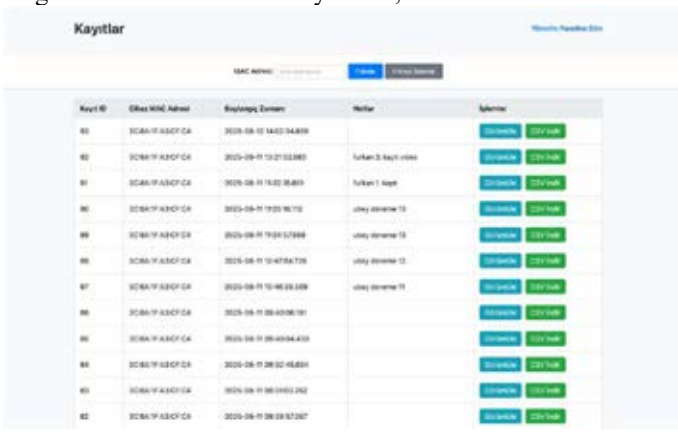
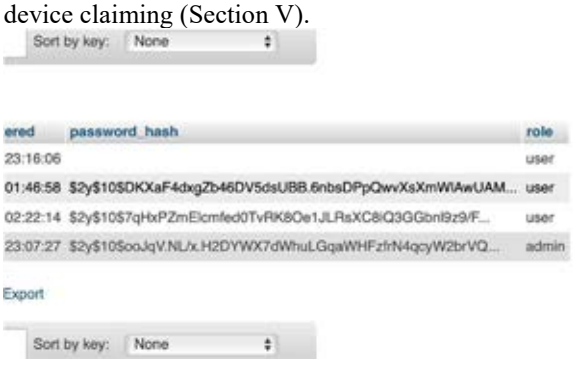


Fig. 4. Records list – filter by MAC; view or download CSV.



E. Security and Privacy

Traffic is protected via TLS; user access follows least privilege via RBAC. Participants are represented by pseudonymous identifiers; audit logs capture access. Future enhancements include mutual-TLS for device authentication and automated device claiming (Section V).



III. STUDIES: PROOF-OF-CONCEPT IMPLEMENTATIONS

A. Cardiovascular Stress Monitoring (GSR/HR)

GSR relies on the fact that human skin contains numerous sweat glands, particularly in areas such as the palms and fingertips, which are innervated by the sympathetic branch of the autonomic nervous system. When a person experiences emotional arousal or stress, sympathetic activation stimulates

these sweat glands to secrete moisture, even at levels too subtle to be visibly detected. [22]

Objective: infer stress responses using galvanic skin response and photoplethysmography-derived heart rate/variability. Data: continuous time-series (e.g., GSR, IR, BPM). Integration: the proto-device publishes per-sample payloads to MQTT; the ingestion service writes samples to measurements; the dashboard provides live plots and session-level notes (e.g., "Stroop start", "Rest"). Outcome: researchers visualize artifacts in real time and export sessions for HRV analysis. [4], [9], [19], [20]

Fig. 5. Record detail view – synchronized GSR and BPM time series (sample session).



B. Eye-Tracking Proof-of-Concept (Dynamic Calibration Approach)

Eye movements reveal aspects of problem solving, reasoning, mental imagery, and search strategies, behavioral studies often use objective methods to examine these relationships during visual tasks by using eye-tracking [23].

This PoC implements a webcam-based, software-only eye-tracking approach with dynamic calibration. During a ~40-second calibration, thousands of (feature, target) pairs are collected while the user tracks a smoothly moving target; each frame is labeled with the target's on-screen (x, y) coordinates and a compact eye feature vector. A LightGBM model is trained per user to map features to gaze coordinates, and a Kalman filter stabilizes predictions. Compared to classical 9/16-point static calibration, the dense, person-specific map reduces error and is more tolerant to head drift. The pipeline runs with Python/OpenCV/MediaPipe and streams outputs to the cloud backend for storage and visualization.

C. Psychomotor Skills Analysis (Finger-tapping)

The finger-tapping test (FTT) provides a simple yet sensitive measure of motor speed and coordination, which are closely linked to both cognitive performance and overall neurological health. Because it can detect subtle motor slowing or asymmetry, the FTT offers an objective, quantifiable metric for tracking changes in motor function over time or in response to interventions. [21]

We collect per-keystroke timestamps to compute simple speed/regularity summaries; detailed clinical analysis is outside the scope of this platform paper.

DISCUSSION

Storing GSR, eye-tracking, and finger-tapping data in a unified cloud-based platform enables seamless integration and large-scale analysis of multimodal physiological and behavioral measures. By combining indicators of GSR, eye tracking, and finger tapping, researchers can develop a more comprehensive understanding of human behavior across sensory, cognitive, and motor domains.

Benefits: (i) extensibility—new devices integrate by topic and payload conventions; (ii) centralization—single storage and portal reduce friction; (iii) collaboration—role-aware remote access for multi-site studies; (iv) low cost—commodity IoT hardware and open-source stack. Limitations: current device onboarding is manual (MAC-based), and high-rate streams rely on careful broker and DB tuning. Threats to validity include small-scale lab deployment and limited diversity of devices so far.

Evaluation plan: measure end-to-end latency (publish to plot), throughput (messages/s), data loss, and setup effort per PoC; report per-role access correctness via test scenarios. (Insert results once experiments are consolidated.)

IV. CONCLUSION AND FUTURE WORK

We introduced a role-based cloud platform unifying heterogeneous IoT physiological/behavioral data and validated it with three distinct PoCs (Stress Measurement, Eye Tracking, Fingertapping). The unified schema and messaging allowed plug-and-play integration without backend changes. Next steps include (i) device claiming for non-technical onboarding (e.g., QR-based claim, pending→verified workflow), and (ii) embedded machine-learning services for on-portal inference (e.g., real-time stress classification), alongside broader multi-site deployment.

V. REFERENCES

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," European Heart Journal, vol. 17, no. 3, pp. 354–381, 1996.

[2] W. Boucsein, Electrodermal Activity, 2nd ed. New York, NY, USA: Springer, 2012.

[3] W. Boucsein, D. C. Fowles, S. Grimnes, G. Ben-Shakhar, W. T. Roth, M. E. Dawson, and D. L. Filion, "Publication recommendations for electrodermal measurements," Psychophysiology, vol. 49, no. 8, pp. 1017–1034, 2012, doi: 10.1111/j.1469-8986.2012.01384.x.

[4] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156–166, 2005, doi: 10.1109/TITS.2005.848368.

[5] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in Proc. 2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction (ACII), 2013, doi: 10.1109/ACII.2013.117.

[6] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," IEEE Transactions on Biomedical Engineering, vol. 63, no. 4, pp. 797–804, 2016, doi: 10.1109/TBME.2015.2474131.

[7] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," Current Cardiology Reviews, vol. 8, no. 1, pp. 14–25, 2012, doi: 10.2174/157340312801215782.

[8] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," Bioengineering, vol. 3, no. 4, p. 21, 2016.

[9] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in Proc. 20th ACM International Conference on Multimodal Interaction (ICMI), 2018, pp. 400–408, doi: 10.1145/3242969.3242985.

[10] D. Kahneman and J. Beatty, "Pupil diameter and load on memory," Science, vol. 154, no. 3756, pp. 1583–1585, 1966, doi: 10.1126/science.154.3756.1583.

[11] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources," Psychological Bulletin, vol. 91, no. 2, pp. 276–292, 1982.

[14] OASIS, "MQTT Version 5.0 – OASIS Standard," Mar. 2019. [Online]. Available: https://www.oasis-open.org/standard/mqtt-v5-0-os/

[15] OWASP, "Password Storage Cheat Sheet," 2025. [Online]. Available: https://cheatsheetseries.owasp.org/cheatsheets/Password_Storage_Cheat_Sheet.html

[16] Internet Research Task Force (IRTF), "Argon2 Memory-Hard Function for Password Hashing and Proof-of-Work Applications," RFC 9106, Sep. 2021. [Online]. Available: https://www.rfc-editor.org/info/rfc9106

[17] PHP Manual, "password_hash()" and "password_verify()," [Online]. Available: https://www.php.net/manual/en/function.password-hash.php and https://www.php.net/manual/en/function.password-verify.php

[18] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, "Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support," Journal of Biomedical Informatics, vol. 42, no. 2, pp. 377–381, 2009, doi: 10.1016/j.jbi.2008.08.010.

[19] N. Milstein, A. Stout, S. J. R. Masters, and A. Kringelbach, "Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in interactive dyadic states," Frontiers in Behavioral Neuroscience, vol. 14, 2020, Art. 148. doi: 10.3389/fnbeh.2020.00148.

[20] Analog Devices/Maxim Integrated, "MAX30102: High-Sensitivity Pulse Oximeter and Heart-Rate Sensor for Wearables," Datasheet, 2019. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/max30102.pdf

[A] S. Abdulmalek, A. Nasir, W. A. Jabbar, M. A. M. Almuhaya, A. K. Bairagi, M. A.-M. Khan, and S.-H. Kee, "IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review," Healthcare, vol. 10, no. 10, p. 1993, 2022. doi:10.3390/healthcare10101993

[B] H. Kurdi and V. Thayananthan, "A Multi-Tier MQTT Architecture with Multiple Brokers Based on Fog Computing for Securing Industrial IoT," Applied Sciences, vol. 12, no. 14, 7173, 2022. doi:10.3390/app12147173

[21] D. Civitello, et al., "How do physiological responses such as respiratory frequency, heart rate, and galvanic skin response (GSR) change under emotional stress?," 2014.

[22] D. Austin, J. McNames, K. Klein, H. Jimison, and M. Pavel, "A statistical characterization of the finger tapping test: Modeling, estimation, and applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 501–507, 2014, doi: 10.1109/JBHI.2014.2313115.

[23] M. L. Mele and S. Federici, "Gaze and eye-tracking solutions for psychological research," *Cognitive Processing*, vol. 13, suppl. 1, pp. 261–265, 2012, doi: 10.1007/s10339-012-0499-z.

Hybrid CNN–LSTM Framework for Automated Detection of Pediatric Heart Murmurs

Asmaa Abduh AL-taib
Dept. Artificial Intelligence and Intelligent Systems
Gümüşhane University,GU
Gümüşhane , Turkey
asmaabdu221@gmail.com

Hayati Türe
Dept.Software Engineering
Trabzon University
Trabzon , Turkey
hayatiture@trabzon.edu.tr

Muhammed Emin Aslan
Dept. Software Engineering
Gümüşhane University,GU
Gümüşhane, Türkiye
measlan.tr@gmail.com

***Abstract*—This study presents a comprehensive deep learning framework for automated heart murmur detection from phonocardiogram (PCG) signals to address the subjective limitations of cardiac auscultation. Using the CirCor DigiScope dataset of 5,272 pediatric heart sound recordings from 1,568 subjects, we systematically evaluated multiple approaches including traditional machine learning models (Logistic Regression, SVM, KNN, Decision Tree, AdaBoost), basic deep learning architectures (standalone CNN and LSTM), and advanced transfer learning models (DenseNet121, EfficientNet B0/B1/B2, ResNet 50/101, MobileNet V2/V3, GhostNet, AttentionCNN). Our hybrid CNN-LSTM architecture combines mel-spectrogram spatial feature extraction with temporal sequence modeling using 128 traditional audio features and comprehensive preprocessing with data augmentation. The CNN-LSTM model achieved superior performance with 84.6% accuracy, 0.84 F1-score, 0.748 precision, and 0.800 recall, significantly outperforming traditional methods (best SVM: 0.716 F1-score) and advanced architectures (best DenseNet121: 0.72 F1-score), demonstrating robust classification with 0.879 AUC. Results indicate that hybrid deep learning architectures effectively automate cardiac auscultation, providing objective diagnostic tools with potential for improved healthcare accessibility and clinical decision-making.**

Keywords—Convolutional Nerural Networks; Deep Learning; Mel Spectrogram; Artificial intelligence; Recurrent Neural Networks; Heart Murmur Detection;Cardiac Auscultation;Long Short-Term Memory

I. *INTRODUCTION*

Cardiovascular diseases (CVDs) are a leading global cause of morbidity and mortality, necessitating early, accurate diagnosis [1]. Cardiac auscultation is a cost-effective screening tool [2], but its reliance on clinician expertise introduces variability, especially where specialists are scarce [3]. Digital phonocardiography (PCG) and AI—particularly ML and DL—enable objective analysis, automating heart sound interpretation[4]. Most AI systems focus on binary murmur detection, but reliable multi-class classification is still needed.

Early automated analysis segmented heart sounds (S1, S2) and extracted features (statistical moments, spectral centroids, energy ratios) for SVM, KNN, and Random Forest classifiers, achieving moderate accuracy but limited in capturing complex PCG patterns [5]. DL models, especially CNNs, learn hierarchical features from raw signals [6], while LSTMs capture temporal dependencies [7]. Hybrid CNN-RNN models combine spatial and temporal learning, enhanced by data augmentation (time stretching, pitch shifting, noise injection) to improve generalization [8].

Multi-class classification frameworks further characterize murmurs (normal, systolic, diastolic) and valvular pathologies, using multi-branch CNNs, RNN-based hidden semi-Markov models, and hierarchical CNNs, achieving high sensitivity and specificity (e.g., 76.3% and 91.4%) [9]. However, clinically graded, fine-grained murmur classification remains limited.

This study presents a hybrid CNN-LSTM framework for automated murmur detection. PCG signals are converted to Mel spectrograms, with the model learning spatial and temporal patterns. Data augmentation improves robustness, demonstrating potential for accurate and interpretable automated cardiac auscultation.

II. THE PCG DATASET

This study utilized the publicly available CirCor DigiScope Phonocardiogram Dataset, extensively annotated and published on PhysioNet by Oliveira et al. (2022) for the PhysioNet/CinC Challenge 2022, serving as a benchmark for automated cardiac auscultation research .The CirCor DigiScope Dataset contains 5,272 PCG recordings from 1,568 pediatric subjects ($0-21$ years) , lasting $4.8-80.4$ s (mean 22.9 s), totaling 33.5 hours. Four auscultation sites capture the heart profile. Murmurs were

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

annotated for timing, shape, pitch, grade (Levine), quality, and location; S1 and S2 sounds were labeled via a semi-supervised approach with three algorithms and expert verification [10].

III. METHODS

This study aims to detect the presence or absence of cardiac murmurs, a task normally requiring expert cardiologists in controlled settings. Using advanced AI, 2D CNNs classify PCG recordings as murmur present or absent, learning hierarchical representations to capture subtle acoustic variations. The methodology covers dataset, preprocessing, model, and evaluation metrics. Fig. 1 shows Mel-scaled spectrograms of normal and murmur signals.

A. Preprocessing

The study analyzed PCG recordings from 942 patients, each with up to four .wav files (aortic, mitral, pulmonary, tricuspid; 5–30 s). Only patients with all four recordings were retained, yielding 574 patients (117 with murmurs, 457 without). All waveforms were standardized to 5 s, with data augmentation applied to balance murmur-positive samples. Long murmur recordings (>15 s) were split into three 5-s segments. Samples from the same patient were kept in the same subset to prevent data leakage. The final dataset included 510 training samples (47% positive), 124 validation (27% positive), and 124 testing samples (24% positive).

Feature Extraction

This study extracted 128 features from PCG signals across time, frequency, cepstrum, and statistical domains for all four auscultation sites to distinguish normal from murmur signals. Key features included the Amplitude Envelope (maximum frame amplitude, mean, median, standard deviation, 75th percentile) [11], Total Energy (overall loudness, computed by summing squares)(1),

$$E_x = \sum_n |x[n]|^2 \tag{1}$$

and Root Mean Squared Energy (less sensitive to outliers) (2)[12]. Additional features comprised Zero-Crossing Rate (signal sign changes, selected via RFE) [13], Skewness and Kurtosis (signal asymmetry and peakedness, selected via RFE), Peak Frequency (highest-amplitude frequency) [14], Onset Detection (S1, S2, and murmur peaks), Band Energy Ratio (low/high-band energy comparison) [15], Spectral Centroid (energy concentration), and the first 13 Mel-Frequency Cepstral Coefficients (MFCCs) capturing the spectral envelope on the Mel scale [16], highlighting their significance across all cardiac sites.

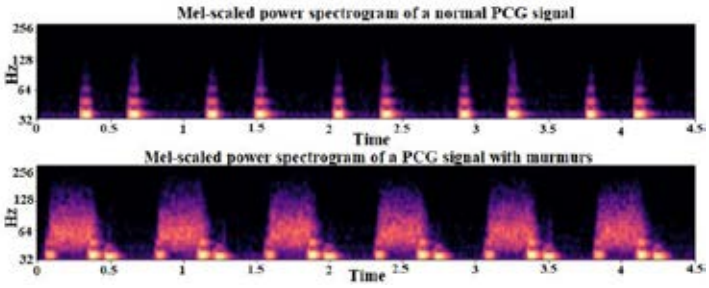


Fig. 1. Mel-scaled power spectrogram of a normal PCG signal. b) Mel-scaled power spectrogram of a PCG signal with murmurs.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}[x(n) - \tilde{x}(n)]^2} \tag{2}$$

B. Heart Murmur Classification

Supervised learning was applied to labeled audio from AV, PV, MV, and TV regions. Feature selection reduced redundancy via Lasso Logistic Regression, ANOVA F-value, and RFE (retaining 20 features) . Data were split into training (510 samples, 47% positive), validation (124, 27%), and testing (124, 24%), with augmentation keeping all samples per patient in one subset.

C. Feature Selection

To improve model performance, three feature selection methods were applied: Lasso Logistic Regression (regularization shrinks coefficients to zero, optimized via cross-validated F1 score), ANOVA F-value (removes features independent of class), and RFE (Recursive Feature Elimination,iteratively removes least important features until 20 remain) . Deep learning models included CNNs for spatial features, LSTMs for temporal dependencies, and CNN-LSTM hybrids. Transfer learning used GhostNet, EfficientNet (B0–B2), ResNet (50,101), MobileNet (V2,V3), DenseNet121, and AttentionCNN for efficiency, scalability, or specialized feature extraction.Training used stratified splits, hyperparameter tuning via EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint, with Adam/RMSprop optimization and Categorical Crossentropy loss. Performance was evaluated with Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrices.

IV. RESULTS

This study evaluates deep learning models for heart murmur detection using Mel spectrograms, demonstrating the effectiveness of various neural networks in classifying heart sounds as Absent, Present, or Unknown murmurs.

A. Dataset Characteristics

The dataset contained 942 patient recordings from four valve positions—Aortic (AV), Mitral (MV), Pulmonary (PV), and Tricuspid (TV)—with 73.8% absent, 19.0% present, and

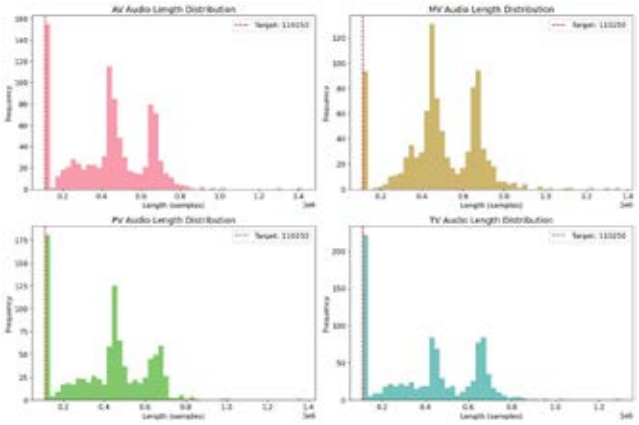


Fig. 2.Audio sample length distribution with target length for normalization

7.2% unknown murmurs. Preprocessing addressed class imbalance.

B. Mel Spectrogram Feature Extraction

Missing data were replaced with zero arrays, and all audio normalized to 5 seconds (Fig. 2). Data augmentation balanced classes. Mel spectrograms (22,050 Hz, 128 bins, 2,048 FFT, 512 hop, 128×216×1) and 14 conventional audio features (e.g., MFCCs, spectral centroid) were extracted.

C. Model Performance Comparison

- *Traditional Machine Learning Models:*

A Dummy baseline underperformed due to imbalance; tuned Logistic Regression, SVM, KNN, Decision Tree, LDA, QDA, and AdaBoost with 84 RFE-selected features improved F1 by ~10%. AV alone performed best, but combining all valves was superior.

- *Basic Deep Learning Models*

CNN, LSTM, and CNN-LSTM were tested; CNN-LSTM achieved the top F1 score of 0.84 (Table I).

- *Advanced Model Architectures*

GhostNet, EfficientNet (B0–B2), ResNet (50, 101), MobileNet (V2, V3), DenseNet121, and AttentionCNN were compared (Table II); DenseNet121 performed best due to dense connectivity.

D. Evaluation

The final CNN–LSTM model, configured with a learning rate of 0.001 and sigmoid activation, attained precision, recall, F1-score, and accuracy of 0.748, 0.800, 0.840, and 0.846, respectively, on the test set, outperforming prior studies as shown in Table III.

Its probabilistic behavior was examined using the ROC curve, where class predictions result from the sign function at the decision boundary, while scikit-learn applies a sigmoid

TABLE I.COMPARISON OF TRADITIONAL MACHINE LEARNING MODELS AND THE CNN-LSTM DEEP LEARNING MODEL PERFORMANCE

Model	Model Selected	Training f1 Score	Validation f1 Score	Testing f1 Score
Logistic Regression	C=0.01, penalty = elasticnet	0.745	0.66	0.704
SVM	C=0.1, kernel=sigmoid	0.69	0.685	0.716
KNN		0.84		0.51
Decision Tree	criterion=entropy, max depth=50, min samples leaf=5 min samples split=20, splitter=random	0.8	0.68	0.49
LDA		0.82		0.51
QDA		0.97		0.30
ADABOOST	algorithm=SAMME, nestimators = 100	0.93	0.68	0.66
CNN-LSTM		0.88	0.82	0.84

TABLE II. PERFORMANCE COMPARISON OF ADVANCED DEEP LEARNING ARCHITECTURES IN HEART MURMUR DETECTION

Model	Test_Accuracy	Test_Loss	Precision	Recall	F1_Score	Parameters	Model_Size_MB
DenseNet121	0.77	0.94	0.70	0.77	0.72	7565897	28.8
AttentionCNN	0.77	0.66	0.76	0.77	0.68	488718	1.86
GhostNet	0.75	0.69	0.56	0.75	0.64	2439341	9.30
EfficientNetB0	0.75	0.71	0.56	0.75	0.64	4839596	18.46
ResNet50	0.75	0.86	0.56	0.75	0.64	26216329	100
MobileNetV3Small	0.75	0.75	0.56	0.75	0.64	1238137	4.72
EfficientNetB1	0.65	0.98	0.66	0.65	0.65	7365264	28.09
MobileNetV2	0.33	1.66	0.62	0.33	0.40	2917449	11.12

transformation to the model's linear output, with parameters A and B estimated via Maximum Likelihood Estimation. This analysis yielded an AUC of 0.879, as depicted in Fig. 3.

TABLE III.COMPARISON OF OUR CNN–LSTM MODEL WITH PREVIOUS HEART MURMUR DETECTION STUDIES

Study	Method	Dataset	Accuracy	F1-score
Our Work	(2025)CNN-LSTM	CirCor DigiScope	84.6%	0.84
Huang et al. (2024)	2D CNN (Multi-task)	CirCor DigiScope	95.19%	-
Reyna et al. (2023)	Various CNN architectures	PhysioNet 2022	80.2%	0.78
Wang et al. (2024)	Branchformer with Uncertainty	PhysioNet 2022	79.8%	0.65
Li et al. (2023)	Neural Architecture Search	PhysioNet 2022	83.4%	0.83
Kumar et al. (2024)	CNN with Mel-spectrogram	PhysioNet 2022	81.8%	0.80

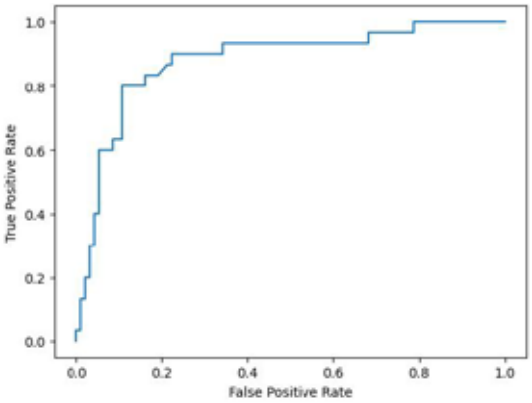


TABLE III.COMPARISON OF OUR CNN–LSTM MODEL WITH PREVIOUS HEART MURMUR DETECTION STUDIES

V. DISCUSSION

This study presents a comprehensive framework for automated heart murmur detection, leveraging hybrid CNN–LSTM and advanced transfer learning models, including GhostNet, EfficientNet, ResNet, MobileNet, DenseNet121, and AttentionCNN. Cardiac sounds were classified as Absent, Present, or Unknown using 5-second Mel spectrograms combined with 14 traditional audio features, addressing class imbalance (73.8% Absent) via data augmentation and normalization. The CNN–LSTM effectively captured both spectral and temporal patterns, achieving an F1-score of 0.84 and outperforming baseline classifiers, while broader evaluation of multiple architectures extends prior work. Strengths include robust preprocessing, multi-dimensional feature representation, stratified splitting, and evaluation using F1-score, ROC-AUC, and visualizations, enhancing clinical relevance. Unlike prior studies limited in scope, this work offers a balanced comparison across models and explicitly addresses critical metrics. Limitations involve dataset imbalance and generalizability, motivating future work on advanced augmentation (SMOTE, GANs), external validation, multimodal clinical integration, Explainable AI, and model compression for real-time deployment. Overall, the framework advances AI-driven cardiovascular diagnostics, enabling earlier murmur detection and broader healthcare accessibility.

REFERENCES

[1] World Health Organization. Cardiovascular diseases (CVDs), 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] ScienceDirect. Heart auscultation—an overview. Accessed from https://www.sciencedirect.com/topics/medicine-and-dentistry/heart-auscultation

[3] ScienceDirect. Extraction and assessment of diagnosis-relevant features for heart sound classification. Accessed from https://www.sciencedirect.com/science/article/abs/pii/S104620232100181X

[4] LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. Nature, 521(7553), 436–444, 2015.

[5] Springer, D. B., Tarassenko, L., & Clifford, G. D. Logistic regression-HSMM based heart sound segmentation. IEEE Transactions on Biomedical Engineering, 63(4), 822–832, 2016.

[6] Maknickas, V., & Maknickas, A. Recognition of normal and abnormal phonocardiographic signals using deep convolutional neural networks and Mel frequency spectral coefficients. Physiological Measurement, 38(8), 1671, 2017.

[7] Bozkurt, B., Germanakis, I., & Stylianou, Y. A study on CNN-based automatic heart sound classification for pathological detection. Computers in Biology and Medicine, 100, 132–143, 2018.

[8] Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In 2016 Computing in Cardiology Conference (CinC), 621–624. IEEE, 2016.

[9] A. Elola et al., IEEE Xplore. Beyond heart murmur detection: Automatic murmur grading from phonocardiogram. IEEE Journal of Biomedical and Health Informatics, 27(8), 3856–3866, 2023. DOI: https://doi.org/10.1109/JBHI.2023.3275039. Accessed from https://ieeexplore.ieee.org/document/10122565

[10] PhysioNet. The CirCor DigiScope Dataset: An open-access pediatric heart sound database for murmur detection, 2022. https://physionet.org/content/circor-heart-sound/1.0.3/

[11] He, L., & Dellwo, V. Amplitude envelope kinematics of speech signal: Parameter extraction and applications. University of Zurich, 2017. https://www.researchgate.net/publication/318275305_Amplitude_envelope_kinematics_of_speech_signal_parameter_extraction_and_applications

[12] Hodson, T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development, 15, 5481–5487, 2022. https://doi.org/10.5194/gmd-15-5481-2022

[13] McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 2015. https://librosa.org/

[14] Erbe, C., Duncan, A., Hawkins, L., Terhune, J. M., & Thomas, J. A. Introduction to acoustic terminology and signal processing. In Exploring Animal Behavior Through Sound: Volume 1, 111–152. Springer, 2022. https://link.springer.com/chapter/10.1007/978-3-030-97540-1_4

[15] ScienceDirect. Audio frame—an overview. Accessed August 10, 2025, from https://www.sciencedirect.com/topics/engineering/audio-frame

[16] M. W. Lakdari, A. H. Ahmad, S. Sethi, G. A. Bohn, and D. J. Clink,ScienceDirect. Mel-frequency cepstral coefficients outperform embeddings from pre-trained convolutional neural networks under noisy conditions for discrimination tasks of individual gibbons. Accessed August 10, 2025, from https://www.sciencedirect.com/science/article/pii/S1574954123004867

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Evaluation of Thermal Damage Thresholds and Coagulation Behavior of 809 nm Diode Laser in Liver Tissue

Oguzhan Efe SAPMAZ[1], Ayse Sena SARP[2,3]
[1]Department of Bioengineering, Bahçesehir University, Istanbul, Turkey
[2]Faculty of Dentistry, Bahçeşehir University, Istanbul, Turkey
[3]Department of Biomedical Engineering, Bahçesehir University, Istanbul, Turkey
efesapmaz@hotmail.com, aysesena.sarp@bau.edu.tr

***Abstract*—Liver is a unique organ which has ability to regenerate itself. However, the liver tumors represent the most prevalent pathologies in this organ. Current surgical techniques for liver treatment suffer from drawbacks have been observed as insufficient application precision, uncontrolled ablation volumes, and post-application bleeding. Laser technology has emerged as a promising alternative due to its effectiveness. However, a significant limitation of lasers has resulted thermal damage which was caused by tissue temperature elevation. Laser light induces multi-layered thermal effects in targeted tissues, including ablation, carbonization, and coagulation, where progresses from inner to outer layers. Selecting an appropriate laser type is crucial based on its penetration depth and the depth of the tumor. Among laser techniques, Laser Interstitial Thermal Therapy (LITT) is widely recognized for its efficacy in liver tumor ablation and leveraging modified laser applicators. This study aims to evaluate the impact of irradiation duration on the liver tissue to establish optimal parameters for 809 nm wavelength diode laser. Initially, irradiation durations leading to coagulation or carbonization were measured, and corresponding energy values were calculated. Subsequently, temperature variations within the tissue were recorded. Findings indicated that the average thresholds for coagulation and carbonization were observed at irradiation times ranging from 0.33 to 1 second at power levels of 0.5 to 1 W, and from 0.34 to 1.31 seconds at 1 to 3 W for the 809 nm laser.**

Keywords— liver; tumor; laser ablation; coagulation; carbonization

I. INTRODUCTION

A healthy liver has ability to regenerate itself as different from other organs [1]. Especially in hepatectomy operations, even if 90% of the liver is removed, the organ can return to its previous state after 3 months and fulfills all its functions without any problems However, long-term bad habits such as alcohol, obesity, smoking and diseases such as hepatitis A, B and C types cause irreversible scars in the liver malignant tumor such as hepatocellular carcinoma (HCC) begins to develop in the cirrhotic liver which loses regeneration capacity and liver failure occurs in last stage. HCC was reported fifth most frequent cancer and third leading cause of cancer-related death, causing more than 800,000 deaths annually in worldwide [2]. Liver metastasis is another malignancy which is risen mostly from colorectal cancer (CRC). Approximately, 50% of patients with CRC were reported as developing liver metastases during the disease [3]. In all these cases, organ transplantation with a healthy liver from a donor maintains last option.

Various ablation methods have been benefited to both stop bleeding after hepatectomy and transplantation and to eliminate liver tumors. While radiofrequency ablation destroys the tumor by delivering 200-500 kHz acoustic waves to the tissue [4]. Microwave ablation is benefited from heat vibration of water molecules at a frequency of 900-2500 MHz [5]. Cryoablation is the process of cooling argon and helium gases to -196 ºC degrees and exposing the diseased area to sudden cold-hot cycles between -35 ºC and 55 °C degrees [6, 7]. Irreversible electroporation is based on positioning cancer cells between electrodes and exposing them to high energy which is aimed to destroy the cell with a voltage between 1000-3000 V entering the cell nucleus [8]. Another method is the injecting ethanol through a syringe needle into the diseased tissue to dehydrate the tumors. It has been observed that all these applications are more effective in tumors below 5 cm size [9].

In the surgical treatment of liver; insufficient single application, uncontrolled ablation volume and bleeding in lesions after application are among their disadvantages of present techiques. In these ablations, the fact that the instruments are not in direct contact with the tumor or diseased tissue and bleeding continues has shown that laser ablation can be used as an alternative.

The laser was first successfully performed in 1960 by Maiman. Later, it began to be used in various military, cosmetic and health fields. Laser light contains a single wavelength; it does not dissipate as it moves away from the source; it differs from other types of light in that its power, direction and frequency remain

979-8-3503-2896-7/23/$31.00 ©2025 IEEE

constant. In laser formation, a trigger source is firstly required. This source activates light particles called photons in the substance called medium. In order to prevent the resulting heat from dissipating around, an optical resonator, which includes the source and medium, is used. The mirrors inside the resonator ensure the reflection of the beam inside, and the final laser beam passes through the semi-reflective mirror and exits the system [10].

In laser-tissue interaction, the laser beam can be reflected, passed through, scattered or absorbed. Hemoglobin, water, melanin, and retinoid play a significant role in the absorption of light. This process depends on the wavelength of the incoming light. For instance, melanin and hemoglobin can absorb light between 500-1000 nm, while retinoid protein tends to absorb light below 200 nm more easily. Water has the feature of absorbing almost every wavelength, although it is mostly above 1000 nm [11].

Various types of laser-tissue interactions occur depending on the wavelength of the laser used. Among these, low wavelength lasers such as excimer lasers are used in photo-ablation. For instance, since retinoid protein is found mostly in the eyes, this laser is used to eliminate corneal defects. In photothermal ablation, lasers with wavelengths above 1000 nm such as Nd:YAG, CO_2, Er:YAG, Ho:YAG are used. When these lasers are applied to the tissue, the temperature on the tissue surface increases from generally 37°C to 300°C, resulting in hyperthermia, coagulation, vaporization (ablation), carbonization and melting, respectively.

Liver contains high amount of water and hemoglobin, thus CO_2, Argon and Nd:YAG lasers are used to destroy cancerous or diseased tissues or to separate from healthy tissues. Due to the fact that, laser treatment has demonstrated to reduce the amount of infection caused by the use of scalpels in surgeries. Nd:YAG laser can provide necrosis of the tumors where located deep below the tissue surface [12]. Because the 1064 nm wavelength has the highest optical penetration as 4 mm into the tissue among all wavelengths. In the CO_2 laser with 10600 nm, this effect is more superficial as 0.75 mm [13].

Laser Interstitial Thermo Therapy (LITT) treatment is advantageous laser ablation which based on laser output creating of laser output a three-dimensional coagulation area in the area close to the tumor directly through the tissue. Fiber cables are used to transport the laser beam more easily within the system. In the treatment, the laser beam coming from the source under the guidance of an MRI or CT device travels through the fiber and reaches the target tissue with the help of a transparent catheter previously positioned at the subcutaneous tissue entrance [14]. Generally Nd:YAG and diode laser are preferred over CO_2 lasers in LITT because Nd:YAG laser travels further in the tissue. The fiber tip is positioned within the tumor and the laser exits from this end, ensuring thermo toxic death of the tumor. In practice, it is important to carry out coagulation slowly in order to avoid damage to the surrounding tissues and to avoid possible vaporization and carbonization [15].

Various studies were conducted after 1980 to better understand the type and extent of thermal damage caused by laser on liver tissue. Among these, Nikfarjam *et al*. applied 1064 nm Nd:YAG and 980 nm diode laser 5 mm below the tissue at a certain power and in two varying second parameters. While lower coagulation occurred with the Nd:YAG laser compared to the diode laser at the shorter time parameter, carbonization was observed only as a result of the application of the diode laser at the high time parameter. Previously, Jacques and his colleagues applied the 805 nm diode laser Nd:YAG on dog and rat livers separately at similar powers and for longer periods of time. The penetration of both lasers on the liver tissues was measured and the vaporization and coagulation zones were determined according to the different color parts of the damage. While the diameter and depth of the damage caused by the 805 nm laser were close to each other, the diameter was higher and the depth was almost zero with the 1064 nm laser [16, 17].

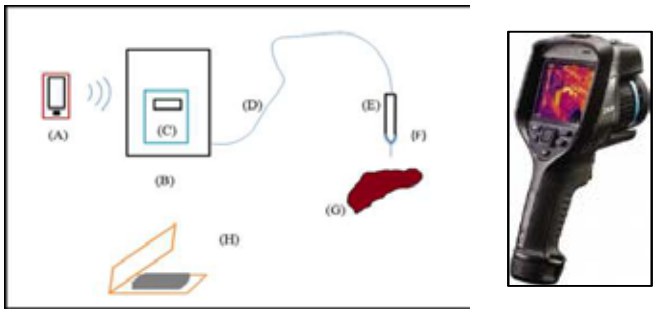
In liver-laser applications, in addition applying laser for certain periods of time to understand coagulation or carbonization damage size, the time until these thermal damage occurrence can be measured to also determine the safe laser and power value. This method was implemented by Yıldız *et al.* using 980 nm diode, 1070 nm YLF and 1940 nm Thulium fiber lasers in various power ranges. As a result of measuring the extent of both types of thermal damage occurring on the server with the help of a caliper and microscope, the found times showed that 980 nm creates high-level thermal damage, 1940 nm is theoretically suitable for ablation because it is absorbed faster on the water side, and 1070 nm is the damage that occurs even at high power density. Since it is close to zero, the 1070 nm has been determined as a safe wavelength [18].

II. MATERIALS AND METHODS

This study aims to evaluate the impact of irradiation duration on liver tissue to establish optimal parameters for 809 nm wavelength diode laser. The experiments were done in two parts. Firstly, the 809 nm diode laser was directly applied on the liver tissue, meanwhile the time was measured until coagulation and carbonization formation occurred on the tissue. In addition to this, the diameter of these thermal damages were measured. In the second part of the experiments, temperature changes were measured with the help of a thermal camera by applying these different power, time and energy parameters, and morphological changes were observed on the tissue surface.

The first part of the experiment of this study was performed at Bahçeşehir University Faculty of Dentistry using SOLASE Dental Diode Laser device which is a circuit consisting of a smartphone nest, a fiber cable

connected to it, a handpiece, a disposable type and a foot pedal connected via Bluetooth. Tip type, type of the operation mode and varying laser powers were adjusted separately for each application in a smartphone.



(1)

(2)

Fig. 1. (1) Portable 809 nm Diode Laser System Circuits; (A) smartphone, (B) laser source, (C) phone nes, (D) fiber cable, (E) handpiece, (F) 400 µm diameter disposable tip, (G) liver tissue sample and (H) foot pedal, (2) Handpiece Thermal Camera

Depending on the capacity of the device, ranges of 1-6 W for 809 nm were used. Finally, by pressing the foot pedal, lasers with 809 nm wavelength was applied to the liver surface. Approximately 10 laser applications were performed at the same power value at all wavelengths.

Measurements of coagulation and carbonization threshold times and diameter measurements of two types of thermal damage occurring in the tissue were made with the help of calipers. The found threshold time data were multiplied by the required values to be applied to obtain the required energy values for the optimum parameter.

After taking the average and standard deviation of the measurements taken with the same parameters, student-t-test ($p \leq 0.05$) was calculated between the measurement groups at any two power values via Excel program to determine whether there was a significant difference between the groups and to find the optimum lasing parameters for that wavelength.

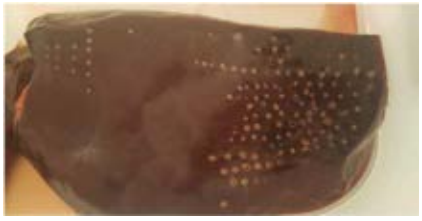


Fig. 2. Thermal damage points after 809 nm laser application

Second part of the experiment was performed in Boğaziçi University Biophotonics Laboratory with using the Flir-E4 brand handpiece thermal camera which was held between THORLAB brand metal holders. In order to make a comparison based on the same energy values, 4 W 6 sec at 809 nm. During the applications, temperature increases on the tissue surface were measured separately for both wavelengths.

In addition, with the same wavelength laser, regardless of the threshold times, using the same power values as in the previous experiment, between 1-6 W, the liver tissue was first injected with sequential energy values as similar as possible, with increasing time values for each power value, for example, both 3 W and 4 W. The energies of the applied lasers were calculated as 6 j, 12 j, 24 and 36 j, and these time parameters were determined according to the purpose.

III. RESULTS

At the 809 nm wavelength used in this study, the average coagulation threshold times decreased from 14.96 seconds to 1.62 seconds as the power increased from 1 W to 6 W in Fig. 4. Although the laser was applied for 40 seconds at 1 W to determine the average carbonization threshold time, no measurement was recorded because carbonization was not observed. For power levels ranging from 2 W to 6 W, the onset times for carbonization decreased from 12.52 seconds to 5.13 seconds in Fig. 3. Generally, a decrease in these times was observed for both types of thermal damage with increasing power. This is expected, as higher laser power facilitates faster heat transfer to the tissue. The required energy values, as shown in Table I, demonstrated a decreasing trend for coagulation, ranging from 23.1 J to 9.74 J with increasing power. In contrast, the energy values for carbonization ranged from 20.4 J to 38.06 J, showing no significant change with increasing power from 2W.

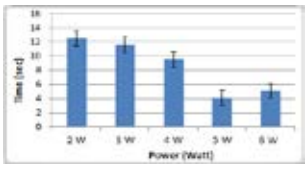


Fig. 3. Measurements of carbonization times in applications with 809 nm laser

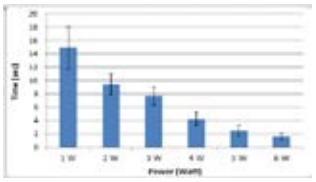


Fig. 4. Measurements of coagulation times in applications with 809 nm laser

TABLE I. REQUIRED ENERGY DATA FOR 809 NM LASER APPLICATION

Power (Watt)	**Energy Values (Joule)**	
	Coagulation	***Carbonization***
1	14.96	-
2	18.93	25.06
3	23.1	33.53
4	17.1	38.06
5	12.66	20.41
6	9.74	30.8

In the 809 nm laser application, the t-test values for the times of 9.46 and 2.53 seconds at 2 W and 5 W, respectively, and 7.69 and 1.62 seconds at 3 W and 6 W, respectively, were found to be below 0.05, indicating a significant difference between these time measurements. However, the carbonization threshold times, which were 12.52, 11.17, and 9.51 seconds at 2 W, 3 W, and 4 W, respectively, showed the opposite trend with t-test values of 0.08 and 0.06. The only significant difference was observed between 4 W and 5 W, with times of 9.51 seconds and 4.08 seconds, respectively, and a t-test value of $1.3*10^{-6}$. Generally, considering the t-test values and the trend that threshold times decrease with increasing power, it can be concluded that the difference between coagulation and carbonization threshold times increases inversely as the applied power decreases.

The coagulation diameters of 809 nm laser in the tissue was found at wider range than carbonization diameter ranges ,which were below 1 mm, at same applied time and power values. For instance when coagulation diameters were found between 1.66-2.67 mm at 6, 9 and 12 sec laser application at 2 W, carbonization were found between 0.66-1.36 mm. Only at 4 W, diameter ranges of coagulation and carbonization thermal damage was at almost same width as 2.08-2.41 and 0.75-1.05 mm respectively at same applied seconds. In both thermal damage diameter ranges, there was a decrease of up to 4 W and an increase of 6 W with increasing power. Although lower application times were used at 6 W, the higher diameter range showed that the power parameter rather than time is more effective in diameter size.

IV. CONCLUSION

In this study, we identified the optimal safe wavelength and power settings for laser applications by measuring thermal damage threshold times, temperature changes, and damage diameters in tissue. The 809 nm wavelength was found to have significantly high threshold times and caused low damage to liver tissue at the same duration and power parameters. This supports the theory that the absorption coefficient of water is lower at the 809 nm wavelength.

Future research could involve modifying the time intervals to milliseconds, microseconds, or nanoseconds in continuous-wave (CW) applications and examining temperature changes during laser application to determine the onset time. This could help in finding appropriate parameters to prevent excessive tissue irradiation or burning of liver tissue during surgeries. Additionally, understanding the coagulation threshold times for lasers, as determined in this study, may assist in evaluating the duration of liver tumor surgeries and choosing between laser ablation methods such as LITT or hepatectomy.

REFERENCES

[1] Gilgenkrantz, H., &, Hortet, A. C. (2018). Understanding liver regeneration. *The American Journal of Pathology,* 188*(6),* 1316–1327. doi: 10.1016/j.ajpath.2018.03.008

[2] Li D, Kang J, Golas BJ, Yeung VW, Madoff DC. Minimally invasive local therapies for liver cancer. Cancer Biol Med. 2014 Dec;11(4):217-36. doi:10.7497/j.issn.2095 -3941.2014.04.001.

[3] Zhou, H., Liu, Z., Wang, Y. *et al.* Colorectal liver metastasis: molecular mechanism and interventional therapy. *Sig Transduct Target Ther* **7**, 70 (2022). https://doi.org/10.1038/s41392-022-00922-2

[4] Agrawal, S., Agarwal, S., Arnason, T., Saini, S., & Belghiti, J. (2014). Management of hepatocellular adenoma: Recent advances. Clinical Gastroentology and Hepatology, 13(7), 1221–1230. doi: 10.1016/j.cgh.2014.05.023

[5] Yu, H., & Burke, C. T. (2014). Comparison of percutaneous ablation technologies in the treatment of malignant liver tumors. Semin Intervent Radiol, 31(2), 129–137. doi: 10.1055/s-0034-1373788.

[6] Dick, E. A. (2002). Ablative therapy for liver tumours. Gut, 50(5), 733–739. doi: 10.1136/gut.50.5.733

[7] Gannon C. J., & Curley, S. A. (2005). The role of focal liver ablation in the treatment of unresectable primary and secondary malignant liver tumors. Seminars in Radiation Oncology, 15(4), 265–272. doi: 10.1016/j.semradonc.2005.04.004

[8] Brown, R. E., Martin, R. C. G., & Scoggins, C. R. (2011). Ablative therapies for colorectal liver metastases. Surgical Oncology Clinics of North America, 20(2), 259–271. doi: 10.1016/j.soc.2010.11.005.

[9] Sindram, D., Lau, K. N., Martinie, J. B., & Iannitti, D. A. (2010). Hepatic tumor ablation. Surgical Clinics of North America, 90(4), 863–876. doi: 10.1016/j.suc.2010.04.014

[10] Romanos, G. E. (2021). Laser fundamental principles. Advanced laser surgery in dentistry, (pp. 1–40). Hoboken, NJ: Wiley-Blackwell

[11] Boechat, Á. (2016). Biophotonics. M. A. Issa, & B. Tamura (Eds.), Lasers, lights and other technologies (pp. 1–47). Heidelberg: Springer.

[12] Gibson, K. F., & Kernohant, W. G. (1993). Lasers in medicine-a review. Journal of Medical Engineering & Technology, 17(2), 51–57. doi: 10.3109/03091909309008351

[13] Welch, A. J., & Gemert, M. J. C. (1995). Optical-thermal response of laser irradiated tissue. New York: Plenum Press

[14] Vogl, T. J., Farshid, P., Naguib, N. N. N., Darvishi, A., Bazrafshan, B., Mbalisike, E., ... Zangos, S. (2014). Thermal ablation of liver metastases from colorectalcancer: Radiofrequency, microwave and laser ablation therapies. La Radiologia Medica, 119(7), 451–461. doi: 10.1007/s11547-014-0415-y

[15] Curley, S. A., Cusack, J. C., Tanabe, K. K., & Ellis, L. M. (2002). Advances in the treatment of liver tumors. Current Problems in Surgery, 39(5), 461–571. doi: 10.1067/msg.2002.122810

[16] Nikfarjam, M., Malcontenti, W. C., & Christophi, C. (2005). Comparison of 980- and 1064-nm wavelengths for interstitial laser thermotherapy of the liver. *Photomedicine and Laser Surgery, 23*(3), 284–288. doi:10.1089/pho.2005.23.284

[17] Jacques, S. L., Rastegar, S., Motamedi, M., Thomsen, S. L., Schwartz, J. A., Torres, J. H., & Mannonen, I. (1992). Liver photocoagulation with diode laser (805 nm) versus Nd:YAG (1064 nm). *Laser-tissue Interaction III, 1646*, 107–117. doi: 10.1117/12.137451

[18] Yıldız, F. (2012). In vitro investigation and comparison of NIR-IR lasers effects in liver tissue (Master's thesis). Available from YÖK Ulusal Tez Merkezi database

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Preliminary Evaluation of the Passive Rehabilitation Phase of the REACT System for Tendon Therapy

Habibe Asran*, Zikra Alfaris*, Elif Hocaoglu†
*Living Robotics Laboratory, Biomedical Engineering and Bioinformatics,
Istanbul Medipol University, Istanbul, Türkiye
†Living Robotics Laboratory, School of Engineering and Natural Sciences
Istanbul Medipol University, Istanbul, Türkiye
Research Institute for Health Sciences and Technologies (SABITA),
Istanbul, Türkiye
Email: ehocaoglu@medipol.edu.tr

***Abstract*—Tendon injuries, particularly in the hand, require early and controlled rehabilitation to restore joint mobility while preventing adhesion and minimizing the risk of complications, such as suture rupture or gap formation. Robotic tendon therapy systems offer a promising alternative through programmable and repeatable motion delivery, yet many current designs are hindered by bulk and mechanical complexity, limiting their practicality in clinical use. Shape memory alloy (SMA) based actuators represent a promising solution for tendon-driven exoskeletons owing to their lightweight construction, high force-to-weight ratio, and silent operation, features that are particularly advantageous for wearable and portable rehabilitation devices. However, the nonlinear, hysteretic, and thermally dependent nature of SMA behavior poses significant control challenges, necessitating detailed experimental characterization and modelling. This study presents the tendon-driven *Rehabilitation Exoskeleton for Assisted Coordination of Tendon Therapy (REACT)* system, incorporating experimentally derived force–temperature and current–temperature models of SMA units into a current-based position control framework for the passive rehabilitation phase. The controller incorporates these models into the forward dynamics of the finger to achieve accurate torque generation and trajectory tracking within the physiological range of motion. Simulation results indicate high-fidelity reference tracking, with a steady-state error of approximately -8×10^{-3} rad, thereby confirming the feasibility of SMA-based actuation for safe, precise joint positioning in early-stage tendon rehabilitation.**

***Keywords*—Assistive tendon therapy, hand exoskeleton, shape memory alloys, current-based position control**

I. Introduction

Tendon injuries to the fingers significantly impair hand function, restricting mobility and daily activities [1]. Surgical repair is usually required to rejoin the tendon ends and restore joint function [2]. Following surgery, either active or passive, early rehabilitation is strongly recommended to prevent adhesions and promote tendon healing. Effective rehabilitation aims to restore tendon function and joint mobility while avoiding complications such as stiffness, rupture, or gap formation [3]. Conventional rehabilitation, however, relies heavily on therapist supervision, precise movement execution, and repetitive training factors that are difficult to sustain over long recovery periods.

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

Robotic rehabilitation has therefore emerged as a promising complement to traditional therapy. Robots provide consistent and repeatable motion, enable quantitative progress tracking, and allow therapy to be tailored to the patient's needs [4], [5]. Several robotic hand and finger exoskeletons have been introduced for rehabilitation, with promising outcomes [6], [8], [9]. Yet, most designs remain bulky, noisy, or difficult to personalize, limiting portability and tendon-specific therapy. This underscores the need for lightweight, silent, and wearable solutions.

SMAs have been explored as a promising alternative, combining high force-to-weight ratio and compactness with quiet operation, though they require careful control to address nonlinear behavior [14], [15], [16]. A lack of adaptability in the control of SMAs can either increase the risk of tendon rupture through excessive force or, conversely, restrict recovery by failing to provide adequate mobilization [7], [8], [9]. This makes it important to control SMA actuators in a way that is safe and reliable for tendon rehabilitation. However, SMA actuators also exhibit hysteresis and temperature-dependent characteristics, which pose additional challenges for stable and predictable control. Addressing these issues often requires mathematical modeling and experimental characterization to better understand their behavior and guide controller design [14], [15], [16].

Prior studies on SMA-based actuation robotic systems have explored relatively simple control schemes, such as pulse-width modulation (PWM) with proportional–derivative (PD) feedback and feedforward trajectory planning. These approaches demonstrate that SMAs can achieve reliable motion tracking when combined with sensor feedback. However, their performance is fundamentally constrained by the thermal dynamics of the material, which slow both activation and recovery. This problem has been addressed by integration of active cooling or antagonistic actuator arrangements to improve cycle speed[10], [11].

In addition to controller design, effective control of SMA actuators critically depends on precise thermal and mechanical characterization of the material. The force output of SMA springs depends strongly on temperature through changes

in the shear modulus, meaning that accurate knowledge of transformation temperatures such as austenite finish temperature is essential. Temperature evaluation techniques such as differential scanning calorimetry and dynamic thermal analysis are therefore widely used to identify these properties and enable consistent mapping between electrical input, temperature response, and mechanical force [12], [13]. Together, these studies underline both the promise and the challenge of using SMAs in rehabilitation devices.

To explore these challenges, we propose the REACT system, a wearable robotic exoskeleton specifically designed for tendon rehabilitation. The REACT system leverages SMA spring actuators in combination with dynamic system modeling, experimental mechanical characterization, and sensor-based feedback within an adaptive control framework. In this paper, we present simulation results that demonstrate the feasibility of this approach for tendon rehabilitation. These findings provide an initial step toward evaluating adaptive control strategies for safe and responsive assistance, which will require further validation beyond simulation in future work.

The remainder of this paper is structured as follows: Section II introduces the system design, SMA actuator characterization, and control methodology. Section III reports the results and analyzes system performance. Section IV concludes with a summary and outlines directions for future work.

II. Methodology

The REACT system is a tendon-driven wearable robotic exoskeleton designed for finger tendon rehabilitation. Actuated by SMA springs, it controls the range of motion (ROM) of the metacarpophalangeal (MCP) joint. This paper presents the implementation of a position controller for the passive phase of therapy, where the exoskeleton moves the finger through flexion and extension without patient effort, reducing the risk of tendon re-rupture. The control strategy integrates SMA characterization and forward dynamics modelling to achieve smooth, accurate, and safe joint actuation within the prescribed ROM.

A. System Design

The REACT exoskeleton is a tendon-driven, wearable device designed for finger tendon rehabilitation. As shown in Fig. 1, it features two actuation units responsible for flexion and extension, respectively. Each unit contains antagonistically arranged SMA springs: one contracts to produce the necessary actuation force, while the opposing spring returns the system to its initial position. This arrangement enables precise bidirectional control of finger movement without relying on passive elastic elements.

Force transmission is achieved through a tendon-driven mechanism, where cables connect the SMA actuators to adjustable phalanx rings via Bowden cables. A low-friction slider between the SMA-based actuator unit and tendon cables minimizes force loss, ensuring accurate and efficient motion control. To overcome the SMA's inherent slow cooling and

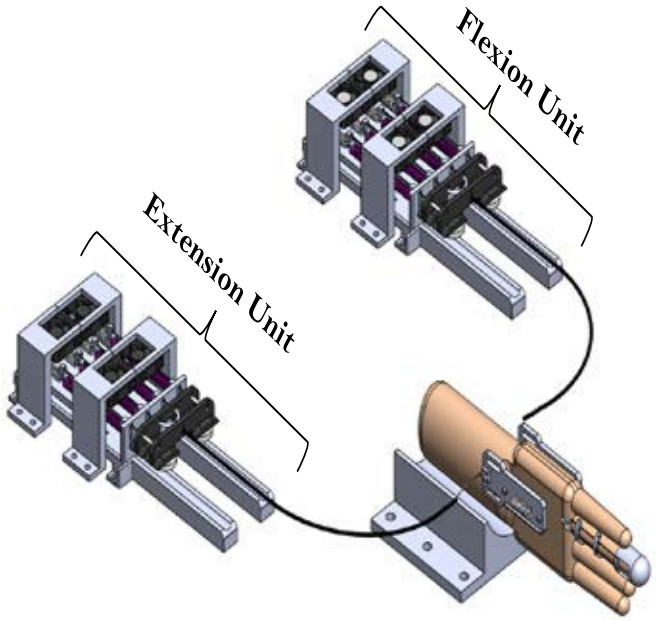


Fig. 1: Solid model of REACT: SMA-based actuation units and wearable module

improve actuation speed, compact cooling fans are integrated to facilitate rapid thermal dissipation.

The system's compact and lightweight design promotes patient comfort and portability. This mechanical configuration optimizes force transmission and establishes the essential foundation for precise position control of the SMA actuators, which is critical for accurate finger joint positioning during rehabilitation.

B. Control Architecture for Passive Therapy

In the REACT system, position control is implemented by modulating the contraction of SMA springs, which actuate the tendon-driven mechanism to achieve the desired finger joint angle. Achieving precise position control requires an understanding of how the finger's motion responds to actuation forces. To this end, a forward dynamics model of the finger is developed, enabling the desired joint position to be expressed as the torque demand for the SMA actuators.

Accurate torque generation requires a detailed understanding of SMA actuator behavior. For this purpose, experimentally derived models of the SMA springs are developed to characterize their actuation response under different operating conditions. These models are integrated into the control block to predict and regulate actuator output during position control.

The overall control process, illustrated in Fig. 2, begins with the desired MCP joint position (q_{ref}) as input. The position error (e), calculated from the difference between the target and measured angles (q_{actual}) via an IMU sensor, is processed by a PID controller. The controller output is passed through the finger's forward dynamics model, which computes the torque required to achieve the commanded motion. The equations

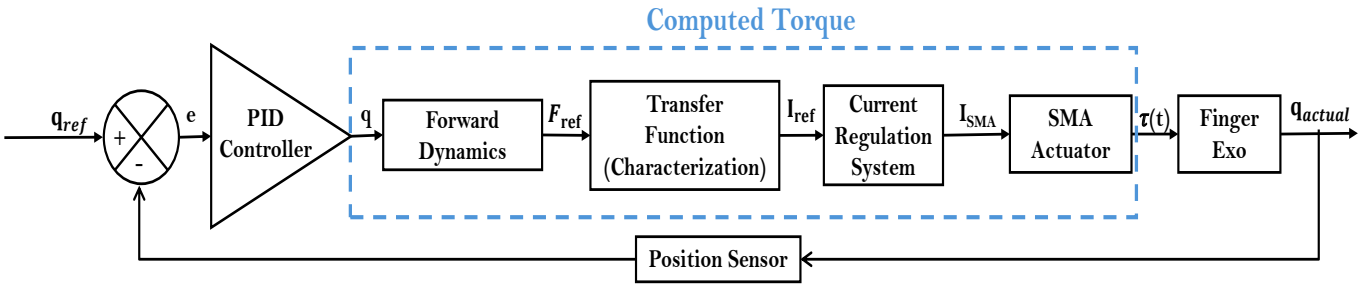


Fig. 2: Representation of the control block diagram of the REACT system for the passive therapy of the patients

governing this model follow the standard manipulator dynamics formulation given in Eq. 1.

$$\tau = M(\theta)\ddot{\theta} + C(\theta,\dot{\theta})\dot{\theta} + G(\theta), \tag{1}$$

,where symbol $M(\theta)$ represents the inertia matrix, $C(\theta,\dot{\theta})$ captures Coriolis and centrifugal effects, and $G(\theta)$ accounts for gravitational torques. For integration with the SMA characterization model, this torque is represented as an equivalent linear tendon force using the relationship given in Eq.2.

$$\tau = F_{\text{ref}} \cdot R \tag{2}$$

, where symbols R and F_{ref} are the joint moment arm and force, respectively. Here, the force acquired from the SMA actuator is employed to generate the desired joint torque to actuate the finger.

The required torque is then mapped to an SMA force and converted into a target temperature using the experimentally obtained force–temperature curve. The Joule heating model subsequently determines the necessary electrical current (I_{ref}), which is regulated by a dedicated current control loop. The actuated SMA contracts to generate the desired tendon force, rotating the MCP joint to the reference position. Continuous sensor feedback ensures accurate tracking and compensates for any discrepancies due to actuator nonlinearities or external disturbances.

C. SMA Characterization

Accurate control of the REACT system relies on the mathematical modelling of SMA actuator behavior. Since the performance of SMA actuator units is highly dependent on both temperature and electrical current, an experimental characterization is carried out to systematically establish the relationships among input current, actuator temperature, and the resulting force output.

The characterization involved measuring SMA spring force output over a range of temperatures while ensuring operation within the identified elastic limits to avoid plastic deformation. A fourth-order polynomial model is obtained to represent the force–temperature relationship:

$$\begin{aligned} F(T) = {} & 1.06 \times 10^{-5} T^4 - 1.96 \times 10^{-3} T^3 \\ & + 1.30 \times 10^{-1} T^2 - 3.36T + 29.7 \end{aligned} \tag{3}$$

, where $F(T)$ is the output force [N], and T is the SMA temperature [°C].

To link electrical input to thermal response, the SMA heating rate is measured for various applied currents. The resulting temperature–current relationship follows the Joule heating principle and is fitted to a linear model:

$$\frac{\Delta T}{\Delta t} = 0.1986\, I^2 - 0.3515 \tag{4}$$

where $\frac{\Delta T}{\Delta t}$ is the average temperature rise rate [°C/s], and I is the applied current [A].

These experimentally derived models are integrated into the control architecture (Fig. 2) to determine the required temperature and corresponding current needed to generate the torque demanded by the forward dynamics model, ensuring accurate and repeatable finger positioning during rehabilitation.

III. Implementation and Simulation Results

The position control scheme for the passive therapy phase is implemented in the MATLAB/Simulink environment, integrating the experimentally derived SMA characterization models with the forward dynamics of the finger exoskeleton to assess control accuracy. A sinusoidal reference trajectory, ranging from 0° to 90°, is used to represent the physiological ROM of the finger joint, simulating the target movement pattern for rehabilitation.

Figure 3 presents the simulation results for reference tracking, tracking error, and a zoomed-in view of the steady-state error. The top plot shows that the controller accurately follows the desired joint position, with the actual trajectory (solid line) closely overlapping the reference (dashed line) throughout the motion cycle. The middle plot depicts the tracking error over time, revealing a rapid initial transient followed by a small and stable error profile. The bottom plot highlights the steady-state error, which remains approximately -8×10^{-3} rad (about $-0.46°$), confirming the controller's high tracking accuracy in the passive phase.

The simulation demonstrates that the integration of the SMA actuator model within the control loop enables accurate torque generation for position regulation, despite the nonlinear and temperature-dependent nature of SMA behavior. The use of the forward dynamics model ensures that the control law accounts for the mechanical dynamics of the finger exoskeleton, while

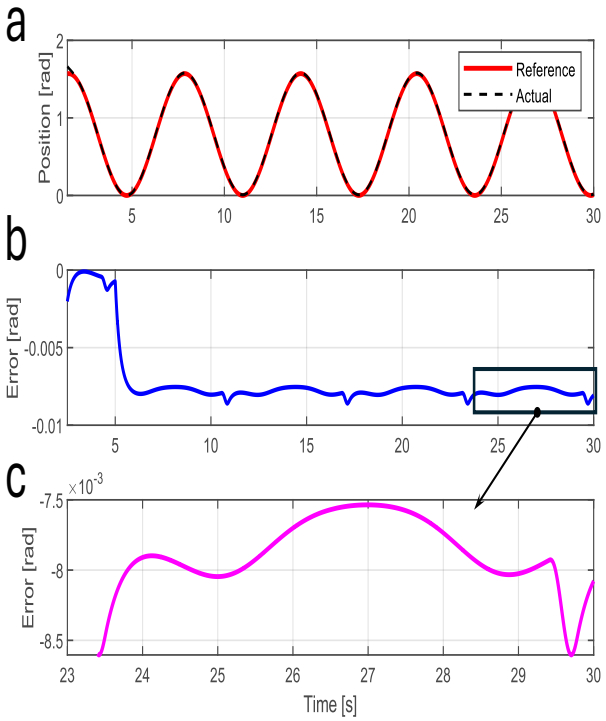


Fig. 3: Simulation results of the passive phase position controller. (a) Reference and actual MCP joint position tracking. (b) Tracking error over the full simulation time. (c) Zoomed-in view of the steady-state error, showing a consistent offset of approximately -8×10^{-3} rad.

the characterization-based transfer functions translate desired forces into appropriate current inputs for each SMA actuator.

IV. Conclusion

This study successfully demonstrates the implementation of a current-based position control strategy for the REACT tendon-driven finger exoskeleton for the purpose of the passive rehabilitation phase. The SMA characterization model derived experimentally is used in the forward dynamics of the finger; the control of the physical system effectively compensates for the nonlinear and temperature-dependent behavior of SMA actuator units. Simulation results demonstrate precise tracking performance within the physiological range of motion, with steady-state errors maintained in the milliradian range. These findings validate the feasibility of SMA-based actuation for delivering precise, reliable, and safe joint movements in tendon rehabilitation. Future work will involve implementing position control based on the dynamic model in the physical prototype of the REACT system, followed by experimental evaluations with healthy participants and patients. These trials will assess the system's performance under clinical rehabilitation conditions during the passive rehabilitation phase, ensuring its safety, accuracy, and suitability for early-stage tendon therapy.

Acknowledgment

This project has been supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2209B Industry-Oriented Research Project Support Programme.

References

[1] M. V. Radomski and C. A. Trombly Latham, *Occupational Therapy for Physical Dysfunction*, 7th ed., 2014.
[2] J. T. Dennerlein, "Finger flexor tendon forces are a complex function of finger joint motions and fingertip forces," *Journal of Hand Therapy*, vol. 18, no. 2, pp. 120–127, 2005. doi: 10.1197/j.jht.2005.01.011.
[3] R. El Hawary, W. D. Stanish, and S. L. Curwin, *Rehabilitation of Tendon Injuries in Sport*, vol. 24, 1997.
[4] K. Nykänen, *The Effectiveness of Robot-Aided Upper Limb Therapy in Stroke Rehabilitation: A Systematic Review of Randomized Controlled Studies*, 2010.
[5] G. Kwakkel, B. J. Kollen, and H. I. Krebs, "Effects of Robot-Assisted Therapy on Upper Limb Recovery after Stroke: A Systematic Review," *Neurorehabilitation and Neural Repair*, pp. 111–121, Mar. 2008.
[6] T. Du Plessis, K. Djouani, and C. Oosthuizen, "A review of active hand exoskeletons for rehabilitation and assistance," *Robotics*, vol. 10, 2021. [Online]. Available: https://doi.org/10.3390/robotics10010010
[7] L. Marchal-Crespo and D. J. Reinkensmeyer, "Review of control strategies for robotic movement training after neurologic injury," *Journal of NeuroEngineering and Rehabilitation*, vol. 6, no. 1, p. 20, Jun. 2009. [Online]. Available: https://doi.org/10.1186/1743-0003-6-20
[8] I. H. Ertas, E. Hocaoglu, and V. Patoglu, "AssistOn-Finger: An underactuated finger exoskeleton for robot-assisted tendon therapy," *Robotica*, vol. 32, no. 8, pp. 1363–1382, 2014.
[9] I. H. Ertas, E. Hocaoglu, D. E. Barkana, and V. Patoglu, "Finger exoskeleton for treatment of tendon injuries," in *2009 IEEE International Conference on Rehabilitation Robotics*, Kyoto, Japan, 2009, pp. 194–201, doi:10.1109/ICORR.2009.5209487.
[10] S. Kim, Y.-L. Park, and R. J. Wood, "A shape memory alloy-based tendon-driven actuation system for wearable robotic hand applications," *Smart Materials and Structures*, vol. 22, no. 5, p. 055018, 2013, doi:10.1088/0964-1726/22/5/055018.
[11] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft robotic grippers," *Advanced Materials*, vol. 30, no. 29, p. 1707035, 2018, doi:10.1002/adma.201707035.
[12] C.-C. Lan and M.-C. Chen, "Evaluation of the characteristics of a shape memory alloy spring actuator," *Sensors and Actuators A: Physical*, vol. 189, pp. 434–441, 2013, doi:10.1016/j.sna.2012.10.031.
[13] K. Otsuka and C. M. Wayman, *Shape Memory Materials*. Cambridge, UK: Cambridge University Press, 1998.
[14] D. K. Soother, J. Daudpoto, and B. S. Chowdhry, "Challenges for practical applications of shape memory alloy actuators," *Materials Research Express*, Jul. 2020. [Online]. Available: https://doi.org/10.1088/2053-1591/aba403
[15] S. H. Liu, T. S. Huang, and J. Y. Yen, "Tracking control of shape-memory-alloy actuators based on self-sensing feedback and inverse hysteresis compensation," *Sensors (Basel)*, vol. 10, no. 1, pp. 112–127, Jan. 2010, doi: 10.3390/s100100112.
[16] R. Pillai, G. Murali, and M. Gopal, "Modeling and simulation of a shape memory alloy spring actuated flexible parallel manipulator," *Procedia Computer Science*, vol. 133, pp. 895–904, 2018, doi: 10.1016/j.procs.2018.07.104.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Benchmarking Deep Neural Networks for Lung Nodule Classification in LUNA25

M. Taha Gökcan
Department of Computer Engineering
Yıldız Technical University,
Istanbul, Türkiye
taha.gokcan@std.yildiz.edu.tr

Songül Varlı
Department of Computer Engineering
Yıldız Technical University,
Istanbul, Türkiye
svarli@yildiz.edu.tr

***Abstract*—Early and accurate classification of pulmonary nodules as benign or malignant is critical for improving lung cancer survival rates, particularly when tumors are detected at an asymptomatic stage. In this study, we present a comprehensive benchmark of 2D and 3D deep learning models for malignancy risk estimation on the LUNA25 dataset. We explore a wide range of architectures, from classical CNNs like ResNet to modern models such as ConvNeXt. Furthermore, we assess the impact of Focal Loss versus Binary Cross-Entropy and propose a custom 3D ResNet model that outperforms other models. Our findings highlight that simple 3D architectures, when carefully optimized, offer significant gains in performance.**

***Keywords*—Lung CT, Pulmonary Nodule, LUNA25, Focal Loss**

I. Introduction

Lung cancer is among the leading cancers worldwide in terms of morbidity and mortality. According to the 2020 Global Cancer Statistics, it accounts for approximately 18% of all cancer cases and is associated with 1.8 million deaths globally [1]. While the five-year survival rate for locally advanced disease remains below 5%, detecting the tumor when it is small and asymptomatic can raise the five-year survival rate above 60% [2]. Therefore, early diagnosis and timely treatment are critical for improving patient survival. The primary early indicator of lung cancer is the pulmonary nodule. The complex morphology of pulmonary nodules, their variety in size and shape (ranging from 3 mm to 30 mm in diameter), and their location surrounded by background tissue make the automatic classification of these nodules a challenging task. Early diagnosis relies largely on analyzing these nodules in computed tomography (CT) images.

Accurate classification of pulmonary nodules on CT images as benign or malignant is essential for assessing lung health. However, slice-by-slice inspection of CT data, the complexity and multivariate nature of diagnostic criteria, and the fact that some benign and malignant nodules share similar visual characteristics make this task challenging for radiologists. The constantly increasing volume of medical imaging in hospitals further elevates workload, thus increasing the risk of errors. To address these challenges and provide radiologists with an objective decision-support tool, the development of computer-aided diagnosis (CAD) systems has great importance.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

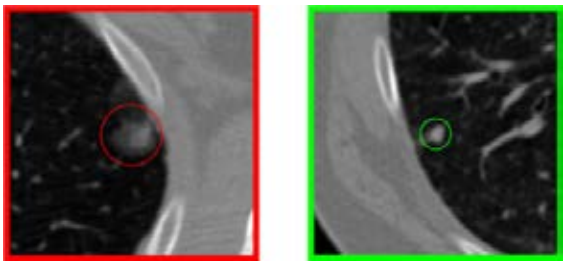


Fig. 1: A malignant and benign nodule from LUNA25 Dataset

In pulmonary nodule classification, convolutional neural networks (CNNs) remain the most widely adopted deep-learning approach because they learn hierarchical features directly from CT data. Both 2D CNNs and 3D CNNs such as ResNet, DenseNet GoogleNet, Xception are used, while attention mechanisms and Vision Transformers (ViTs) have recently gained popularity for modeling long-range context [1].

Some approaches proposed in recent years will be explored. To counter data scarcity and class imbalance, MSCS-DeepLN [3] ensembles lightweight ResNet over multi-scale inputs. To mitigate imbalance, it incorporates an AUC-based loss function as a penalty term. 3D SAACNet [2] is designed to use volumetric spatial features for benign/malignant classification by utilizing asymmetric convolutions within the segmentation-attention network and then combining deep features through a gradient boosting machine. FocalNeXt [4] merges ConvNeXt-style feature extraction with focal-modulation attention from FocalNet, targeting efficient global-context modeling in high-resolution CT volumes. Using a widened inverted-ResNet backbone augmented with 3D Global Coordinate Attention, GC-WIR [5] aims to improve accuracy with relatively few parameters and stable training dynamics. The PARE [6] model mimics radiologist reasoning by parsing spatial context and recalling learned prototypes. It utilizes stacked Vision Transformer (ViT) blocks. A U-Net based segmentation module parses the nodule's spatial context, while a prototype recall module stores learned cases as references for comparative analysis. Combining multiple self-attention heads with a 3D RTConvBlock to capture local–global dependencies while

TABLE I: Comparison of Focal Loss and Binary Cross-Entropy in Custom 3D ResNet Model

Fold	AUC	Sens@Spec	Spec@Sens	Malignant F1	Benign F1
1	0.837	0.336	0.343	0.428	0.862
2	0.871	0.446	0.472	0.386	0.894
3	0.861	0.471	0.358	0.380	0.870
4	0.874	0.367	0.468	0.447	0.879
5	0.868	0.414	0.548	0.445	0.907
BCE Loss Avg.	0.862	0.407	0.438	**0.417**	**0.882**
1	0.846	0.367	0.476	0.452	0.888
2	0.886	0.500	0.637	0.381	0.893
3	0.865	0.471	0.416	0.353	0.836
4	0.881	0.458	0.589	0.444	0.879
5	0.881	0.360	0.619	0.447	0.906
Focal Loss Avg.	**0.872**	**0.431**	**0.547**	0.416	0.880

keeping computation practical, MSA-Net [7] is a recent transformer-based alternative for 3D nodule volumes.

In this context, the LUNA25 Challenge [8] focuses on classifying pulmonary nodules from low-dose chest CT scans as benign or malignant. In addition, a reader study is planned on an isolated data set to compare the relative performance of human experts and artificial intelligence models. The aim of the competition is to strengthen the potential for concrete contributions to clinical decision support workflows by improving the performance gains achieved in various architectures.

The LUNA25 challenge establishes standardized 2D and 3D baselines for the task of malignancy-risk estimation. The 3D component of the challenge utilizes an inflated-3D (I3D) architecture based on I3DR-Net [9], a model designed for nodule detection. While, a 2D ResNet18 [10] model is employed for the 2D stream.

II. Materials and Methods

A. Dataset

The primary dataset was LUNA25 from the National Lung Screening Trial (NLST), containing 2,120 patients and 4,069 low-dose chest CT studies, with 555 malignant and 5,608 benign nodules. We employed 5-fold cross-validation with patient-level partitioning to prevent leakage, meaning that all nodules from a given patient were confined to a single fold. Except for the test fold, other folds allocated 80% of cases to training and 20% to validation. Two nodules, one benign and one malignant, are shown in the Figure 1.

Nodules were cropped into 64×64×64 voxel cubes, clipped to [-1000, 400] Hounsfield Units, rescaled to the 0–1 range, and augmented with random rotations and translations. To overcome data imbalance, malignant nodules were sampled more frequently during training.

B. Experimental Setup

Training was conducted for up to 50 epochs with a batch size of 32. We used the Adam optimizer with a weight decay of 0.0005. Early stopping terminated training if the validation AUC did not improve for 10 consecutive epochs. As loss function, Binary Cross-Entropy (BCE) and Focal loss [11] were examined. Focal Loss improves classification performance in imbalanced datasets by down-weighting easy negatives and forcing the model to focus on harder, informative

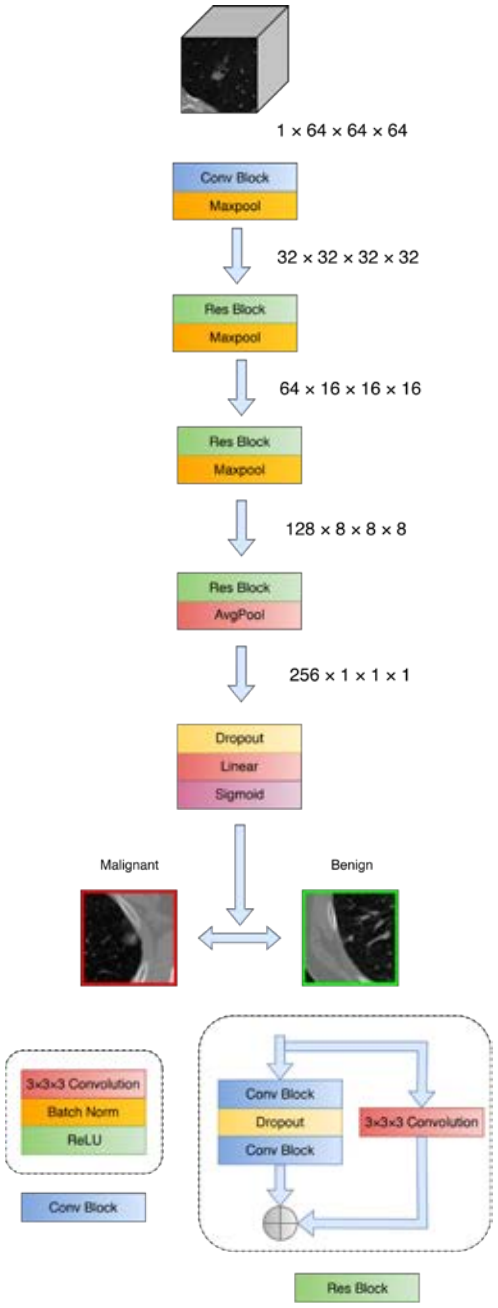


Fig. 2: Custom 3D ResNet Model

TABLE II: Results of 2D Models

Model	# Params	AUC	Sens@Spec	Spec@Sens	Malignant F1	Benign F1
ResNet18	11.2	0.807	0.276	0.379	0.356	0.877
ResNet34	21.3	0.810	0.328	0.366	0.332	0.846
ResNet50	23.5	0.808	0.261	0.375	0.326	0.822
EfficientNet-B0	4.0	0.826	0.295	0.417	0.365	0.858
EfficientNet-B1	6.5	0.818	0.308	0.353	0.348	0.838
EfficientNet-B2	7.7	0.810	0.299	0.382	0.327	0.826
DenseNet121	7.5	0.790	0.305	0.290	0.347	0.885
EfficientNet-V2-S	20.2	0.826	0.330	0.380	0.360	0.857
MobileNet-V3-Small	1.5	0.809	0.288	0.362	0.358	0.848
ConvNeXt-Tiny	27.8	**0.858**	**0.406**	**0.505**	**0.411**	**0.901**
ConvNeXt-Small	49.5	0.759	0.192	0.280	0.303	0.836
ConvNeXt-V2-Tiny	27.8	0.828	0.287	0.467	0.339	0.835

examples. Since Focal loss was the most effective, it was utilized in all setup. The initial learning rate was 0.0001 and was annealed to 0.00001 using a cosine scheduler.

Our ResNet-based model begins with a convolutional block comprising a 3×3×3 convolution, Batch Normalization, and ReLU. This is followed by three stacked ResNet blocks. Each block contains two convolutional Conv Blocks separated by Dropout (0.2), wrapped by a residual skip connection. When channel dimensions differ, the skip path uses a 3×3×3 convolution, and outputs are combined by element-wise addition. Spatial resolution is progressively reduced via 2×2×2 max-pooling. The final feature map is compressed using adaptive average pooling, after which a Dropout–Linear–Sigmoid head produces the benign/malignant probability.The proposed model can be seen in the Figure 2.

Beyond our model, we benchmarked a range of off-the-shelf encoders in both 2D and 3D. For 2D, ResNet-18 [10] served as the official LUNA25 baseline. We also evaluated other ResNet variants, EfficientNet [12], EfficientNet-V2 [13], DenseNet [14], MobileNet-V3 [15], ConvNeXt [16], ConvNeXt-V2 [17], and Xception [18]. For 3D, beside the I3D baseline model, we assessed ResNet, EfficientNet, ConvNeXt, ConvNeXt-V2, and Xception [18]. The 3D versions were obtained by replacing 2D convolution and normalization layers with their 3D versions while preserving the original configurations as closely as possible.

C. Evaluation Metrics

To assess the performance of each model in predicting lung nodule malignancy, we employ the following evaluation metrics:

- **Area Under the ROC Curve (AUC):** Measures the likelihood of correctly ranking positive samples higher than negatives.
- **Sensitivity at 95% Specificity (Sens@Spec):** Represents the true positive rate (sensitivity) when the model is constrained to a fixed false positive rate of 5%.
- **Specificity at 95% Sensitivity (Spec@Sens):** Indicates the true negative rate (specificity) when sensitivity is maintained at 95%.
- **F1 Score:** The harmonic mean of precision and recall, offering a single measure that balances both false positives and false negatives.

For all employed 2D and 3D architectures, the final performance metrics are reported as the average results obtained from 5-fold cross-validation.

III. Results and Discussion

To evaluate models performance, 5-Fold cross-validation was employed. The final metrics were obtained by averaging the results across all folds based on the performance metrics introduced in the previous section.

Table I presents a detailed comparison between Focal Loss and Binary Cross-Entropy (BCE) loss, along with the corresponding results for each fold. As can be seen from the table, Focal Loss consistently achieved higher AUC scores across all folds and yielded nearly a 1-point improvement in the overall AUC compared to BCE. Importantly, it maintained comparable F1 scores, indicating that its enhancement in discriminative ability has not resulted in any loss. Moreover, the notable improvement in the Spec@Sens metric demonstrates the model's strengthened ability to make high-confidence predictions.

Performance results for 2D models are summarized in Table II. To identify optimal model complexity, various configurations of ResNet and EfficientNet architectures were evaluated. The baseline model ResNet-18 achieved an AUC of 0.807, establishing an initial benchmark. While more complex ResNet variants did not show substantial gains in AUC, their F1 scores tended to decline, suggesting a degree of overfitting. EfficientNet architectures exhibited a trend in which increasing parameter counts led to diminished performance. EfficientNet-B0, the smallest variant, achieved the best AUC score among its group (0.826). Notably, EfficientNet-V2-S, despite having approximately five times more parameters than EfficientNet-B0, failed to outperform it. MobileNet-V3-Small delivered competitive performance with very low complexity, achieving results comparable to ResNet-18, highlighting its suitability for lightweight deployment. In contrast, DenseNet-121 yielded unsatisfactory outcomes in this setting. Among ConvNeXt models, the lower-complexity variant outperformed its deeper counterparts, with ConvNeXt-Tiny reaching an the best score of 0.858 AUC. However, ConvNeXt-V2-Tiny did not show further improvement.

Motivated by the effectiveness of lightweight 2D models, their 3D counterparts were subsequently explored. Results are presented in Table III. The I3D baseline model achieved

TABLE III: Results of 3D Models

Model	# Params	AUC	Sens@Spec	Spec@Sens	Malignant F1	Benign F1
I3D	12.2	0.811	0.334	0.330	0.393	0.911
CNN	1.2	0.811	0.324	0.348	0.357	0.853
ResNet	33.2	0.821	0.317	0.375	0.376	**0.916**
EfficientNet-B0	4.7	0.833	0.342	0.354	0.383	0.882
Xception	22.4	0.828	0.338	0.364	0.386	0.889
ConvNeXt-Tiny	31.3	0.772	0.236	0.248	0.312	0.812
ConvNeXtV2-Atto	4.2	0.835	0.302	0.451	0.377	0.869
Custom ResNet	4.6	**0.872**	**0.431**	**0.547**	**0.416**	0.880

an AUC of 0.811. In the CNN architecture, where standard convolutional blocks replaced the ResNet blocks, performance remained comparable in terms of AUC, though F1 scores were quite lower. The 3D ResNet variant improved the AUC by 1.4% compared to 2D ResNet18, while EfficientNet-B0's 3D version showed a 0.6% gain, both supporting the advantage of volumetric feature extraction.

Interestingly, despite its strong 2D performance, the 3D version of ConvNeXt-Tiny performed poorly. In contrast, ConvNeXt-V2-Atto, a simpler variant, delivered significantly better results in 3D, suggesting that reduced complexity may benefit learning. Xception also outperformed the ResNet model.

The proposed Custom ResNet model, achieved the highest AUC score of 0.872, outperforming the baseline by nearly 6% and the second-best model (ConvNeXt-V2-Atto) by approximately 4%. Furthermore, improvements in both Sens@Spec and Spec@Sens by around 10% indicate that the model is particularly reliable when operating under high specificity or sensitivity constraints. Although some models demonstrated higher F1 scores for benign cases, likely due to overfitting, the malignant F1 score of the Custom model was notably superior. It is also worth noting that the best 2D ConvNeXt-Tiny outperformed most of the 3D models.

IV. Conclusion

This study provides a systematic evaluation of 2D and 3D deep learning models for lung nodule malignancy classification using the LUNA25 dataset. Our experiments demonstrate that ConvNeXt-Tiny 2D architecture achieve strong baseline performance, while deeper models are generally prone to overfitting. When extended to 3D, compact models continue to outperform their 2D versions.We show that using Focal Loss in place of standard Binary Cross-Entropy leads to consistent improvements in AUC and more confident predictions.

Our proposed Custom 3D ResNet model achieves the best overall performance, surpassing other architectures. These findings validate the effectiveness of architecture optimization and loss function design in building reliable, clinically useful AI tools for early lung cancer detection.

References

[1] A. Naik and D. R. Edla, "Lung nodule classification on computed tomography images using deep learning," *Wireless personal communications*, vol. 116, no. 1, pp. 655–690, 2021.

[2] Z. Guo, J. Yang, L. Zhao, J. Yuan, and H. Yu, "3d saacnet with gbm for the classification of benign and malignant lung nodules," *Computers in Biology and Medicine*, vol. 153, p. 106532, 2023.

[3] X. Xu, C. Wang, J. Guo, Y. Gan, J. Wang, H. Bai, L. Zhang, W. Li, and Z. Yi, "Mscs-deepln: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks," *Medical Image Analysis*, vol. 65, p. 101772, 2020.

[4] T. Gulsoy and E. B. Kablan, "Focalnext: A convnext augmented focalnet architecture for lung cancer classification from ct-scan images," *Expert Systems With Applications*, vol. 261, p. 125553, 2025.

[5] W. Wang, S. Yin, F. Ye, Y. Chen, L. Zhu, and H. Yu, "Gc-wir: 3d global coordinate attention wide inverted resnet network for pulmonary nodules classification," *BMC Pulmonary Medicine*, vol. 24, no. 1, p. 465, 2024.

[6] J. Zhang, X. Ye, J. Zhang, Y. Tang, M. Xu, J. Guo, X. Chen, Z. Liu, J. Zhou, L. Lu *et al.*, "Parse and recall: Towards accurate lung nodule malignancy prediction like radiologists," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 199–209.

[7] J. Pan, L. Liang, P. Sun, Y. Liang, J. Zhu, and Z. Chen, "Msa-net: multiple self-attention mechanism for 3d lung nodule classification in ct images," *BMC Medical Imaging*, vol. 25, no. 1, p. 193, 2025.

[8] D. Peeters, B. Obreja, N. Antonissen, C. Jacobs *et al.*, "The luna25 challenge: Public dataset and reader-study for lung nodule malignancy risk estimation," https://luna25.grand-challenge.org/, LUNA25 Grand Challenge, 2025, [Online]. Accessed: Jun. 2025.

[9] I. W. Harsono, S. Liawatimena, and T. W. Cenggoro, "Lung nodule detection and classification from thorax ct-scan using retinanet with transfer learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 3, pp. 567–577, 2022.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[12] M. Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, pp. 6105–6114, 2019.

[13] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[15] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.

[16] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

[17] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 16 133–16 142.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

Design and Implementation of a Phonocardiography System Combining ECG and Stethoscope Features

Ziya Telatar
Biomedical Engineering, Baskent University,
ztelatar@baskent.edu.tr

Onur Koçak
Computer & Biomedical Engineering, Baskent University,
okocak@baskent.edu.tr

Cansel Ficici
Electrical and Electronics Engineering, Ankara University,
ogretmenoglu@ankara.edu.tr

Abstract*— Cardiovascular diseases (CVDs) remain the leading cause of global mortality, underscoring the need for cost-effective and reliable diagnostic and monitoring systems. Conventional stethoscopes and single-lead electrocardiogram (ECG) devices, still prevalent in clinical settings, lack digital storage and analytical capabilities, thereby limiting both diagnostic accuracy and educational value. In this study, we present the design and implementation of an integrated phonocardiography (PCG)–ECG system that combines the functional features of a stethoscope and an ECG within a single portable device. The system employs an ESP32-WROOM-32 microcontroller for synchronous acquisition of cardiac sound and ECG signals at 1 kHz, incorporating front-end microphone modules, an AD8232 analog front end, and a custom battery management subsystem. Signal conditioning was achieved through band-pass and notch filtering, with subsequent R-peak detection and S1–S2 heart sound identification. Data were transmitted to a remote server via a REST API, enabling secure storage, labeling, and server-side analysis. Experimental results demonstrated high-fidelity detection of S1 and S2 heart sounds from the PCG signal and robust R-peak detection from the ECG signal, with strong temporal alignment between electromechanical cardiac events. These findings confirm the feasibility of the proposed device for real-time monitoring, clinical education, and potential telemedicine applications, offering a locally developed and export-capable solution to current gaps in digital stethoscope technology.

Keywords—Phonocardiography (PCG), Electrocardiogram (ECG), Biomedical Signal Processing, Digital Stethoscope, Wearable Health Monitoring, IoT-based Medical Devices

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain among the leading causes of mortality worldwide, accounting for approximately 18 million deaths annually, according to data from the World Health Organization (WHO) [1]. In 2017, an estimated 126.5 million individuals were living with heart disease, and this number is projected to rise substantially, with annual deaths expected to reach 23 million by 2030 [2]. CVD prevalence is particularly high in regions such as the United States, Central Europe, North Africa, and the Middle East, posing a significant public health challenge that demands both urgent medical intervention and continuous patient monitoring [3]. In Türkiye, factors such as demographic growth, increased health tourism, and improvements in healthcare accessibility have placed considerable strain on cardiology clinics and emergency departments, creating challenges in resource allocation and service delivery [4].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Despite significant advances in healthcare technology, many existing diagnostic and monitoring systems still exhibit critical limitations. These include the inability to effectively utilize multiple data types simultaneously and the lack of advanced algorithms capable of interpreting complex signals that integrate mechanical and electrical cardiac activity [5]. Conventional stethoscopes and electrocardiogram (ECG) devices, still widely used in clinical practice, are predominantly analog and typically lack data storage capability. As a result, valuable clinical information is often lost, as these devices rely heavily on subjective interpretation by healthcare professionals. This limitation not only affects clinical diagnostics but also hinders medical education, where the ability to review and analyze recorded data is essential [6].

In response to these challenges, recent research has emphasized the development of innovative systems integrating phonocardiography (PCG) and ECG measurements into a single platform. Such systems enable the simultaneous monitoring of mechanical heart sounds and electrical cardiac activity, generating digital data that can be stored, analyzed, and utilized for both real-time clinical decision-making and educational purposes. These advancements have the potential to improve rapid diagnosis, continuous patient monitoring, data management, and the quality of training provided to medical students and healthcare professionals [7-8].

A notable example is a 2025 study introducing a portable, integrated PCG–ECG system that demonstrated significant improvements in signal-to-noise ratio (SNR) and enhanced extraction of electromechanical parameters critical for cardiac health assessment. This device represents a step forward in wearable cardiac health monitoring, combining high-fidelity data acquisition with user-friendly portability [9]. Furthermore, these developments align with national strategies to strengthen healthcare infrastructure through domestic innovation. The proposed integrated system offers a unique, locally produced solution with high export potential, addressing gaps in current digital stethoscope technologies and contributing significantly to Türkiye's healthcare capabilities [9, 10, 11].

Current study began with research analyzing the performance of various digital and analog microphones used in medical applications. The scope then expanded to include the acquisition of single-channel electrocardiogram (ECG) signals. A circuit model was developed to synchronously acquire both signals at a sampling frequency of 1 kHz. Using an ESP32

Wroom-32 microcontroller, basic signal processing algorithms were applied to the acquired signals to produce clean, noise-reduced data. Subsequently, R-peak detection was performed on the ECG signal, followed by the acquisition of S1, S2, S3, and S4 heart sounds. Through the developed interface, these signals can be processed and statistically evaluated, with the aim of enabling their use in the detection of specific cardiological conditions.

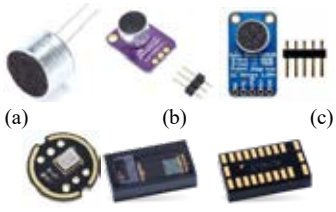
This study, unlike previous works in the literature, integrates a stethoscope with a three-channel ECG device to enable the simultaneous acquisition of both signal and sound at the same sampling frequency. In this way, it is demonstrated that clinical data collection in experimental studies can be facilitated with the help of a lower-cost device. Furthermore, the development stages of a cost-effective design using modern electronic systems are also presented in this study.

II. MATERIALS AND METHODS

A. *Microphone Configurations Integrated into the Device Hardware*

A microphone transduces ambient acoustic pressure into an electrical signal. The preamplifier conditions this signal and feeds it to the microcontroller, whose digital output reflects amplitude variations of the surrounding sound; higher digital values indicate higher sound pressure, and lower values indicate lower sound pressure.

To characterize front-end options, we procured and evaluated multiple microphone types commonly used in medical applications, including MAX-series modules, INMP-series digital MEMS microphones, analog electret capsules, MEMS devices, and SMD-package microphones employed in industrial medical equipment. The signal performance of each was examined. The analog electret capsule is shown in Fig. 1(a). The MAX4466 pre-amplified microphone module (Fig. 1(b))—selected based on datasheet sensitivity and noise specifications—has been widely used in prior medical prototypes. The MAX9814 module with AGC is shown in Fig. 1(c). A 12-bit microcontroller ADC was programmed to acquire analog outputs from these modules, and recordings were obtained in both quiet and noisy environments. The INMP441 digital MEMS microphone (Fig. 1(d)) was interfaced via the microcontroller's I²S peripheral. Figures 1(e)–1(f) depict an SMD digital microphone IC (MAX81650) of the type also found in industrial medical devices (e.g., Masimo and Philips systems). Sampling rates of 800 Hz and 1 kHz were employed for preliminary trials.



(a) (b) (c)

(d) (e) (f)

Figure 1. Microphone Integrated Circuit Models

B. *Cardiac Sound Bandwidth and Sampling Method*

Typical phonocardiographic frequency content is concentrated in the low-frequency range , S1: 20–45 Hz, S2: 50–70 Hz and S4: < 20 Hz

By the Nyquist criterion, the sampling frequency must satisfy $f_s \geq 2f_{max}$, with $f_{max} = 70$ Hz, the minimum is $f_s \geq 140$ Hz. For clinical studies, a practical margin of 4x or 5x the maximum frequency is preferred. In practice: 500 Hz adequately captures S1, S2, and S4, 1,000 Hz yields clearer clinical recordings, and 2,000 Hz allows detection of higher-frequency phenomena (e.g., murmurs).

During stethoscope-based acquisition, the desired signal is often contaminated by hand-borne vibrations and ambient acoustic noise. To mitigate these effects, signals were filtered using a combination of band-pass and notch responses. Specifically, the filter configuration passes 0–48 Hz and 52–98 Hz (implemented in 4 Hz passband segments) to suppress the 50 Hz mains component while preserving cardiac sound content.

C. *Single-Lead ECG Acquisition and Microprocessor*

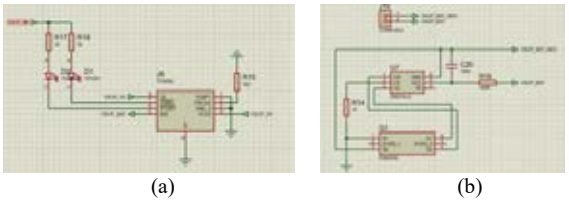
ECG signals were acquired in a single-lead configuration using the AD8232 analog front end, and R-peak detection was implemented for heart rate estimation. A compact PCB layout was designed by integrating the AD8232 with the ESP32-WROOM-32U module. While initial data transfer was performed via USB serial communication, the design objective was to enable wireless connectivity by linking the device to a local network. The ESP32-WROOM-32U supports both Wi-Fi and Bluetooth RF interfaces; to ensure reliable transmission and avoid data loss, the Transmission Control Protocol (TCP) was selected at the transport layer for Wi-Fi communication.



Figure 2. AD8232 Single-Lead ECG Sensor

D. *Battery Management System*

Following signal acquisition studies, the battery control and charging subsystem was implemented. The TP4056 integrated circuit was employed as the charge controller, ensuring proper charging of the battery with the correct voltage and current levels. The TP4056 is one of the most widely used ICs for single-cell battery systems in commercial applications.



(a) (b)

Figure 3. (a) Battery Charge Circuit , (b) Battery Protection Circuit

For battery protection, the DW01A integrated circuit was selected. Specifically designed for Li-ion cell protection, this IC functions primarily as a voltage comparator. When the battery voltage drops to a level that may cause damage, the DW01A disables its output, thereby preventing current draw from the circuit. Current interruption at the output is achieved through the integrated FS8205A dual MOSFET.

E. Data Base Management

Data recording operations were implemented in Kotlin and can be performed at varying time intervals. During each session, the device buffers sensor data locally and, upon completion, transmits the data to a remote server via a REST API. Each record is stored with mandatory patient information entered through the user interface, along with a timestamp of the recording date, and is displayed accordingly. When previously recorded data are accessed, the user can trigger server-side analysis by selecting the *Analyze* option within the application. The resulting analysis files are stored with an appended "-analyzed" suffix and timestamp, and results are immediately displayed to the user. The same file can be revisited at later times, with the analysis data preserved. The analysis process includes labeling of S1–S2 heart sounds and ECG R-waves sequentially within the dataset, after which results are transmitted back to the Kotlin application. All communication through the REST API is carried out using JSON objects. On the server side, JSON-formatted records are stored on a dedicated secondary virtual disk provisioned specifically for data storage.

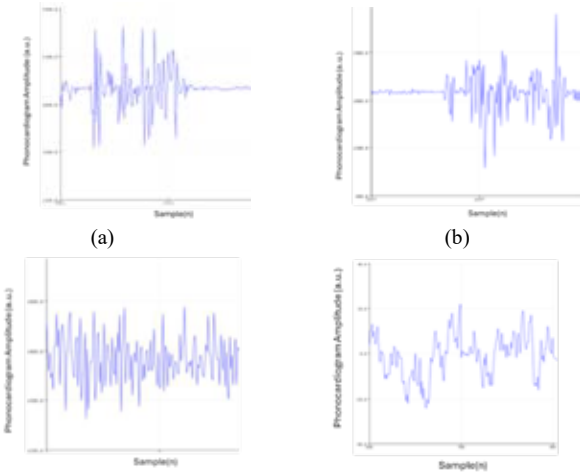
III. RESULT AND DISCUSSION

An electret microphone was selected due to its analog output capability. Measurements performed with amplifier circuits commonly used in MAX-series microphone modules are shown in Figure 4. These analog values were digitized by the microcontroller and subsequently converted into voltage levels. The analog-to-digital conversion (ADC) outputs were transformed to corresponding voltage amplitudes using the formula provided in Equation (1), where the reference voltage of the ESP32-WROOM-32 microcontroller is specified.

$$V_{Out} = \left(\frac{ADC\,Value}{2^{12}-1}\right) \cdot V_{ref} \quad (1)$$

To enhance the quality of the acquired acoustic signals and suppress noise, an operational amplifier (op-amp)–based filtering stage was proposed. Low-pass and high-pass filters were designed to remove unwanted frequency components: the low-pass filter attenuates high-frequency noise, while the high-pass filter suppresses low-frequency interference.

Some recordings contain PCG2, AUX1, and AUX2 channels, which generally exhibit signals of very low amplitude—at or near the quantization noise level. However, when these channels are audibly monitored and visually inspected, they are observed to capture weak background sounds and electronic device noise present in the environment.



(a) (b) (c) (d)

Figure 4. Acoustic signal acquired using an analog microphone (a), signal acquired using the MAX4466 microphone (b), noise signal acquired in a quiet environment using the MAX9814 microphone (c), and noise signal acquired in a quiet environment using the INMP441 microphone (d).

Although not directly usable for primary analysis, these channels may still be of interest to researchers working in signal processing. Specifically, they can serve as auxiliary data for the design of adaptive noise cancellation systems or multi-channel blind/partially blind source separation algorithms.

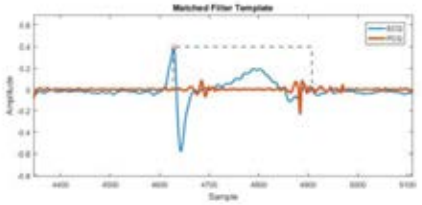
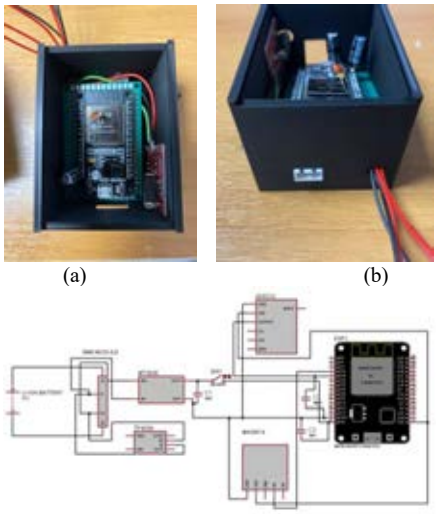


Figure 5. Matched filter template selection based on R-peak position.



(a) (b) (c)

Figure 6. Placement of the controller, phonocardiography module, and single-lead ECG system (a); battery slot positioned on the system (b); and block diagram of the complete system (c).

The developed integrated PCG–ECG monitoring system was successfully assembled and tested. The electronic components were enclosed within a custom-designed housing, ensuring proper placement of all modules and protection of the system.

Internal wiring and system integration were carried out in accordance with the design schematic, in which the ESP32 development board was employed as the central processing unit.

The power supply was provided by three 3.7 V Li-Po batteries connected through a BMS module, enabling portability and stable operation. Following assembly, the system was tested using ECG and heart sound measurements. The electronic wiring diagram was followed to ensure correct connections and maintain stable sensor performance. The test results demonstrated that S1 and S2 heart sounds were accurately detected from the PCG signal, and R-peaks were reliably identified from the ECG signal. Figure 7 presents the block representation of the system along with sample signals acquired from the device.



Figure 7. Critical System Design

The acquired waveforms demonstrated high signal integrity and temporal alignment between mechanical (PCG) and electrical (ECG) cardiac events, thereby confirming the system's real-time monitoring capability.

In Figure 8 simultaneously acquired phonocardiogram and ECG signals are presented. The R-peaks, along with the S1 and S2 heart sounds, are clearly identified on the synchronized waveforms.



Figure 8. Synchronously acquired heart sounds and ECG signals

IV. CONCLUSION AND FUTURE WORKS

Single-channel ECG and heart sound signals, synchronously acquired at a 1 kHz sampling frequency, were tested on different subjects to observe the performance of the study. The device's LVD and EMC test performance was also tested and verified by an independent testing institution. At this stage, the interface and database performance tests will be monitored together with clinical trials. With the data to be obtained thereafter, studies on automated prediction and disease diagnosis will be scientifically investigated. It should be noted that the AD8232 sensor used in this study provides research-grade ECG results rather than clinical-grade accuracy. Therefore, in future works, alternative acquisition methods and advanced sensor technologies will be considered to improve signal quality and reliability for clinical applications.

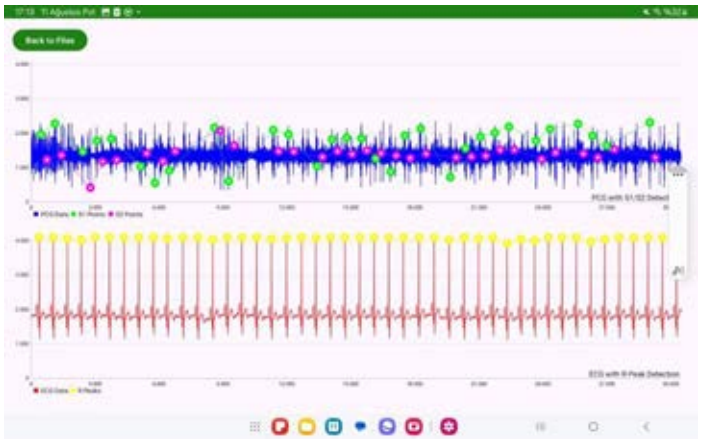


Figure 9. Other Participant Scores

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO Fact Sheets, Jun. 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] A. Aminorroaya, et al., "Global, regional, and national quality of care of ischaemic heart disease from 1990 to 2017: A systematic analysis for the Global Burden of Disease Study 2017," Eur. J. Prev. Cardiol., vol. 29, no. 2, pp. 371–379, Mar. 2022, doi: 10.1093/eurjpc/zwab066.
[3] M. Vaduganathan, et al., "The global burden of cardiovascular diseases and risk," J. Am. Coll. Cardiol., vol. 80, no. 25, pp. 2361–2371, Dec. 2022, doi: 10.1016/j.jacc.2022.11.005.
[4] C. Yıldız, et al., "Characteristics of patients attending a cardiology outpatient clinic: A focus on the Turkish healthcare system," Int. J. Cardiovasc. Acad., 2024.
[5] S. Karimi, A. J. Shah, G. D. Clifford, and R. Sameni, "Cross-learning between ECG and PCG: Exploring common and exclusive characteristics of bimodal electromechanical cardiac waveforms," arXiv, Jun. 11, 2025. [Online]. Available: https://arxiv.org/abs/2506.10212
. [Accessed: Aug. 11, 2025].
[6] Türkiye İstatistik Kurumu (TÜİK), "Sağlık istatistikleri," Ankara, Türkiye, 2022. [Online]. Available: https://data.tuik.gov.tr
[7] A. Watkins, et al., "Digital stethoscopes in clinical practice: Technology update," Med. Devices: Evid. Res., vol. 12, pp. 181–191, 2019.
[8] M. H. Mehdi, et al., "Phonocardiography and ECG integration for cardiac diagnosis," Biomed. Signal Process. Control, vol. 55, p. 101627, 2020.
[9] S. Yılmaz, et al., "Design of a local integrated phonocardiography and ECG measurement system for education and clinical use," Turk. J. Electr. Eng. Comput. Sci., vol. 30, no. 2, pp. 345–358, 2022.
[10] A. H. Ertugrul, X. Zhang, Y. Wang, and L. Chen, "A novel wearable device integrating ECG and PCG for cardiac health monitoring," Microsyst. Nanoeng., vol. 11, no. 1, pp. 1–12, Jan. 2025. [Online]. Available: https://www.nature.com/articles/s41378-024-00858-3
[11] C.-H. Lin, Y.-H. Lin, C.-C. Tsai, and Y.-C. Chen, "Electrocardiogram and phonocardiogram monitoring system for cardiac auscultation," IEEE Trans. Biomed. Circuits Syst., vol. 13, no. 6, pp. 1635–1644, Dec. 2019, doi: 10.1109/TBCAS.2019.2953064.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Development of Proprioceptive and Vibrotactile Training Modes for BALANSens Platform

Emad-ud-din SHEIKH
Living Robotics Laboratory,
Electrical, Electronics and Cyber Systems Engineering,
Istanbul Medipol University, Istanbul, Türkiye

Hafiz Muhammad BILAL
Living Robotics Laboratory,
Electrical, Electronics and Cyber Systems Engineering,
Istanbul Medipol University, Istanbul, Türkiye

Pınar KAYA
Physiotherapy and Rehabilitation Department
Istanbul Medipol University
Istanbul, Türkiye

Elif HOCAOGLU
Living Robotics Laboratory
School of Engineering and Natural Science
Istanbul Medipol University, Istanbul, Türkiye
Research Institute for Health Sciences and Technologies (SABITA)
Email: ehocaoglu@medipol.edu.tr

***Abstract*—Recent studies have shown that training plantar tactile perception (vibrotactile sensation) and ankle joint position awareness (proprioceptive sensation) can substantially improve balance and gait rehabilitation. Building on these findings, we developed two training modes for the BALANSens robotic platform, a 3-RPS lower-limb rehabilitation manipulator designed in alignment with the Integrated Balance Rehabilitation (I-BaR) framework. In the proprioceptive training mode, the robot guides the subject's ankle to a target position and back, after which the subject reproduces the movement without visual feedback. In the vibrotactile training mode, one of the motors stimulating the plantar surface is randomly activated, and the subject identifies the stimulated region. This paper presents the design and functional evaluation of these training modes with healthy subjects. Results indicate that angular deviation (with and without VR) in proprioceptive training, along with accuracy and response times in vibrotactile training, serve as reliable quantitative performance metrics. Clinical trials are planned to further solidify this claim of potential of sensory feedback–based interventions for rehabilitation.**

***Index Terms*—Proprioception, multi-modal sensory feedback, virtual reality environment, robotic rehabilitation**

I. Introduction

Postural stability and walking balance rely heavily on accurate sensory feedback from the feet and ankle proprioception. Among chronic stroke patients, ankle proprioception is among the strongest predictors of balance impairment [1]. The cutaneous afferent system of the plantar surface provides critical information for maintaining upright stance and coordinating gait, influencing both static and dynamic stability. Disruption of this sensory input, whether due to neurological injury, ageing, or disease (such as Parkinson's or multiple sclerosis), can significantly impair balance control and increase fall risk.

Proprioceptive training, aimed at enhancing the body's awareness of limb position and movement, has been shown through meta-analyses of randomized controlled trials to markedly improve balance performance, gait speed, trunk control, and functional mobility in individuals recovering from stroke [2], [3]. In fact, even a single session of proprioceptive training with feedback has shown to improve movement accuracy and to induce functional connectivity changes in the brains of chronic stroke patients [4].

Among various sensory enhancement strategies, vibrotactile stimulation has emerged as a promising method for augmenting feedback during rehabilitation. Systematic reviews indicate that vibrotactile cues, delivered to areas including the plantar surface, can enhance both static and dynamic postural performance by supplementing the central nervous system's sensory input; however, more high-quality randomized controlled trials are needed to ascertain efficacy [5].

Alongside these sensory-based interventions, robotic rehabilitation systems have gained popularity for their ability to deliver precise, repeatable, and adaptable therapy. Lower-limb robotic devices, including those based on 3-DOF parallel manipulators, have been developed to support targeted ankle rehabilitation by enabling controlled multi-axis movement. Within the broader field of rehabilitation robotics, evidence continues to support the integration of robotic platforms into therapy programs to promote neuroplasticity and enhance sensorimotor recovery [6].

Although many systems have investigated the efficacy of these three aspects individually (i.e. proprioceptive [7], plantar vibrotactile and robot-based training), but no system, to the best of our knowledge, has incorporated these modes within a single device. This limits the potential benefits of a multi-modal approach. I-BaR framework [8], [9] provides a structured approach for combining different modes of balance-related rehabilitation in robotic platforms. In addition to the aforementioned modes it also proposes engaging visual feedback in the form of a serious Virtual Reality (VR) game. This has been shown to be significantly effective in improving rehabilitation efficacy among people with Parkinson's Disease [10]. As a result many commercially available rehabilitation

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

robots utilize VR in some form [6]. Examples include Lokomat and Hunova. Addressing these gaps could enable more comprehensive rehabilitation strategies that improve both balance and gait.

In light of the aforementioned discussion, a critical gap persists in the current state of the art: only a limited number of systems incorporate plantar vibrotactile stimulation, ankle proprioceptive training, and virtual reality within a unified robotic platform. Addressing this gap presents a unique opportunity to advance a synergistic rehabilitation paradigm that simultaneously augments sensory feedback and actively engages the neuromuscular system through precise, robot-assisted motor execution. This work introduces such an integrated approach, implemented on the BALANSens platform, which incorporates multiple rehabilitation modalities [11].

II. Methods

A. System Description

The BALANSens robotic platform is a rehabilitation device developed in accordance with the I-BaR framework [8]. It is structured around three sequential stages to provide comprehensive balance and gait rehabilitation: (i) ankle preparation, (ii) balance training, and (iii) stepping rehabilitation.

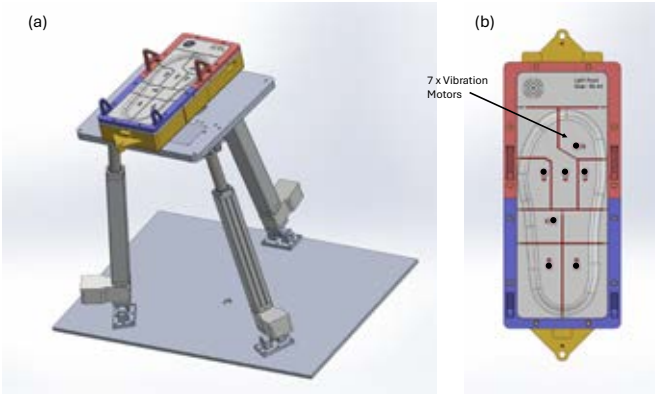


Fig. 1. The BALANSens Robotic Platform (a) For the ankle preparation phase, the 3-RPS manipulator has only one foot assembly attached to the end-effector (b) Top view of the foot assembly

In the ankle preparation phase, the 3-RPS parallel robot is mounted with one foot assembly as shown in Fig. 1a. The detailed design of the robotic platform and the VR system has been published elsewhere [11], [12], but it is important to restate here that the foot platform of the device is divided into 7 regions. Each region has a coin-type vibration motor (5 VDC, 8 mm diameter) mounted on rubber pads in it. This motor is protruding very slightly (1-2 mm) above the surface of the foot platform. This gap ensures that when the subject places his foot on top of the foot platform, the vibration motors will be in direct contact with the desired plantar regions. This construction enables the incorporation of proprioceptive and vibrotactile training into the same product.

B. Proprioceptive Training

The proprioception mode operates in two phases. In the first phase, the robotic platform moves the subject's ankle joint to a predefined position. This target position, entered by the user through the interface (Fig. 2a), is specified by two parameters: the inversion–eversion angle (α) and the dorsiflexion–plantarflexion angle (β). Afterwards, the foot is returned to the neutral position, marking the start of the second phase. In this phase, impedance control is activated, rendering the manipulator effectively "transparent." When forces are applied to any of the seven regions of the foot platform, the corresponding torques in the α (inversion–eversion) and β (dorsiflexion–plantarflexion) directions are computed. These forces are simultaneously visualized on the lower-left corner of the user interface as a blue–yellow–red color map. The manipulator then moves according to the calculated torque, allowing the subject to perceive the platform as following the natural motion of the foot with minimal resistance. The details of the torque computation and impedance control scheme are provided in an earlier publication [13].

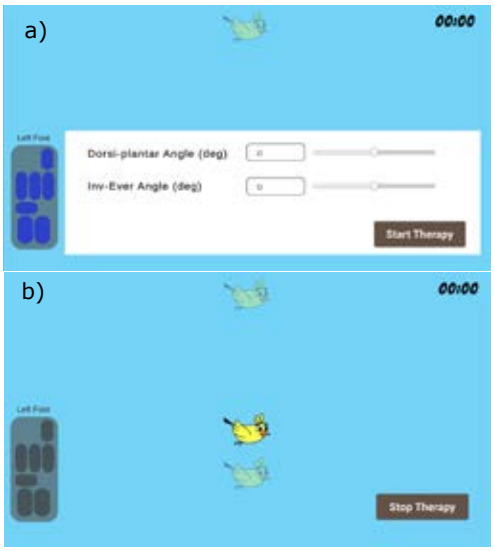


Fig. 2. VR Interface for the Proprioceptive Training Mode. (a) The user can input the desired target position. The input menu is hidden when training starts. (b) A faded image of the bird is shown at the initial zero position and at the 20-degree dorsiflexion position.

The BALANSens platform uses a serious VR game to complement the training. A bird avatar is shown in the game that follows the position of the foot in the 2D plane (Fig. 2); It moves up-down for dorsi-plantar motion and left-right for inversion-eversion motion.

C. Vibrotactile Training

In the vibration sensing mode the goal is to enhance the subject's sense of touch on the sole of his foot. So, for subjects with severely compromised tactile sensation, vibration therapy is provided first. In this therapy all 7 vibration motors turn on at the same time as per the on/off time durations and intensity specified by the user (Fig. 3a). This procedure can enable the

subject to develop new neurological pathways, associating the vibration with the felt sensation.

If the subject's sensation is in an improved state then Vibrotactile training is employed by turning only one vibration motor on at random. The subject then has to identify which vibration motor turned on. This encourages the subject to develop a stronger mental connection with the individual plantar regions. After assessing the sensation the subject must click on the corresponding region on the screen (Fig. 3b). The regions also display the force applied by the subject on that region, in the form of a blue-yellow-red colour map. Then the next vibration motor turns on at random, and the cycle continues until the specified training time has elapsed.

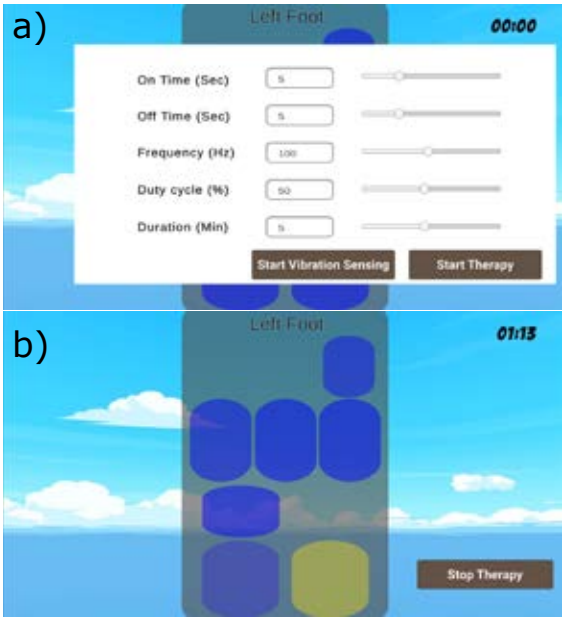


Fig. 3. Graphical User Interface for the Vibrotactile Training Mode. (a) Desired parameters are input on this screen. (b) Once the training begins, the subject may click on the corresponding plantar region shown on the screen.

At the end of the session, the subject's accuracy (the number of correctly identified iterations) and the subject's response time is displayed. This quantitative feedback acts as motivation for the subject and can encourage him to stick with the rehabilitation program. Furthermore, if the subject's results seem positive, then the intensity of the vibration motor may be reduced to increase the difficulty level of the training.

III. Functional Evaluation and Results

A series of trials is conducted on two healthy male individuals to validate the system's design and assess the suitability of the performance metrics for the two training modes. The studies involving human participants are reviewed and approved by the University Research Ethics Council of Istanbul Medipol University under the E-66291034-772.02-4494 number. The participants provided their written informed consent to participate and data sharing in this study. The age

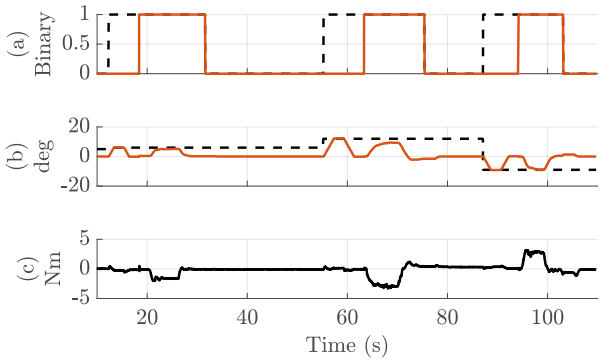


Fig. 4. Results for the proprioceptive training (Blind Trial) (a) The dashed line shows the start flag and the solid line shows the impedance flag. (b) The dashed line illustrates the target angle, and the solid line shows the angle of the manipulator. (c) The line represents the torque generated by the subject.

of the subjects are 29 and 33 years, and they do not report any history of lower limb injuries or falls.

A. Proprioception Training Results

The trial involving 3 repetitions is conducted, with one subject performing the training blind, while the other is allowed to look at the VR screen. The trials were conducted only in the dorsi-plantar direction for evaluation purpose. Results for the blind trial are shown in Fig. 4. At the moment that the user sets the desired target angle β and clicks on the start button, the start flag is set to 1 (Fig. 4a - dashed line). Consequently, the manipulator starts moving towards the target position and back, at a steady pace (Fig. 4b - solid line). Here, it can be noted that the variation in torque is minimal (Fig. 4c). Next, the impedance flag is set to 1 (Fig. 4a - solid line) and the subject applies some torque to the manipulator, as is observed in this relevant region (Fig. 4c), which in turn is responsible for enabling the motion of the manipulator.

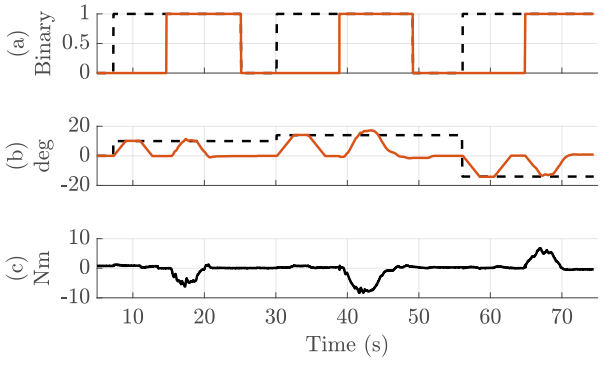


Fig. 5. Results for the proprioceptive training with VR. (a) Three iterations were conducted in this trial, like the blind trial. (b) The error between the target angle (dashed line) and observed angle (solid line) is significantly lower than in the blind trial. (c) Torque values are higher than in the blind trial, signifying calculated and confident movement.

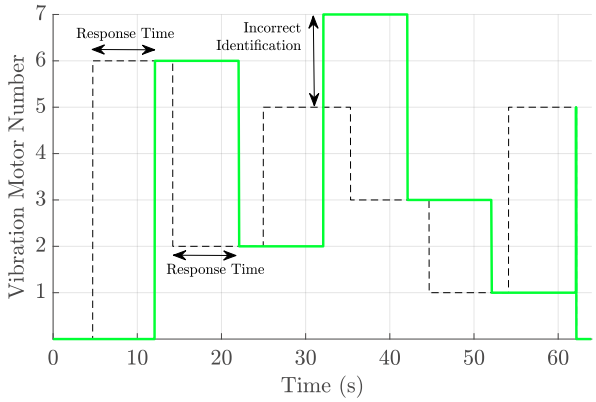


Fig. 6. Results of Vibrotactile training. The dashed line represents the motor number that is turned on, and the solid line represents the motor number chosen by the subject.

The average angular deviation from the target is observed to be 12.8% for the blind trial, whereas for the trial in which VR is used, it is 11.9% (Fig. 5). The angular deviation when returning back to the neutral position is 0.99° for the first trial and 0.82° for the second trial. Moreover, the maximum absolute torque applied by the subject is 3.3 Nm for the blind trial and 8.3 Nm in the VR trial, signifying confident and quick movement of the manipulator to the desired angle in the latter case.

B. Vibrotactile Training Results

For this trial, the on and off times of the motors are set to 5 seconds each, and the test duration is set to 60 seconds. Therefore, the trial consists of a total of 6 on/off cycles. Both subjects play the one minute trial once, and the combined accuracy in this mode is about 91.7% (11 out of 12 correctly identified), and the average response time is 8.01 seconds and 7.44 seconds for the first and second subject, respectively (Fig. 6). It is worth noting that the duty cycle is set to 50% for the first subject and 70% for the other subject.

IV. Conclusion

According to recent meta-reviews, very few lower-limb rehabilitation devices adopt a holistic approach that integrates multi-modal sensory feedback training (particularly proprioceptive and vibrotactile training) alongside traditional therapy, despite strong neurological evidence supporting their effectiveness. This work addresses this gap by developing and implementing such an integrated solution on the BALANSens robotic platform, in line with the Integrated Balance Rehabilitation (I-BaR) framework, to enable systematic clinical evaluation.

Validation tests are performed with healthy subjects to evaluate the system's readiness for patient trials. The platform computes quantitative performance metrics for both proprioceptive and vibrotactile training modes, with results monitored across multiple sessions. An important consideration in proprioceptive rehabilitation is that individuals with impaired proprioception often rely excessively on visual feedback to determine limb position. To mitigate this, subjects are restricted from viewing their ankle during training. A serious VR game is incorporated in our implementation of the training to enhance engagement, particularly during prolonged or repetitive therapy. However, while VR can increase motivation, it may also provide indirect visual cues, potentially reducing the intended proprioceptive challenge.

Overall, the integration of proprioceptive and vibrotactile training modes within a single rehabilitation framework shows strong potential to improve balance and gait therapy outcomes. Future work includes clinical trials with participants from diverse demographics to further substantiate the benefits of this form of multi-modal therapy.

Acknowledgment

This study is supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under Grant No. 122E246.

References

[1] J.-E. Cho and H. Kim, "Ankle proprioception deficit is the strongest factor predicting balance impairment in patients with chronic stroke," *Arch. Rehabil. Res. Clin. Transl.*, vol. 3, no. 4, p. 100165, 2021.

[2] R. W. Apriliyasari, P. V. Truong, and P.-S. Tsai, "Effects of proprioceptive training for people with stroke: A meta-analysis of randomized controlled trials," *Clinical Rehab.*, vol. 36, pp. 431–448, apr 2022.

[3] Y. Mao, Z. Gao, H. Yang, and C. Song, "Influence of proprioceptive training based on ankle-foot robot on improving lower limbs function in patients after a stroke," *Front. Neurorobot.*, vol. 16, p. 969671, 2022.

[4] S. Vahdat, M. Darainy, A. Thiel, and D. J. Ostry, "A single session of robot-controlled proprioceptive training modulates functional connectivity of sensory motor networks and improves reaching accuracy in chronic stroke," *Neurorehab Neural Repair*, vol. 33, pp. 70–81, jan 2019.

[5] S. De Angelis, A. A. Princi, F. Dal Farra, G. Morone, C. Caltagirone, and M. Tramontano, "Vibrotactile-based rehabilitation on balance and gait in patients with neurological diseases: A systematic review and metanalysis," *Brain Sciences*, vol. 11, no. 4, 2021.

[6] X. Zhang, Z. Yue, and J. Wang, "Robotics in lower-limb rehabilitation after stroke," *Behavioural Neurology*, vol. 2017, no. 1, p. 3731802, 2017.

[7] C. M. L. Hughes, P. Tommasino, A. Budhota, and D. Campolo, "Upper extremity proprioception in healthy aging and stroke populations, and the effects of therapist- and robot-based rehabilitation therapies on proprioceptive function," *Front. Hum. Neurosci.*, vol. 9, 2015.

[8] T. Ersoy, P. Kaya, E. Hocaoglu, and R. Unal, "I-BaR: integrated balance rehabilitation framework," *Frontiers in Neurorobotics*, vol. 18, 2024.

[9] H. M. Bilal, E. Hocaoglu, R. Unal, and P. Kaya, "An overview of integrated balance rehabilitation (I-BAR) framework," in *International Conference on NeuroRehabilitation*, pp. 560–564, Springer, 2024.

[10] S. Fernandes, B. Oliveira, S. Sacadura, C. Rakasi, I. Furtado, J. P. Figueiredo, R. S. Gonçalves, and A. C. Martins, "The effectiveness of virtual reality in improving balance and gait in people with parkinson's disease: A systematic review," *Sensors*, vol. 25, no. 15, 2025.

[11] T. Ersoy, E. Hocaoglu, P. Kaya, and R. Unal, "Design of BALANSens: Functional evaluation in ankle preparation phase," *Journal of Bionic Engineering*, vol. 21, no. 6, pp. 2893–2912, 2024.

[12] E. Hocaoglu, M. Beyaz, M. A. T. Ahmed, A. Gurluk, R. Unal, and P. Kaya, "sEMG-based assessment of engagement in VR-integrated BALANSens ankle rehabilitation," in *2025 International Conference On Rehabilitation Robotics (ICORR)*, pp. 1673–1678, 2025.

[13] H. M. Bilal, M. Beyaz, R. Unal, and E. Hocaoglu, "Impedance control for ankle preparation phase in BALANSens platform," in *2025 International Conference On Rehabilitation Robotics (ICORR)*, pp. 490–495, 2025.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Extraction and Characterization of Chitin and Chitosan from Shrimp Shell and Squid Pen Waste: Application in Biofilm Production

Yaraslau Krotau, Duygu Geçkin, Gizem Ayna Duran
Dept. of Biomedical Engineering
Izmir University of Economics
Izmir, Turkey
yaraslau.krotau@std.ieu.edu.tr, duygu.geckin@ieu.edu.tr, gizem.duran@ieu.edu.tr

***Abstract*—The objective of this study was to extract and characterize α-chitosan from Penaeus Vannamei shrimp shell waste and β-chitosan from Loligo Vulgaris squid pen waste using a chemical extraction method, thereby providing a theoretical basis for the improved development and utilization of chitosan-based biomedical materials. Extraction involved sequential deproteinization, demineralization, and deacetylation, followed by structural characterization using Fourier Transform Infrared (FT-IR) spectroscopy. FT-IR spectra confirmed the successful conversion of chitin to chitosan and revealed characteristic bands corresponding to O-H/N-H stretching, amide groups, and glycosidic linkages. The degree of deacetylation (DD) was calculated as 77.82% for α-chitosan and 94.45% for β-chitosan, indicating that β-chitosan undergoes deacetylation more readily due to its parallel chain arrangement and weaker intermolecular hydrogen bonding. These structural differences significantly influence physicochemical properties such as solubility and crystallinity, which are directly linked to biological activity. Given that higher DD enhances erythrocyte and platelet aggregation, the high-DD β-chitosan obtained from squid pens demonstrates strong potential for use in hemostatic gel biomaterials. Furthermore, biofilms produced from both α- and β-chitosan underline their promise for biomedical applications, particularly in wound healing and surgical hemostasis.**

***Keywords*—Chitin/Chitosan, degree of deacetylation, biofilm, biomaterial applications**

I. Introduction

Globally, shellfish processing generates an estimated 12–14 million tons of waste annually, of which approximately 20–40% is composed of chitin, a naturally occurring biodegradable polymer [1][2].

Chitin, also known as poly(β-(1→4)-N-acetyl-D-glucosamine), is a valuable raw material for the commercial production of chitosan. It is primarily sourced from marine waste such as squid pens, shrimp shells, and crab shells, and is recognized as the second most abundant natural biopolymer after cellulose [3].

Chitin and chitosan differ mainly in their degree of deacetylation (DD) and solubility. Partial deacetylation of chitin in an alkaline medium, where acetyl groups are removed and replaced with amino groups (NH_2), yields chitosan, a linear polysaccharide with a greater application potential [4].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Chitosan consists of two types of sugar units: glucosamine (deacetylated) and N-acetylglucosamine (acetylated). The proportion of acetylated units, known as the degree of acetylation (DA), is an important parameter used to describe and classify chitosan [5][6].

α, β, and γ are the three primary polymorphic structures of chitin, which can be found in fibrous, crystalline forms. The α-form, which is primarily found in crab and shrimp shells, is structurally different from β-chitin, which is primarily found in squid pens, and γ-chitin, which is found in yeast and fungi. Because of its weaker intermolecular hydrogen bonds and parallel polymer chains, β-chitin is more open to deacetylation during the chitosan production process and exhibits higher reactivity and affinity in a variety of physical and chemical treatments [7].

Chitosan's unique biological properties, such as its non-toxicity, biocompatibility, solubility in acidic media, and anti-inflammatory activity, make it incredibly valuable in biomedical applications [8]. It is a versatile and promising bioactive polymer, according to recent studies. Since it is the main derivative of chitin, its free amino groups make it extremely susceptible to chemical modification, which allows the synthesis of a variety of derivatives while maintaining biodegradability. [9]. These readily synthesized and commercially available derivatives have been applied to drug delivery [10], wound healing [11], and tissue engineering [12]. Apart from their biomedical applications, chitin and chitosan have also been used in food, textiles, cosmetics, water treatment, pharmaceuticals, and agriculture [13]. A key biomedical use of chitosan is in hemostatic agents that aid wound care and promote healing, with several products already approved by the United States Food and Drug Administration (USFDA) [5].

Karahaliloğlu et al. (2017) [14] evaluated the performance of active nano/microbilayer hemostatic dressings in a diabetic rat femoral artery bleeding model. The hemostatic dressing containing chitosan and bacterial cellulose, reinforced with protamine sulfate and coated with silk fibroin (SF-coated PS/BC/CTS), significantly reduced bleeding time to 80 ± 0.3 seconds compared to standard gauze ($p < 0.005$).

Recently, drug-loaded chitosan nanoparticles have been

widely studied for cancer therapy. Asiri et al. [15] reported that glutaraldehyde-conjugated chitosan nanoparticles significantly reduced HCT-116 colorectal cancer cell viability in a dose-dependent manner and altered cell morphology.

This study aimed to prepare different types of chitosan and evaluate their application in producing biopolymeric films. Chitosan was extracted from shrimp shell and squid pen waste using a chemical extraction method. β-Chitin was successfully isolated and characterized from squid pens, while α-chitin was obtained from shrimp shells. The resulting chitosan powders were characterized using FT-IR spectroscopy. Afterwards, biofilms were produced from both α- and β-chitosan with the objective of creating a hemostatic and a blood-clotting biomaterial for surgical applications, utilizing chitosan —a biopolymer well-documented- for its antibacterial and hemostatic activities.

II. Experimental Procedure

A. Materials

The squid pens and the raw shrimp shell waste were collected from a local fish market 'Mola Balıkçılık' in İzmir, Turkey. All chemicals and solvents used in this study were analytical grade or the highest available purity and all solutions were freshly prepared using distilled water.

B. Sample Preparation

Chitosan isolation includes 3 basic stages, namely deproteinization, demineralization and deacetylation. Figure 1 shows the chitin and chitosan extraction stages and biofilm production from shrimp and squid waste shells. Before extracting the chitin and chitosan, the shells must be free from undesired substances, so collected wastes were washed with running water and distilled water than dried in an oven at 40 °C for 24h. The dried samples were grounded before starting further extraction.

1) Deproteinization: The process of deproteinization involves breaking the chemical bonds that hold chitin and proteins together, which causes the biopolymer to depolymerize when exposed to an alkaline solution. In order to extract the proteins from the shells, chemical deproteinization is a necessary step [16].

2) Demineralization: Demineralization is the process of removing minerals from shells and pens, mostly calcium carbonate. Usually, demineralization is accomplished by acid treatment with HCl. Because demineralization entails the breakdown of calcium carbonate into water-soluble calcium salts and the release of carbon dioxide, as indicated in [8] by the (1):

$$2\,\mathrm{HCl} + \mathrm{CaCO_3} \longrightarrow \mathrm{CaCl_2} + \mathrm{H_2O} + \mathrm{CO_2}\uparrow \qquad (1)$$

3) Deacetylation: The deacetylation process converts chitin into chitosan by removing acetyl groups through treatment with concentrated NaOH, leaving free amine groups ($-\mathrm{NH_2}$) [17].

C. Production of Chitosan

First step of extraction of chitin was deproteinization. The dried shell and squid pen fragments are immersed in a 5% NaOH solution ratio of 1:10 (w/v) and heated at 70°C for 2.5 hours with stirring to remove residual proteins and fats. The sample is then subjected to repeated washing with distilled water until the pH reaches a neutral level of approximately 7. The second step of extraction was demineralization which is the deproteinized material is then subjected to treatment with 1 M hydrochloric acid (HCl) at room temperature for a period of 60 minutes. This process serves to eliminate calcium carbonate and other mineral components. The material was then filtered, and the washing process was repeated with distilled water until a neutral pH was achieved. The extracted chitin was subjected to deacetylation using a 50% (w/v) NaOH solution at 100 °C for 6h, with a solid-to-solution ratio of 1:20 (w/v), to remove acetyl groups. Following the reaction, the material was filtered and repeatedly washed with distilled water until reaching a neutral pH. The resulting chitosan was then dried and prepared for subsequent biofilm production.

D. Preparation of Chitosan Bio-Films

For each sample, 0.2 g of chitosan was dissolved in 1% (v/v) acetic acid ($\mathrm{CH_3COOH}$) under constant stirring for 1–2 hours, after which the solution was poured into flat, flexible plastic molds. Glycerin is used as a plasticizer to obtain a film with good mechanical properties. The samples were then left to dry at room temperature until solid films were formed.

III. Characterization of α- and β-Chitosan

A. FT-IR Analysis and Deacetylation Degree of Produced Chitosan

FT-IR was used to characterize the obtained chitosan powder in order to identify functional groups and use the baseline method to determine the DD. The analysis was also performed to observe structural differences between the two types of chitosan. Infrared spectra of chitosan were recorded using a Perkin Elmer Frontier™ FT-IR spectrometer (Perkin Elmer, USA) with attenuated total reflectance (ATR) mode. Each spectrum was obtained by averaging 32 scans at a resolution of 4 cm^{-1} over the range 400–4000 cm^{-1} covering both the functional group region and the fingerprint region. The amide III band at 1320 cm^{-1} was used as the analytical band, with 1420 cm^{-1} selected as the reference band for determining the DD. This approach minimizes interference from water, residual acids, or alkalis. DD was calculated from the absorption ratio of the 1320 to 1420 cm^{-1} peaks (A_{1320}/A_{1420}) for both α- and β-chitosan. Brugnerotto et al. (2001) [18] reported that this ratio provides the lowest error and is highly sensitive to the chemical composition of chitin and chitosan, regardless of technique, state, or secondary structure. Moreover, this method has been widely used in other studies and shown to yield reliable results consistent with alternative approaches [19]. The degree of deacetylation was calculated from the obtained FT-IR spectra from (2) proposed by Brugnerotto et al (2001) [18] :

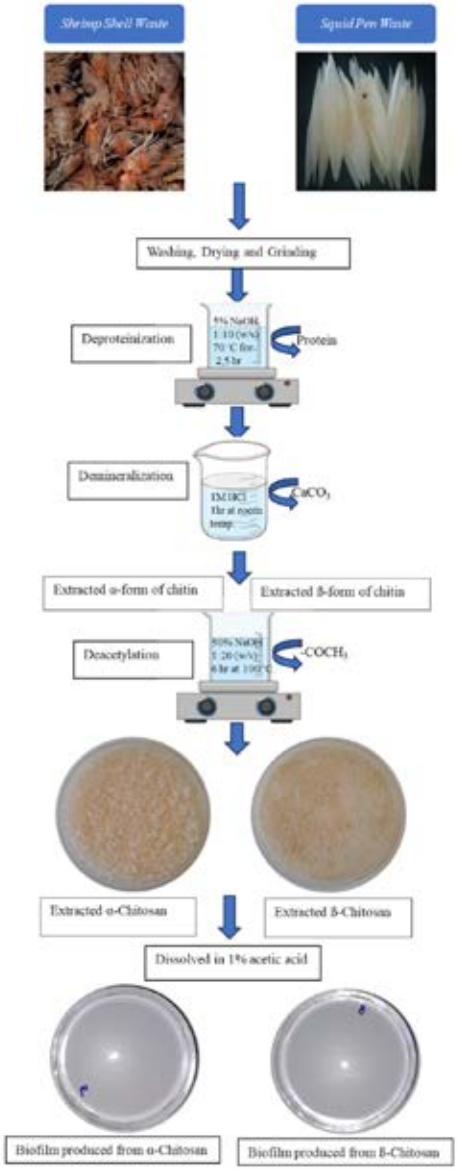


Fig. 1: Schematic representation of the extraction and production process of chitin and chitosan from shrimp shell and squid pen waste.

$$DA(\%) = [((A_{1320})/(A_{1420}) - 0.3822)/0.03133] \quad (2)$$

DA: Degree of Acetylation

$$DD(\%) = 100 - DA(\%) \quad (3)$$

The A_{1320} band reflects N-acetylation, while the A_{1420} band serves as a reference for comparing D-glucosamine and N-acetylglucosamine units [18].

IV. Results and Discussions

A. FT-IR Characterization

Figure 2 shows the FT-IR spectra of the produced α- and β-chitosan, both of which closely match typical literature profiles. Both samples display the characteristic broad O–H/N–H

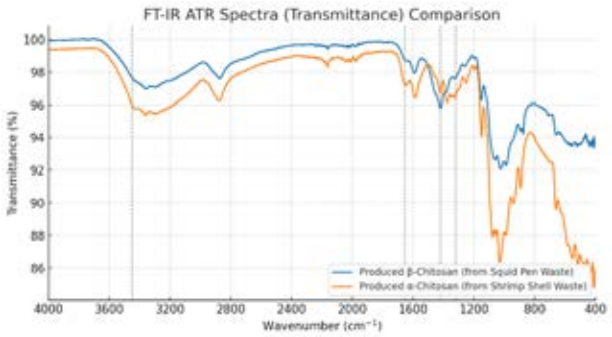


Fig. 2: FT-IR Spectra of α- and β-Chitosan

stretching band at $\sim$ 3400–3200 cm^{-1}, indicative of hydrogen bonding in the chitosan structure. The amide I band ($C=O$ of residual N-acetyl groups) appears at 1658.78 cm^{-1} for α-chitosan and 1655 cm^{-1} for β-chitosan, with greater intensity in α-chitosan, suggesting higher residual acetylation. Similarly, the amide II band at $\sim$ 1560 cm^{-1} (N–H bending) is more intense in α-chitosan, reinforcing this observation. The CH_3 stretching vibration at $\sim$ 2920 cm^{-1} is present in both samples but slightly stronger in α-chitosan, consistent with a more ordered/crystalline matrix. CH_2 bending at 1420 cm^{-1} is relatively higher in β-chitosan, whereas the CH_3 bending at $\sim$ 1380 cm^{-1} is slightly stronger in α-chitosan; these bands are commonly referenced in DD estimations. Both spectra also show the glycosidic β-(1$\rightarrow$4) linkage signal near 895 cm^{-1}, confirming the expected polysaccharide backbone.

Overall, β-chitosan (from squid pens) exhibits weaker amide I and II bands than α-chitosan (from shrimp shells), indicating a higher degree of deacetylation, while α-chitosan retains more acetyl groups and displays features typically associated with greater crystallinity.

B. Degree of Deacetylation

The degree of deacetylation is the most crucial factor defining a chitosan sample. Chitosan is generally defined as partially or fully deacetylated chitin, with DD values typically above 50% [20]. In this study, the DD of chitosan extracted from shrimp shells was calculated as 77.82%, whereas chitosan obtained from squid pens exhibited a markedly higher DD of 94.45%, based on FT-IR analysis. These results confirm that β-chitosan (from squid pens) undergoes deacetylation more readily than α-chitosan (from shrimp shells). This observation is consistent with the FT-IR spectra, where β-chitosan showed weaker amide I and II bands, reflecting a lower residual acetyl content. The higher reactivity of the β-form, linked to its parallel chain arrangement and weaker hydrogen bonding, explains its greater susceptibility to deacetylation. As reported in the literature, the DD is strongly influenced by the morphology and origin of chitin, which in turn significantly affects key properties of chitosan such as solubility, hydrophilicity, crystallinity, and biological response [21], [22].

V. Conclusion

α- and β-chitin were extracted from shrimp shell and squid pen waste and subsequently converted into α- and β-chitosan through deacetylation. The chemical composition of the obtained chitosan was characterized by FT-IR, which showed DD values of 77.82% for α-chitosan and 94.45% for β-chitosan. FT-IR spectra indicated that β-chitosan exhibited weaker amide I and II bands than α-chitosan, confirming a higher DD(%), while α-chitosan retained more acetyl groups and displayed features associated with higher crystallinity. These differences are attributed to the parallel chain arrangement and weaker intermolecular hydrogen bonding of β-chitin, which make it more reactive. Since the degree of deacetylation directly affects hemostatic performance by promoting erythrocyte and platelet aggregation [23], high-DD β-chitosan from squid pens demonstrates strong potential for developing hemostatic gel biomaterials. In addition, the chitosan biofilms produced in this study further highlight the biomedical potential of chitosan, particularly in wound healing and hemostatic applications.

References

[1] Gincy Mathew, Rajeev Sukumaran, Raveendran Sindhu, Parameswaran Binod, and Ashok Pandey, "Green remediation of the potential hazardous shellfish wastes generated from the processing industries and their bioprospecting," *Env. Tech. Inno.*, vol. 24, pp. 101979, 09 2021.

[2] Nilesh Nirmal, Didem Demir, Seda Ceylan, Sameer Ahmad, Gulden Goksen, Pankaj Koirala, and Gioacchino Bono, "Polysaccharides from shell waste of shellfish and their applications in the cosmeceutical industry: A review," *Int. Jour. Bio. Macro.*, vol. 265, pp. 131119, 2024.

[3] Marguerite Rinaudo, "Chitin and chitosan: Properties and applications," *Progress in polymer science*, vol. 31, no. 7, pp. 603–632, 2006.

[4] Inmaculada Aranaz, Andrés R. Alcántara, Maria Concepción Civera, Concepción Arias, Begoña Elorza, Angeles Heras Caballero, and Niuris Acosta, "Chitosan: An overview of its properties and applications," *Polymers*, vol. 13, no. 19, 2021.

[5] Alessandro Pellis, Georg Guebitz, and Gibson S Nyanhongo, "Chitosan: Sources, processing and modification techniques," *Gels*, vol. 8, 2022.

[6] Zhangrun Han, Yangyang Zeng, Hong Lu, and Lijuan Zhang, "Determination of the degree of acetylation and the distribution of acetyl groups in chitosan by hplc analysis of nitrous acid degraded and pmp labeled products," *Carbohydrate Research*, vol. 413, 03 2015.

[7] Ruimin Li, Wenkui Song, Shijia Huang, Chuyi Liu, Mingbo Li, and Leilei Sun, "Synthesis, characterization, and preliminary analysis of squid pen trypsin hydrolysates and chitosan microcapsules," *Int. Jour. Bio. Macro.*, vol. 26, no. 7, 2025.

[8] Islem Younes and Marguerite Rinaudo, "Chitin and chitosan preparation from marine sources. structure, properties and applications," *Marine Drugs*, vol. 13, pp. 1133–1174, 03 2015.

[9] Harsh Yadav, Rishabha Malviya, and Niranjan Kaushik, "Chitosan in biomedicine: A comprehensive review of recent developments," *Carbohydrate Polymer Tech. and App.*, vol. 8, pp. 100551, 2024.

[10] Rajalakshmi Rukmangathen, Indira Muzib Yallamalli, and Prasanna Raju Yalavarthi, "Formulation and biopharmaceutical evaluation of risperidone-loaded chitosan nanoparticles for intranasal delivery," *Drug Dev. and Ind. Phar.*, vol. 45, no. 8, pp. 1342–1350, 2019.

[11] Milena Ignatova, Nevena Manolova, and Iliya Rashkov, "Novel antibacterial fibers of quaternized chitosan and poly (vinyl pyrrolidone) prepared by electrospinning," *European Polymer Journal*, vol. 43, no. 4, pp. 1112–1122, 2007.

[12] Jennifer L Vondran, *Fabrication, optimization, and characterization of carboxymethylated chitosan nanofiber mats for cartilage regeneration applications*, Drexel University, 2007.

[13] Surinder Kaur and Gurpreet Singh Dhillon, "The versatile biopolymer chitosan: potential sources, evaluation of extraction methods and applications," *Crit. Rev. in Micro.*, vol. 40, no. 2, pp. 155–175, 2014.

[14] Zeynep Karahaliloğlu, Murat Demirbilek, İbrahim Ulusoy, Berrak Gumuskaya, and Emir Denkbas, "Hemostatic activities of nano/microporous bilayer dressings in a femoral artery bleeding rat model," *Jour. App. Poly.r Sci.*, vol. 133, 04 2016.

[15] Sarah Mousa Asiri, Firdos Alam Khan, and Ayhan Bozkurt, "Synthesis of chitosan nanoparticles, chitosan-bulk, chitosan nanoparticles conjugated with glutaraldehyde with strong anti-cancer proliferative capabilities," *Artificial cells, nanomedicine, and biotechnology*, vol. 46, pp. 1152–1161, 2018.

[16] Ahmed Hosney, Sana Ullah, and Karolina Barčauskaitė, "A review of the chemical extraction of chitosan from shrimp wastes and prediction of factors affecting chitosan yield by using an artificial neural network," *Marine Drugs*, vol. 20, no. 11, 2022.

[17] A. Domard, "ph and c.d. measurements on a fully deacetylated chitosan: application to cuii—polymer interactions," *International Journal of Biological Macromolecules*, vol. 9, no. 2, pp. 98–104, 1987.

[18] J. Brugnerotto, Jaime Lizardi-Mendoza, Francisco Goycoolea, Waldo Argüelles-Monal, Jacques Desbrieres, and Marguerite Rinaudo, "An infrared investigation in relation with chitin and chitosan characterization," *POLYMER*, vol. 42, pp. 3569–3580, 04 2001.

[19] Joydeep Dutta and Priyanka, "A facile approach for the determination of degree of deacetylation of chitosan using acid-base titration," *Heliyon*, vol. 8, no. 7, pp. e09924, 2022.

[20] Machodi Mathaba and Michael Olawale Daramola, "Effect of chitosan's degree of deacetylation on the performance of pes membrane infused with chitosan during amd treatment," *Membranes*, vol. 10, no. 3, 2020.

[21] Baha Eddine Abdelmalek, Assaâd Sila, Anissa Haddar, Ali Bougatef, and Mohamed Ali Ayadi, "β-chitin and chitosan from squid gladius: Biological activities of chitosan and its application as clarifying agent for apple juice," *International Journal of Biological Macromolecules*, vol. 104, pp. 953–962, 2017.

[22] Hakima El Knidri, Raja Belaabed, Abdellah Addaou, Ali Laajeb, and Ahmed Lahsini, "Extraction, chemical modification and characterization of chitin and chitosan," *International Journal of Biological Macromolecules*, vol. 120, pp. 1181–1189, 2018.

[23] Munezza A. Khan and Mohammad Mujahid, "A review on recent advances in chitosan based composite for hemostatic dressings," *International Journal of Biological Macromolecules*, vol. 124, pp. 138–147, 2019.

Biyomedikal Mühendislikte Üretim ve Regülasyonlar: Mevzuat, Kalite ve Gelecek Perspektifleri

Manufacturing and Regulations in Biomedical Engineering: Legislation, Quality, and Future Perspectives

Selen Koçak
Kalite Müdürü, Marla Teknoloji Müh. San. Ve Tic. Ltd. Şti.
selen@marlateknoloji.com.tr

Onur Koçak
Bilgisayar & Biyomedikal Mühendisliği Bölümü, Başkent Üniversitesi,
okocak@baskent.edu.tr

Arif Koçoğlu
TBMYO, Biyomedikal C.T., Başkent Üniversitesi,
arif@baskent.edu.tr

Ziya Telatar
Biyomedikal Mühendisliği Bölümü, Başkent Üniversitesi,
ztelatar@baskent.edu.tr

Özetçe—**Biyomedikal mühendislik, insan sağlığını korumak ve geliştirmek amacıyla tıbbi cihazların tasarımı, geliştirilmesi, üretimi ve pazara sunulmasını kapsayan disiplinlerarası bir mühendislik dalıdır. Bir tıbbi cihazın elektronik, mekanik ve yazılım tasarımlarının yanı sıra klinik uygulamaları da hem ulusal hem de uluslararası regülasyonlar ve standartlar kapsamında sıkı şekilde denetlenmektedir. Bu makalede, biyomedikal üretim süreçleri tasarımdan kalite kontrolüne kadar teknik detaylarla ele alınmakta; Avrupa Birliği Tıbbi Cihaz Regülasyonu (MDR), ABD Gıda ve İlaç Dairesi (FDA) ve Türkiye İlaç ve Tıbbi Cihaz Kurumu (TİTCK) çerçevesinde regülasyonların uygulanma biçimi incelenmektedir. Ayrıca yapay zekâ tabanlı cihazlar, kişiselleştirilmiş tıbbi cihazlar ve dijital regülasyon gibi geleceğe yönelik eğilimler tartışılmaktadır.**

Anahtar Kelimeler — tıbbi cihaz üretimi; sağlıkta yapay zeka; biyouyumluluk testleri; CE-MDR yönetmeliği; risk yönetimi yaklaşımı

Abstract— Biomedical engineering is an interdisciplinary field that encompasses the design, development, manufacturing, and market introduction of medical devices, with the primary objective of protecting and improving human health. The electronic, mechanical, and software designs of a medical device, along with good clinical practices, are strictly regulated under both national and international standards and regulatory frameworks. This paper addresses biomedical production processes in detail, from design to quality control, and examines the implementation of regulations within the frameworks of the European Union Medical Device Regulation (MDR), the U.S. Food and Drug Administration (FDA), and the Turkish Medicines and Medical Devices Agency (TİTCK). In addition, emerging trends such as AI-based devices, personalized medical devices, and digital regulation are discussed.

Keywords— medical device manufacturing; artificial intelligence in healthcare; biocompatibility testing; CE-MDR regulation; risk management approach

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

I. GENEL BAKIŞ

Biyomedikal mühendislik, mühendislik prensipleri ile tıp ve biyoloji bilgisini birleştirerek insan sağlığına yönelik yenilikçi çözümler üretir. Bu alan hem yüksek teknoloji hem de katı mevzuat uyumu gerektiren bir sektörün temelini oluşturmaktadır.

Medikal cihaz pazarı, 2025 yılı için dünya genelinde yaklaşık 572 milyar dolar büyüklüğe ve 2032 yılı için 886 milyar dolar büyüklüğe ulaşması beklenmektedir [1]. Bu pazarın dinamik yapısı, teknolojik gelişmeler ve regülasyon değişiklikleriyle büyük ölçüde şekillenmektedir. Özellikle Avrupa Birliği'nin MDR (Medical Device Regulation – Regulation (EU) 2017/745) düzenlemesi ve ABD'nin FDA 21 CFR 820 yönetmelikleri, üretim süreçlerine doğrudan yön veren çerçeveler sunmaktadır [2].

Hasta güvenliği, biyouyumluluk ve performans doğrulaması, üretim sürecinde regülasyonların getirdiği en kritik unsurlardır. Bu nedenle, biyomedikal mühendisliğin üretim boyutu, yalnızca teknik değil aynı zamanda hukuki ve etik boyutları da içeren çok katmanlı bir yapıya sahiptir [3].

II. BİYOMEDİKAL ÜRETİM SÜREÇLERİ

A. Tasarım ve Ar-Ge Süreçleri

Medikal cihaz geliştirme süreci, tasarım girdilerinin (Design Inputs) belirlenmesi ile başlar. Bu girdiler; klinik ihtiyaçlar, kullanıcı gereklilikleri, ilgili regülasyon ve standartlar dikkate alınarak hazırlanmaktadır. Girdileri oluşturulurken takip edilmesi gereken ve evrensel olarak kabul görüş yöntemler bulunmaktadır.

URS (User Requirement Specification): Kullanıcı beklentilerinin tanımlandığı dokümandır. Genellikle kullanım kitapçığı ile karıştırılmaktadır. Fakat doğrudan sağlık sistemi uygulayıcıları son kullanıcılarından soru-cevap ya da anket

metodolojisi ile elde edilen verilerden oluşturulmalıdır. Şekil 1'de "URS" evrensel logosu yer almaktadır.



Şekil 1. "URS" evrensel görseli [4].

Tasarım çıktıları (Design Outputs): Teknik çizimler, malzeme listeleri, üretim talimatları şeklinde belgelenmelidir. Ürünün elektronik bir tıbbi cihaz mı yoksa sadece enjeksiyon temelli bir üretime tabi olacak sarf malzeme mi olduğu belirlenir ve teknik dosya buna göre oluşturulur [5-6].

Tasarım incelemeleri (Design Reviews): Sürecin periyodik değerlendirilmesini kapsar. Eğer bu cihaz sınıf IIA/B risk değerlendirmesi kapsamında bir elektrikli tıbbi cihaz olan diyaliz cihazı ise (Şekil 2) cihaz tasarımı ve üretim süreci; ISO 13485:2016 kalite yönetim sistemi, ISO 14971 [7] risk yönetimi ve IEC 60601-1/60601-1-2 elektriksel güvenlik ve EMC normlarına uygun şekilde yapılandırılmalıdır.



Şekil 2 Diyaliz sistemi içerisinde yer alan "diyaliz makinası" [8].

Tasarım doğrulama (Verification): Tasarım doğrulama planları; elektriksel testler, yazılım simülasyonları, sensör kalibrasyonları, su ve elektrolit analizleri, kullanıcı testleri ve sterilizasyon uyumluluğu başlıklarında gerçekleştirilecektir.

Geçerlilik (Validation): Ürünün klinik gereklilikleri karşıladığının test edilmesi işlemleridir. Özellikle yeni regülasyonların devreye girmesi ile (MDR) klinik çalışmalar ile birlikte cihaz validasyonu ve klinik performans gözlemlenmesi birçok tıbbi cihaz için zorunlu bir hal almıştır.

Tasarım ve Ar-Ge süreci boyunca MDR ve FDA gibi regülasyonlar dikkate alınır. Risk yönetimi, kalite kontrol ve izlenebilirlik, üretim sürecine önceden entegre edilmektedir. Dolayısıyla daha kavramsal tasarım aşamasında kalite birimi ve ar&ge birimi ISO 14971 risk yönetimi standartlarına uygun olarak riskleri belirlemek zorundadır [7]. Bu aşama sadece mühendislik düzeyinde değil sistemi kullanacak klinisyen ve diğer sağlık profesyonelleri ile de doğrulanmalıdır.

AB MDR (2017/745) Madde 10(9), tasarım sürecinde risk yönetiminin ISO 14971 standardına göre uygulanmasını zorunlu kılmaktadır. Ayrıca biyouyumluluk değerlendirmesi için ISO 10993 serisi standartlara uyulmalıdır. Bu yaklaşım, cihazın pazara sunulmadan önce tüm güvenlik ve performans gerekliliklerini karşılamasını sağlamaktadır. Bir tıbbi cihazın hasta ile temas eden aktif olmayan bölümleri de var ise mutlaka uzun dönem biyouyumluluk riskleri gözetilmelidir. Ayrıca tasarım aşamasında yapılan validasyon ve prototip testleri, üretim sırasında ortaya çıkabilecek hataları en aza indirmeye yardımcı olan işlemlerin başında yer alır [9]. Şekil 3'te günümüzde tıbbi cihaz belgelendirmede sıklıkla kullanılan standart ve regülasyonların logoları görülmektedir.



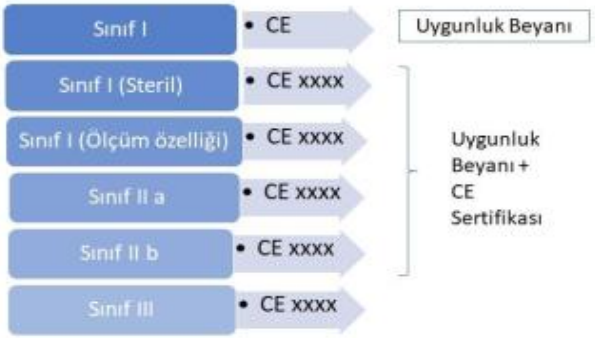
(a) (b)

(c)

Şekil 3. MDR 2017/745 Tıbbi Cihaz Yönetmeliği (a), ISO 10993 Biyouyumluluk Standardı – Hastaya temas eden parçaların biyouyumluluğun gözlemlenmesi (b)ISO 14971 Risk Yönetimi Standardına göre tasarlanacak tıbbi cihazın risk analizi gerçekleştirilir (c).

B. Üretim Süreci

Üretim süreci, ürünün risk sınıfına, kullanılacağı klinik alana ve üretim teknolojisine göre şekillenir. MDR yönetmeliğine göre Sınıf I'den Sınıf III'e ilerledikçe tıbbi cihazın canlı üzerinde kullanımında yaşamsal risk oluşturma faktörü artar ve artan riske göre belgelendirme süreci ve denetimi gereksinimleri detaylanır. Şekil 4'te günümüzde geçerli olan tıbbi cihaz sınıfları ve belgelendirme yöntemleri görülmektedir.



Şekil 4. Tıbbi cihaz risk sınıfları

Tıbbi cihazlar elektronik bileşenlere sahip elektrikli sistemler olabilecekleri gibi doğrudan insan vücuduna implante edilebilen ve metalik özellikli malzemelerden de oluşabilir. Dolayısıyla tıbbi cihaz endüstrisinde her zaman tek tip bir üretici modelinden söz edilemez. Yukarıdaki örnekte olduğu gibi bir diyaliz makinası üreticisi diyalizat sıvısı transferi, sızdırmazlığı, ısı ve elektrolit izleme sensörleri ve elektroniği, arayüz ve gömülü yazılım uygulamaları, sesli ve görsel alarm

sistemi, veri tabanı yönetimi, elektronik kart dizgisi ve kaçak akım testleri gibi fabrika seviyesinde tasarım ve kontrollerden sorumludur.

Tek kullanımlık plastik enjeksiyon malzeme üreticileri ve metal işleme ürünleri üreten diğer fabrikalarda yukarıdakine benzer tasarım gereksinim analizleri ile uzmanlıklarını arttırabilirler.

Genel olarak tıbbi cihaz üreticileri;

- Elektronik kart üretimi ve PCB dizgi işlemleri,
- Yazılım geliştirme işlemleri,
- Sensör elektroniği geliştirme ve tasarım işlemleri,
- CNC İşleme ile metal ve polimer komponentlerde yüksek hassasiyet sağlama teknikleri,
- Enjeksiyon Kalıplama – Tek kullanımlık sarf malzemelerin yüksek hacimli üretilmesi modelleri,
- 3D Baskı (Additive Manufacturing) – Kişiye özel implant ve cerrahi kılavuz üretim çalışmaları,
- Yüzey Modifikasyonu – Plazma kaplama, anodizasyon, antimikrobiyal yüzeyler elde etme ile malzeme mühendisliği

çalışmalarına odaklanmışlardır [10-11].

Günümüzde özellikle malzeme tabanlı üretim, çoğunlukla temiz oda (cleanroom) şartlarında yapılmaktadır. Temiz oda sınıfı, ürün tipine göre ISO 14644-1 standardına göre belirlenmektedir (ör. ISO Class 7). Bu durumda üreticiye birde temiz oda kurulumu ve 24 saat esaslı işletilmesi maliyeti oluşmaktadır [12].

Üretimde kullanılacak malzemeler, cihazın türüne göre değişmekle birlikte, ISO10993 serisi standartlara uygun olarak biyouyumlu olmalıdır. Bu durumda tedarikçilerin sertifikaları ve malzeme belgeleri kalite yönetim sistemi kapsamında izlenmek zorundadır. Malzeme seçimi kadar kritik tedarikçi seçimi ve yönetimi de artık MDR Regülasyonunun çok önemli bir parçasıdır.

Steril cihazlar için üretim sonrası sterilizasyon, hasta güvenliği açısından kritik öneme sahiptir. Kullanılan yöntemler (EO - etilenoksit, gamma, buhar) ISO 11135, ISO 11137 ve ISO 17665 standartlarına göre validasyon testlerine tabi tutulmalıdır [13].

MDR ve/veya FDA gereklilikleri, üretim sürecine önceden entegre edilmelidir. Risk yönetimi üretim protokollerinde takip edilir ve malzeme, proses ve ürün güvenliği sürekli denetlenmektedir. ABD mevzuatında 21 CFR 820.70 "Production and Process Controls" kapsamında, her üretim adımının validasyonu ve kayıt altına alınması zorunludur [14-15]. Bu kayıtlar Device History Record (DHR) dosyasında saklanmaktadır. Tahmin edileceği üzere MDR yönetmeliği kapsamında benzer zorunluluklar CE belgelendirmeye tabi üreticilere de zorunludur [5-6].

C. Kalite Kontrol ve Testler

Üretim sürecinin tamamlanmasının ardından kalite kontrol ve testler, cihazın regülasyonlarla ve teknik spesifikasyonlara uyumlu ve güvenli olduğunu doğrulamak için uygulanmaktadır.

In-Process kontroller : Üretim sırasında yapılan kontroller, hataların bant sistemi üzerinde tespit edilmesini sağlamaktadır. Örneğin; boyut ölçümler, mekanik dayanıklılık testleri, üretim sıcaklık ve nem kontrolleri gibi fabrika seviyesinde yapılacak izlenebilir testleri kapsamaktadır.

Nihai ürün testleri : Cihaz tamamlandıktan sonra yapılan testlerdir. Tasarım çıktıları ve regülasyon gereklilikleri ile uyumu doğrulamaktadır. Fabrika içinde yapılır fakat regülasyonun kategorisine göre bağımsız test kuruluşu hizmeti de gerekebilir. Nihai ürünlere uygulanan testlerden bazıları;

- Biyouyumluluk Testleri – ISO 10993 serisi testler (sitotoksisite, duyarlılık, genotoksisite vb.).
- Sterilizasyon Validasyonu – ISO 11135 (ETO), ISO 11137 (gama ışını), ISO 17665 (buhar).
- Stabilite ve Raf Ömrü Testleri – ASTM F1980'e göre gerçek zamanlı ve hızlandırılmış yaşlandırma testleri.
- IEC 60601-1: Elektriksel güvenlik
- IEC 60601-1-2: Elektromanyetik uyumluluk (EMC)

Olası üretim hataları, CAPA (corrective and Preventive Action) sistemi ile analiz edilir ve giderilir. Ayrıca UDI (Unique Device Identification) sistemi, ürünün tüm üretim ve tedarik sürecinin izlenebilirliğinin CE bölgesi ülkelerin sistemleri üzerinden de sağlanır. Bu sayede geri çağırma veya güvenlik uyarıları gerektiğinde hızlı müdahaleler sağlık otoritesi tarafından kolaylıkla uygulanabilmektedir [16-17].

III. REGÜLASYONLAR VE STANDARTLAR

Medikal regülasyonlar, biyomedikal cihazların yalnızca üretim sürecinde değil, tasarım, klinik değerlendirme, pazara sunma ve kullanım aşamalarında da güvenli ve etkili olmasını AB (Avrupa Birliği) Pazarında garanti eden yasal çerçevelerdir.

Bu regülasyonlar, cihazın yaşam döngüsünü bütüncül olarak kapsar ve üreticilere belirli yükümlülükler getirir [16-17]. Regülasyonlar, cihazın piyasaya sürülmeden önce belirli gereklilikleri karşılamasını şartını koşmaktadır.

Klinik veri toplama : Cihazın güvenliği ve etkinliğini göstermek için prototip testleri ve klinik çalışmalar üretici sorumluluğunda yürütülür. PMCF (post market clinic follow) raporları hazırlanmalı ve vijilanslar takip edilmelidir.

Etiketleme ve kullanıcı bilgilendirmesi: Kullanıcıya yönelik güvenlik ve kullanım talimatları regülasyonlara uygun şekilde hazırlanmalıdır. Özellikle MDR ile etiket hazırlama ve UDI beyanı takip edilmelidir.

Sınıflandırma ve NB (Nodified Body) onaylanmış kuruluş : Cihazın risk sınıfı belirlenir ve ilgili regülasyon otoritelerine bildirim yapılır. Onaylanmış kuruluş seçimi NANDO üzerinden gerçekleştirilir ve sözleşme yolu ile ilgili ülke otoritesinin takibine girilir.

Üretim bölümünde detaylıca ele alınan hasta güvenliği, biyouyumluluk ve performans doğrulaması, regülasyon açısından da temel kriterlerdir. Bu unsurların vurgulanması, cihazın pazara sunulmadan önce güvenli ve etkili olmasını temin edebilmektedir.

Türkiye'de Tıbbi Cihaz Yönetmeliği, AB MDR ile uyumlu şekilde yürütülmektedir. TİTCK (Türkiye Tıbbi Cihaz ve İlaç Kurumu) mevzuat uygulama, denetim ve ürün kayıtlarını yönetmektir. Ayrıca NB'lerin atanması ve denetçilerin atanması hususlarını AB adına takip eder. UBB (Ürün Takip Sistemi) – ise pazara sunulan cihazların izlenebilirliğini sağlamaktadır.

IV. GELECEK EĞİLİMLER

Tıbbi cihaz endüstrisi teknolojik ilerlemeler ve regülasyonların evrimi ile hızlı bir dönüşüm içerisindedir. Gelecekteki eğilimler, üretim süreçlerinden regülasyon uygulamalarına kadar geniş bir yelpazeyi kapsayacaktır.

Özellikle kişiselleştirilmiş medikal cihazlar ve implantlar için üretim süreçlerini hızlandırmakta ve maliyetleri düşürmektedir. 3D baskı ile üretim, kişiye özel implant üretimini arttıracaktır. Bunun yanında MDR ile birlikte oluşan "implant kart" uygulamaları ve üreticilerin doğrudan hastayı takip edip izlemesi gereksinimi ortaya çıkacaktır.

IoT (Internet of Things) ve sensör teknolojileri ile üretim süreçlerini gerçek zamanlı olarak izleyerek kalite kontrolün ve izlenebilirliğin artacağı ön görülmektedir.

Blockchain ve veri yönetimi ile medikal cihazların tedarik zinciri ve kullanıcı verileri için güvenli kayıtlar oluşacaktır. Üretim ve regülasyon süreçleri kişiselleştirilmiş tıp ve hasta odaklı tasarım ile daha da entegre olacağı düşünülmektedir. Ayrıca yapay zekâ uygulamalarının hem üretim tarafında kullanılması hem de ürünlerin içerisinde elde edilen verilerden erken teşhis amaçlı kullanılması farklı regülasyon modellerinin gelişmesine de katkı sunacaktır.

V. SONUÇ VE ÖNERİLER

Biyomedikal cihaz üretimi, teknolojik gelişmeler ve regülasyonlar tarafından şekillenen karmaşık bir süreçtir. Bu makalede aktarılmaya çalışıldığı üzere tasarım, ar-ge, üretim, kalite kontrol, regülasyon entegrasyonu, tıbbi cihaz güvenliği, birbirine bağlı aşamalardır. Üretim sürecinde regülasyonların uygulanması, hasta güvenliği, biyouyumluluk ve performans doğrulamasını daha izlenebilir kılmaktadır. Kalite kontrol ve testler de cihazın standartlara uygunluğunu doğrulamaktadır.

Gelecekte, kişiselleştirilmiş tıp, dijitalleşme ve ileri üretim teknolojileri ile tıbbi cihaz üretim ve tıbbi cihaz regülasyon süreçlerinin daha da entegre olması beklenmektedir. Regülasyonlar, tasarım aşamasından itibaren üretim sürecine entegre edilmelidir. Üretici, risk yönetimini ve kalite süreçlerini mevzuata göre şekillendirmeli; her değişiklik için yeniden doğrulama yapmalıdır. Sonuç olarak, biyomedikal cihaz endüstrisinde üretim ve regülasyonlar birbirini tamamlayan süreçler olarak ele alınmalı ve geleceğe dönük olarak teknolojik yenilikler ile sürekli geliştirilmelidir.

KAYNAKLAR

[1] Fortune Business Insights. (2025). Medical devices market size, share & industry analysis. Retrieved September 17, 2025, from https://www.fortunebusinessinsights.com/industry-reports/medical-devices-market-100085.
[2] TİTCK, Tıbbi Cihaz Yönetmeliği, Türkiye İlaç ve Tıbbi Cihaz Kurumu, Ankara, 2021.
[3] ISO/ASTM 52900:2021, Additive manufacturing — General principles — Fundamentals and vocabulary, ASTM International and ISO, 2021.
[4] Google Images, "Medical device regulatory scheme infographic," Available: https://images.app.goo.gl/XeYknSHF8CVv4VxJ8 [Accessed: Aug. 18, 2025].
[5] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, Human–Computer Interaction, 3rd ed., Pearson, 2004.
[6] Y. Rogers and H. Sharp, Interaction Design: Beyond Human–Computer Interaction, 5th ed., Wiley, 2019.
[7] ISO 14971:2019, Medical devices — Application of risk management to medical devices, International Organization for Standardization, Geneva, 2019.
[8] Google Images, "Medical device sterilization processes," Available: https://images.app.goo.gl/VKnsy3rWuz37nEm58 [Accessed: Aug. 18, 2025].
[9] ISO 10993 series, Biological evaluation of medical devices, International Organization for Standardization.
[10] IEC 62443 series, Industrial communication networks — IT security for networks and systems, International Electrotechnical Commission.
[11] National Institute of Standards and Technology (NIST), Cybersecurity Framework, Gaithersburg, MD, 2018.
[12] ISO 14644-1:2015, Cleanrooms and associated controlled environments — Part 1: Classification of air cleanliness by particle concentration, International Organization for Standardization, Geneva, 2015.
[13] ISO 11135, ISO 11137, ISO 17665, Sterilization of health-care products standards, International Organization for Standardization.
[14] ASTM F1980-21, Standard Guide for Accelerated Aging of Sterile Barrier Systems for Medical Devices, ASTM International, West Conshohocken, PA, 2021.
[15] U.S. Food and Drug Administration, 21 CFR Part 820 – Quality System Regulation, FDA, 2020.
[16] Regulation (EU) 2017/745 of the European Parliament and of the Council, Medical Devices Regulation (MDR), 2017.
[17] ISO 13485:2016, Medical devices — Quality management systems — Requirements for regulatory purposes, International Organization for Standardization, Geneva, 2016.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Wearable Soft Actuator for Restoring Finger Adduction–Abduction

Yazan KANJO
Department of Biomedical Engineering
Istanbul Medipol University
Istanbul, Türkiye

Abed Alrahman AL SAYYED
Department of Biomedical Engineering
Istanbul Medipol University
Istanbul, Türkiye

Elif HOCAOGLU
School of Engineering and Natural Sciences
Istanbul Medipol University, Istanbul, Türkiye
Research Institute of Health Sciences and Technologies (SABITA)
E-mail: ehocaoglu@medipol.edu.tr

***Abstract*—Finger adduction and abduction are essential for daily activities such as typing, grasping, and manipulating small objects, yet these movements are often impaired by neuromuscular disorders or injuries, significantly reducing quality of life. Conventional rehabilitation strategies frequently overlook lateral finger motions or lack the adaptability to accommodate the full range of movement. To address this limitation, we present a portable, low-cost, and user-friendly soft pneumatic rehabilitation device specifically designed to restore finger adduction and abduction. The elastomer-based actuator, dorsally mounted across the phalanges, enables controlled MCP joint rotation through regulated inflation–deflation cycles while ensuring natural alignment and automatic return to the neutral position. In contrast to existing systems that restrict movement by positioning actuators between fingers, the proposed design prioritizes comfort and physiological motion. Experimental validation with ten healthy participants demonstrates an average angular displacement of 36°, reliable inflate–hold–deflate cycle completion within 5 seconds, and peak lateral force generation of 8 N at a stable operating pressure of 80 kPa, sufficient to induce finger motion without discomfort. With position feedback integrated into an on–off control scheme, the device achieves consistent and repeatable motion patterns suitable for unsupervised rehabilitation. By targeting lateral finger motion recovery, the proposed system fills a critical gap in hand therapy and represents a practical alternative to rigid exoskeletons, advancing wearable rehabilitation technology.**

***Keywords*—Soft robotics, hand rehabilitation, soft actuators, spasticity**

I. Introduction

Hand spasticity and impaired motor control are highly prevalent among patients with neuromuscular disorders and stroke survivors, with studies reporting occurrence rates exceeding 70% of post-stroke patients [1], [2]. Such impairments significantly reduce hand functionality and independence in daily life. Even though lateral movements, more especially, adduction and abduction, are essential for coordinated hand function, they are frequently disregarded in traditional rehabilitation techniques. The majority of the repeated manual exercises used in current therapies tend to be time-consuming, physically taxing for physiotherapists, and challenging to perform consistently over extended periods of time [3]. These drawbacks highlight the necessity of an easily accessible, reproducible, and successful rehabilitation program that focuses on finger adduction and abduction in particular.

The coordinated activity of muscles, tendons, and nerves creates the intricacy of finger motion, and injury or a lack of neural function can result in stiffness or improper movements [4], [5]. Finger adduction and abduction are frequently affected by neurological conditions, including stroke, spinal cord injury, or multiple sclerosis, as well as traumatic injuries, such as tendon or nerve damage [6]. Stroke alone affects over 65% of patients with upper extremity disabilities [7]. In addition, injuries such as tendon laceration, nerve damage, or bone fractures can limit natural motion [8]. Physical and occupational therapy are examples of traditional rehabilitation techniques that involve repetitive manual exercises that can be time-consuming, reliant on the availability of therapists, and sensitive to patient variability [7]. Robotic rehabilitation systems, especially rigid ones such as Hand Mentor, Amadeo, and HEXORR, have been introduced to address these challenges but come with drawbacks such as high cost, limited mobility, and potential discomfort [9], [10], [11], [12], [13]. Soft robotic systems have become a viable substitute in recent years because of their safety, lightweight design, and ability to adjust to human biomechanics [14]. Soft pneumatic actuators have been shown to increase grip strength and range of motion in devices like the Harvard Soft Exosuit [15]. Because of their flexibility and durability, elastomeric materials such as Dragon Skin, Mold Max, and Elastosil have been widely used in actuator types [16], [17]. These actuators' morphology and structure, whether pleated, ribbed, or cylindrical, have an impact on their bending capabilities and tip force [18]. Rapid and personalized prototyping is made possible by the frequent use of 3D-printed molds or directly printed soft materials in the fabrication of soft robotic systems [19], [20]. Although open-

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

loop control techniques are basic [21], closed-loop control approaches that use sensors like rotary potentiometers or IMUs offer more dependable and flexible performance, particularly when it comes to accurately simulating MCP joint movements [22], [23]. The abduction and adduction motion at the MCP joints, which is frequently underrepresented in conventional rehabilitation technologies, must be addressed by these systems.

Despite significant advances in soft robotics, there remains a notable scarcity of finger rehabilitation systems specifically designed to address adduction and abduction movements. Most existing solutions reported in the literature are limited either to isolated thumb motion or primarily focus on flexion–extension, leaving critical lateral finger movements underrepresented in current therapeutic approaches. Actuators positioned between the fingers in many current designs limit the full range of motion and the ability to return to the neutral hand position. Furthermore, soft actuators that have been produced thus far either have complicated fabrication and integration issues or require high inflation pressures. These drawbacks highlight the need for a novel approach that restores the fingers' natural lateral mobility by combining flexible movement, enough force production, and an ideal positioning plan.

This work suggests a soft pneumatic actuator-based rehabilitation device for regaining finger adduction and abduction movements in order to solve these issues. By positioning the actuator directly above the phalanges, regulated air inflation and deflation allow the MCP joints to rotate naturally. The suggested placement approach enables the fingers to be restored to their neutral position, guaranteeing full mobility in treatment, in contrast to traditional designs that insert actuators between fingers and limit the entire range of motion.

II. Materials and Methods

A. System Overview

The soft robotic finger rehabilitation system is designed to meet specific functional requirements based on typical human hand kinematics and therapeutic needs estimated from literature and experimental measurements. The actuator must achieve a minimum range of motion of 0°–25° for finger abduction and adduction[24], with a speed of 10-30 degrees per second[25]. The average required force to manipulate the index and middle fingers is 5-7 N [26]. The actuator is simulated on ANSYS to verify the meeting of the desired performance, as shown in Fig. 1(a,b).

The kinematic and dynamic analyses are performed to establish the relationships among joint motion, fingertip position, orientation, and required torque. The index finger is modelled as a single rigid link with a revolute joint at the MCP, representing adduction–abduction movement. Based on the dynamic analysis, a second-order transfer function of the finger is derived. Similarly, the actuator model is evaluated and simulated in ANSYS to characterize the relationship between input pressure, angular deflection, and output force. The resulting step responses, presented in Fig.1 (c,d), are used to approximate the actuator behavior with a second-order transfer function.

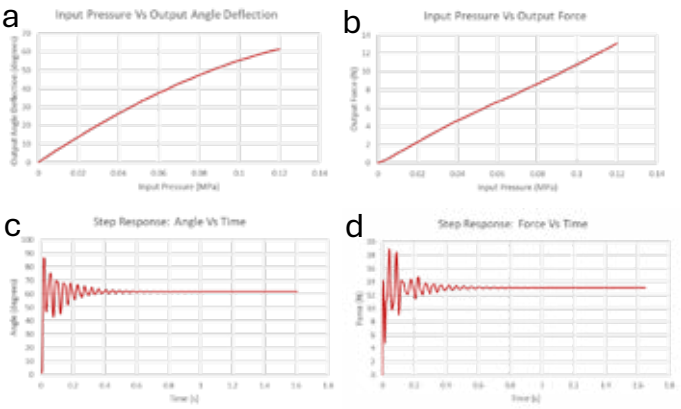


Fig. 1: ANSYS simulation results for the soft actuator: (a) Output angle deflection vs. input pressure, (b) Output force vs. input pressure, (c) Step response of angular deflection over time, and (d) Step response of force over time. The linear relationships in (a) and (b) validate the actuator's quasi-static performance, while the transient responses (c) and (d) were used to derive the actuator's second-order transfer function for dynamic modelling.

A specially made soft pneumatic actuator that produces finger adduction and abduction forms the basis of the rehabilitation system. SolidWorks is used to model the actuator, taking into account the typical human hand geometry, namely the separation between the index and middle fingers' MCP and PIP joints (see Fig. 2). Its form is designed to accommodate the fingers' natural curvature while allowing for controlled lateral movement. Natural finger motion during inflation and an unhindered return to the neutral position during deflation are made possible by the actuator's placement on the dorsal side of the phalanges. This arrangement facilitates full-range adduction and abduction workouts without limiting other finger movements.

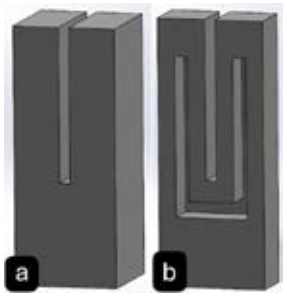


Fig. 2: Solid design of the soft actuator unit showing (a) an isometric perspective and (b) a cross-sectional view highlighting the internal channels.

The actuator is fabricated using two-part molds (upper and lower sections) coated on the inner surfaces with a thin layer of grease to facilitate demolding. A total of 20 mL of

SILASTOSIL A/B (10 mL of component A and 10 mL of component B) is mixed in a test tube using a vortex mixer for 3–4 min, followed by degassing in a vacuum chamber for 4 min to eliminate entrapped air bubbles. The degassed silicone is poured into the prepared molds, with any residual surface bubbles removed using a fine needle. The molds are cured at ambient conditions for 24 h, after which the two parts are demolded, joined with an additional layer of mixed and degassed silicone, and gently pressed together. A secondary curing period of 24 h yields the final assembled actuator.

The system integrates pneumatic and electrical components. Airflow is regulated by a 5/2 solenoid valve and a 12 V DC mini pump, each powered by separate 12 V DC sources. Pump and valve actuation are controlled via two 5 V relay modules. Force sensors are positioned between the actuator wall and the finger connectors for real-time force measurement, while two rotary potentiometers are attached to the index and middle fingers to provide continuous angular displacement monitoring. The assembled solid model of the system, positioned above the hand, is shown in Fig. 3.

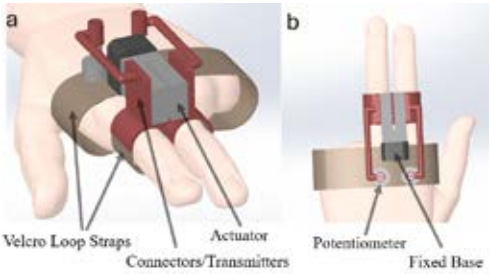


Fig. 3: Three-dimensional model illustrating the system implementation, from (a) an isometric view and (b) a top view

B. Experimental Evaluation

The control cycle consists of four sequential states: inflation, hold, deflation, and wait. During the inflation phase, both the pump and 3-way valve are activated until the target angle (35°) is reached, as determined by the differential measurement of the finger potentiometers. In the hold phase, the pump is switched off via relay control while the valve remains closed, maintaining internal actuator pressure for a preset duration. Upon completion of the hold interval, the system transitions to the deflation phase, in which both pump and valve are deactivated and the stored air is released through the valve exhaust. A short wait phase is introduced before the cycle restarts, triggered when the measured finger angle returns to a predefined threshold. The control loop is illustrated in Fig. 4

III. Results

The connectors connect the fingers to the potentiometers that are fixed on the dorsal side of the hand. The fixed base stabilizes the actuator on the hand, preventing disturbing movements throughout the exercises. The connectors are attached to Velcro bands, which are fixed around the hand and fingers, ensuring controlled motion by the actuator. Force

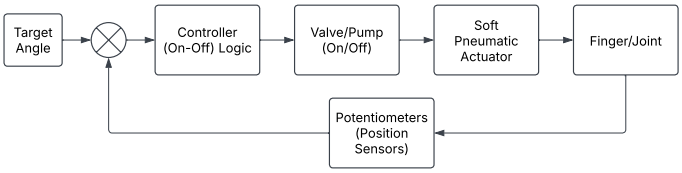


Fig. 4: Closed loop control block diagram of the soft pneumatic actuator unit

sensors are attached to the sides of the attachments where the actuator's surface pushes the connectors. The entire prototype is shown in Fig. 5.

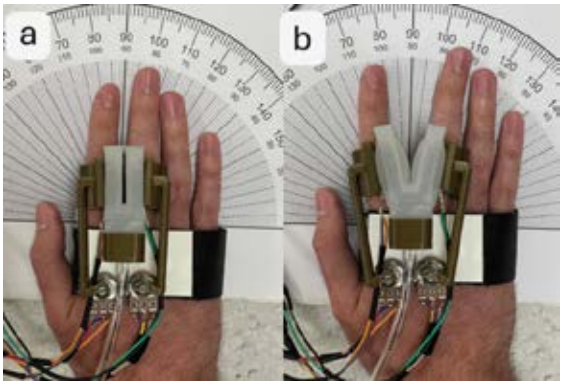


Fig. 5: Physical prototype of the soft rehabilitation system, showing the overall assembly in (a) deflation state and (b) inflation state

The calibrated potentiometers record the angle for each finger and send the data to the Arduino, where it subtracts the angles from each other to obtain the angle between both fingers, the abduction angle. The force sensors on each finger record the resistance difference due to the applied force and convert the resistance to output force through calibration. The results provide the behavior of the system's response to being cycled on and off 80 kPa of pressurized air.

The system is tested on ten people, and data for angle and force are recorded. These results demonstrate the average system's behavior applied to ten different persons. Hand stiffness differs among different individuals. The maximum angle is recorded as 38 degrees for the person with the least hand stiffness, and the minimum angle is recorded as 35 degrees, as this angle is set as the threshold for phase switching. The average angle is measured as 36 degrees, as demonstrated in Fig. 6(a). For the force recordings, the maximum force recorded is 11 N for the highest hand stiffness, and the minimum force value is recorded as 5 N for the lowest hand stiffness, averaging at 7.5-8 N, as shown in Fig. 6(b).

IV. Discussion

This study presents the development of a soft pneumatic actuation system designed to facilitate the rehabilitation of finger adduction and abduction. Following a review of prior designs in the literature, elastomeric actuators are selected

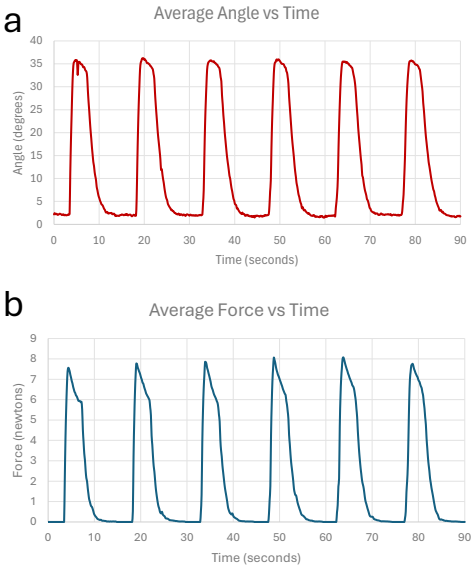


Fig. 6: Average system performance provided from cyclic actuation across ten participants: (a) Angle vs. time, showing a mean angle of 36° (b) Force vs. time, with an average output of 7.5–8 N reflecting variations in hand stiffness.

owing to their repeatability, flexibility, and capacity to deliver controlled motion. To guide the design process, both kinematic and dynamic analyses are performed, establishing the relationships between MCP joint rotation, fingertip displacement, and the torque required for functional movement. These analyses inform the geometric and mechanical design of the actuator, which is tailored to average human finger dimensions, with particular attention to the MCP–PIP joint spacing.

Experimental evaluation demonstrates that the actuator generates sufficient force to induce finger movement and achieves the desired angular displacement. These findings confirm the feasibility of the proposed actuator as a safe and efficient means of assisting finger rehabilitation. The integration of soft pneumatic actuation with real-time feedback control highlights the potential of this approach to provide cost-effective and adaptable rehabilitation solutions.

V. Conclusion

In this study, a soft pneumatic actuator mechanism is designed and implemented to assist in the rehabilitation of finger adduction and abduction. The proposed design shows consistent behavior concerning repeatability, angular motion, and force generation. Effective cyclic actuation is made possible by the system's on-off control, guaranteeing controlled and secure rehabilitative movements. The efficacy of the strategy is validated by the implementation of the prototype, which verified that each of the design and performance requirements is satisfied.

Acknowledgment

This study was supported by TÜBİTAK 2209-A Undergraduate Research Support Program.

References

[1] H. Zeng, J. Chen, Y. Guo, and S. Tan, "Prevalence and risk factors for spasticity after stroke: A systematic review and meta-analysis," *Front. Neurol.*, vol. 11, p. 616097, Jan. 2021, doi: 10.3389/fneur.2020.616097.
[2] M. J. Schinwelski, E. J. Sitek, P. Waż, and J. W. Sławek, "Prevalence and predictors of post-stroke spasticity and its impact on daily living and quality of life," Neurol. Neurochir. Pol.*, vol. 53, no. 6, pp. 449-457, Dec. 2019.
[3] B. H. Dobkin, "Clinical practice: Rehabilitation after stroke," *N. Engl. J. Med.*, vol. 352, no. 16, pp. 1677–1684, Apr. 2005.
[4] J. F. Soechting and M. Flanders, "Flexibility and repeatability of finger movements during typing: Analysis of multiple degrees of freedom," *J. Comput. Neurosci.*, vol. 4, no. 1, pp. 29–46, Jan. 1997.
[5] F. S. Cromwell and J. Bear-Lehman, *Clinical Mechanics of the Hand*, 3rd ed. St. Louis, MO, USA: Mosby, 2012, pp. 183–184.
[6] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *Lancet*, vol. 377, no. 9778, pp. 1693–1702, May 2011.
[7] G. Kwakkel, B. J. Kollen, and H. I. Krebs, "Effects of robot-assisted therapy on upper limb recovery after stroke: A systematic review, "Neurorehabil. Neural Repair, vol. 22, no. 2, pp. 111–121, Mar. 2008.
[8] J. B. Tang, "Tendon injuries across the world: Treatment," Injury, vol. 37, no. 11, pp. 1036–1042, Nov. 2006.
[9] T. Proietti, V. Crocher, A. Roby-Brami, and N. Jarrasse, "Upper-limb robotic exoskeletons for neurorehabilitation: A review on control strategies," IEEE Rev. Biomed. Eng.*, vol. 9, pp. 4–14, 2016.
[10] C. N. Schabowsky, S. B. Godfrey, R. J. Holley, and P. S. Lum, "Development and pilot testing of HEXORR: Hand EXOskeleton Rehabilitation Robot," J. NeuroEng. Rehabil., vol. 7, no. 1, p. 36, Jul. 2010.
[11] S. J. Housman, K. M. Scott, and D. J. Reinkensmeyer, "A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis,"Neurorehabil. Neural Repair, vol. 23, no. 5, pp. 505–514, Jun. 2009.
[12] J. Iqbal, N. G. Tsagarakis, and D. G. Caldwell, "A multi-DOF robotic exoskeleton interface for hand motion assistance," in Proc. IEEE Eng. Med. Biol. Soc. (EMBS), Boston, MA, USA, 2011, pp. 1575–1578.
[13] P. Heo, G. M. Gu, S. Lee, K. Rhee, and J. Kim, "Current hand exoskeleton technologies for rehabilitation and assistive engineering," Int. J. Precis. Eng. Manuf., vol. 13, no. 5, pp. 807–824, May 2012.
[14] P. Polygerinos, Z. Wang, K. C. Galloway, R. J. Wood, and C. J. Walsh, "Soft robotic glove for combined assistance and at-home rehabilitation," Robot. Auton. Syst., vol. 73, pp. 135–143, Nov. 2015.
[15] H. K. Yap, J. H. Lim, F. Nasrallah, J. C. H. Goh, and R. C. H. Yeow, "A soft exoskeleton for hand assistive and rehabilitation application using pneumatic actuators with variable stiffness," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), Seattle, WA, USA, 2015, pp. 4967–4972.
[16] Wacker Chemie AG, Munich, Germany. [Online]. Available: https://www.wacker.com
[17] Smooth-On, Inc., Macungie, PA, USA. [Online]. Available: https://www.smooth-on.com
[18] A. D. Marchese, R. K. Katzschmann, and D. Rus, "A recipe for soft fluidic elastomer robots, "Soft Robot., vol. 2, no. 1, pp. 7–25, Mar. 2015.
[19] M. Tiboni and D. Loda, "Monolithic PneuNets soft actuators for robotic rehabilitation: Methodologies for design, production and characterization, "Actuators, vol. 12, no. 7, p. 299, Jul. 2023.
[20] M. Schaffner et al., "3D printing of robotic soft actuators with programmable bioinspired architectures, "Nat. Commun., vol. 9, no. 1, p. 878, Mar. 2018.
[21] B. Mosadegh et al., "Pneumatic networks for soft robotics that actuate rapidly, "Adv. Funct. Mater., vol. 24, no. 15, pp. 2163–2170, Apr. 2014.
[22] H. K. Yap, H. Y. Ng, and C.-H. Yeow, "High-force soft printable pneumatics for soft robotic applications, "Soft Robot., vol. 3, no. 3, pp. 144–158, Sep. 2016.
[23] Z. Wang, J. Zhang, X. Li, and Y. Chen, "Closed-loop control of a 3D printed soft actuator with integrated flex sensors and SMA wires,"J. Phys. Conf. Ser., vol. 2716, no. 1, p. 012050, 2024.
[24] M. Li et al., "A 3D-printed soft hand exoskeleton with finger abduction assistance," in Proc. 16th Int. Conf. Ubiquitous Robots (UR), pp. 319-322, 2019.
[25] J. Malesevic et al., "Beagle—a kinematic sensory system for objective hand function assessment in technology-mediated rehabilitation, "IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 1817–1826, 2021.
[26] A. Completo et al., "The silicone metacarpophalangeal joint arthroplasty: An in-vitro analysis," Clin. Biomech., vol. 110, p. 106120, 2023.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Obstructive Sleep Apnea Detection from Single-Lead ECG Using Signal Processing and Machine Learning

Sona Semsari Parapari
Dept. of Electrical and Electronics Engineering
Istanbul Kültür University
Istanbul, Turkey
s.semsari@gmail.com

Esra Saatçi
Dept. of Electrical and Electronics Engineering
Istanbul Kütür University
Istanbul, Turkey
esra.saatci@iku.edu.tr

***Abstract*—Obstructive Sleep Apnea (OSA) is a common sleep disorder characterized by the temporary cessation of breathing during sleep. Polysomnography (PSG), considered the gold standard for OSA diagnosis, is costly and labor-intensive due to the requirement of recording multiple physiological signals. In this study, a non-invasive method for OSA detection using only single-lead Electrocardiogram (ECG) signals is proposed. The ECG signals were denoised using a bandpass filter and the Savitzky–Golay filter to reduce noise and artifacts. R-peaks were detected using the Pan–Tompkins algorithm, and R-R interval time series were derived. ECG-derived respiration (EDR) signals were extracted in three forms: QRS area variation, QRS slope, and R-peak amplitude. In addition to time and frequency domain features, cardiopulmonary coupling (CPC) features based on Continuous Wavelet Transform (CWT) were also obtained. Classification was performed using the Random Forest algorithm, and feature selection was conducted via the minimum redundancy maximum relevance (mRMR) method. The model was applied to identify the presence or absence of apnea events on a per-minute basis during sleep. Using the top 10 selected features, an accuracy of 88.99%, a sensitivity of 0.878, and a specificity of 0.898 were achieved. The results indicate that this ECG-based approach may serve as an effective alternative for OSA detection.**

Keywords — Obstructive Sleep Apnea; EDR; R-R intervals; CWT; Random Forest; Machine Learning

I. INTRODUCTION

Sleep apnea is a sleep disorder characterized by recurrent episodes of complete or partial cessation of breathing during sleep, which is prevalent and may lead to significant adverse health outcomes. It is broadly classified into three subtypes: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea (MSA). OSA arises from physical obstruction of the upper airway, whereas CSA results from a deficiency in the neural signals that activate the respiratory muscles [1]. MSA, on the other hand, is characterized by the coexistence of both obstructive and central events within the same episode or during the same sleep period. Among these, OSA is recognized as the most common subtype.

OSA causes recurrent apnea, oxygen desaturation, and sleep fragmentation, leading to inadequate and non-restorative sleep. This condition has been associated with cardiovascular diseases, cognitive impairments, car accidents, and a marked decline in quality of life. Globally, approximately one billion individuals are affected by OSA when using a diagnostic threshold of five or more events per hour. The prevalence of OSA is known to increase with age and is strongly associated with obesity rates [2].

Polysomnography (PSG) is regarded as the gold standard for the diagnosis of OSA; however, its high cost, time-consuming nature, and limited accessibility necessitate the development of more practical and widely available diagnostic alternatives [3]. In this context, various approaches have been proposed, including home sleep apnea testing (HSAT) systems, smart wearable technologies, pulse oximetry, and automated detection methods based on single-lead ECG signals [4]. Among these, ECG-based apnea detection methods stand out due to their low cost, non-invasive nature, and applicability to large patient populations, and they can also be integrated with other diagnostic approaches to enhance practicality and improve accuracy.

In 1984, Guilleminault et al. identified a potential association between heart rate variability observed in ECG signals and sleep apnea, thereby establishing the groundwork for subsequent research on ECG-based apnea detection [5]. Since then, the performance of various algorithms has been evaluated using signal processing techniques such as Fourier analysis [6] and wavelet transform [7]. These studies demonstrated that variations in R–R intervals are effective in distinguishing between apneic and non-apneic periods. For example, classifiers such as Support Vector Machines (SVM) and k-Nearest Neighbors (kNN) have achieved accuracies of up to 89% [8]. More recently, multi-task learning frameworks that process ECG and respiratory effort signals simultaneously have yielded promising results in predicting the Apnea–Hypopnea Index (AHI) [9].

One of the key components in this field is heart rate variability (HRV), which represents the natural variation in R–R intervals and serves as an indicator of autonomic nervous system activity. Reduced HRV levels have been reported to be associated with OSA, and classifiers based on HRV have achieved accuracies of up to 85% [10]. Similarly, the EDR signal, obtained from morphological variations in the QRS complex without the need for direct respiratory measurement, has been considered an effective metric for apnea detection [11].

Another prominent approach is cardiopulmonary coupling R–R intervals and EDR time series. CPC enables the assessment of

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

sleep quality and apnea detection, offering complementary capabilities in situations where HRV or EDR alone may be insufficient. Recently, spectral features obtained using the SST-CPC method proposed by Wang et al. have been applied in apnea classification through various machine learning algorithms [13].

Despite these advances, there remains a need for methods that are simultaneously accurate, low-cost, and practical for home use. To address this, we propose an approach for detecting OSA using only R–R intervals and EDR signals derived from single-lead ECG recordings. By integrating signal processing, CPC analysis, and machine learning, our method aims to provide a portable and accessible alternative to existing diagnostic tools, while offering improved applicability in real-world home environments.

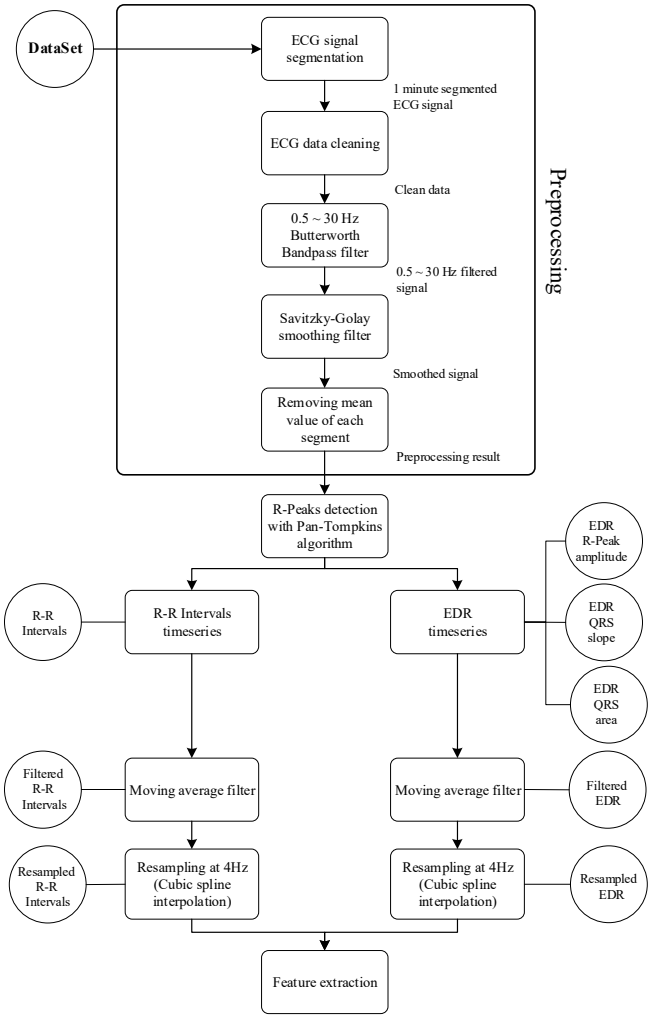


Figure 1: Preprocessing and feature extraction block diagram.

II. MATERIALS AND METHODS

In this section, the database and the signal processing methods employed are described. Figure 1 presents the preprocessing and feature extraction steps in a flow diagram. The subsequent section details the machine learning methodology.

A. Dataset

Real ECG data were obtained from the publicly available Physionet Sleep Apnea dataset [14] (http://www.physionet.org/physiobank/database/apnea-ecg/). The dataset consists of a total of 70 single-lead ECG recordings, which are evenly divided into training and test sets, each containing 35 recordings. The recordings have a sampling frequency of 100 Hz. All recordings were manually annotated on a per-minute basis, indicating whether each minute corresponded to a normal or apnea-affected period. The dataset comprises a total of 34,278 labeled minutes, including 13,064 minutes affected by apnea and 21,214 minutes without apnea.

B. Preprocessing

During the preprocessing stage, the dataset was initially segmented into one-minute intervals. Subsequently, to exclude segments containing high-amplitude, transient artifacts, segments with excessive amplitude within each recording were removed, and a 300-minute clean segment was selected from the middle portion of each recording. Following this, a zero-phase Butterworth bandpass filter with a passband of 0.5–30 Hz was applied to eliminate noise in both low and high-frequency ranges. Finally, a Savitzky–Golay filter was employed to smooth the signal. Figure 2 illustrates the effect of the filtering procedures on a representative one-minute signal segment.

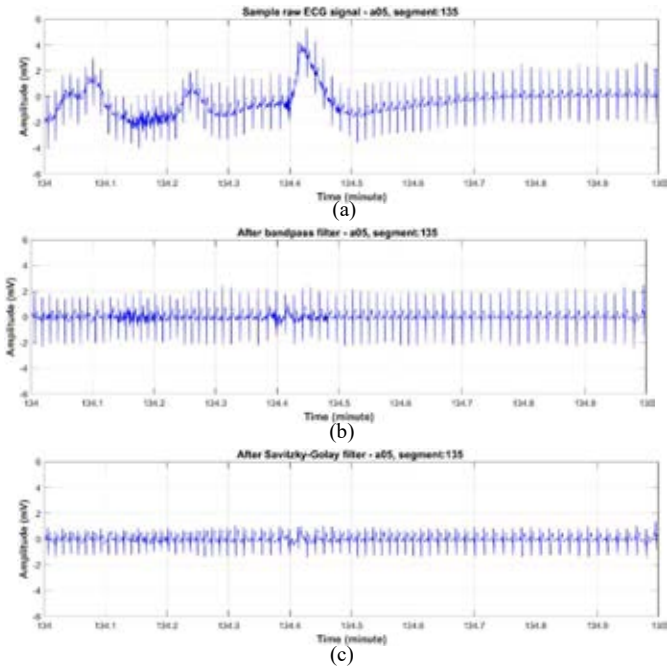


Figure 2:(a) Raw ECG signal, (b) bandpass filtered (c) smoothed ECG signal

C. R-R Intervals and EDR

Following preprocessing, R-peaks were automatically detected using the Pan–Tompkins algorithm, and the intervals between consecutive peaks were computed as R–R intervals. Based on the detected R-peaks and preprocessed ECG signals, the EDR time series were extracted in three forms: QRS area variation, QRS slope, and R-peak amplitude. Both the R–R intervals and EDR signals were then resampled at 4 Hz using cubic spline interpolation. The resulting signals are illustrated in Figure 3.

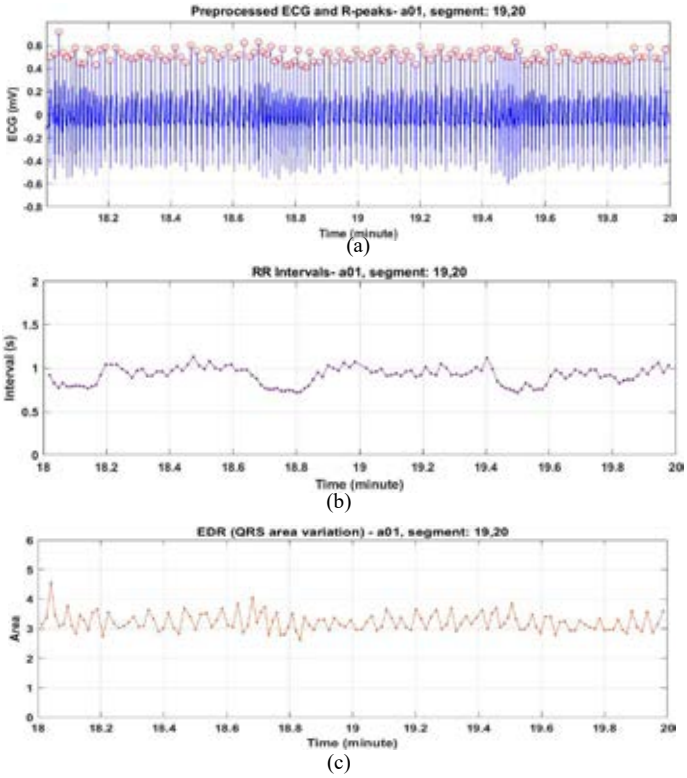


Figure 3: Sample a01, segment 19 and 20, (a) detected R-peaks on preprocessed ECG signal, (b) R-R intervals (c) EDR (QRS area variation)

D. Feature Extraction

For the detection of OSA using ECG signals, a comprehensive set of 33 baseline features reflecting both cardiac and respiratory dynamics was extracted. To improve classification performance, CPC features, representing the interaction between R–R intervals and ECG-derived respiration (EDR) signals, were additionally incorporated. In this context, the continuous wavelet transform (CWT) was employed to compute CPC features from three different EDR derivation methods, resulting in 27 features in total (9 per method). Consequently, 60 features were extracted for use in classification.

E. Machine Learning

In this study, a machine learning–based Random Forest algorithm was employed for apnea classification. This method reduces variance and enhances model generalizability by constructing a large number of decision trees and aggregating their outputs [15]. After partitioning the dataset into 80% training and 20% testing subsets, model evaluation was conducted using cross-validation. Classification was performed on three distinct feature sets: (1) time–frequency based features, (2) continuous wavelet transform–based cardiopulmonary coupling (CWT-CPC) features, and (3) a combined feature set. During each cross-validation iteration, feature selection was carried out using the Minimum Redundancy Maximum Relevance (mRMR) algorithm. As illustrated in Figure 4, overfitting occurred when the TopK value exceeded 20. Accordingly, TopK values of 10 and 15 were selected, yielding six classification results across the three feature sets. For the case of 15 features selected from the combined feature set, the following features were identified: SDSD, mean EDR area, peak-to-peak, mean ECG, standard deviation of EDR area, SD2, SD ratio, standard deviation of EDR slope, LF/HF (CPC-slope), LF/HF (CPC-area), PSE, minimum RR, minimum ECG, standard deviation of RR, and mean EDR slope.

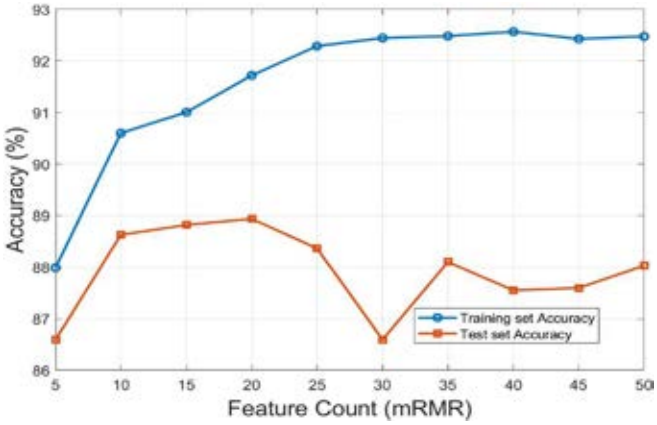


Figure 4. Variation of cross-validation accuracy on the training set (blue curve) and test set accuracy red curve) with respect to the number of features.

III. RESULTS AND DISCUSSION

Tables I and II summarize the classification results for TopK values of 10 and 15. The combined feature set consistently yielded the best performance, achieving a test accuracy of 88.99%, sensitivity of 0.878, and specificity of 0.898 at TopK=15. This indicates that integrating CPC features with time–frequency features enhances the discrimination between apneic and non-apneic segments.

Table I. Classification results for $TopK = 10$

Criteria	Time Frequency Features	CWT-CPC Features	Combined Features
Training Accuracy	90.13%	79.83%	90.19%
Test Accuracy	86.76%	80.20%	88.77%
Sensitivity (Recall)	0.8329	0.8072	0.8895
Specificity	0.4702	0.3628	0.8867

Table II. Classification results for $TopK = 15$

Criteria	Time Frequency Features	CWT-CPC Features	Combined Features
Training Accuracy	90.44%	80.40%	90.94%
Test Accuracy	85.64%	81.42%	88.99%
Sensitivity (Recall)	0.7918	0.8246	0.8779
Specificity	0.54	0.3443	0.897

As shown in Table II, the highest classification performance was achieved when the combined feature set was used at TopK= 15.

A comparative evaluation with existing studies is presented in Table III. The proposed method outperformed several conventional machine learning approaches [10,13,16]. While deep learning methods, such as the CNN-based model by Thompson [17], achieved higher accuracy (95.28%), the proposed method demonstrated competitive performance with the advantages of reduced computational complexity, interpretability, and suitability for portable applications.

Table III. Comparison of similar studies and their results

No	Author	Reference	Dataset	Feature	Method	Accuracy (%)	Sensitivity (Recall)	Specificity
1	Wang	[13]	PhysioNet Apnea-ECG Dataset	R-R intervals/ EDR/SST-CPC	Decision tree	82.49	0.781	0.757
2	Tuncer	[16]		CWT	Random forest	82.43	N/A	N/A
3	Mendez	[10]		HRV	kNN	85.55	0.839	0.885
4	Liu	[19]		ECG signals	Random forest	88.4	0.87	N/A
5	Thompson	[17]		ECG signals	1D-CNN	95.28	0.961	0.973
6	Nabil	[18]		ECG signals	1D-CNN	93	0.95	N/A
-	Proposed Method	-		R-R intervals EDR (amplitude, slope, area) CWT-CPC	Random forest	88.99	0.878	0.898

IV. CONCLUSION

In this study, a machine learning–based approach for obstructive sleep apnea detection was proposed using R–R intervals and EDR signals from single-lead ECG recordings. A comprehensive feature set was constructed by combining time–frequency features with CWT-based cardiopulmonary coupling (CPC) features, followed by mRMR-based feature selection and Random Forest classification. The combined feature set achieved the highest performance, with an accuracy of 88.99%, sensitivity of 0.878, and specificity of 0.898, outperforming several conventional methods. Although deep learning may provide higher accuracy, the proposed method demonstrated practical benefits in terms of computational efficiency, interpretability, and feasibility for portable and home-based applications.

Despite these strengths, limitations remain due to the inherent constraints of machine learning techniques and the relatively small and limited PhysioNet Apnea-ECG dataset, which may not fully represent variability across populations. Moreover, since CPC features had tiny improvements in results, their inclusion may be less beneficial in future work. Moving forward, research should focus on larger and more diverse clinical datasets, including hospital-sourced data, and the integration of advanced deep learning architectures. These efforts aim to enable reliable, efficient, and widely deployable diagnostic tools for the early detection and management of sleep apnea in both clinical and home-care environments.

ACKNOWLEDGMENT

This work is partially supported by the Scientific Research Projects Coordination Unit of Istanbul Kültür University.

References

[1] M. K. Moridani, M. Heydar, and S. S. Jabbari Behnam, "A reliable algorithm based on combination of EMG, ECG and EEG signals for sleep apnea detection," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, 2019, pp. 256–262, doi: 10.1109/KBEI.2019.8734992.

[2] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *Am. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, 2013, doi: 10.1093/aje/kws342.

[3] K. H. Levin and P. Chauvel, "Jessica Vensel Rundo and Ralph Downey III," in *Clinical Neurophysiology: Basis and Technical Aspects*, vol. 160. Elsevier, 2019, p. 381.

[4] I. M. Rosen *et al.*, "Clinical use of a home sleep apnea test: An updated American Academy of Sleep Medicine position statement," *J. Clin. Sleep Med.*, vol. 14, no. 12, pp. 2075–2077, 2018.

[5] C. Guilleminault, S. Connolly, R. Winkle, K. Melvin, and A. Tilkian, "Cyclical variation of the heart rate in sleep apnoea syndrome: Mechanisms, and usefulness of 24 h electrocardiography as a screening technique," *Lancet*, vol. 323, no. 8369, pp. 126–131, 1984, doi: 10.1016/S0140-6736(84)90062-X.

[6] M. Drinnan, J. Allen, P. Langley, and A. Murray, "Detection of sleep apnoea from frequency analysis of heart rate variability," in *Comput. Cardiol. 2000*, vol. 27, Cambridge, MA, USA, pp. 259–262, doi: 10.1109/CIC.2000.898506.

[7] A. Hossen, "A soft decision algorithm for obstructive sleep apnea patient classification based on fast estimation of wavelet entropy of RRI data," *Technol. Health Care*, vol. 13, no. 3, pp. 151–165, 2005.

[8] B. Yilmaz, M. H. Asyali, E. Arikan, S. Yektin, and F. Ozgen, "Sleep stage and obstructive apneaic epoch classification using single lead ECG," *Biomed. Eng. Online*, vol. 9, 2010.

[9] T. Osa, Y. Morita, T. Tsuji, *et al.*, "A multi-task learning model using RR intervals and respiratory effort to assess sleep disordered breathing," *Biomed. Eng. Online*, vol. 23, p. 16, 2024, doi: 10.1186/s12938-024-01240-0.

[10] M. O. Mendez, D. Ruini, O. Villantieri, M. Matteucci, T. Penzel, and A. Bianchi, "Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 6105–6108.

[11] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," in *Comput. Cardiol.*, vol. 12, 1985, pp. 113–116.

[12] R. J. Thomas, J. E. Mietus, C.-K. Peng, and A. L. Goldberger, "An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep," *Sleep*, vol. 28, no. 9, pp. 1151–1161, Sep. 2005.

[13] Y. Wang, W. Shi, and C. H. Yeh, "A novel measure of cardiopulmonary coupling during sleep based on the synchrosqueezing transform algorithm," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, pp. 1790–1800, Apr. 2023.

[14] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Proc. Comput. Cardiol.*, vol. 27, 2000, pp. 255–258.

[15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[16] E. Tuncer, "Development of new machine learning based algorithm for the diagnosis of obstructive sleep apnea from ECG data," *J. Comput. Sci. Res.*, vol. 5, no. 3, pp. 15–21, 2023.

[17] S. Thompson, P. Fergus, C. Chalmers, and D. Reilly, "Detection of obstructive sleep apnea using features extracted from segmented time series ECG signals using a one-dimensional convolutional neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2020, pp. 1–8.

[18] R. I. Nabil et al., "End-to-end sleep apnea identification from raw ECG signals utilizing 1D convolutional neural networks," in Proc. Int. Conf. Quantum Photon., Artif. Intell., Netw. (QPAIN), Rangpur, Bangladesh, 2025, pp. 1–6.

[19] X. Liu and W. Liu, "Innovative machine learning approaches for real-time detection of obstructive sleep apnea using ECG signals in medical wireless systems," in Proc. Cross Strait Radio Sci. Wireless Technol. Conf. (CSRSWTC), Macao, China, 2024, pp. 1–3.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Classifying Age-Specific Stages of Cervical Vertebral Maturation Using Deep Feature Analysis

Banu Akkuş*, Shireen Al-Ofary*, Hamza Osman İlhan*, Yasin Yaşa†
*Yıldız Technical University, Istanbul, Turkey
Email: banu.akkus@std.yildiz.edu.tr, shireen.alofari@std.yildiz.edu.tr, hoilhan@yildiz.edu.tr
†Ordu University, Ordu, Turkey
Email: yasinyasa@odu.edu.tr

***Abstract*—The cervical vertebrae play a critical role in the structural and functional integrity of the human body. Their morphology is closely linked to spinal pathologies, injuries, and developmental conditions. This study investigates the classification of cervical vertebral maturation (CVM) stages using three convolutional neural networks (CNNs): ShuffleNet, AlexNet, and DenseNet. A dataset of 1018 cephalometric radiographs from patients aged 6 to 56 years was analyzed and categorized into three age-based stages. Each image was annotated to ensure accurate labeling of C2, C3, and C4 vertebrae. Features extracted by the CNNs were further processed using Principal Component Analysis (PCA) and classified with Support Vector Machines (SVM). The highest performance was achieved with ShuffleNet features combined with PCA and SVM, reaching an accuracy of 72.69%.**

***Keywords*—Deep Feature Analysis, SVM, PCA, Cervical Vertebrae Maturation (CVM)**

I. Introduction

The timing of orthodontic or orthopedic treatment is critical for addressing craniofacial skeletal abnormalities. An accurate assessment of skeletal maturity helps identify the optimal treatment period. Rather than relying on chronological age, radiographic methods such as cervical vertebral maturation (CVM) staging on lateral cephalometric radiographs offer a more reliable estimate of skeletal development [1]. Numerous studies have validated the reliability of CVM staging for estimating growth potential [2], [3], [4].

Cervical vertebrae are key anatomical structures that support head movement and spinal protection. In CVM analysis, morphological changes in C2, C3, and C4, particularly at their superior and inferior borders, are critical indicators of development. These vertebrae are distinct for their small size and the presence of transverse foramina, which carry blood to the brain. Age estimation holds significant value in forensic science for identifying deceased individuals and ensuring legal clarity on age status [5]. It also plays a vital role in medicine, supporting orthodontic diagnosis and treatment planning during growth spurts, and aiding the detection of metabolic and endocrine disorders. Methods include evaluating skeletal and dental development, as well as secondary sexual characteristics.

The powerful pattern recognition ability of the human brain has inspired artificial intelligence techniques aimed at automating tasks such as age, gender, and race classification using medical images [6]. Machine learning, especially deep learning, has become an effective tool in estimating age based on bone morphology, and has also been applied for chronological age estimation from lateral cephalometric radiographs [7].

This study uses lateral cervical radiographs to classify vertebral development stages of C2, C3, and C4 in individuals aged 6 to 56 years in three age categories. Three pre-trained CNN models, ShuffleNet, AlexNet, and DenseNet, were used to extract features. These features were refined using Principal Component Analysis (PCA) and classified using both neural networks and Support Vector Machines (SVM), with SVM yielding the most accurate results.

II. Related Work

Atici et al. developed a deep learning approach using directional filters to emphasize vertebral borders in cephalometric X-ray images. Using 1,018 radiographs, their model achieved an accuracy of 84.63% for the classification of five classes of CVM and 75.11% for the classification of six classes [2].

Seo et al. compared six CNN models (ResNet18, MobileNet-v2, ResNet-50/101, Inception-v3, and Inception-ResNet-v2) for CVM stage classification using Grad-CAM for visualization. All models exceeded the accuracy 90%, with Inception-ResNet-v2 showing the highest performance [3].

Kim et al. evaluated an automated bone age assessment system using hand radiographs of 200 patients (ages 3–17). Despite the limited sample size, the system showed 69.5% concordance with the reference ages, supporting the feasibility of deep learning in clinical settings [8].

Rahimi et al. trained a transfer learning model on 890 cephalograms, focusing on vertebrae C2–C4 to classify stages of CVM into three and six categories. Their model achieved 82.8% accuracy for the three-class problem and 61.62% for six-class CVM classification [4].

Al-Ofary and Ilhan applied transfer learning to classify colon and lung cancer datasets. They leveraged pre-trained lightweight convolutional neural networks, namely ShuffleNet, AlexNet, and DenseNet. In three separate scenarios, the first step involved using pre-trained models individually to classify the datasets. Subsequently, the features extracted by these

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

pre-trained networks served as feature extractors, and these extracted features were fed into a Support Vector Machine (SVM) classifier. Lastly, feature selection was performed using Principal Component Analysis (PCA) on the extracted features from the pre-trained models to reduce the number of features and enhance classification performance. Using PCA-based feature selection with pretrained CNNs, the highest classification accuracies—99.93% for colon and 97.92% for lung cancer images—were achieved [9]. This study demonstrated the effectiveness of the deep feature analysis of CNNs in biomedical images.

III. Materials and Methods

Three pretrained CNN models were utilized to extract features, which were then classified after significant features were selected using PCA. This process was carried out using a data set comprising cephalometric images of the C2, C3 and C4 cervical vertebrae. Patients were grouped into age categories according to the developmental status of these vertebrae.

A. Dataset Information

In this study, cephalometric X-ray images of cervical vertebrae C2, C3, and C4 were used to classify patients according to their stage of vertebral maturation. The data set was gathered from the Faculty of Dentistry of Ordu University. In addition, the data set consisted of an Excel table containing patient details such as name, age, sex, and vertebrae maturation stage. Ethical approval for this study was granted by the Clinical Research Ethics Committee of Ordu University under Decision No: 58.

The data set comprised three classes (Class 1, Class 2, and Class 3), which were classified according to the maturation stage of the vertebrae. In the CVM literature, six stages (Stage 1–Stage 6) are defined to describe vertebral maturation. However, due to class imbalance and the limited number of samples in certain stages, these six stages were grouped into three broader categories in this study: Stage 1 and Stage 2 were combined as Class 1, Stage 3 and Stage 4 were combined as Class 2, and Stage 5 and Stage 6 were combined as Class 3. This grouping was adopted to ensure a balanced distribution of samples across classes while maintaining clinical relevance in the classification. However, the data set exhibited an imbalance, with Class 1 containing 280 images, Class 2 containing 655 images, and Class 3 containing 2090 images. To rectify this imbalance, the data set was split into five cross-validations, with each class having both training and test sets. The training data set was augmented from Class 1 to Class 3 to ensure a balanced representation in all classes.

The images were already cropped around the vertebral region, focusing exclusively on the regions of C2, C3 and C4. However, the resolution of the images was low, which posed limitations in obtaining high-resolution results. Example images for each class are shown in Fig. 1.

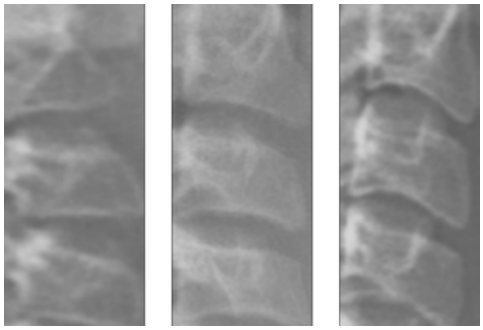


Fig. 1: Example images for each class - 8 years old class 1 (left), 14 years old class 2 (middle), 19 years old class 3 (right)

B. Transfer Learning

The study of cervical vertebrae is vital for diagnosing spinal conditions through CT and X-ray imaging, which is essential for fields like physical therapy, neurology, and orthopedics. This research employed three CNN architectures—ShuffleNet, AlexNet, and DenseNet201—to extract vertebral maturation data, utilizing transfer learning with pretrained model weights from other domains.

These models were chosen due to their complementary architectural characteristics: ShuffleNet is known for its computational efficiency and suitability for mobile and embedded devices; AlexNet introduced fundamental deep learning concepts and remains a baseline for CNN performance; and DenseNet201 ensures feature reuse through dense connections, leading to improved information flow and parameter efficiency. By comparing these architectures, the study aimed to explore how model depth and connectivity influence feature extraction and classification performance in medical imaging tasks.

ShuffleNet improves efficiency with group convolution and channel shuffle, AlexNet provides a simple baseline with fundamental CNN concepts, and DenseNet201 enhances feature reuse through dense connections.

These models were selected due to their balance of computational efficiency (ShuffleNet), foundational relevance and historical significance in CNN research (AlexNet), and feature reuse capabilities with low computational cost (DenseNet). This allows a comparison across different network complexities and architectural innovations. For the training process, hyperparameters were set empirically based on preliminary experiments. Specifically, the learning rate was set to 0.001, batch size to 32, and the number of epochs to 100. No automated hyperparameter optimization method such as grid search or Bayesian optimization was applied. These values were selected to ensure model convergence while avoiding overfitting.

C. Principal Components Analysis (PCA)

PCA was applied to reduce the dimensionality of the extracted features while preserving 95% of the variance. This step ensured that essential information was retained and computational complexity reduced before classification.

D. Performance Metrics

Evaluation of the model incorporated several metrics, namely accuracy, precision, recall, and the F1 score. These

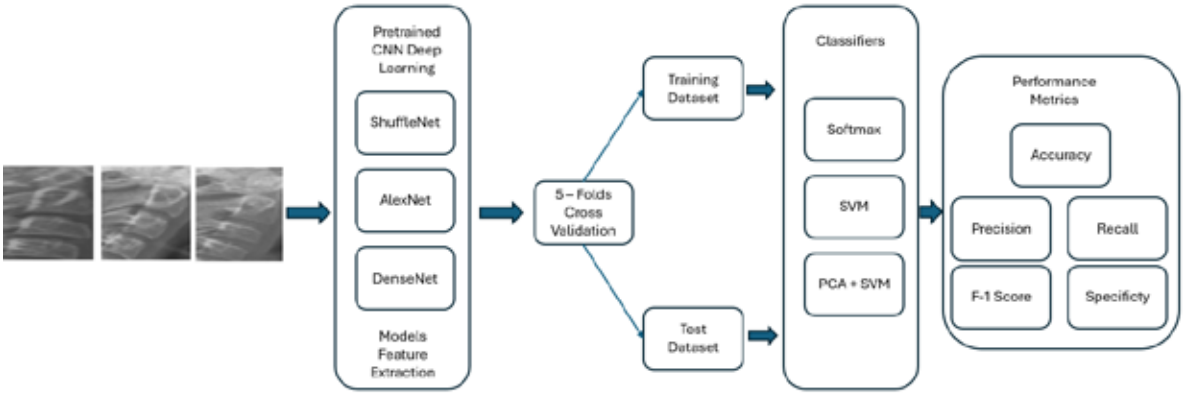


Fig. 2: Flowchart of the study

metrics were derived from an analysis of true positives, true negatives, false positives, and false negatives within the dataset. Notably, the F1 score proves particularly advantageous when it comes to balancing precision against recall, a critical consideration in the comparative assessment of various models.

E. Methodology

In this study, three convolutional neural networks were used to classify a set of cephalometric X-ray images of cervical vertebrae C2, C3, and C4 and categorize a group of patients by their ages based on the development of these vertebrae into three categories. Each category covers a range of ages from 6 years to 56 years.

In the first stage, three convolutional neural networks were used to classify the data set. The features were extracted within these networks. As no modifications were made to the network, features were taken from the last convolutional layer and the data were classified into three categories.

In the second stage, only the features extracted in the previous step were utilized to classify the data differently using classical machine learning, specifically employing the Support Vector Machine (SVM) technique. To compare the results of applying deep learning and classical machine learning, SVM demonstrated its ability to classify rapidly and with high accuracy, particularly given that the features were obtained from convolutional neural networks, renowned for their effectiveness as feature extractors.

In the third step, Principal Component Analysis (PCA) was applied to extract only the significant features from the images. PCA selected the important features from images without affecting the original data by transforming a new dataset containing the important features; this technique helps reduce time, not wasting effort on irrelevant or redundant features in the image and focusing on the most informative ones.

These features were then reused for classification using the SVM technique, and the best results were obtained when this approach was applied in the final scenario. The general flow chart of the study is shown in Figure 2.

IV. Experimental Results

The Google Colab environment was utilized for all experiments with the Python Keras library. Table I presents the accuracy scores of the models under the three scenarios on the cervical vertebrae dataset.

TABLE I: The classification performances of the models under the three scenarios: i) Individual CNN Results, ii) SVM results over CNN derived Features, iii) PCA applied reduced Feature classification results

Accuracy (%)	***SoftMax***	***SVM***	***SVM + PCA***
ShuffleNet	68.32	69.70	72.69
AlexNet	68.73	64.36	71.04
DenseNet	69.14	70.19	71.16

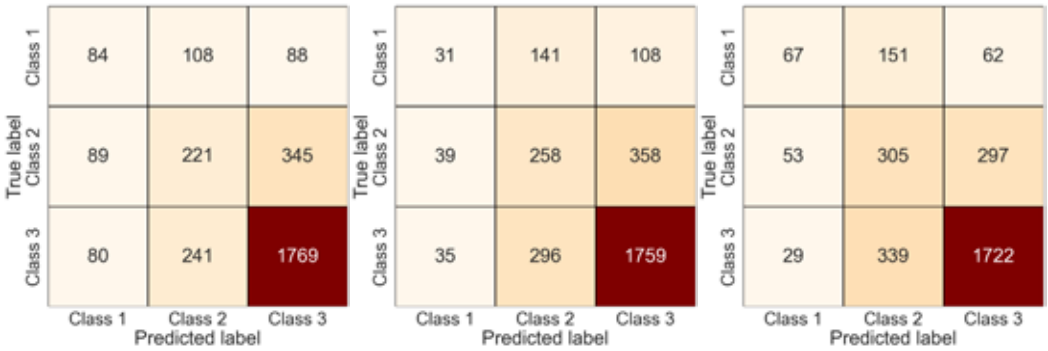
Upon examining the individual performance of pre-trained networks, based on the SoftMax classification results presented in Table I, ShuffleNet, AlexNet, and DenseNet, achieved accuracy scores of 68.32%, 68.73%, and 69.14% for cervical vertebrae classification, respectively. The confusion matrices for the first scenario is presented in Figure 3.

In the second experimental case, the feature extractor was a network, and then the SVM classifier was supplied with the features. The accuracy scores for ShuffleNet, AlexNet and DenseNet are 69.7%, 64.36%, and 70.19%, respectively.

In the last case, uninformative features were eliminated using PCA. Then, similar to the second case, the condensed feature sets were provided to SVM. The highest scores in this scenario were attributed to categorization performance, which achieved accuracy rates of 72.69%, 71.04%, and 71.16%, respectively.

TABLE II: SVM Classification Results of PCA reduced ShuffleNet Features

%	***Precision***	***Recall***	***F1-Score***	***Support***
Class 1	68.32	69.70	72.69	280
Class 2	68.73	64.36	71.04	655
Class 3	69.14	70.19	71.16	2090
Overall Accuracy	72.69			



(a) DenseNet-201 (b) ShuffleNet (c) AlexNet

Fig. 3: Confusion matrices of individual network classification results.



Fig. 4: Confusion Matrix of Best Result (SVM classification over PCA reduced ShuffleNet Features)

Figure 4 presents the confusion matrices of the best outcomes that ShuffleNet obtained using SVM and PCA. Furthermore, Table II lists additional performance metrics, including precision, recall, and F1 scores derived from the corresponding confusion matrix.

V. Conclusion

In this study, three deep learning networks—ShuffleNet, AlexNet, and DenseNet—were used to classify the developmental stages of cervical vertebrae (C2, C3, C4) in the cervical spine for a group of patients. The objective was to observe how these vertebrae develop with age. Due to limitations in the number of patients in certain age groups, the age ranges were grouped. These networks were employed as pre-trained models using transfer-learning techniques.

In the first study, convolutional neural networks (CNN) functioned as classifiers for radiographic images of the cervical vertebrae and each model showed strong classification performance, as indicated by SoftMax scores. Then, only the features extracted from the images by the pre-trained models were collected and entered into a Support Vector Machine (SVM) classifier. Finally, Principal Component Analysis (PCA) was applied to the features extracted from the networks to reduce their dimensionality. Following this feature selection process, the reduced feature sets were provided to the SVM to improve classification accuracy. Remarkably, the final scenario yielded the best results when SVM with PCA was applied, using features extracted by ShuffleNet, reaching an overall accuracy of 72.69% (see Table I and Table II).

These findings highlight that the proposed framework offers not only competitive accuracy but also practical advantages. In particular, the use of ShuffleNet with PCA and SVM demonstrates a lightweight and computationally efficient solution, which is valuable for real-world clinical settings where fast, low-cost, and resource-friendly analysis is required. Future work will fix data imbalance/resolution limits, add tuning and AUC, and test decision-layer classification.

References

[1] E.-G. Kim, I.-S. Oh, J.-E. So, J. Kang, V. N. T. Le, M.-K. Tak, and D.-W. Lee, "Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network," J. Clin. Med., vol. 10, no. 22, pp. 5400, 2021.

[2] S. F. Atici, R. Ansari, V. Allareddy, O. Suhaym, A. E. Cetin, and M. H. Elnagar, "Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters," PLoS One, vol. 17, no. 7, pp. e0269198, 2022.

[3] H. Seo, J. Hwang, T. Jeong, and J. Shin, "Comparison of deep learning models for cervical vertebral maturation stage classification on lateral cephalometric radiographs," J. Clin. Med., vol. 10, no. 16, pp. 3591, 2021.

[4] H. Mohammad-Rahimi et al., "Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study," Korean J. Orthod., vol. 52, no. 2, pp. 112–122, 2022.

[5] A. Schmeling, A. Olze, W. Reisinger, and G. Geserick, "Age estimation of living people undergoing criminal proceedings," The Lancet, vol. 358, no. 9276, pp. 89–90, 2001.

[6] Y.-C. Guo, M. Han, Y. Chi, H. Long, D. Zhang, J. Yang, Y. Yang, T. Chen, and S. Du, "Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images," Int. J. Legal Med., vol. 135, pp. 1589–1597, 2021.

[7] N. Liu, "Chronological age estimation of lateral cephalometric radiographs with deep learning," arXiv preprint arXiv:2101.11805, 2021.

[8] J. R. Kim et al., "Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency," Am. J. Roentgenol., vol. 209, no. 6, pp. 1374–1380, 2017.

[9] S. Al-Ofary and H. O. Ilhan, "Classification of PCA based Reduced Deep Features by SVM for Diagnosing Lung and Colon Cancer," Proc. 5th Int. Congr. Human-Computer Interact., Optimization Robot. Appl. (HORA), pp. 1–5, 2023.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Patch-Based Deep Learning Approaches for Automated 3D Skull Segmentation in CT Imaging

Fatih Ekrem ONAT[2,3] *fonat@yildiz.edu.tr*
Yiğit KAHRAMAN[2] *yigit.kahraman@std.yildiz.edu.tr*
Berk ARIDAĞ[2] *berk.aridag@std.yildiz.edu.tr*
Melih Fırat BUDAK[2] *firat.budak@std.yildiz.edu.tr*
Caner BULUT[2] *caner.bulut@std.yildiz.edu.tr*
Berke APAYDINLI[2] *berke.apaydinli@std.yildiz.edu.tr*
Abdülkadir GÜNAY[4] *abdulkadir.gunay@std.yildiz.edu.tr*
Yunus Emre ÇAKMAKLI[2] *emre.cakmakli@std.yildiz.edu.tr*
Nuri Serdar BAŞ[1] *nuriserdarbas@gmail.com*
Hamza Osman İLHAN[3] *hoilhan@yildiz.edu.tr*
Görkem SERBES[2] *gserbes@yildiz.edu.tr*
Mihrigül EKŞİ ALTAN[2,4] *meksi@yildiz.edu.tr*

[1]*Sağlık Bilimleri Üniversitesi Kanuni Sultan Süleyman Eğitim ve Araştırma Hastanesi, İstanbul, Türkiye*
[2]*Biyomedikal Mühendisliği, Yıldız Teknik Üniversitesi, İstanbul, Türkiye*
[3]*Bilgisayar Mühendisliği, Yıldız Teknik Üniversitesi, İstanbul, Türkiye*
[4]*Makine Mühendisliği, Yıldız Teknik Üniversitesi, İstanbul, Türkiye*

***Abstract*—Accurate and efficient skull segmentation from cranial CT scans is a critical prerequisite for computer-aided cranial implant design, surgical planning, and postoperative assessment. In this study, we comparatively evaluate three state-of-the-art deep learning architectures (U-Net, Attention U-Net, and UNETR) for automated 3D skull segmentation. Experiments were conducted on a clinically curated novel dataset of 120 CT volumes encompassing diverse age groups and anatomical variations. Both patched and non-patched inference strategies were investigated to assess their impact on segmentation accuracy and robustness. Quantitative evaluation using a subject-based five-fold cross-validation protocol revealed that patched inference consistently outperformed non-patched processing, with Attention U-Net achieving the highest mean Dice Similarity Coefficient. Qualitative analysis demonstrated that patched inference produced anatomically complete and visually coherent skull reconstructions, particularly in challenging pediatric cases. The proposed approach shows strong potential for integration into clinical workflows, providing high-fidelity 3D reconstructions suitable for patient-specific cranial implant manufacturing.**

***Index Terms*—Skull segmentation, Cranial CT, Deep learning, Attention U-Net, UNETR, Patch-based inference**

I. Introduction

Skull segmentation (SS) in computed tomography (CT) imaging refers to automated extraction of cranial bone structures—such as the calvaria, facial bones, and sutures—from volumetric scans. This task underpins clinical workflows such as cranial implant design and reconstruction, where accurate skull delineation directly affects implant fit and surgical outcomes [1]. Besides, segmentation is an essential component in craniofacial and orthodontic planning, improving anatomical landmark localization and measurement consistency [2].

Image-processing techniques, including thresholding and active contour level-set, have been applied for bone extraction in CT. However, these approaches often underperform in the presence of noise, artifacts, or in pediatric scans with incomplete sutural fusion [3]. The advent of deep learning, especially convolutional neural networks (CNNs) like U-Net, has significantly advanced the field, thanks to their robust localization and efficient learning from limited data [4, 5].

Due to privacy regulations and the sensitivity of cranial imaging, particularly in children, publicly annotated SS datasets remain scarce. Most existing studies rely on small, private collections, making reproducibility and benchmarking challenging. For instance, Ghadimi et al. [8] proposed a method for segmenting and reconstructing neonatal skulls using coupled level sets on a dataset of 18 CT scans, highlighting the limited scale of available annotation resources. While later efforts such as Xi et al. [9] introduced a larger pediatric CT cohort, only few datasets contain manually labeled skull masks suitable for supervised learning. In SS process, transformer-based models such as UNETR have emerged to address the limitation of local receptive fields in CNNs by capturing long-range spatial dependencies in 3D volumes while retaining multi-scale context through a U-Net-like structure [6]. Attention-augmented CNNs, like Attention U-Net, further enhance SS performance by dynamically focusing on contextually relevant anatomical regions during feature fusion [7].

In cranial implant planning, accurate SS is a critical first

This study was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under the ARDEB 1001 - Scientific and Technological Research Projects Funding Program, with the project number 123E494.

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

step. Precise delineation ensures reliable reconstruction, supports anatomically consistent implant design, and reduces manual corrections. Thus, SS robustness directly impacts the success of the implant pipeline, especially where structural precision is essential. Our study conducts a comparative evaluation of three deep learning methods (U-Net, Attention U-Net, and UNETR) for SS. The models were analyzed in terms of their architectural design, spatial context modeling capabilities, and segmentation accuracy across a heterogeneous dataset encompassing a broad range of anatomical variations.

II. Materials and Methods

A. Dataset Construction and Preparation

A total of 120 cranial CT scans were collected from Bağcılar Training and Research Hospital under an approved institutional ethics protocol (Approval No: 2024/01/14/014). The dataset comprises both pediatric and adult cases, capturing a wide range of anatomical variability. All scans were initially annotated by an experienced radiologist and subsequently reviewed by a second senior radiologist to ensure clinical validity of the ground truth masks for bone structures. This dataset was used for training, validation, and comparative evaluation of segmentation models throughout the study. The demographic distribution of the dataset, including age statistics and gender ratios, is summarized in Table I.

TABLE I
DEMOGRAPHIC PROPERTIES OF THE DATASET.

Variable	Property	Value
Count	# Samples	120
Age (year)	min	0.33
	max	85
	mean	31.26
	median	26.5
	std.	24.89
Sex	Male (M)	66
	Female (F)	54

B. Data Augmentation

To improve model generalization and reduce overfitting, we applied a set of medical image-specific data augmentation. Training augmentations included spatial normalization orientation standardization, isotropic resampling, intensity scaling, and foreground cropping. In augmentation, firstly, spatial transformations including random cropping of $32 \times 32 \times 32$ patches with balanced positive-negative sampling, flipping along each spatial axis (x, y, z), random 90-degree rotations, and intensity shifting were applied. Secondly, the images were resized into $192 \times 192 \times 128$ without being divided into patches. All transformations were applied on-the-fly during training. Validation data underwent the same normalization pipeline without augmentation to ensure consistency across evaluations.

C. Segmentation Methods

Three state-of-the-art deep learning architectures (U-Net, Attention U-Net, and UNETR) were employed for 3D cranial CT segmentation. To investigate the effect of inference strategy on performance, each model was evaluated under two settings: patched inference, where the CT volume is processed in smaller subvolumes to preserve local anatomical detail, and non-patched inference, where the full volume is segmented in a single pass. This comparison enabled assessment of how localized context modeling impacts skull boundary delineation and overall SS quality.

1) U-Net Architecture: U-Net is a CNN that follows an encoder-decoder structure, originally designed for biomedical image segmentation [4]. Its simplicity, symmetry, and use of skip connections make it an effective architecture for segmenting anatomical structures such as the skull, where precise delineation of bony boundaries is required. The encoder path of U-Net consists of successive convolutional blocks, each containing two 3×3 convolution layers with ReLU activations, followed by 2×2 max pooling operations for spatial downsampling. The decoder mirrors this structure using 2×2 transposed convolutions to progressively upsample the feature maps. Skip connections directly link encoder and decoder feature maps at corresponding resolutions, facilitating the recovery of spatial details lost during downsampling. U-Net was selected in this study due to its strong localization capabilities, efficient learning on limited data, and proven effectiveness in biomedical segmentation tasks involving high-resolution structural boundaries. Skip connections reintroduce high-frequency boundary cues essential for thin cortical bone, making U-Net a strong baseline for cranial structures.

2) Attention U-Net Architecture: Attention U-Net extends the original U-Net architecture by introducing *Attention Gates (AGs)* into the skip connections [7]. These gates learn to suppress irrelevant feature responses and highlight salient structures relevant to the target task during feature fusion. This mechanism proves particularly beneficial in cranial segmentation tasks, where anatomical boundaries often exhibit low contrast and irregular morphology due to patient-specific anatomical variations or pathology.

3) UNETR Architecture: UNETR is a hybrid encoder-decoder architecture that integrates a Vision Transformer (ViT) as the encoder backbone within a U-Net-like framework [6]. Unlike conventional CNN encoders, UNETR leverages self-attention mechanisms to model long-range dependencies across volumetric inputs, which is particularly valuable for segmenting large or discontinuous cranial structures. The input volume is divided into non-overlapping patches of size $P \times P \times P$. Each patch is flattened, projected, and combined with positional embeddings to form a sequence of tokens. This sequence is processed by transformer blocks composed of multi-head self-attention and multilayer perceptron layers. The outputs from selected transformer layers are reshaped into spatial feature maps and integrated into the decoder through skip connections. The decoder follows the classical U-Net structure with transposed convolutions for upsampling.

D. Training Configuration and Evaluation Protocol

All deep learning models were trained using the AdamW optimizer with a learning rate of 10^{-4} and a weight decay

of 10^{-5}. The combined Dice and Cross-Entropy loss function was employed to supervise voxel-wise segmentation outputs, balancing overlap-based and probabilistic accuracy metrics. Each model was trained for 15,000 iterations with a batch size of 1 on 3D volumetric CT inputs. Training was performed on a workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM), 64 GB system RAM, and an Intel i9-13980HX CPU. Performance was evaluated using the Dice Similarity Coefficient (DSC) as the primary segmentation metric. A subject-based five-fold cross-validation protocol was implemented, ensuring that all scans from a patient were assigned exclusively to either the training, validation, or test set within each fold. This strategy prevented data leakage between folds and provided a realistic assessment of model generalizability across diverse patient anatomies. Final performance was reported as the average Dice score across all folds.

III. Experimental Results

Table II summarizes the quantitative performance of the evaluated segmentation models (U-Net, Attention U-Net, and UNETR) across both patched and non-patched inference settings. The primary evaluation metric was the DSC, reported as mean, standard deviation, and range statistics over the test set. Across all models, patched inference consistently outperformed non-patched inference, yielding higher mean DSC values and reduced variance. This improvement is likely attributable to the reduced memory footprint and increased local context consistency in smaller subvolumes, which can enhance boundary delineation for fine anatomical structures.

TABLE II
DSC STATISTICS FOR DIFFERENT MODELS AND SEGMENTATION TYPES.

Model	Type	DICE				
		Mean	Std	Min	Median	Max
Attention U-Net	Patched	**0.9335**	**0.0279**	0.8135	0.9367	0.9777
	Non-Patched	0.7836	0.0604	0.6106	0.7982	0.8691
U-Net	Patched	0.9211	0.0357	0.7549	0.9293	0.9710
	Non-Patched	0.7865	0.0538	0.6378	0.8031	0.8645
UNETR	Patched	0.9239	0.0406	0.7560	0.9341	0.9775
	Non-Patched	0.7812	0.0630	0.6059	0.7981	0.8678

Attention U-Net achieved the highest mean DSC (0.9335 ± 0.0279), followed by UNETR (0.9239 ± 0.0406) and U-Net (0.9211 ± 0.0357). The improvement of Attention U-Net over the baseline U-Net suggests that the attention gating mechanism effectively suppresses irrelevant features and emphasizes boundary-relevant structures—particularly beneficial for cranial bone contours where local intensity transitions can be subtle. In the non-patched setting, all models experienced a performance drop, with mean DSCs ranging between 0.7812 and 0.7865. U-Net achieved the highest score (0.7865), followed closely by Attention U-Net (0.7836) and UNETR (0.7812). This aligns with findings in prior studies [6, 7], where non-patched inference was shown to dilute local contrast information due to the dominance of global context. Fig. 1 visually confirms these trends, illustrating both the superiority of patched inference and the incremental gains from attention mechanisms. Notably, the performance bars demonstrate that patched models not only achieve higher mean DSC but also maintain more stable performance across the dataset.

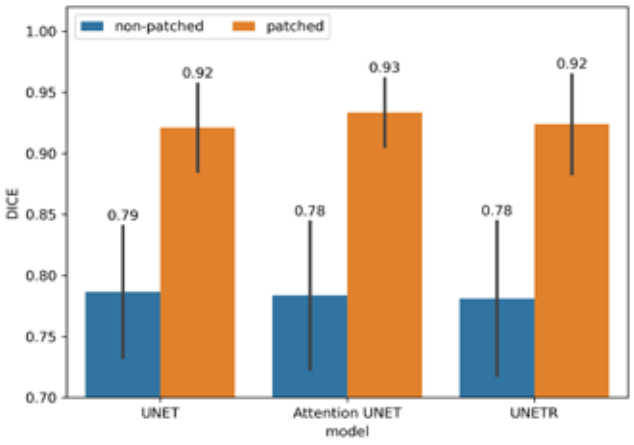


Fig. 1. The performance of all three models. Processing the images as patches has given significantly better performance than whole volumes.

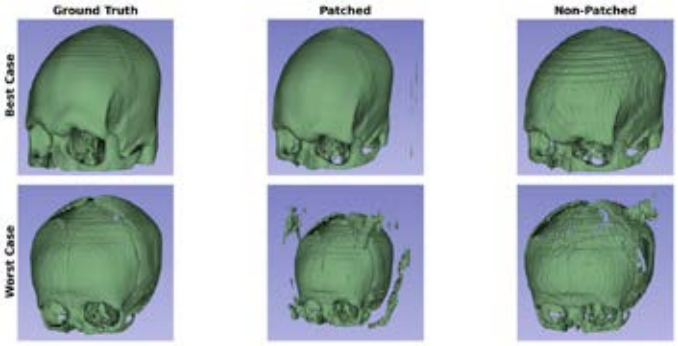


Fig. 2. The best case and the worst case for Attention UNET, compared to ground truth.

Given that Attention U-Net achieved the highest overall performance, qualitative results for this model are presented in Fig. 2. The best-case example corresponds to a 1-year-old infant scan, where the patched inference output exhibits near-perfect alignment with the ground truth. Cranial sutures and curved anatomical surfaces are reconstructed with high fidelity, and no visible over- or under-segmentation is observed. This demonstrates the model's ability to capture fine structural details when bone mineralization is relatively complete and sutures are well-defined. In contrast, the worst-case example involves a 4-month-old neonatal scan, where skull bones are thinner, sutures are wider, and fontanelles remain open. These characteristics result in reduced CT radiodensity contrast and more complex local intensity patterns, which present a significant challenge for automated SS. Additionally, in both patched and non-patched inferences, the presence of the infant's hand within the scan field further degraded segmentation quality by introducing non-cranial structures with similar intensity values. This led to false-positive regions and irregular boundaries in the reconstructed skull, particularly in the non-patched

setting where global context overshadowed fine local details. Patched inference partially mitigated these issues by providing the model with higher-resolution local context and reducing the influence of irrelevant anatomical regions.

IV. Discussion and Conclusion

The results demonstrate the effectiveness of deep learning methods for automated SS. Across three architectures (U-Net, Attention U-Net, and UNETR) and two inference strategies (patched and non-patched), all models produced anatomically coherent and clinically relevant skull masks. The patched Attention U-Net achieved the highest performance, delivering precise cranial boundary delineation while minimizing segmentation errors. Results indicate that, with suitable inference strategies, deep models can provide robust outputs for applications such as cranial implant design and surgical planning.

Among the evaluated architectures, the patched Attention U-Net achieved the highest mean DSC, confirming that attention mechanisms effectively suppress irrelevant background responses while enhancing boundary delineation. Although the transformer-based UNETR also demonstrated competitive results, the attention-augmented CNN maintained a slight advantage in this application, likely due to the localized and well-defined nature of cranial bone structures. These findings are consistent with prior work in cranial imaging: for instance, Xi et al. [9] reported a DSC of 0.96 in pediatric cranial CT segmentation using a ResU-Net, underscoring the feasibility of achieving near-expert accuracy. In our experiments, patched inference results (DSC $>$ 0.93) confirm that CNN-based architectures with attention remain highly competitive, combining accuracy with computational efficiency. Qualitative analysis further highlighted the benefits of patch-based inference, which produced more complete and coherent segmentations in anatomically challenging cases such as thin cranial bones, wide sutures, or open fontanelles. Even in neonatal scans containing non-cranial structures (e.g., an infant's hand within the CT field), patched inference reduced false positives by leveraging finer local context. Patched inference consistently yielded more coherent reconstructions by allowing models to focus on fine boundaries (e.g., sutures, thin cortices) that were often blurred in non-patched predictions. The performance gain arises because patches preserve local contrast, reduce the downsampling needed for non-patched inputs, and—with overlap and blending—mitigate block artifacts at tile borders.

Beyond segmentation benchmarking, the findings of this work highlight the potential of automated skull reconstruction frameworks in clinical cranial implant manufacturing. In trauma, tumor resection, or congenital defects, accurate 3D skull models enable the design of patient-specific implants with precise anatomical fit. From a clinical perspective, reliable and consistent SS forms the foundation for applications such as implant design, surgical planning, and postoperative assessment. Our results indicate that patch-based strategies, combined with attention mechanisms, offer a practical balance between accuracy and computational efficiency, making them suitable for integration into such workflows.

A potential refinement to enhance clinical applicability is Connected Component Analysis (CCA), which retains only the largest connected structure in the predicted mask. By removing smaller noise-induced components, CCA produces cleaner reconstructions, reducing false positives and improving the reliability of outputs for applications such as implant generation. Regarding the limitations, the dataset, although diverse in age range and anatomical variability, remains limited in size. Additionally, only three architectures were evaluated; exploring more recent hybrid or transformer-based variants could provide further insights. Finally, while DSC is a robust metric, future studies could incorporate surface distance or shape similarity measures to better assess clinical applicability. Future work may also explore hybrid CNN–Transformer decoders and assess whether the proposed patch-based strategy transfers to other modalities (e.g., MRI). In conclusion, patch-based inference combined with attention-enhanced architectures, particularly Attention U-Net, provides accurate and robust SS. The approach achieves high Dice scores and generates anatomically precise reconstructions, making it well-suited as the initial step in cranial implant design.

References

[1] M. Wodzinski, M. Daniol, M. Socha, *et al.*, "Deep learning-based framework for automatic cranial defect reconstruction and implant modeling," in *AutoImplant Challenge, MICCAI*, 2022.

[2] M. Gillot, "Automatic multi-anatomical skull structure segmentation of CBCT volumes using UNETR and MONAI," *PLOS ONE*, 2022.

[3] L. Zhou, A. Toga, Z. Xue, "An overview of bone segmentation in clinical computed tomography," *arXiv:1504.00221*, 2015.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[5] G. Litjens, T. Kooi, B. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[6] A. Hatamizadeh, V. Nath, Y. Tang, et al., "UNETR: Transformers for 3D Medical Image Segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2022, pp. 574–584.

[7] O. Oktay, J. Schlemper, L. Le Folgoc, et al., "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[8] S. Ghadimi, H. A. Moghaddam, R. Grebe, and F. Wallois, "Skull segmentation and reconstruction from newborn CT images using coupled level sets," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 563–573, 2016.

[9] F. Xi, L. Tu, F. Zhou, Y. Zhou, J. Ma, and Y. Peng, "Automatic segmentation and quantitative analysis of brain CT volume in 2-year-olds using deep learning model," *Frontiers in Neurology*, vol. 16, art. no. 1573060, 2025.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

HbA1c Estimation with Machine Learning and Implementation on Diabetes Self-Management Platform Based on CGM Measurements

Şaban Can Tiryaki
Dept. of Computer Engineering
Kocaeli University
İzmit, Türkiye
245112019@kocaeli.edu.tr

Adnan Kavak
Dept. of Computer Engineering
Kocaeli University
İzmit, Türkiye
akavak@kocaeli.edu.tr

Alpaslan Burak İnner
Dept. of Computer Engineering
Kocaeli University
İzmit, Türkiye
binner@kocaeli.edu.tr

***Abstract*— Diabetes self-management requires accurate monitoring of long-term blood glucose levels, typically measured by HbA1c. However, traditional laboratory tests provide delayed information and cannot capture real-time trends. This study presents a real-time approach based on Gradient Boosting based machine learning models to predict HbA1c from Continuous Glucose Monitoring (CGM) data to enable personalized diabetes management. CGM data set, i.e., a multimodal resource CGMacros dataset is used in this study. Data augmentation techniques including SMOTE for regression and jittering are employed to improve model learning. Simulation results show that among the models applied, CatBoost model resulted in the highest accuracy, robustness, and reliability in HbA1c prediction. This model is also implemented on a diabetes self-management mobile platform referred to as AIDCARE system via FastAPI, PostgreSQL, and Docker.**

Keywords*— *CGM; CatBoost; Diabetes Self-management; HbA1c Estimation; Machine Learning; XGBoost

I. INTRODUCTION

Accurate assessment of long-term glycemic control, most commonly evaluated through glycated hemoglobin (HbA1c) levels, is fundamental to effective diabetes management. However, conventional laboratory assessments offer only intermittent in-vivo measurements and may fail to capture short-term glycemic variability. Continuous Glucose Monitoring (CGM) measurements based HbA1c estimation may overcome this limitation by delivering frequent, real-time glucose data, thus facilitating more individualized and precise diabetes care strategies. [1].

Recent research has employed machine learning approaches, including Random Forest, XGBoost, and Support Vector Machines (SVM), to predict HbA1c levels from Continuous Glucose Monitoring (CGM) data, demonstrating potential in capturing the nonlinear relationships between glucose dynamics and HbA1c outcomes [2]. Nonetheless, limitations persist, particularly with respect to small sample sizes, class imbalance, and the insufficient exploitation of temporal features.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

In [3], a few-shot learning architecture is proposed for long-term HbA1c prediction [3]. While the model demonstrated reasonable predictive capability, it lacked real-time implementation and mobile platform deployment and did not employ data augmentation techniques such as SMOTE.

An explainable machine learning framework for real-time hypoglycemia and hyperglycemia prediction focusing on short-term glucose events rather than HbA1c prediction is presented [4]. A feature-based machine learning model is introduced in [5] for real-time hypoglycemia prediction. These studies did not incorporate mobile-friendly deployment or fast API-based services for real-time diabetes self-management. A deep transfer learning model is explored to improve blood glucose prediction in [6]. While this approach improved short-term glucose estimation, it did not specifically target HbA1c prediction based on short-term CGM data and also lacked mobile or real-time deployment capabilities.

This study proposes a real-time machine learning framework for predicting HbA1c from Continuous Glucose Monitoring (CGM) data, utilizing data augmentation strategies to mitigate class imbalance and extreme glucose values. Gradient boosting–based models, specifically XGBoost and CatBoost, are employed owing to their proven effectiveness in modeling nonlinear associations and handling high-dimensional, noisy biomedical data. The proposed gradient boosting framework is further integrated into AIDCARE, a mobile-enabled, AI-powered diabetes self-management platform, to facilitate real-time HbA1c prediction and personalized diabetes monitoring.

The remainder of this paper is structured as follows. Section II provides a detailed description of the dataset utilized in this study. Section III outlines the HbA1c estimation problem and reports the corresponding simulation results. Section IV presents the implementation of the HbA1c prediction framework within the AIDCARE platform. Finally, Section V provides concluding remarks.

II. Dataset Description

This study used the CGMacros dataset [7], a multimodal resource from the Sansum Diabetes Research Institute (2021–2024), which includes CGM data, meal information, physical activity, demographics, blood tests, and gut microbiome profiles. The dataset consists of 45 participants—15 healthy, 16 prediabetic, and 14 with type 2 diabetes—aged 18–69, each monitored for ~10 days using two CGM devices and a Fitbit Sense. From CGM data, over 30 statistical and time-based features were extracted, including mean, standard deviation, coefficient of variation, area under the curve, trend coefficient, time in range, and daily max–min averages. Missing values were imputed using Iterative Imputer to preserve multivariate relationships. To improve model training and account for extreme HbA1c values, the target variable was transformed using Power Transformer (Yeo-Johnson) and standardized. Additionally, the dataset was augmented via regression-adapted SMOTE and jittering by increasing the sample size and enhancing learning for high-risk individuals [8].

III. HbA1c Estimation Using Gradient Boosting Based Machine Learning Models

Let the input feature vector be $X=[x_1 \ x_2 \ \dots \ x_N]$ which is extracted from Libre Glucose data, the input feature data set is Xs after Power Transformer and Standard Scaler preprocessing. The target variable is

$$y = log\ (1 + HbA1c). \qquad (1)$$

The predicted HbA1c, $\widehat{y_t}(i)$ for sample i at iteration t is given by

$$\widehat{y_t}(i) = \widehat{y_{t-1}}(i) + \eta \cdot f_t(\boldsymbol{X}_s(i)), \qquad (2)$$

where η is the learning rate, $f_t(.)$ is the decision tree output. The optimization objective at iteration t is

$$C_t = \sum_i L\big(y_t, \widehat{y_t}(i)\big) + \Omega(f_t), \qquad (3)$$

where L is the loss function (in this study, RMSE) and $\Omega(f_t)$ is the regularization term controlling the model complexity.

To estimate HbA1c from CGM-derived features, two nonlinear regression models namely XGBoost, and CatBoost were applied. These models were chosen for their ability to capture nonlinear relationships and handle high-dimensional, noisy biomedical data

XGBoost regressor is a gradient boosting method in which sequential trees iteratively correct errors from previous iterations. Each new tree is fitted to the residuals of the previous ensemble to improve accuracy while regularization and learning rate tuning help reduce overfitting. XGBoost builds each tree sequentially, using sampling strategies to improve diversity and generalization. Specifically, the subsample parameter controls the fraction of rows (observations) used to build each tree, while colsample_bytree specifies the fraction of features (columns) sampled for each tree. These strategies increase tree diversity, reduce overfitting, and make the model effective for complex, high-dimensional datasets.

CatBoost Regressor is a gradient boosting method that handles categorical variables automatically and reduces overfitting through ordered boosting, early stopping, and L2 regularization. It supports fast and stable learning on mixed data types.

After M iterations, the final predicted log-transformed HbA1c value is converted back to the original scale,

$$\widehat{HbA1c_i} = exp\ \big(\widehat{y_M}(i)\big) - 1 \qquad (4)$$

The simulations were conducted in a Python environment using the Scikit-learn, XGBoost, and CatBoost libraries applied on CGMacros dataset. Model parameters were tuned by optimizing the learning rate, tree depth, and regularization terms. Data preprocessing steps are applied using Power Transformer and Standard Scaler. Data augmentation with regression-adapted SMOTE and Jittering expanded the dataset in order to improve model learning capability, especially for especially for high HbA1c values.

In the model training, the RMSE loss function and gradient boosting techniques were applied The XGBoost and CatBoost models were tuned to optimize prediction performance. XGBoost model was adjusted for learning rate, tree depth, and row/column sampling ratios to balance accuracy and generalization. CatBoost achieved the best performance with 300 iterations, a depth of 7, a learning rate of 0.05, and an L2 regularization term (L2_leaf_reg) of 1. XGBoost and CatBoost regression models were evaluated on the log-transformed HbA1c values using RMSE, MAE, and R^2.. The evaluation results are summarized in Table 1.

CatBoost achieved the smaller RMSE and MAE, and the highest R^2 , demonstrating superior predictive accuracy and robustness for HbA1c estimation from CGM-derived features. These results highlight the effectiveness of combining data augmentation techniques with gradient boosting models for real-time HbA1c prediction, supporting their potential integration into personalized diabetes management systems.

TABLE I. Evaluation Results for HbA1c Estimation

Model	Measures		
	RMSE	***MAE***	***R^2***
CatBoost	0.1540	0.1183	0.9806
XGBoost	0.1811	0.1438	0.9732

IV. Implementation on AIDCARE-Diabetes Self Management Platform

The findings of our theoretical investigation into the effectiveness of Gradient Boosting Models motivated their implementation within AIDCARE platform, a real-time, AI-powered, and mobile-assisted diabetes self-management platform currently under development [9]. The system consists of seven main modules: Risk, Measurement, Recommendations, Exercise, Diet, Reminders, and Education. The mobile screenshot of the AIDCARE application is shown in Figure 1. With its user-friendly mobile interface and integrated notification system, the AIDCARE application simplifies the daily health management process for patients with diabetes. Through a dedicated web-based management panel for doctors, medical staff can instantly monitor patients' glucose measurements, blood pressure values, insulin dosages, and other health data. This enhances doctor–patient interactions, accelerates checkups, and improves the efficiency of healthcare services.

One of the most distinctive features of the system is its ability to not only collect data but also analyze it to provide meaningful health recommendations. With its rule-based and AI-driven analytical infrastructure, AIDCARE offers diabetes risk prediction, blood glucose trend forecasting, insulin dosage recommendations, and personalized health advice. In the AIDCARE application, when the user clicks the "Calculate HbA1c Value" button under "Risk module", the system displays the predicted HbA1c value. The application then determines the corresponding risk level and informs the patient.

The HbA1c prediction using CatBoost model is implemented as a modular, three-layer architecture, illustrated in Figure 4. Users interact via desktop or mobile clients, while the FastAPI application runs inside a Docker container. The API exposes a /Results endpoint for retrieving current or historical predictions. Real-time HbA1c predictions are automatically triggered when new glucose measurements arrive in the PostgreSQL database. An asynchronous background process continuously polls the database every 10 seconds, retrieves new data, extracts statistical and temporal features, and applies the same preprocessing pipeline used during model training (PowerTransformer and StandardScaler).

The preprocessed features are input to the pre-trained CatBoost model to generate log-transformed HbA1c predictions, which are converted back to the original scale. Predicted values are classified into risk categories—Normal, Prediabetes, and Diabetes—and stored in the database. Users can access their predictions and associated risk information. This design ensures reproducibility, real-time performance, and scalable deployment suitable for clinical or research applications.

The HbA1c prediction using CatBoost model is implemented as a modular, three-layer architecture, illustrated in Figure 4. Users interact via desktop or mobile clients, while the FastAPI application runs inside a Docker container. The API exposes a /Results endpoint for retrieving current or historical predictions. Real-time HbA1c predictions are automatically triggered when new glucose measurements arrive in the PostgreSQL database. An asynchronous background process

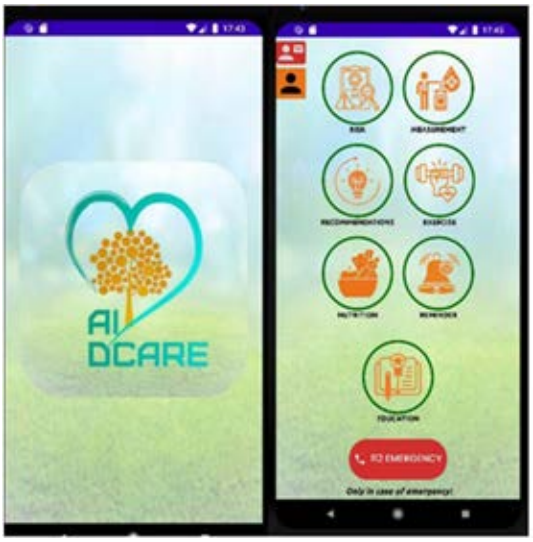


Fig.1 AIDCARE Mobile Application for Diabetes Management



Fig.2 Screenshot of HbA1C Prediction Implemented on AIDCARE Application.

continuously polls the database every 10 seconds, retrieves new data, extracts statistical and temporal features, and applies the same preprocessing pipeline used during model training (PowerTransformer and StandardScaler). The preprocessed features are input to the pre-trained CatBoost model to generate log-transformed HbA1c predictions, which are converted back to the original scale. Predicted values are classified into risk categories—Normal, Prediabetes, and Diabetes—and stored in the database. Users can access their predictions and associated risk information. This design ensures reproducibility, real-time performance, and scalable deployment suitable for clinical or research applications.

V. CONCLUSION

This study presented a real-time machine learning framework, grounded in gradient boosting models, for predicting HbA1c levels from Continuous Glucose Monitoring (CGM) data. The incorporation of data augmentation techniques, including regression-based SMOTE and jittering, enhanced the overall learning performance. Among the evaluated models, CatBoost achieved superior results, demonstrating higher accuracy, robustness, and reliability in HbA1c prediction.

The HbA1c prediction system was further deployed within a modular, containerized architecture utilizing FastAPI and PostgreSQL, with compatibility across both web and mobile platforms. Predictions were automatically initiated upon the insertion of new records into the database, thereby enabling structured risk assessments to support real-time diabetes self-monitoring.

The proposed approach highlights the potential for non-invasive, continuous HbA1c monitoring, offering a scalable and reproducible solution suitable for integration into personalized diabetes management platforms.

ACKNOWLEDGMENT

The study presented in this paper is conducted within the AIDCARE project funded by Health Institutes of Türkiye (TUSEB), Turkish Ministry of Health under grant TUSEB 33987.

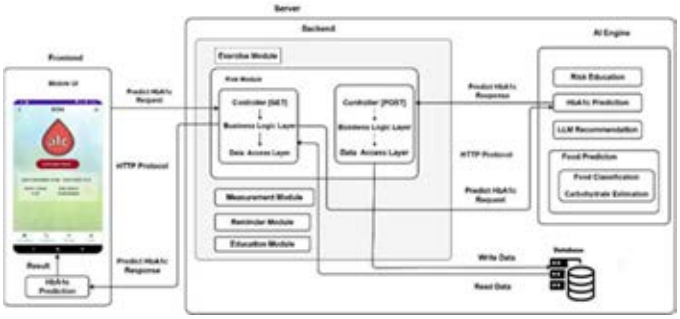


Fig.3 Implementation Methodology of HbA1C Prediction on AIDCARE Platform

REFERENCES

[1] H. Chehregosha, M.E. Khamseh, M. Malek, F. Hosseinpanah, and F. Ismail-Beigi, "A view beyond HbA1c: role of continuous glucose monitoring," Diabetes Therapy, vol. 10 (3), pp. 853-863, 2019.

[2] T. Elhadd, R. Mall, M. Bashir, J. Palotti, L. Fernandez-Luque, F. Farooq, D. Al Mohanadi, Z. Dabbous, R.A. Malik , and A.B. Abou-Samra, "Artificial intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during ramadan (The PROFAST–IT Ramadan study)," Diabetes Research and Clinical Practice, vol. 169, 108388, 2020.

[3] M. Qaraqe, A. Elzein, S. Belhaouari, M.S. Ilam, and G. Petrovski, "A novel few shot learning derived architecture for long-term HbA1c prediction," Scientific Reports, vol.14(1), 482, 2024.

[4] C. Duckworth, M.J. Guy, A. Kumaran, A.A. O'Kane, A. Ayobi, A. Chapman, P. Marshall , and M. Boniface, "Explainable machine learning for real-time hypoglycemia and hyperglycemia prediction and personalized control recommendations," Journal of Diabetes Science and Technology, vol. 18(1), pp. 113-123, 2024.

[5] D. Dave, D.J. DeSalvo, B. Haridas, S. McKay, A. Shenoy, C.J. Koh, M. Lawley, M. Erraguntla, "Feature-based machine learning model for real-time hypoglycemia prediction," Journal of Diabetes Science and Technology, vol. 15(4), pp. 842-855, 2021.

[6] Y. Deng, L. Lu, L. Aponte, A.M. Angelidi, V. Novak, G.E. Karniadakis, and C.S. Mantzoros, "Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients.," NPJ Digital Medicine, vol. 4(1), 109, 2021.

[7] R. Gutierrez-Osuna, D. Kerr, B. Mortazavi and A. Das, "CGMacros: A scientific dataset for personalized nutrition and diet monitoring (version 1.0.0)," PhysioNet, 2025, https://doi.org/10.13026/3z8q-x658

[8] A. Géron, (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media Inc., 2022.

[9] Ö.F Üçer, A. Kavak , Y. Demirhan, M. Uzun, B. Savas, A.B. İnner, Ö. Alkan, B. Çetinarslan, Z. Cantürk, A. Selek, E. Gezer, U. Yiğit, A. Aktaş, A.T. Fırat, M.A. Demirtaş, G. Okandan, K. Ünlü, S. Ünersoy, and Ö. Çakır, "Implementation of a decision support mechanism on a mobile platform using clinical evidence based dynamic insulin dosage adjustment for artificial intelligence enabled diabetes care (AIDCARE) system," MDPI Engineering Proceedings, vol 104(27), 2025.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Multi-modal Biosignal and Media Content Fusion for Human Interest Prediction Using HGNNs

Teoman Berkay Ayaz, Akhan Akbulut
Department of Computer Engineering, Istanbul Kültür University, Istanbul, Turkey
Email: 1800004169@stu.iku.edu.tr; a.akbulut@iku.edu.tr;

***Abstract*—Prediction of consumer interest in advertisements is an increasingly significant challenge in the interdisciplinary field of Neuromarketing. Traditional neuromarketing studies often rely on electroencephalographic (EEG) signals or facial expression data, in addition to self-reported responses to estimate user preferences. While this information can provide rich insight into human decision-making, it often suffers from limited availability in real-world deployment scenarios, in addition to causing overfitting when employing predictive solutions for user interest. To overcome these challenges, our study presents a novel multi-modal data fusion approach working with a hierarchical deep learning pipeline to overcome overfitting and availability problems. By representing emotion signals, consumer demographic information, and advertisement semantic information on a heterogeneous graph, our study uses an emotion intensity regressor and a hybrid Heterogeneous Graph Neural Network (HGNN) model working consecutively to effectively predict user interest in advertisements. The empirical results demonstrate robust performance on limited data with an accuracy of 0.71 and an F1-score of 0.82.**

***Keywords*—Neuromarketing, Graph Data Structure, Graph Neural Networks, Multi Modal Data Fusion, Consumer Behavior Prediction**

I. Introduction

In recent years, the interdisciplinary field of neuromarketing emerged as a research domain to understand human decision making in response to stimuli. Neuromarketing typically leverages methods more commonly found in domains like biomedical engineering such as, signal acquisition, image processing, and statistical modeling to capture the link between cognitive and emotional mechanics responsible for consumer behavior [1]. NeuroBioSense dataset [2] provides a multi-modal corpus of participant demographics, wearable sensor signals, facial expression recordings, and interest ratings for a diverse array of products and brands, making a significant contribution to this field.

Despite the richness of data, predicting customer interest in advertisements remain a difficult problem. Emotional signals are strong predictors of engagement, yet oftentimes deployment scenarios will have no observed emotional signals. Even in scenarios where emotion signals are available, they can often lead to overfitting, rendering any predictive solution for this task unusable. Furthermore, neuromarketing data often comes in heterogeneous and complex forms. This creates a necessity for a solution to handle the context-dependent engagement between advertisements and customers.

979-8-3315-5565-8/25/$31.00 © 2025 IEEE

In this paper, we propose a multi-modal fusion of media content with biosignals and participant demographics to create a heterogeneous customer-advertisement graph for the prediction of user interest with Heterogeneous Graph Neural Networks (HGNN). The study leverages the NeuroBioSense dataset as a testbed to construct a heterogeneous graph with participant nodes, connected to advertisement nodes to achieve a machine learnable form for the multi-modal fusion of data. We convert natural language data and video data into machine-learnable embeddings using pre-trained encoders. We then train and evaluate a hybrid HGNN model based on Graph Attention (GAT), advanced linear networks, and internal feature gating mechanisms. Additionally, we compare our model against more commonly used tabular models to provide a more comprehensive evaluation. Lastly, our study provides an emotion intensity predictor that works on customer demographics and advertisement information to predict and generate the emotion signals, which are often unavailable at inference time, overcoming the availability and overfitting problems. The fourfold contributions of our study can be summarized as:

- A multi-modal feature integration pipeline which combines demographic information, media content, and self-reported emotional signals.
- A signal predictor aiming to provide biosignal data that is often unavailable at inference time, allowing for higher-performing model generalization.
- A hybrid HGNN architecture that leverages GAT, internal feature gating mechanisms, and advanced linear networks for improved predictive performance.
- An empirical demonstration of the effectiveness of the proposed approach on the NeuroBioSense dataset, achieving considerable predictive performance.

II. Related Works

Neuromarketing as a field of study often leverages methods borrowed from the Biomedical Engineering Domain. Majority of studies done in Neuromarketing often leverage biosignals recorded from volunteers in participants. These studies use neural signals and advanced machine learning to predict user preferences and interest in specific ad campaigns and commercials.

One relevant study conducted by Alimardani et al. [3] in 2021, compared traditional machine learning (ML) algorithms

to deep learning based methods, specifically Convolutional Neural Networks (CNNs). Their approach leveraged these predictive methods to predict user preferences on the basis of electroencephalographic (EEG) signals on two independent EEG datasets. The researchers found that CNN architecture performed considerably better in predicting movie ratings based on EEG signals, showcasing the potential of neural networks in neuromarketing tasks. A similar study done in 2022 by Ullah et al. [4] leveraged EEG signal analysis to predict product preferences among users, using deep learning methods and support vector machines (SVM). Their findings demonstrated that using EEG information for consumer behavior prediction is an effective solution. Similarly, Derdiyok and Patlar Akbulut [5] proposed a biosignal-based emotion-oriented video summarization framework that utilizes physiological signals such as EDA, BVP, and accelerometer data, demonstrating the effectiveness of emotion-driven approaches in multimedia analysis. A recent survey by Patlar Akbulut [6] provides a comprehensive overview of emotion recognition methods based on physiological signals, highlighting the role of biosignals as a reliable source for affective computing applications

Multi-modal data fusion is another field of study investigated by researchers. Studies investigating this curious subject will often leverage a heterogeneous graph data structure to represent the various types of data in a machine-learnable form. One such study conducted by Silva et al. in 2022 [7], leveraged an unsupervised HGNN model for hit song prediction by fusing together a multi-modal dataset. The researchers included musical data like lyrics and artist information. Following this, the researchers formulated a one-class prediction task and found promising results as opposed to other strategies when it comes to representing similar data.

This study proposes a novel approach, which works by fusing multi-modal data to create a machine-learnable heterogeneous graph. By combining participants' recorded emotions as biosignals, demographic information, and advertisement information from both videos and textual data, our study builds an informative heterogeneous graph. Building on this information, the method trains two consecutive back-to-back models to form a production-ready pipeline for predicting user interest.

III. Methodology

As mentioned our methodology combines multi-modal data by the usage of heterogeneous graphs and pre-trained encoders to process multimedia data in a machine-learnable format. The initial stage is the processing of the NeuroBioSense Dataset [2] to extract relevant features and combine the data into a tabular dataframe. Following this, an emotion signal regressor is trained to make the biosignals available for the interest prediction stage without overfitting. Following this, the graph construction is done, and an HGNN model is trained to predict participant interest.

A. NeuroBioSense Dataset

The NeuroBioSense dataset [2] provides multidimensional data designed to study emotional and cognitive responses of consumers to advertisements. The dataset contains physiological signals, facial expressions, and demographic information of participants, presenting a rich source for emotion detection, consumer behavior, and signal processing studies, making it a suitable dataset for this study. The dataset contains 58 unique participants, each participant watching 10 or 15 ads based on the advertisement category. Each participant's age, sex, and emotional responses are stored in the dataset, allowing for demographic and biosignal information to be included in the study. Outside of the presented dataset, the advertisement videos were collected to process alongside the related participant information. One detail regarding the collection of videos is that only 31 of the 35 presented ads were collected, as the 4 remaining videos were removed from their sources. Hence, the instances related to those 4 ads were omitted from this study.

1) Multi-Modal Data Fusion & Preprocessing: The preprocessing for this study was done in three main stages: participant demographic info processing, advertisement info processing, and emotion signal processing. The first and last of these stages leverage standard preprocessing applications such as one-hot encoded binning for participant age, one-hot encoding for gender, and one-hot encoded emotion intensities for 7 recorded participant emotions.

In the dataset and collected videos, there are three types of available information: numeric advertisement information (duration, fps, etc.), advertisement videos, and textual information (advertisement title and category descriptions). The numeric information was preprocessed using min max scaling as a standard approach. Aside from this, to convert the natural language data belonging to these ads into machine-learnable, unidimensional representations, the approach leverages the all-MiniLM-L6-v2 pre-trained sentence transformer [8], [9]. For both types of information, this pre-trained encoder was used to generate a 384-dimensional representation. Finally, for the extraction of video-level embeddings, the approach employs the Contrastive Language–Image Pre-training (CLIP) model [10]. Each video was uniformly sampled into 12 frames across the full duration, enabling even coverage for all available clips. The frames were then encoded using the CLIP image encoder (ViT-B/32). The resulting embeddings were then normalized and aggregated by mean pooling to acquire a 512-dimensional representation for all available video clips. After conversion, all acquired representations are concatenated to form advertisement representations.

Following the preprocessing, all of the available data was fused into a heterogeneous graph with two node types (participant and advertisement), one edge type (participant, watches, ad) with emotions as edge attributes, and user interest as the edge label and prediction target. One noticeable detail in the pipeline is that the insertion of edge attributes and graph creation is done using the predictions made by the emotion

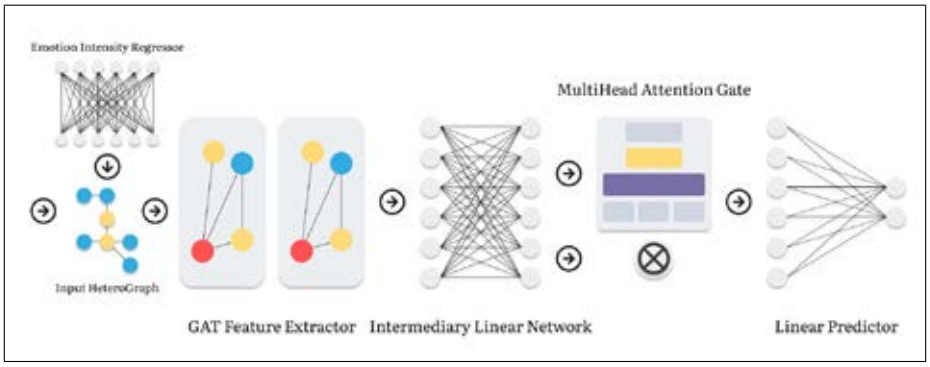


Fig. 1: Architecture of the proposed hierarchical model, combining the Emotion Intensity Regressor and the HGNN predictor

intensity predictor. Given that emotion signals can often act as standalone predictors of user interest, they can often lead to overfitting in addition to being unavailable in most production environments. The emotion intensity predictor closes this prevalent gap and makes it possible to incorporate these signals into a production-ready pipeline for further predictive tasks.

B. Hierarchial Model Architecture

To address the gap caused by missing or unreliable biosignals, the study proposes a hierarchical architecture composed of two consecutive models. The initial approximates an emotion signal based on demographic and advertisement information, generating a continuous feature stream when genuine biosignals are unavailable. The latter uses these approximated signals on a heterogeneous graph, and makes supervised user interest predictions based on this data.

1) Emotion Intensity Predictor: As mentioned, the users emotional response, such as shock or joy, is often unavailable in production environments. Given this, any approach aiming real-world deployment needs to overcome and fill this gap to make these biosignals available at inference time. To close this gap, this problem can be approached as a data imputation problem. Researchers working in the Artificial Intelligence and Data Science domains will often use ML models to impute the missing fields in their datasets by treating the column with missing values as a prediction target. The NeuroBioSense dataset and collected videos provide key features that can be used for this downstream prediction task. By using a concatenated set of participant demographic representations and advertisement representations, we can formulate an input feature set of X. By treating the one-hot encoded signal intensities as our prediction target y, we can formulate this problem as a multidimensional regression problem and train a model to approximate these missing signals.

For this multidimensional regression problem, our study employs a fully connected XNet model to approximate these signals. As opposed to the traditional Multi-Layer Perceptron (MLP) architecture with Rectified Linear Unit (ReLU) activation terms, the XNet architecture uses Cauchy Activation terms, which are based on the Cauchy Integration Theorem [11]. Recent studies have demonstrated that XNet architecture provides higher performance in both predictive and regressive tasks [12]. In addition to its reported predictive advantages, XNet was chosen as it provides smoother nonlinear transformations and stronger resistance to overfitting compared to conventional MLPs, making it more suitable for the limited-sample, high-dimensional setting of this study. This study employs this architecture to provide a better proof of concept for the proposed methodology.

2) Hybrid HGNN Predictor: As mentioned, the HGNN model does not use the original recorded emotion signals but instead relies on the approximations generated by the emotion signal predictor. As shown in Figure 1, the HGNN model is fed through two channels: the initial being the demographic and advertisement info, which is also fed into the emotion intensity regressor, and the latter being the outputs of the emotion intensity regressor. Once the emotion intensity regressor predicts the signal, the HGNN pipeline starts to make its predictions.

The HGNN model comprises 4 key components as shown in Figure 1. The initial component is the GAT-based feature extractor to process the graph-based features. The implemented version leverages the dynamic attention mechanism proposed by Brody et al. [13] to provide further components with richer latent representations. These representations are then processed using an XNet-based intermediary network to transform them for the Multi-Head Attention-based internal feature gating mechanism. The internal gating mechanism is integrated to learn and only pass through informative signals, making it easier for the predictor head to make more accurate predictions. The last key component in the HGNN architecture is an XNet-based predictor to make edge label predictions.

To ensure reliable prediction boundaries for the HGNN model, the study employs a threshold tuning process, as opposed to relying on a fixed 0.5 threshold. To prevent collapse onto a single class of predictions to minimize loss, an F1-score-based threshold was tuned, preventing an inflated trivial all-positive solution.

C. Model Training & Experimental Setup

The preprocessing of the dataset yields a 592-row dataframe where each row represents a distinct edge on the graph structure. Given the relative low count of instances, the study leverages a 3-fold cross validation approach for the emotion

intensity regressor, and a 2:1 train-test split for the entirety of the study, with cross-validation done solely on the training dataset and further testing on the held-out test set.

IV. Experimental Results

A. Emotion Intensity Regressor

The emotion intensity regressor yields a mean R^2 score of 0.2159 with a standard deviation of 0.0384. The low variance across folds showcase that learning on different folds is relatively consistent. The 0.2923 R^2 score on the held out test set, shows that, despite the relatively low size of the dataset, it can still learn to approximate about 30% of the important signals contained within the dataset, especially when considering only demographic information available, such as the gender and age of a given participant.

TABLE I: Performance of the emotion intensity regressor

Testing Phase	MSE	MAE	R^2 Score
Fold 1	0.0315	0.0907	0.1623
Fold 2	0.0302	0.0867	0.2504
Fold 3	0.0343	0.0911	0.2350
Test	0.0289	0.0829	0.2923

B. Interest Predictors

When we inspect Figure 2, we can see that the HGNN model outperforms all of its tabular counterparts in both accuracy and recall, with scores of 0.7010 and 0.9412 for each metric. The high recall showcases that it performs significantly better when it comes to predicting an interested user, which is reflected onto the F1-score 0.8153, the highest observed among the model selection. Given this, in comparison to the XGBoost, LightGBM, Random Forest, and MLP-based models, a noticeable tradeoff in precision. Whereas majority of the model selection yields a precision of approximately 0.76, the HGNN model yields 0.7191. Although this tradeoff is reflected as a higher F1-score and accuracy, it still showcases that the HGNN model falls short when it comes to predicting negative samples, indicating a biased learning process.

V. Conclusion

The study proposed a novel multi-modal data fusion approach and a hierarchical deep learning pipeline that integrates participant demographics, advertisement media content, and biosignals into a heterogeneous graph representation for predicting user interest. The proposed approach aims to close the gap between academic neuromarketing studies and production-ready environments, paving the way for deployment-ready predictive neuromarketing solutions. The experimental results on the NeuroBioSense dataset showcases that the emotion intensity predictor can learn to approximate meaningful, emotional responses from demographic and advertisement data, enabling the use of these signals in downstream tasks. Furthermore, the proposed hybrid HGNN architecture showcases considerable predictive performance given the constraints of the dataset size. Overall, this study provides a valuable proof of concept

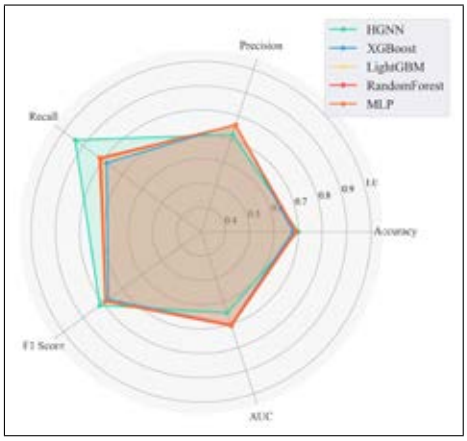


Fig. 2: Comparison of model performance across HGNN and baseline methods

for the use of heterogeneous graph learning and multi-modal data fusion to advance neuromarketing and consumer behavior prediction.

VI. Acknowledgements

This research is supported by Istanbul Kültür University under ULEP-2025-2026.

References

[1] S. Saban and E. Dağdevir, "Biomedical signal processing methods for neuromarketing: A comparative study," in *5th International Conference on Applied Engineering and Natural Sciences*, 2023.
[2] B. Kocaçınar, P. İnan, E. N. Zamur, B. Çalşimşek, F. P. Akbulut, and C. Catal, "Neurobiosense: A multidimensional dataset for neuromarketing analysis," *Data in Brief*, vol. 53, p. 110235, 2024.
[3] M. Alimardani and M. Kaba, "Deep learning for neuromarketing; classification of user preference using eeg signals," in *12th Augmented human international conference*, 2021, pp. 1–7.
[4] A. Ullah, G. Baloch, A. Ahmed, A. B. Buriro, A. Junaid, B. Ahmed, and S. Akhtar, "Neuromarketing solutions based on eeg signal analysis using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, 2022.
[5] S. Derdiyok and F. P. Akbulut, "Biosignal based emotion-oriented video summarization," *Multimedia Systems*, vol. 29, no. 3, pp. 1513–1526, 2023.
[6] F. P. Akbulut, "Evaluating the effects of the autonomic nervous system and sympathetic activity on emotional states," *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, vol. 21, no. 41, pp. 156–169, 2022.
[7] A. C. M. d. Silva, M. P. S. Gôlo, and R. M. Marcacini, "Unsupervised heterogeneous graph neural network for hit song prediction through one class learning," *Anais*, 2022.
[8] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.
[9] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," *Advances in neural information processing systems*, vol. 33, pp. 5776–5788, 2020.
[10] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.
[11] X. Li, Z. Xia, and H. Zhang, "Cauchy activation function and xnet," *Neural Networks*, vol. 188, p. 107375, 2025.
[12] X. Li, Z. J. Xia, and X. Zheng, "Model comparisons: Xnet outperforms kan," *arXiv preprint arXiv:2410.02033*, 2024.
[13] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.

3D U-Net-Based Deep Learning Approach for Pediatric Cardiovascular MRI Segmentation

Musab Ahmed Khan Umair
Artificial Intelligence
Istinye University
Istanbul, Turkey
musabkhan2003@gmail.com

Irem Yilmaz
Biomedical Engineering
Yildiz Technical University
Istanbul, Turkey
irem.yilmaz@std.yildiz.edu.tr

Omer Faruk Sahin
Bioengineering
Marmara University
Istanbul, Turkey
umarfaruksahin@outlook.com

Ahmed Boray
Computer Engineering
Istinye University
Istanbul, Turkey
ahmedegy263@gmail.com

Mohamad Humam Alzaeim
Biomedical Engineering &
Bioinformatics Istanbul
Medipol University
Istanbul, Turkey
mohamad.Alzaeim@std.medipol.edu.tr

Kevser Banu Kose
Biomedical Engineering Department
Istanbul Medipol University
Istanbul Turkey
kbkose@medipol.edu.tr

Fatma Zeynep Gungoren
Radiology
The Anadolu Medical Center
Istanbul, Turkey
fzgungoren@yahoo.com

Ruqaiyah Mirza
Software engineering
Istinye University
Istanbul, Turkey
ruqaiyah.mirza@stu.istinye.edu.tr

Beyza Nur Yazıcı
Computer Engineering
Istinye University
Istanbul, Turkey
beyzanur.yazici@stu.istinye.edu.tr

Senol Piskin*
Mechanical Engineering Department
Istinye University
Istanbul, Turkey
*Corresponding author:
senol.Piskin@istinye.edu.tr

JS Femilda Josephin
Computer Engineering Department
Istinye University
Istanbul, Turkey
femilda.bai@istinye.edu.tr

Ayham Pezri
Biomedical Engineering
Medipol University
Istanbul, Turkey
ayham.pezri@std.medipol.edu.tr

Nedim Yildirim
Sev American College
İstanbul,Turkey
ny.nedimyildirim@gmail.com

Ibrahim Faress
Software Engineering
Istinye University
Istanbul, Turkey
ibrahim.faress@stu.istinye.edu.tr

***Abstract*— Accurate segmentation of cardiovascular structures in pediatric patients is essential for clinical evaluation, but variability in imaging data and the labor-intensive nature of manual segmentation remain major challenges. The given paper provides a reproducible deep-learning model intended to counteract these issues. With our deep learning approach, cardiac MRI data is standardized by (i) converting anatomical labels into a four-class mask (involving background, aorta, chambers, pulmonary structures), (ii) aligning images and masks on coordinates using nearest-neighbor resampling, and (iii) automatically cropping the region-of-interest (ROI). To handle class imbalance, we trained a 3D U-Net with the Focal Tversky Loss function, using both public and private clinical datasets. To improve robustness and reduce outlier errors, we applied model ensembling by averaging predictions from independently trained networks. By automating preprocessing and segmentation, our approach provides a pediatric cardiac analysis tool that reduces manual workload and increases reproducibility.**

Keywords*— *3D U-Net, Cardiac Segmentation, Pediatric MRI, Deep Learning, Ensemble Methods, Focal Tversky Loss, Automated Preprocessing, ROI Cropping

I. INTRODUCTION

A. Problem Statement

variability, especially when dealing with small, complex pediatric anatomies. These challenges are further compounded by low-resolution or artifact-prone scans common in clinical datasets, especially in young children who may require sedation or who cannot remain still during imaging. Traditional segmentation pipelines struggle with inconsistent image quality, label imbalance, and unreliable region delineation. Thus, there is a critical need for robust, automatic segmentation methods tailored to the anatomical and imaging-specific constraints of pediatric cardiac MRI.

To address these challenges, we present a fully automated segmentation approach using a 3D U-Net–based multi-class segmentation trained on both open and clinical datasets. This approach enables reliable extraction of critical cardiovascular structures.

B. Our Contribution

A novel deep learning model is proposed in this study to perform multi-class segmentation of pediatric cardiovascular structures in three-dimensional magnetic resonance imagery. The pipeline unifies the anatomical labels and resamples the scans and masks with nearest neighbor resampling in 3D Slicer to ensure consistency. A subsequent automatic ROI cropping step uses padding of relevant cardiac regions to make the input size smaller and speed up computation. When small and imbalanced structures need to be accurately segmented, training a 3D U-Net with the Focal Tversky Loss is effective. To improve robustness, an ensemble inference was applied where the predictions of a variety of individually trained models are averaged. The proposed approach generalizes well across diverse anatomies and imaging conditions. By automating preprocessing and segmentation, it significantly reduces manual workload and supports scalable deployment.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

II. Related Work

Recent developments in cardiovascular anatomical segmentation have been largely driven by hybrid atlas-learning pipelines and deep learning methods. Bruns et al. [1] provide a comprehensive study using CT data, detailing dataset cataloging, preprocessing (e.g., intensity normalization, cropping), and evaluation metrics such as Dice and Hausdorff scores. The authors emphasize persistent challenges like scanner-based domain shifts, class imbalance in small structures, and annotation limitations, proposing standardized benchmarks and external validation as remedies. Complementing this, Alnasser et al. [2] introduce a multi-class statistical atlas-based whole-heart segmentation (WHS) framework for CT and MRI, demonstrating the enduring value of atlas priors when combined with data-driven refinement to enhance boundary consistency across substructures.

References

S.No	Journal/Authors	Combined Description
1	Alnasser TN, Abdulaal L, Maiter A, Sharkey M, Dwivedi K, Salehi M, Garg P, Swift AJ & Alabed S (Frontiers in Cardiovascular Medicine)	This 2024 review evaluated 18 CT-based studies using various deep learning models like ResNet50, U-Net, and nn-UNet. It reported strong segmentation performance (e.g., LA 0.88, LV 0.91) and assessed study quality using the CLAIM checklist.
2	Bruns S, Wolterink JM, van den Boogert TPW, Runge JH, Bouma BJ, Henriques JP, Baan J, Viergever MA, Planken RN & Išgum I (Computers in Biology and Medicine)	In 2022, this study used 3D CNNs to segment 4D cardiac CT images from 1,509 patients. The models achieved high Dice scores around 0.89 and demonstrated >90% clinical usability.
3	Ghosh TK, Hasan MK, Roy S, Alam MA, Hossain E & Ahmad M (IEEE Access)	This 2021 work proposed a probabilistic atlas-based segmentation method for CT and MRI heart images, using Bayesian EM and CNN variants, achieving improved accuracy and reduced volume error.
4	Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W & Rueckert D (Frontiers in Cardiovascular Medicine)	This 2020 review covered over 100 studies on cardiac image segmentation, primarily using CNNs on datasets like ACDC and MM-WHS
5	Zhuang X (Journal of Healthcare Engineering)	This 2013 review discussed early automatic segmentation using atlas-based and deformable models, reporting modest Dice scores (~0.4–0.5) and highlighting challenges in generalization.
6	Karimi-Bidhendi et al.	A GAN-augmented fully convolutional network was developed for pediatric CHD CMR segmentation, addressing limited annotated datasets. The model achieved Dice scores up to **91%**, outperforming U-Net and cvi42, and showed strong agreement with manual segmentation.
7	Yoshida A, Kondo Y, Yoshimura N, Kuramoto T et al. (Measurement Sensors)	The 2021 study applied a modified U-Net on pediatric CT scans, showing excellent whole-heart segmentation (DSC > 0.95) and reasonable chamber-level performance (0.78–0.88).

From a learning-centered viewpoint, Ghosh et al. [3] explore CNN-based heart segmentation, highlighting attention modules, multi-scale feature aggregation, and U-Net variants that effectively handle complex anatomy such as thin-walled ventricles and great veins. The existing study also emphasizes strategies to improve geometric fidelity, including 3D/2.5D contexts, patch-based training for memory efficiency, and post-processing techniques like connected-component filtering. Chen et al. [4] expand segmentation to cardiac radiography, where low soft-tissue contrast and overlapping projections present unique challenges. The researchers have also underlined the importance of robust augmentation, domain adaptation, and uncertainty estimation to minimize clinical errors. Song et al. [5] address segmentation in pediatric congenital heart disease, employing topology-aware losses, shape constraints, and tailored approaches for rare morphologies under limited training data. Across all studies [1–5], unresolved issues remain: achieving generalization across sites, consistent performance on small or atypical anatomies, and maintaining rigorous, externally validated reporting.

III. DATASETS

In the proposed study, two datasets have been utilized, namely, an open-source pediatric cardiac dataset and a combined dataset created by augmenting the public data with local clinical scans. The open-source data was used to develop early models and to serve as an external benchmark. Each of the cases was anonymized, transformed into NRRD, and named in a standardized scheme. The concatenated dataset includes local data gathered using institutional approvals, which cover different imaging conditions and pediatric anatomies. To make all scans consistent, the same preprocessing routines were employed, i.e., format conversion and label standardization (0: background, 1: aorta, 2: chambers, 3: pulmonary structures).

IV. METHODOLOGY

A.Data Preprocessing

Patients' cases consisted of a volumetric scan and its multi-label segmentation mask in NRRD format. All the data were preprocessed and cached in the form of NumPy arrays (.npy) to shorten training time and reduce I/O overhead. The trilinear interpolation was used to resize volumes of identical shapes of 128x128x64, and nearest neighbor interpolation was used out of a mask to maintain the accuracy of labels. The intensities were normalized to the range [0,1] using min-max normalization. The same scheme of classes was employed: 0 was adopted to describe background, 1 to describe the aorta, 2 to describe the chambers, and 3 to describe the pulmonary structures. Consistent preprocessing was applied across validation and test cases.

B.Data Augmentation

Training-time augmentation was used to increase model robustness to anatomical variation and the presence of acquisition artifacts, with augmentation implemented in TorchIO. This also involved random spatial changes, including flips, affine deformations (scaling, rotation, translation), and elastic deformations to model changes of enlistment. Scanner-related variations were modelled using such intensity augmentations as gamma correction, random bias field, and Gaussian noise. These augmentations were applied stochastically to every batch during training, improving the model's generalizability across varied datasets.

C. Model Training & Evaluation

The model utilizes a 3D U-Net architecture for multi-class cardiac segmentation, featuring a symmetric encoder-decoder structure with skip connections to combine semantic and detailed features. The encoder includes three downsampling blocks, each with two 3x3x3 convolutions, batch normalization, ReLU, and 2x2x2 max pooling. A bottleneck with two convolutional layers and 256 filters connects the encoder and decoder. The decoder mirrors the encoder's structure, using transposed convolutions for upsampling. At every stage, encoder outputs are concatenated with the corresponding upsampled features from the decoder. The output layer applies a 1x1x1 convolution with four channels (background, aorta, chambers, and pulmonary vessels), and a Softmax layer transforms the logits into per-voxel class likelihoods.

$$Softmax(z_i) = \frac{e^{zi}}{\Sigma_{j=1}^{C} e^{zj}} \quad (1)$$

A softmax layer is used at the network output to convert logits into class probabilities. Here, C=4 is the total number of classes.

$$TI = \frac{TP + \epsilon}{TP + \alpha \cdot FN + \beta \cdot FP + \epsilon} \quad (2)$$

To address the challenges of class imbalance and indistinct boundaries in cardiac imaging, the model was trained using a **Focal Tversky Loss (FTL)** function.

$$L_{FTL} = (1 - TI)^{\gamma} \quad (3)$$

This loss function is built upon the Tversky Index (TI), which generalizes the Dice score by weighting false negatives and false positives differently ($\alpha = 0.7$, $\beta = 0.3$). The FTL includes a focusing parameter that places greater emphasis on difficult-to-classify areas, helping the model improve accuracy on underrepresented structures.

The 3D U-Net model was trained with a batch size of 2 for 1000 epochs, utilizing early stopping based on validation loss to prevent overfitting and optimize training time. Auto mixed-precision (AMP) was employed to reduce GPU memory consumption and accelerate convergence. Data loaders were multi-threaded to ensure efficient data processing during training and validation.

Model performance was assessed using three standard segmentation metrics: the **Dice Similarity Coefficient (DSC)** to measure spatial overlap, the **Jaccard Index (IoU)** to provide a stricter comparison of regions, and **Pixel Accuracy** to measure overall voxel-wise correctness. These metrics were calculated globally and on a per-class basis to provide a comprehensive evaluation.

D. Ensemble Approach

An ensemble learning strategy was chosen to enhance the stability of the segmentation and to decrease the variance. The 3D U-Net model of the specific architecture was trained multiple times with various sets of random initialization. When performing inference, consensus prediction was obtained by averaging, voxel-wise, the softmax probability maps output by each model. The ensemble solution minimized the possibility of outlier errors that could arise due to each model and enhanced the general reliability of segmentation.

V. EXPERIMENTS & RESULT

A. Qualitative Results

In order to analyze the visual quality of our test segmentation models, we provide two visual comparisons. The two figures, a and b, can prove together that data augmentation and longer training can positively affect segmentation performance

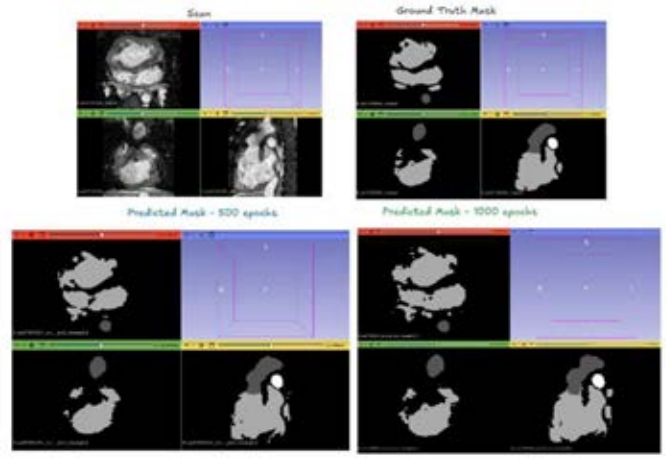


Figure 1: Multi-planar visual comparison of model predictions at 500 vs. 1000 training epochs.

Figure 1 shows a multi-planar view of Patient 55, with the results of the 500-epoch trained model and the 1000-epoch trained model. The latter model, trained for an increased amount of time, displays an accurate adherence to the boundaries and less noise. Overall, these two figures highlight how combining augmentation with long-term training enhances segmentation accuracy and reliability.

VI. DISCUSSION & CONCLUSION

This study shows that a reproducible, uniform preprocessing pipeline with a 3D U-Net as an ensemble yields good results on multi-class pediatric cardiovascular segmentation. The pipeline is easier to use and frees up far more processing resources, and the model reports good performance when applied to a standardized open-source dataset. It has high scores in the aorta and the chambers, which is indicative of good segmentation of well-defined structures.

There were two principal problems. First, the small number of cases and lack of diversity in both the combined and the open-source datasets were limited, constraining the model from learning different kinds of examples and impacting the performance across various clinical scans. Second, the pulmonary structures were the weakest group of structures. Although Focal Tversky Loss has been used to overcome the problem of class imbalance, segmentation of small, branching vessels is still challenging. An alternative direction is to investigate attention mechanisms or shape-aware priors in the future.

The most important contribution is the automatic pipeline of preprocessing that normalizes labels, resamples geometry, and crops ROIs, decreasing input size and training time, and resulting in consistency. This work introduces an end-to-end automated cardiovascular MRI segmentation pipeline, which uses standardized preprocessing to then use a 3D U-Net ensemble trained with Focal Tversky Loss to deliver high precision on published data, addressing the issues of label contradictions and inefficient computation.

VII. ACKNOWLEDGMENT

This work was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under the 1001 Grant Program (Project No. 122E488).

VIII. AFFILIATIONS

This study was conducted in collaboration with Istinye University, Istanbul Medipol University, and Anadolu Medical Center.

REFERENCES

[1] Bruns, S., Wolterink, J. M., van den Boogert, T. P., Runge, J. H., Bouma, B. J., Henriques, J. P., ... & Išgum, I. (2022). Deep learning-based whole-heart segmentation in 4D contrast-enhanced cardiac CT. *Computers in biology and medicine*, *142*, 105191.

[2] Alnasser, T. N., Abdulaal, L., Maiter, A., Sharkey, M., Dwivedi, K., Salehi, M., ... & Alabed, S. (2024). Advancements in cardiac structures segmentation: a comprehensive systematic review of deep learning in CT imaging. *Frontiers in cardiovascular medicine*, *11*, 1323461.

[3] Ghosh, T. K., Hasan, M. K., Roy, S., Alam, M. A., Hossain, E., & Ahmad, M. (2021). Multi-class probabilistic atlas-based whole heart segmentation method in cardiac CT and MRI. *IEEE Access*, *9*, 66948-66964.

[4] Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., & Rueckert, D. (2020). Deep learning for cardiac image segmentation: a review. *Frontiers in cardiovascular medicine*, *7*, 25.

[5] Song, Y., Ren, S., Lu, Y., Fu, X., & Wong, K. K. (2022). Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. *Computer Methods and Programs in Biomedicine*, *220*, 106821.

[6] Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A., & Jafarkhani, H. (2020). Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of cardiovascular magnetic resonance*, *22*(1), 80.

[7] Yoshida, A., Lee, Y., Yoshimura, N., Kuramoto, T., Hasegawa, A., & Kanazawa, T. (2021). Automated heart segmentation using U-Net in pediatric cardiac CT. *Measurement: Sensors*, *18*, 100127.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Structure-Aware V-Net Framework for 3D Multi-Class Segmentation of Pediatric Cardiovascular Structures

Irem Yilmaz
Biomedical Engineering
Yildiz Technical University
Istanbul, Turkey
irem.yilmaz@std.yildiz.edu.tr

Omer Faruk Sahin
Bioengineering
Marmara University
Istanbul, Turkey
umarfaruksahin@outlook.com

Mohamad Humam Alzaeim
Biomedical Engineering & Bioinformatics
Istanbul Medipol University
Istanbul, Turkey
mohamad.Alzaeim@std.medipol.edu.tr

Beyza Nur Yazici
Computer Engineering
Istinye University
Istanbul, Turkey
beyzanur.yazici@stu.istinye.edu.tr

Doruk Gunveren
Cemberlitas Anatolian High School
Istanbul, Turkey
gunverendoruk@gmail.com

Sakip Onder
Biomedical Engineering
Yildiz Technical University
Istanbul Turkey
sonder@yildiz.edu.tr

Kevser Banu Kose
Biomedical Engineering
Istanbul Medipol University
Istanbul Turkey
kbkose@medipol.edu.tr

Fatma Zeynep Gungoren
Radiology Department
The Anadolu Medical Center
Istanbul, Turkey
fzgungoren@yahoo.com

Musab Ahmed Khan Umair
Artificial Intelligence
Istinye University
Istanbul, Turkey
musabkhan2003@gmail.com

Senol Piskin*
Mechanical Engineering
Istinye University
Istanbul, Turkey
*Corresponding author:
senol.Piskin@istinye.edu.tr

JS Femilda Josephin
Computer Engineering
Istinye University
Istanbul, Turkey
femilda.bai@istinye.edu.tr

Ruqaiyah Mirza
Software Engineering
Istinye University
Istanbul, Turkey
ruqaiyah.mirza@stu.istinye.edu.tr

Fatmanur Aydogdi
Zuhtu Kurtulmus Anatolian High School
Istanbul, Turkey
nurraydogdi21@gmail.com

***Abstract*—This research shows a structure-aware deep learning model. The model segments 3D multi-class pediatric cardiovascular anatomy, focusing on small complex structures, like the pulmonary arteries, also it is based upon a modified V-Net architecture. The multi-center dataset includes expert-annotated pediatric Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) volumes. The model underwent training through utilization of this specific dataset. For resolving class disequilibrium in conjunction with morphological variance, investigators incorporated a supplementary decoder division for specialization within pulmonary artery segmentation. The pipeline represents standardized preprocessing measures like resampling, normalization, also cropping, furthermore it represents advanced augmentation techniques. Evaluation was performed using Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) metrics. This assessment revealed that the partitioning of minute anatomy progressed noticeably. The proposed architecture shows support is offered for preoperative planning in congenital heart disease (CHD) cases because structure-specific modifications are able to improve segmentation accuracy plus consistency in pediatric cardiovascular imaging.**

***Keywords*—V-Net; pediatric cardiovascular segmentation; 3D medical imaging; pulmonary artery; deep learning; class imbalance**

I. Introduction

Congenital heart diseases (CHDs) exist among the most typical birth anomalies, so pediatric morbidity as well as mortality are largely triggered [1]. Although CT and MRI provide detailed views of cardiovascular anatomy, manual 3D segmentation is time-consuming and subject to inter-observer variability [2]. Deep learning models, notably V-Net, are frequently utilized for segmenting volumetric tasks [3][4]. However, conventional architecture struggles to segment small, detailed regions such as the pulmonary arteries due to class imbalance and thin anatomical structure [5].

The proposed study puts forth an adjusted V-Net architecture for pulmonary artery segmentation using a supplementary decoder. The model performs multi-class segmentation and structure-specific segmentation. It was trained on a collected, multi-center dataset including pediatric CT and MRI scans. An effective preprocessing pipeline was utilized, including resampling, normalization, cropping, and TorchIO-centered augmentation [6]. Assessment using Dice and IoU metrics affirmed performance in augmented reality. This became improved in a way that was most obvious in smaller anatomical regions. The structure-aware design improved segmentation accuracy. For CHD diagnosis along with surgical planning, benefit could come from it as well.

II. Background

A. Challenges in Pediatric Cardiovascular Imaging

Pediatric cardiovascular structures remain arduous for one to visualize because of diminutive organ sizes, elevated heart rates, and also intrinsic anatomical variance. Segmentation is necessary for effective surgical planning and reliable diagnosis. In CHD cases, segmentation that is accurate is critical. Pediatric imaging is, however, often obstructed by motion artifacts, low signal-to-noise ratios, and limited contrast, making it difficult to delineate small structures such as the pulmonary arteries [7].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

B. Limitations of Conventional Segmentation Techniques

Manual segmentation by clinicians is common yet untenable for large datasets and time-consuming. Since conventional techniques often fail for complex pediatric anatomy, inaccurate boundaries and distorted segmentation happen in low-contrast regions [8].

C. V-Net for 3D Medical Image Segmentation

V-Net like networks have exhibited strong performance during 3D medical image segmentation. V-Net conserves spatial detail via skip connections coupled with residual blocks, intended for volumetric data. Contemporary augmentations have exhibited refined precision within juvenile cardiac data sets [9]. They represent attention gates and expanded convolutions.

D. Architectural Solutions to Class Imbalance

Segmentation models often grapple with class imbalance; wherein larger structures dominate training while smaller pulmonary arteries are underrepresented. Refining model susceptibility to these underrepresented classifications without impairing overall efficacy has demonstrated merit through the use of structural models [10].

III. Method

Data compilation officially began once the institutional ethics committee approved this study's procedure. The completed procedures are described in each of the following sections.

A. Dataset and Cohort Description

The dataset used in the proposed work comprises anonymized pediatric cardiovascular cases obtained from four medical centers and also included contrast-improved CT and cardiac MRI scans that represented various CHD phenotypes. Each volume was annotated by at least two expert radiologists. Therefore multi-class ground truth masks for the aorta and heart chambers and pulmonary arteries were generated. Preparation of uniforms occurred because the scan settings differed.

B. Preprocessing and Data Augmentation

Consistent input was ensured via cardiac region truncation and volume resampling to 1×1×1 mm^3 voxel spacing. Additionally, volumes did undergo intensity standardization to have a zero mean and unit variance and resizing of them to 128×128×64 since the researchers had used center cropping or padding. Data were augmented in real-time with random affine transformations, gamma correction, and elastic deformations using TorchIO library. The model gained some extrapolation abilities by way of these enhancements. The model also gained in resilience.

C. Data Anonymization and Management

Adhering to protocols approved by the Institutional Review Board (IRB), patient data were initially anonymized, also data were stored securely under the General Data Protection Regulation (GDPR). Metadata were tracked using a structured database. The database included patient age as well as modality and institution plus segmentation status in addition to preprocessing progress.

D. Labeling Protocol and Auxiliary Masking

Segmentation masks for multiple classes designated each scan in this manner: 0 signified background, 1 signified aorta, 2 signified heart chambers, 3 signified pulmonary arteries. For supervising a specialized decoder segment inside the model, an ancillary binary mask separating the pulmonary artery was fashioned. This helped the model focus on learning small anatomical structures more effectively.

E. 3D Surface Model Generation

SimpleITK and 3D Slicer transform anticipated segmentation results into STL surface models for qualitative assessment utilizing the marching cubes algorithm. These models were visually validated through anatomical correctness and spatial precision by comparing them across sagittal, coronal, and axial planes to the ground truth masks.

IV. Network Architecture

To address the challenges intrinsic to segmenting small complex cardiovascular structures along with it, a modified V-Net architecture in pediatric imaging was developed. This model utilizes a three-dimensional encoder–decoder framework, plus it presents four encoding and decoding strata. The architecture strides also transposes convolutions then activates ReLU. It also normalizes the instances also it skips the connections so as to preserve spatial information for an improved gradient flow.

For improvement regarding the partitioning of some pulmonary arteries, they incorporated such a subordinate decoder division. An exclusive Dice loss for the pulmonary artery class supports autonomous optimization of the auxiliary branch. This independent enhancement betters output on obscure areas since it shows a shared encoder with the core decoder. Because it combines Dice with cross-entropy in multi-class segmentation the central decoder employs a hybrid loss. Adam with cosine annealing optimized the training across 1000 epochs. A GPU such as an NVIDIA RTX 3090 was used for the training. Toward improved generalizability, researchers utilized real-time data enrichment through TorchIO that included affine and intensity transformations, and early stopping occurred on validation loss.

V. Results

A. Segmentation Performance

The dual-head V-Net model was trained on a multi-center pediatric dataset over wide-ranging epochs, as it was evaluated through the use of standard segmentation metrics including Intersection over Union (IoU) and Dice Similarity Coefficient (DSC). The results demonstrated strong overall performance especially for background and heart chambers as anatomical structures well-defined and large. On account of its small size in conjunction with class imbalance, the pulmonary arteries were still the most challenging region whereas the aorta showed moderate segmentation accuracy. Importantly, the additional decoder branch performed better at segmenting these underrepresented and morphologically complex structures. The model can improve in consistency and accuracy. These results stress its promise in heart imaging for children.

B. Training and Validation Performance

Training and validation losses gradually converged over the 1000-epoch training cycle. The hybrid loss function fostered an improved convergence speed. This function curtailed overfitting, coupled with structure-specific supervision through the auxiliary decoder. Figure 1 displays the training curves, and it incorporates a heatmap of class-specific segmentation measures that depict uniform efficacy throughout all anatomical targets.

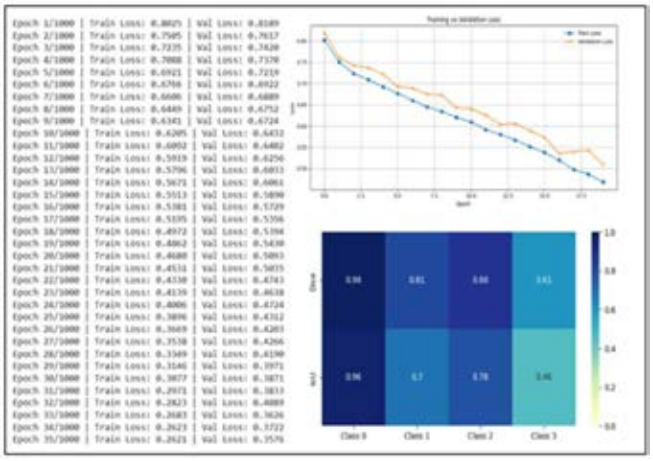


Fig 1. *Training vs. validation loss curve and class-wise segmentation heatmap (Dice and IoU) for the multi-class model.*

C. Visual and Structure-Specific Assessment

Segmentations were visually evaluated by comparing the predicted and the ground-truth masks in axial, coronal and sagittal planes. This model precisely circumscribed global cardiovascular architecture with smaller conduits like pulmonary arteries as depicted within Figures 2 and 3. Within representative instances, manual annotations displayed large spatial congruence alongside the predicted masks like patient 1 and patient 2. The supplementary decoder helped defining refined borders along with diminishing inaccuracies throughout detailed regions. It proved important for conserving great anatomical detail, which accentuated this contribution.

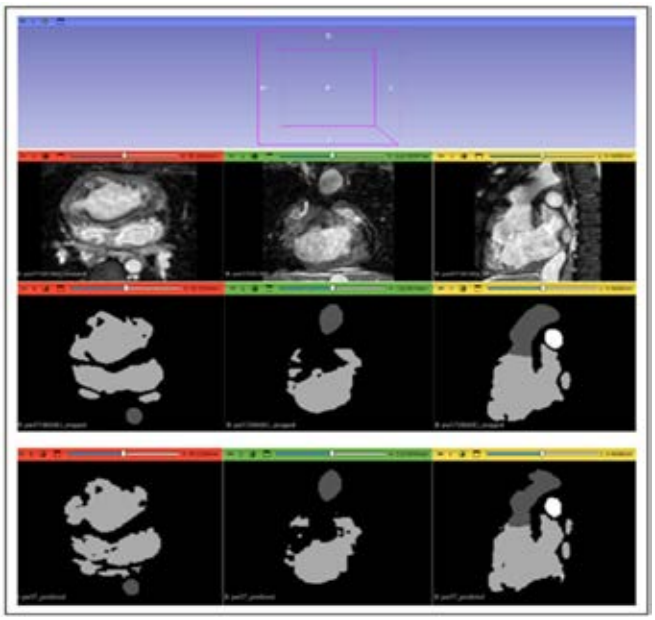


Fig 2. *Multi-plane segmentation results for patient 1, showing ground truth and predicted masks.*

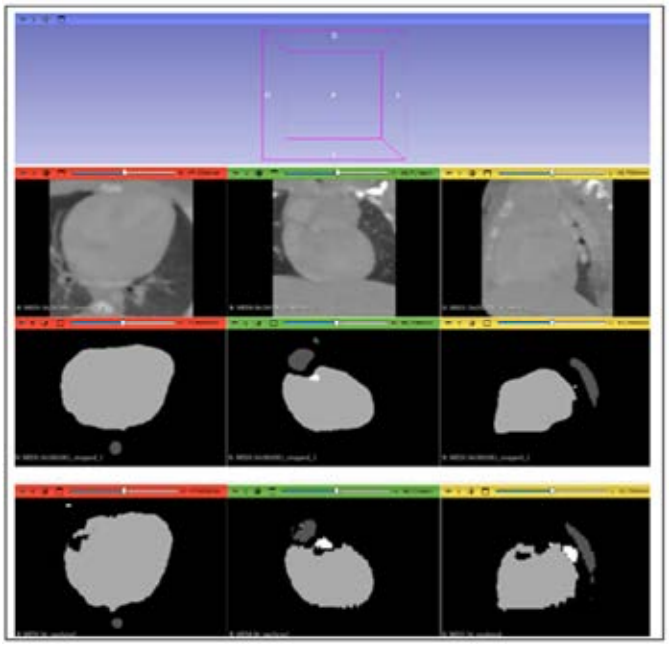


Fig 3. *Multi-plane segmentation results for patient 2, showing ground truth and predicted masks.*

VI. Discussion and Conclusion

This study introduces a structure-aware V-Net architecture for multi-class segmentation of pediatric cardiovascular structures, in which is tailored with particular emphasis on anatomically small complex regions like the pulmonary arteries. The dual-head design has within it an auxiliary decoder that targets at the pulmonary artery. This design addresses class imbalance as well as anatomical variability since they do present key challenges in pediatric imaging. The framework improves generalization and class-wise accuracy by integrating standardized preprocessing and a hybrid loss function along with TorchIO-based data augmentation.

The results showed consistent IoU and Dice scores for the aorta and heart chambers, which reflects strong segmentation. Pulmonary artery segmentation, which is typically underrepresented in training data, was greatly improved via the auxiliary decoder. Steady convergence over 1000 epochs confirmed effective learning. Because predicted masks showed clear boundaries and structural integrity across all planes, qualitative results further demonstrated anatomical accuracy. Surface reconstructions of 3D confirmed the model could capture vascular morphology precisely.

Through this automated approach, clinical workflows may be supported especially for CHD. Manual labor and inter-observer variability may decrease. Planning before surgery may be improved in fact, even in complex anatomies too. Dataset diversity and annotator variability are current limitations. To address these limits, future work should validate performance via multi-center studies, expand CHD subtypes, and incorporate deeper attention-based architectures.

VII. Acknowledgment

This work was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under the 1001 Grant Program (Project No. 122E488).

VIII. AFFILIATIONS

This study was conducted in collaboration with Istinye University, Istanbul Medipol University, and Anadolu Medical Center.

REFERENCES

[1] Karimi-Bidhendi, S., Arafati, A., Cheng, A. L., Wu, Y., Kheradvar, A., & Jafarkhani, H. (2020). Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance*, *22*(1), 80. https://doi.org/10.1186/s12968-020-00678-0

[2] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., & Shen, D. (2021). Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE reviews in biomedical engineering*, *14*, 4–15. https://doi.org/10.1109/RBME.2020.2987975

[3] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, *18*(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

[4] Ma, X., Song, H., Jia, X., & Wang, Z. (2024). An improved V-Net lung nodule segmentation model based on pixel threshold separation and attention mechanism. *Scientific reports*, *14*(1), 4743. https://doi.org/10.1038/s41598-024-55178-3

[5] Cui, L., Xu, M., Liu, C., Liu, T., Yan, X., Zhang, Y., & Yang, X. (2025). Towards Reliable Healthcare Imaging: A Multifaceted Approach in Class Imbalance Handling for Medical Image Segmentation. *Interdisciplinary sciences, computational life sciences*, *17*(3), 614–633. https://doi.org/10.1007/s12539-025-00726-2

[6] Pérez-García, F., Sparks, R., & Ourselin, S. (2021). TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer methods and programs in biomedicine*, *208*, 106236. https://doi.org/10.1016/j.cmpb.2021.106236

[7] Leiner, T., Bogaert, J., Friedrich, M. G., Mohiaddin, R., Muthurangu, V., Myerson, S., Powell, A. J., Raman, S. V., & Pennell, D. J. (2020). SCMR Position Paper (2020) on clinical indications for cardiovascular magnetic resonance. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance*, *22*(1), 76. https://doi.org/10.1186/s12968-020-00682-4

[8] Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical image analysis*, *79*, 102444. https://doi.org/10.1016/j.media.2022.102444

[9] Li, H., Li, D., Dong, Q., Han, X., & Dong, S. (2025). Multi-view semi-supervised attention network for 3D cardiac image segmentation. *Frontiers in cardiovascular medicine*, *12*, 1461774. https://doi.org/10.3389/fcvm.2025.1461774

[10] Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, *95*, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Aortic-Flow Reference Shaping via Q-Learning for CF-LVAD Support

Mahmud Karno*, Mert yigit*, Ahmed Alhajyounis†, Kamuran Kadipasaoglu*
*Biomedical Engineering Department.
Yildiz Technical University, İstanbul, Turkey
mahmud.karno@std.yildiz.edu.tr, Yigitmert1@hotmail.com, kamuran@yildiz.edu.tr
†Control and Automation Engineering Department.
Yildiz Technical University, İstanbul, Turkey
ahmed.alhajyounis@std.yildiz.edu.tr

***Abstract*—Continuous-flow left ventricular assist devices (CF-LVADs) operate at fixed speeds, suppressing pulsatility, prolonging aortic-valve closure, and impairing root washout, thereby elevating thrombotic risk. To address these limitations, we propose a Q-learning–based framework that synthesizes a smooth, bandwidth-limited aortic-flow reference without explicit plant modeling. The reference is parameterized by waveform coefficients optimized offline through interactions with a cardiovascular simulator, guided by a composite reward that penalizes deviations in mean arterial pressure, encourages periodic aortic valve opening, prevents suction events, and maintains sufficient and proper flow. In simulation, the learned waveform maintained mean arterial pressure between 60–90 mmHg, left-ventricular pressure between 5–70 mmHg, pump flow between 2–5 L/min, ventricular volume between 50–140 mL, and a pulsatility index around 2. Importantly, the aortic valve opened approximately 6 times per 20 s, improving washout while avoiding suction. These results underscore the potential of reference-level learning to restore physiological pulsatility and preserve hemodynamic safety in CF-LVAD support.**

***Keywords*—Q-learning, LVAD, Aortic flow reference.**

I. Introduction

Left ventricular assist devices (LVADs) are well-established therapeutic options for patients with advanced heart failure, providing long-term circulatory support either as a bridge to transplantation or as destination therapy [1]. Their ability to restore cardiac output, improve end-organ perfusion, and enhance quality of life has driven widespread clinical adoption. Among the available technologies, continuous-flow LVADs (CF-LVADs) have gained particular prominence due to their compact size, mechanical simplicity, and superior durability compared with pulsatile-flow designs [2]. Nevertheless, the fixed-speed operation of CF-LVADs inherently suppresses physiological pulsatility, which can lead to prolonged aortic valve closure, reduced or irregular valve motion, and diminished aortic root washout [2]. These hemodynamic changes have been associated with adverse events such as aortic insufficiency, thrombosis, and vascular remodeling [3].

In the literature, multiple strategies have been proposed to mitigate these effects by shaping aortic-flow reference signals. For example, Dobarro et al. employed pre-programmed pulsatility patterns to impose periodic variations in pump speed, thereby promoting intermittent aortic valve opening and improving aortic-root washout [4]. More recently, Fernández-Zapico et al. demonstrated the feasibility of reinforcement learning for physiological modulation in mechanical circulatory support, using it to modulate LVAD motor speed so as to achieve left-ventricular ejection and periodic aortic-valve opening highlighting the promise of data-driven approaches for LVAD reference-signal design [5].

In earlier studies, we utilized a seven-compartment lumped-parameter cardiovascular system (CVS) simulation model [6] in Simulink to evaluate candidate aortic flow reference waveforms for CF-LVAD operation. This platform was employed to characterize the hemodynamic impact of different pulsatility patterns, assess their influence on aortic valve dynamics, and identify key parameters affecting mean arterial pressure and pulsatility under varying preload, afterload, and contractility conditions. These studies provided quantitative evidence of the trade-offs between root washout, pressure stability, and overall pump performance, and revealed that fixed-pattern modulation often failed to maintain optimal hemodynamics across different physiological states.

Q-learning, a model-free reinforcement learning method, offers a promising framework for optimizing aortic flow reference signals in CF-LVADs. Unlike conventional algorithms, Q-learning learns optimal state–action mappings through repeated interaction with the environment, maximizing a cumulative reward over time. This enables the simultaneous pursuit of multiple, potentially conflicting objectives [7], such as maintaining mean arterial pressure within 80–100 mmHg, preventing suction by preserving adequate left ventricular volume, sustaining physiological pulsatility, and promoting periodic aortic valve opening, while adapting to changes in preload, afterload, and contractility.

Prior work has largely focused on direct pump-speed modulation or on prescribing aortic-flow references with heuristic, preprogrammed, or model-based schemes; to our knowledge, none has framed the task as synthesizing the aortic-flow reference via Reinforcement Learning (RL). We instead formulate reference design at the flow level, anchoring the objective in clinically salient endpoints (periodic aortic-valve opening, effective aortic-root washout, a target pulsatility index, and

979-8-3315-5565-8/25 $31.00 © 2025 IEEE

improved ventricular unloading) while leaving low-level actuation to the pump controller. Trained across variations in preload, afterload, and contractility, the approach explicitly negotiates the trade-offs among mean arterial pressure regulation, pulsatility index preservation, valve-opening duty cycle, and root washout, yielding a smooth, bandwidth-limited reference.

In this formulation, a model-free Q-learning agent interacts with a numerical cardiovascular (CVS) simulator to identify the reference waveform from sampled transitions, without requiring an explicit plant model during learning. The resulting reference balances pressure targets with physiologic pulsatility and addresses limitations of fixed-pattern and model-dependent schemes, while remaining fixed at deployment.

II. MATERIALS AND METHODS

A. Simulation environment

All experiments were run in MATLAB/Simulink [R2024b]. The cardiovascular environment is a seven-compartment lumped-parameter model with ideal-diode valves and time-varying ventricular elastance. The plant executes with a fixed step time of 10^{-4} second. MATLAB/Simulink R2024b with MathWorks Reinforcement Learning Toolbox installed; all plant dynamics resided in Simulink, while the toolbox was used only to host the Q-learning agent (details in Section D). The same topology and baseline parameters were used unless otherwise stated. Pressures, flows, and volumes are reported in mmHg, mL/s, and mL, respectively.

B. System Architecture

Overall, the methodology was organized as a closed reinforcement-learning loop centered on reference synthesis. At each decision step, the agent maintained and updated a value function (details in section D.2) over aortic-flow references $Q_{Ao,ref}$ and produced a candidate reference waveform (details in section C). This reference was then supplied to a controller, which generated the corresponding LVAD command summarized as Q_{LVAD}; the commanded actuation drove the cardiovascular system (CVS) model. The CVS yielded observations from which scalar rewards were computed according to predefined hemodynamic objectives (details in section D.3). Since the agent and plant operated at different sample times, plant measurements were decimated and synchronized through a rate-transition stage before forming the next observation vector. The observations (details in section D.1) and rewards were then fed back to the agent (details in section D.4) which updated its value estimates and issued a new action that modified the aortic-flow reference parameters for the subsequent step. In this way, the cycle iterated continuously to align the learned reference with the desired physiological targets.

C. Parametric Encoding of the Aortic-Flow Reference

We parameterized the aortic-flow reference (1) as a superposition of ten sinusoids with an additional bias:

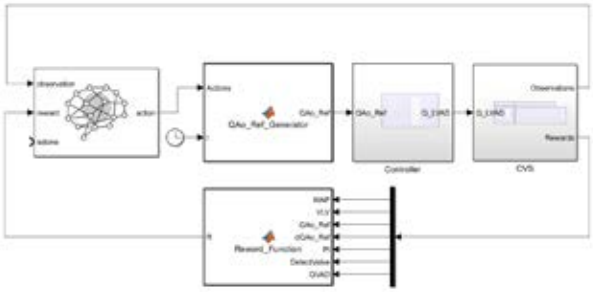


Fig. 1: Block Diagram of the System Architecture.

$$Q_{Ao,ref} = b_0 + \sum_{k=1}^{10} A_k sin(\omega_k t + \varphi_k). \quad (1)$$

The coefficient set consisted of the amplitudes A_k, angular frequencies ω_k, and phases φ_k, for $k = 1 \dots 10$ together with the DC term b_0 yielding 31 scalar parameters in total. These parameters were treated as the agent's actions (details in section C.4) and were learned offline through interaction with the numerical cardiovascular simulator.

D. RL design algorithm

1) Observation space: The policy received a seven-dimensional observation vector at each decision step: mean arterial pressure (MAP), left-ventricular volume V_{LV}, aortic-valve-open flag (0/1), pulsatility index (PI), aortic-flow $(Q_{Ao,ref})$, aortic-flow derivative $(dQ_{Ao,ref}/dt)$, LVAD-flow (Q_{LVAD}). MAP and V_{LV} were read directly from the CVS. $Q_{Ao,ref}$ and $dQ_{Ao,ref}/dt$ were fed back from the agent's current reference to provide context for subsequent decisions. The valve-open flag was generated by evaluating the physical condition $P_{LV} > P_{Ao}$, and $Q_{LV} > 0$. It was emitted as a unit pulse on each rising event. PI was computed over a sliding window as the moving (max-min) of Q_{LVAD} divided by the moving mean. Q_{LVAD} was taken from the controller output. Before being provided to the policy normalized to the interval $[-1, 1]$.

2) Action space: The policy emitted a continuous action vector $a_t \in \mathbb{R}^k$ that parameterized an aortic-flow reference generator. Elements of a_t set (i) a DC bias (mean flow level) and (ii) waveform-shape parameters (e.g., amplitudes/phases). Each element were constrained to fixed bounds to preserve physiological plausibility and maintain training stability. The baseline coefficient b_0 and all phase terms φ_k used global limits in every experiment: $b_0 \in [40100]$ and $\varphi_k \in [-\pi, \pi]$. The remaining coefficients followed scenario-specific intervals that matched the operating regime, as summarized in Table 1. These bounds were enforced at every update step. After convergence, the learned coefficient vector defined the aortic-flow reference waveform used in subsequent evaluations.

3) Reward function: At each decision step the agent received a scalar reward composed of seven additive terms, $R_T = R_1 + \cdots + R_7$, designed to (i) keep the learned aortic-flow reference within physiological bounds, (ii) regulate arterial pressure, (iii) avoid suction by preserving left-ventricular filling, (iv) keep pump flow in a safe operating window, (v)

TABLE I: Scenario-specific bounds for reference-waveform coefficients

k	$A_k^{\min}$	$A_k^{\max}$	$\omega_k^{\min}$	$\omega_k^{\max}$
1	-2	+2	0.5	10
2	-2	+2	0.5	10
3	-10	+10	0.1	5
4	-10	+10	0.1	5
5	-3	+3	0.1	3
6	-3	+3	0.1	3
7	-8	+8	0.5	8
8	-8	+8	0.5	8
9	-4	+4	0.1	3
10	-4	+4	0.1	3

promote periodic aortic-valve opening, (vi) pulsatility index, and (vii) limit aggressive upstrokes in the reference.

i. Aortic-flow reference shaping, $R_1(Q_{Ao,ref})$: This term (2) encouraged a strictly positive, physiologic mean aortic flow with a soft target around 5 L/min and penalties for non-physiologic values:

$$R_1 = \begin{cases} 5Q_{A0,ref}, & Q_{A0,ref} < 0, \\ -5, & Q_{A0,ref} = 0, \\ Q_{A0,ref}/5, & 0 < Q_{A0,ref} < 5, \\ 5/Q_{A0,ref}, & 5 < Q_{A0,ref} < 8, \\ -Q_{A0,ref}, & Q_{A0,ref} \geq 8. \end{cases} \quad (2)$$

This structure rewarded flows near 5 L/min, discouraged zeros/negatives, and penalized excessive outputs.

ii. Arterial-pressure regulation, $R_2(MAP)$: MAP was rewarded within 80–100 mmHg and penalized outside (3):

$$R_2 = \begin{cases} -0.3, & MAP = 0, \\ -0.3/MAP, & 0 < MAP < 80, \\ MAP/240, & 80 \leq MAP < 100, \\ MAP, & MAP \geq 100. \end{cases} \quad (3)$$

iii. LV-volume safeguard, $R_3(V_{LV})$: To avoid suction, small ventricular volumes were penalized (4) while physiologic volumes (≈30-150 ml) were modestly rewarded:

$$R_3 = \begin{cases} 0.2V_{LV}, & V_{LV} < 0, \\ -0.2, & V_{LV} = 0, \\ -0.2/V_{LV}, & 0 < V_{LV} < 30, \\ V_{LV}/1500, & 30 \leq V_{LV} < 150, \\ 0, & V_{LV} \geq 150. \end{cases} \quad (4)$$

iv. LVAD-flow window, $R_4(Q_LVAD)$: Pump flow was encouraged to remain within 20-150 ml/s (≈1-9 L/min) and penalized otherwise (5).

$$R_4 = \begin{cases} -0.1, & Q_{LVAD} = 0, \\ -0.1/Q_{LVAD}, & 0 < Q_{LVAD} < 40, \\ Q_{LVAD}/4000, & 40 \leq Q_{LVAD} < 140, \\ -Q_{LVAD}/1000, & Q_{LVAD} > 140. \end{cases} \quad (5)$$

v. Aortic-valve opening incentive, $R_5(ValveOpen)$: To promote periodic washout and physiologic pulsatility as specified (6);

$$R_5 = \begin{cases} 0.1, & ValveOpen = 1, \\ -0.1, & ValveOpen = 0. \end{cases} \quad (6)$$

vi. Pulsatility index shaping, R_6: PI was computed as (7) [8]:

$$\frac{\max_{\tau\in[t-10,\,t]} Q_{\mathrm{LVAD}}(\tau) \;-\; \min_{\tau\in[t-10,\,t]} Q_{\mathrm{LVAD}}(\tau)}{\frac{1}{10}\int_{t-10}^{t} Q_{\mathrm{LVAD}}(\tau)\,d\tau} \quad (7)$$

This term promoted physiologic pulsatility (8) while discouraging both near-aperiodic and overly large oscillations:

$$R_6 = \begin{cases} -PI10^{-3}, & PI < 0 \; OR \; PI > 8, \\ -10^{-3}, & 0 \leq PI < 2, \\ PI10^{-3}, & 2 \leq PI \leq 8, \end{cases} \quad (8)$$

vii. Slope penalty, $R_7(dQ_{Ao,ref}/dt)$: To enforce a bandwidth-limited, smooth aortic-flow reference and suppress chatter, we penalized the instantaneous slope symmetrically for up- and down-ramps (9):

$$R_7 = -0.01|dQ_{Ao,ref}/dt| \quad (9)$$

The total reward was the unweighted sum of all components. Each term's scale (constants and piecewise bounds) was selected to keep individual contributions in typical operating ranges, ensuring that pressure regulation, suction avoidance, safe pump operation, and valve opening were simultaneously reflected in the learning signal without any single metric dominating.

4) Agent & training: Training proceeded in episodes of 10 s. The actor and critic were multilayer perceptrons with two hidden layers of 64 units and ReLU activations. Actor outputs were squashed and then affinely rescaled to the action bounds listed in Table 1.

Parameters were optimized with Adam using learning rates of 10^{-4} for the actor and 10^{-3} for the critic, with rate-transition handling and no weight decay. The discount factor was $\gamma = 0.99$.

III. RESULTS

The raw aortic-flow reference exhibited high-frequency transients. We therefore applied a low-pass filter followed by a smoothing operator, which suppressed spurious spikes and yielded a continuous, bandwidth-limited waveform suitable for actuator tracking (Fig. 2).

Simulation with the filtered reference yielded physiologically acceptable dynamics: aortic flow (Q_{Ao}) tracked the prescribed waveform $(Q_{Ao,ref})$ within a 2–5 L/min band. Q_{LVAD} supplied the predominant share of cardiac output, while native ejection reappeared whenever LVAD flow transiently approached 1L/min, indicating coordinated interaction between the assist device and the ventricle (Fig. 3).

At 60 bpm, hemodynamics was stable and aortic pressure remained pulsatile. Ventricular volumes (V_{LV}, V_{RV}) exhibited

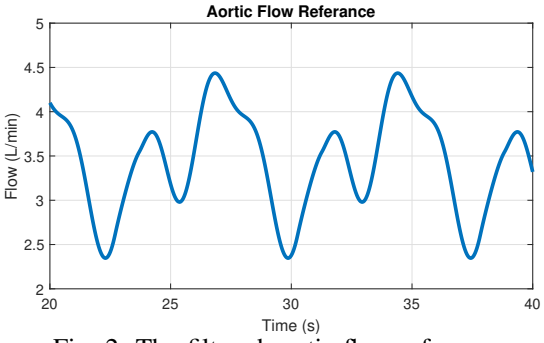


Fig. 2: The filtered aortic-flow reference.

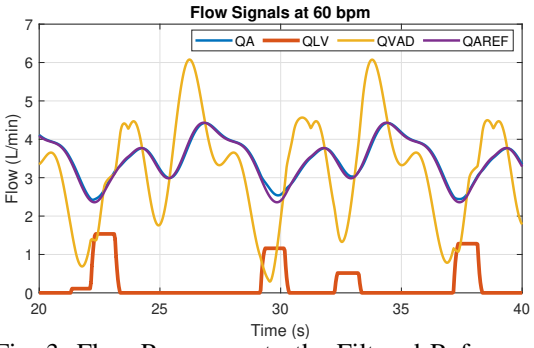


Fig. 3: Flow Responses to the Filtered Reference.

regular filling–ejection cycles: V_{LV} oscillated roughly 70–140 mL with no approach to collapse. No suction events or runaway pressures were observed as shown (Fig. 4).

IV. DISCUSSION & CONCLUSIONS

This study shows that casting aortic-flow reference as a reinforcement-learning problem restored pulsatility while maintaining pressure targets in a numerical CVS model. Optimizing at the reference layer kept the approach controller-agnostic and compatible with existing speed controllers. In simulation, the learned waveform promoted aortic-valve opening and improved root-washout surrogates without suction. These gains resulted from an explicit reward trade-off between pressure regulation and pulsatility/washout rather than ad-hoc speed patterns. Pulsatility was quantified using an LVAD-flow PI rather than an aortic-flow PI. Generally, episodes in which the measured LVAD flow (Q_{LVAD}) approaches zero are a reliable marker of native ventricular ejection, whereas a near-zero aortic-flow reference $(Q_{Ao,ref})$ is merely a command and does not confirm that the pump actually reduced flow at that instant; we therefore considered the LVAD-based PI the more appropriate metric. Baseline speed controller tracked the aortic-flow reference with $\sim 90\%$ fidelity across operating points, sufficient to reveal benefits and limits of reference-level learning. Tighter tracking would likely sharpen predefined hemodynamic objectives but lies outside our scope.

Limitations: (i) learning and evaluation were performed only at 60 bpm; higher rates and rhythm variability may require different optima and reward weights. (ii) The CVS included ideal-diode valves and lumped elastance, omitting valve inertia, coronary coupling, and baroreflex effects. (iii) Evaluation was in silico. Learning the aortic-flow reference is a viable route to reintroduce pulsatility and valve opening while robustly preserving pressure. Future work will broaden training, incorporate aortic-based metrics, explore controller–reference co-design, and validate on hardware and in vivo.

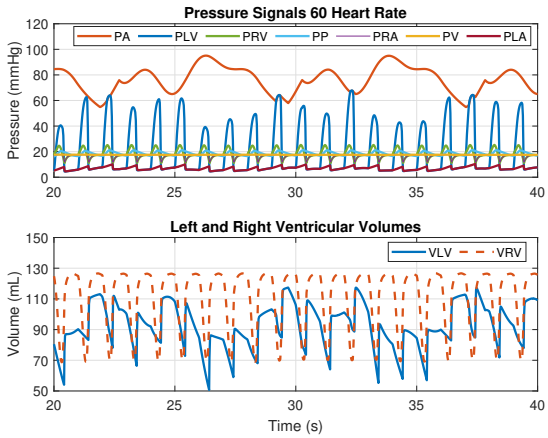


Fig. 4: Pressures and Ventricular Volumes at 60 bpm. Where VLV and VRV are left and right ventricular volumes and PLV, PA, PV, PRA, PRV PP and PLA are left ventricular, aortic, venous, right atrial, right ventricular, pulmonary, and left atrial pressures respectively

ACKNOWLEDGMENT

This work was initiated and completed at the Physiological Control Laboratory with funding from the Scientific and Technological Research Council of Turkey (TUBITAK) 1003 Project (No. 118S098) and Yildiz Technical University Scientific Research Projects Coordination Unit (BAP) (No. FBA-2024-6491)

REFERENCES

[1] National Center for Health Statistics, "Multiple cause of death 2018–2021 on cdc wonder database," https://wonder.cdc.gov/mcd.html?utm=0pt, 2018.

[2] L. W. Miller and J. G. Rogers, "Evolution of left ventricular assist device therapy for advanced heart failure: a review," *JAMA cardiology*, vol. 3, no. 7, pp. 650–658, 2018.

[3] S. A. Rabi and D. A. D'Alessandro, "Suction event after lvad placement," in *Near Misses in Cardiac Surgery*. Springer, 2022, pp. 179–181.

[4] D. Dobarro, M. Urban, K. Booth, N. Wrightson, J. Castrodeza, J. Jungschleger, N. Robinson-Smith, A. Woods, G. Parry, S. Schueler *et al.*, "Impact of aortic valve closure on adverse events and outcomes with the heartware ventricular assist device," *The Journal of Heart and Lung Transplantation*, vol. 36, no. 1, pp. 42–49, 2017.

[5] D. Fernández-Zapico, T. Peirelinck, G. Deconinck, D. W. Donker, and L. Fresiello, "Physiological control for left ventricular assist devices based on deep reinforcement learning," *Artificial organs*, vol. 48, no. 12, pp. 1418–1429, 2024.

[6] A. Alhajyounis, Y. Aydoğan, A. Canbaş, E. Kulakoğlu, U. Kurnaz, L. Soysal, M. Yiğit, D. Arısoy, and K. Kadıpaşaoğlu, "Nonlinear control of a hybrid pneumo-hydraulic mock circuit of the cardiovascular system," 2025.

[7] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," *Supervised and unsupervised learning for data science*, pp. 3–21, 2020.

[8] J. E. Dodge, A. L. Covens, C. Lacchetti, L. M. Elit, T. Le, M. Devries-Aboud, M. Fung-Kee-Fung, G. C. D. S. Group *et al.*, "Preoperative identification of a suspicious adnexal mass: a systematic review and meta-analysis," *Gynecologic oncology*, vol. 126, no. 1, pp. 157–166, 2012.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Deep Learning Framework for B-Mode Ultrasound Image Reconstruction

Mahsa Mikaeili[1]
mahsa.mikaeili@okan.edu.tr
Istanbul Okan University, Mechatronic Engineering, Istanbul, Turkey[1]

Hasan Şakir Bilge[2,3]
bilge@gazi.edu.tr
Gazi University, Biomedical Calibration and Research Center, Ankara, Turkey[2]
Gazi University, Electrical and Electronics Engineering, Ankara, Turkey[3]

***Abstract*— We present a U-Net–based pipeline for B-mode ultrasound image reconstruction that ingests post-processed RF data, predicts a log-compressed image, and then performs an interpolation for display. To identify effective design choices, we compare six configurations formed by two loss functions: mean-squared error and compound MMUAE+TV+gradient loss, and three output activation functions: linear, ReLU, and tanh. Evaluation with PSNR, SSIM, and visual inspection of scan line profiles and B-mode images indicates that the activation function is the dominant factor. Tanh function consistently preserves lesion boundaries, maintains realistic speckle, and avoids dynamic-range saturation; linear function is acceptable but yields softer edges; ReLU degrades contrast owing to negative-value clipping. Switching from MSE to the compound loss function produces only modest changes, suggesting that regularization is secondary to activation choice in this setting. Overall, a tanh-activated U-Net with a simple MSE objective offers a strong accuracy–complexity trade-off for reconstruction.**

Keywords—Ultrasond Image; raw RF data; reconstruction; deep neural network

I. INTRODUCTION

Ultrasound imaging (US) is among the most widely used medical imaging modalities due to its noninvasiveness, lack of ionizing radiation, affordability, and real-time imaging capability. Because ultrasound offers key advantages, it has received sustained research interest, and significant work has targeted better image quality, including higher resolution, improved contrast, and greater signal-to-noise ratio (SNR).

One of the key approaches to enhancing US image quality is beamforming techniques. Conventional beamforming methods include delay-and-sum (DAS), minimum variance beamforming (MVB), and short-lag spatial coherence (SLSC). However, these methods are often limited by reduced image quality and high computational complexity [1].

With the emergence of deep neural networks (DNNs) and their demonstrated high performance in various tasks, including classification, segmentation, and detection, some researchers have adopted this powerful approach for beamforming applications and enhancing US resolution. In [2], the authors classified these approaches into three categories based on the data type utilized: image-based, pre-beamforming, and post-beamforming methods. Among these approaches, image-based methods apply deep neural networks (DNNs) at the final stage of the beamforming process. Pre-beamformed RF–based methods operate directly on raw RF channel data before beamforming. In contrast, post-beamformed methods utilize RF data after beamforming but before envelope detection and logarithmic compression. The block diagram in Fig. 1 illustrates the US image reconstruction process and data categories between stages. In [3], the authors proposed a convolutional neural network (CNN)-based two-step reconstruction method to address the diffraction issue in ultrafast ultrasound imaging. In [4], the authors employ a U-Net architecture to learn the nonlinear mapping between low-quality and high-quality images, focusing on reconstructing high-quality RF images and preserving speckle patterns. Also, in [5], the authors propose an

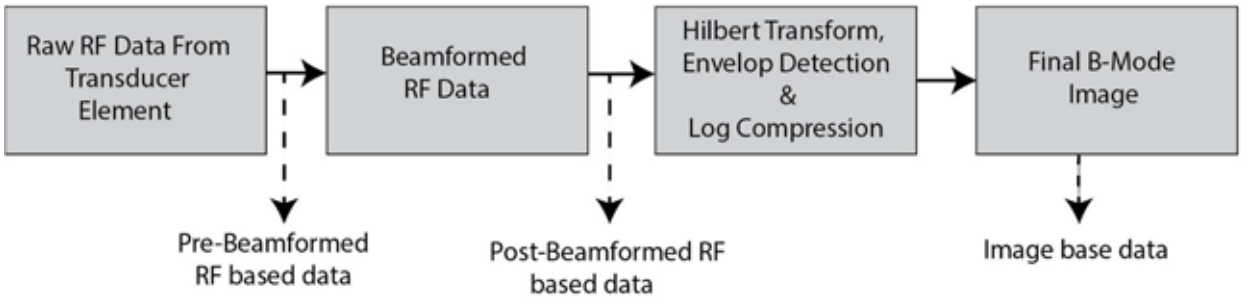


Fig.1. Block diagram of the US image reconstruction process, including data categories between stages.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

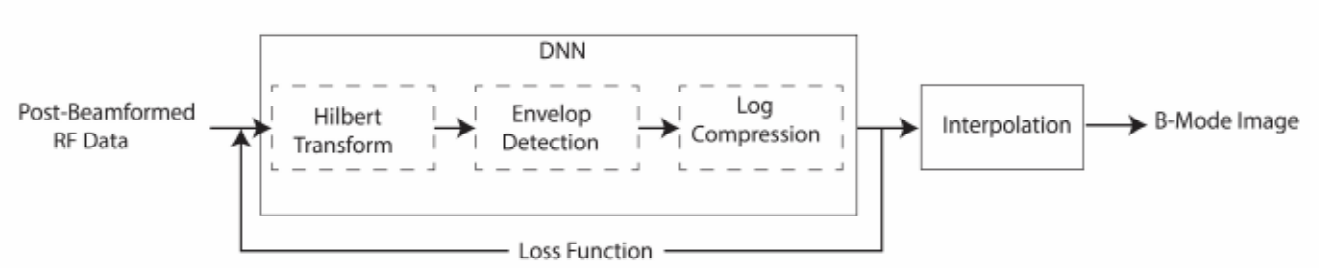


Fig.2. Block diagram of the proposed method. The last processing part of the US image reconstruction process is replaced with a DNN.

end-to-end deep learning framework for plane wave imaging (PWI) that integrates Spatial Matched Filter (SMF)-based beamforming with a residual U-Net for speckle reduction, followed by envelope detection and log compression to enhance image quality. In another study, authors propose a deep learning–based beamforming method for PWI that uses a Fully Convolutional Neural Network with multi-resolution wavelet decomposition and step-by-step training to map RF channel data to high-quality images defined by a sharp Gaussian Point Spread Function (PSF), enabling sharper reconstructions without reducing frame rate [6]. In [7], the authors apply a Generative Adversarial Network (GAN) for multi-focus ultrasound image recovery that maintains frame rate while achieving constant lateral resolution. Their approach learns a mapping from single-focus to multi-focus images using adversarial and Mean Squared Error (MSE) losses, combined with a boundary-seeking strategy to improve training stability.

Evaluated another study is [8]. This work introduces two deep learning–based methods: Inception for DW (Diverging Wave) Network (IDNet), an inception CNN for high-quality diverging wave image reconstruction from few transmissions, and a GAN-based approach for multi-focus ultrasound recovery, enhancing resolution and contrast while maintaining frame rate without requiring raw channel data. Also, in [9], a self-supervised deep neural network that works with an encoder and decoder strategy with sparse regularization-based loss is proposed for ultrafast plane-wave ultrasound image reconstruction from RF data, achieving high-quality results with substantially reduced computational time compared to traditional methods. In [10], the authors propose a neural network–based beamforming framework for ultrasound that uses normalization-independent loss functions to reconstruct speckle-reduced B-mode images from sub-aperture RF signals, demonstrating effective generalization from simulated training to real-world imaging while preserving resolution and detail. Another investigated work is [11]. This work proposes a frequency-domain DNN approach for suppressing off-axis scattering in ultrasound channel data, using short-time Fourier transform processing and per-frequency network training to achieve enhanced contrast and side-lobe suppression compared to conventional beamforming methods. Also, in [12], authors introduce a deep learning-based reconstruction framework that uses an encoder–decoder CNN to generate high-quality B-mode ultrasound images from temporally and spatially sub-sampled channel data. They claim their method can achieve resolution and contrast comparable to advanced beamforming methods while reducing acquisition requirements. In [13], the authors propose a two-stage deep learning framework combining a fully convolutional PixelNet and a conditional GAN to reconstruct high-quality ultrasound images from single plane-wave RF data by reducing angle dependence through learned data transformation.

This study aims to evaluate the performance of a U-Net architecture for reconstructing ultrasound B-mode images while omitting last stages of conventional pipelines, and to systematically assess the impact of the loss function and output activation on reconstruction quality. We benchmark two loss formulations (mean-squared error and a compound MMUAE+TV+gradient loss) and three activations: linear, ReLU, and tanh. The remainder of the paper is organized as follows: Section II details the methodology, network configuration, datasets, and the evaluation protocol for losses and activations; Section III presents and discusses the results; and Section IV concludes the study.

II. Materials and Methods

This study employs a U-Net architecture to reconstruct a US B-mode image. Post-beamformed data undergoes multiple processing stages (Fig.1) in the conventional US image reconstruction process before producing a B-mode image. These stages typically include the Hilbert transform for quadrature signal generation, envelope detection to extract the signal magnitude, logarithmic compression to adjust the dynamic range, and finally, interpolation to improve image resolution. All stages from the Hilbert transform through logarithmic compression are omitted in the proposed approach. Instead, the U-Net directly reconstructs log-compressed data from the post-beamformed data. Finally, an interpolation step is applied to generate the B-mode image. This approach simplifies the US image reconstruction process, replacing the last processing part with a DNN. A schematic representation of the proposed method is shown in Fig. 2.

A. Applied Network

A U-Net architecture was employed to map post-beamformed ultrasound data directly to log-compressed data. The network adopts a standard encoder–decoder configuration, comprising four down-sampling stages in the encoder and four corresponding up-sampling stages in the decoder. Each encoder stage consists of two 3×3 convolutional layers followed by 2×2 max-pooling, with the number of feature channels doubled at each level to progressively capture higher-level representations. The decoder mirrors this process, utilizing up-sampling and

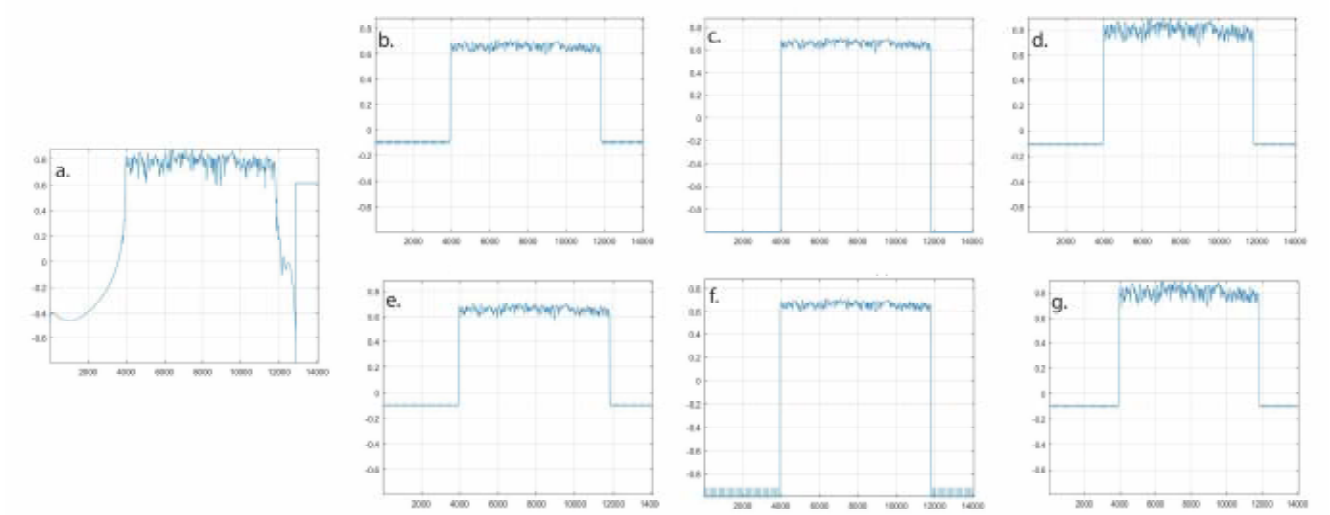


Fig.3. (a) Ground truth signal. A typical scan line reconstructed with (b) MSE loss and Linear activation function, (c) MSE loss and ReLU activation function. (d) MSE loss and Tanh activation function, (e) MMUAE+TV+Grad loss and Linear activation function, (f) MMUAE+TV+Grad loss and ReLU activation function, (g) MMUAE+TV+Grad loss and Tanh activation function.

skip connections to integrate high-resolution spatial information from the encoder, followed by convolutional refinement. To investigate the effect of nonlinearity on the reconstruction task, the model was trained under different activation function configurations, including linear, ReLU, and tanh, applied within the convolutional blocks and at the output layer. The final layer incorporates a 1×1 convolution, where the activation choice determines the output range and intensity scaling. Furthermore, two different loss functions were systematically evaluated. One of these is the conventional mean squared error (MSE), which penalizes voxel-wise intensity discrepancies between predictions and targets, Eq.1

$$L_{MSE}(y_{true}, y_{pred}) = \frac{1}{N}\sum_{i=1}^{N}(y_{true}^{(i)} - y_{pred}^{i})^2 \quad (1)$$

The second loss function is a composite formulation that combines masked mean absolute error (MMUAE), total variation (TV), and gradient similarity (MMUAE+TV+Grad), Eq.2, designed to enforce smoothness while preserving structural edges. This comparative analysis enables a rigorous assessment of the combined impact of loss function design and activation strategy on ultrasound image reconstruction performance.

$$L_{total} = L_{MMUAE} + \alpha L_{TV} + \beta L_{grad} \quad (2)$$

In this formulation, α=0.01 regulates the degree of smoothness enforced by the TV loss, while β=0.01 balances the preservation of structural boundaries.

B. Data Acquisition and Experimental Setup

To generate the data required for training and testing the proposed model, the Field II simulation program [14] was employed. In total, 542 US images were simulated, representing phantoms containing cysts and regions of high scatter density. The diameters of these regions were set to 6, 5, 4, 3, or 2 mm, covering a range of clinically relevant lesion sizes to assess the model's ability to reconstruct structures of varying scales. Each simulated phantom contained one, two, or three structures, which could be reflectors, cysts, or a combination of both, thereby introducing structural diversity and improving the model's generalization capability. This simulation setup was designed to mimic realistic ultrasound imaging conditions by incorporating variations in target size, scatterer distribution, and structural composition, enabling robust evaluation of the proposed reconstruction method.

The simulation parameters were a linear array transducer with a center frequency of 3.5 MHz, 64 active elements, and a scan line of 128. Both input and output datasets were normalized to the range [−1,1] before training to ensure numerical stability. Of the simulated dataset, 100 images were reserved for testing, and the remainder was used for training. Model training was performed with a learning rate of 0.0001 over 25 epochs. The experiments were conducted on a workstation with an Intel Core i7 8th-generation CPU (4 GHz, 32 GB RAM) and an NVIDIA GeForce GTX GPU.

III. Results and Discussion

In this study, by employing the U-Net structure, we evaluate the performance of multiple combinations of loss functions (MSE, MMUAE+TV+Grad) and activation functions (Linear, ReLU, Tanh) in reconstructing a US B-mode image (Fig.3). Based on Fig.3.b, by employing MSE+Linear, the overall shape is recovered, but mild ringing appears near the edges, and the plateau is slightly uneven. However, MSE+ReLU clips negative values, biasing the baseline and flattening transitions, which explains its poor fidelity. Also, in Fig.3.d best matches the reference, preserving edge sharpness and plateau level while keeping outputs bounded. Using the compound MMUAE+TV+Grad loss with Linear (e) yields a profile similar to (b) but a bit smoother on the plateau due to TV regularization, at the cost of slight edge softening. The ReLU variant with the compound loss (f) again suffers from clipping and distorted transitions. Finally,

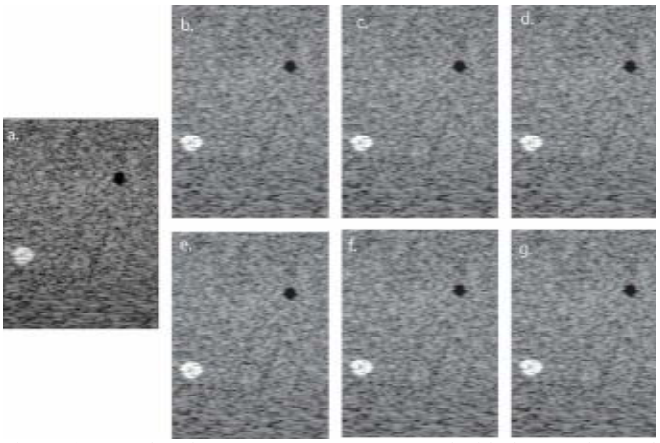


Fig.4. (a) Ground truth Image. Images reconstructed with (b) MSE loss and Linear activation function, (c) MSE loss and ReLU activation function, (d) MSE loss and Tanh activation function, (e) MMUAE+TV+Grad loss and Linear activation function, (f) MMUAE+TV+Grad loss and ReLU activation function, (g) MMUAE+TV+Grad loss and Tanh activation function.

Table. 1. Acquired results with combinations of loss and activation functions.

	MSE			**MMUAE+TV+Grad**		
	Linear	***ReLU***	***Tanh***	***Linera***	***ReLU***	***Tanh***
PSNR	18.281	9.321	19.119	18.251	9.597	19.160
SSIM	0.601	0.552	0.631	0.605	0.551	0.623

MMUAE+TV+Grad+Tanh (g) closely tracks (d)**,** suppressing small ripples with only minimal loss of edge crispness.

Also, Fig.4 and Table 1 demonstrate qualitative and quantitative results of the applied method. The tanh outputs MSE+tanh (d) and MMUAE+TV+Grad+tanh (g) most closely resemble the ground truth: the anechoic cyst interior is uniformly dark, its boundary is crisp with minimal haloing, and the bright point target is compact without blooming. Background speckle looks statistically consistent, which matches their top metrics **(**PSNR ≈ 19.1–19.2 dB, SSIM ≈ 0.623–0.631). The linear variants (b) and (e) recover structure but show slightly elevated background variance and mild edge ringing; cyst edges are a touch softer, and the point target is a bit more spread, consistent with the modestly lower scores**.** The ReLU cases (c) and (f) clip low-intensity regions, compressing dynamic range: speckle appears washed, the cyst shows noticeable fill-in, and edges and point targets lose contrast. This visual degradation aligns with the significant metric drop. Comparing losses within the same activation, MMUAE+TV+Grad mainly smooths small ripples in the speckle (e.g., g vs d, e vs b) but also slightly softens edges, which is why metric changes are minor. Overall, the tanh activation yields the best compromise of edge sharpness, speckle fidelity, and lesion contrast; linear is acceptable; ReLU is unsuitable for B-mode reconstruction because of dynamic-range clipping.

IV. Conclusion

Using a U-Net architecture, we reconstruct B-mode images from post-beamformed RF inputs, predict the log-compressed output, and then apply interpolation to obtain the final B-mode image. Across the examined settings, the activation function was the primary determinant of image quality: the tanh function produced the most faithful reconstructions with crisp boundaries and realistic speckle; the linear function was acceptable but slightly softer; and the ReLU function consistently degraded contrast due to dynamic-range clipping. In contrast, changing the loss from MSE to the compound MMUAE+TV+Grad yielded only marginal differences, reducing minor ripple artifacts while mildly softening edges. For this post-beamformed RF to log-compressed data pipeline followed by interpolation, a tanh-activated U-Net with a simple MSE loss (optionally with light TV regularization) offers the best accuracy–complexity trade-off.

Acknowledgment

This work was supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK) under Grant Number 122E140.

References

[1] Li, Z., Wiacek, A., & Bell, M. A. L. (2020, September). Beamforming with deep learning from single plane wave RF data. In *2020 IEEE International Ultrasonics Symposium (IUS)* (pp. 1-4). IEEE.

[2] Nguon, L. S., Seo, J., Seo, K., Han, Y., & Park, S. (2022). Reconstruction for plane-wave ultrasound imaging using a modified U-Net-based beamformer. Computerized Medical Imaging and Graphics, 98, 102073.

[3] Perdios, D., Vonlanthen, M., Martinez, F., Arditi, M., & Thiran, J. P. (2021). CNN-based image reconstruction method for ultrafast ultrasound imaging. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 69(4), 1154-1168.

[4] Perdios, D., Vonlanthen, M., Besson, A., Martinez, F., Arditi, M., & Thiran, J. P. (2018, October). Deep convolutional neural network for ultrasound image enhancement. In 2018 IEEE International Ultrasonics Symposium (IUS) (pp. 1-4). IEEE.

[5] Mor, E., & Bar-Hillel, A. (2020). A unified deep network for beamforming and speckle reduction in plane wave imaging: A simulation study. Ultrasonics, 103, 106069.

[6] Goudarzi, S., & Rivaz, H. (2022). Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study. Ultrasonics, 125, 106778.

[7] Goudarzi, S., Asif, A., & Rivaz, H. (2020). Fast multi-focus ultrasound image recovery using generative adversarial networks. IEEE Transactions on Computational Imaging, 6, 1272-1284.

[8] Lu, J., Millioz, F., Garcia, D., Salles, S., Liu, W., & Friboulet, D. (2020). Reconstruction for diverging-wave imaging using deep convolutional neural networks. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 67(12), 2481-2492.

[9] Zhang, J., He, Q., Xiao, Y., Zheng, H., Wang, C., & Luo, J. (2021). Ultrasound image reconstruction from plane wave radio-frequency data by a self-supervised deep neural network. Medical Image Analysis, 70, 102018.

[10] Hyun, D., Brickson, L. L., Looby, K. T., & Dahl, J. J. (2019). Beamforming and speckle reduction using neural networks. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 66(5), 898-910.

[11] Luchies, A. C., & Byram, B. C. (2018). Deep neural networks for ultrasound beamforming. IEEE transactions on medical imaging, 37(9), 2010-2021.

[12] Mamistvalov, A., Amar, A., Kessler, N., & Eldar, Y. C. (2022). Deep-learning-based adaptive ultrasound imaging from sub-nyquist channel data. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 69(5), 1638-1648.

[13] Wasih, M., Ahmad, S., & Almekkawy, M. (2023). A robust cascaded deep neural network for image reconstruction of single plane wave ultrasound RF data. Ultrasonics, 132, 106981.

[14] https://field-ii.dk/

COMPARATIVE ERROR ANALYSIS OF 3D- MODELING, -PRINTING AND -SCANNING METHODS USED IN ORTHODONTICS

Rabia Şimal Selamet
Dept. of Mechanical Engineering
Istinye University
Istanbul, Türkiye

Dr. Şenol Pişkin
Dept. of Mechanical Engineering
Istinye University
Istanbul, Türkiye

Abstract— *The use of additive manufacturing in dentistry has exponentially increased, with dental model construction being the most common use of the technology. [1] This requires knowledge of the precision of these models. This is especially important for images used in clinical settings. This project proposal aims to conduct a comparative error analysis of 3D modeling, 3D printing, and 3D scanning methods used in orthodontics to measure the accuracy of 3D images and models used in dentistry and jaw surgery. Anonymized patient images (without identifying information) will be used in the project. The scanned images will be processed to create 3D anatomical models, which will then be printed using a 3D printer. The generated anatomies will then be scanned with a 3D scanner to create a new 3D model. The two resulting models (image processing and scanning results) will be compared by superimposing their axes. This mapping will enable quantitative and qualitative error analysis.*

Keywords— *Orthodontics, Clinical Image Processing, 3D Modeling, 3D Printing, 3D Scanning*

I. Introduction

3D models provide significant benefits in pre-surgical planning, various biomedical research studies, and the education of medical students. With the development of 3D printing technology, its medical applications are expected to expand in the coming years with innovations such as reduced production costs, faster anatomical modeling, and the direct production of organs through bioprinting. [2]
This widespread use raises the question of how precise these models are. Considering that doctors work with models and perform surgeries, precision is a crucial parameter. This project will measure precision and also explore whether modeling is feasible with affordable products.

II. BackGround

Certain steps are used when transferring any object into a three-dimensional environment, such as a chair, car, house, heart, jaw, etc. Since this project will focus on orthodontics and jaw surgery, this section provides information and examples based on medical research.
Similar method, which is explained in the method part applied were in orthognathic surgery for improving the predictability of surgical treatment procedures and outcomes, including data acquisition, image analysis, surgical simulation, surgical intervention, and outcome assessment Lin, HH., Chang, shows that models which used before "LeFort I osteotomy" and "maxillomandibular complex surgery determined ranged from 0.18 to 0.33 mm in the maxilla and from 0.99 to 1.56 mm in the mandible, validation result [3]
Moreover, Narita M, Takaki T used a similar method, but they tested the mean amount of bleeding and mean operating time difference between with 3d models and without 3d models; we tested the deviation analyses between scanning method outcomes [4].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

III. Methods

A. Digital Twin Creation (Imaging, Segmentation, Reconstruction)

Image slices obtained in DICOM format will be compiled into a single directory and uploaded to 3DSlicer (BWH and the Slicer community) for processing. From the clinically obtained datasets (this study will use anonymized scanned patient images. These images are publicly available due to the studies conducted and the permissions obtained. Therefore, no additional ethics committee approval is required; the most appropriate set in terms of sequence and image quality will be determined, and the segmentation phase will begin. The anatomy will be inked on all slices. The inking process will begin with defining the outer borders and continue with automatic inking of the area within the borders. The three-dimensional anatomical structure will be revealed by combining the inked sections. Using this method, two-dimensional pixels will be converted to 3D voxels. The resulting 3D node set will then be saved in STL (stereolithography) file format (this file format is one of the most suitable standard formats for 3D modeling, editing, and 3D printing (folding, additive manufacturing).

B. 3D Printing

The 3D printer detects the file with the '.stl' extension and performs the printing process. 3D printers operate using a production method called additive manufacturing. While many raw materials are used for printing, thermoplastic materials, also known as filaments, are generally used. 3D printers require a three-dimensional model and design to function. Drawings created with a CAD (Computer Aided Design) program or objects scanned with a 3D scanner are exported with the '.stl' extension. [5]
3D printing technologies; [6]

- Stereolithography (SLA)
- Selective Laser Sintering (SLS)
- Fused Deposition Modeling (FDM)
- Multi Jet Fusion (MJF)

- PolyJet
- Direct Metal Laser Sintering (DMLS)

The printing process will be carried out using our university's printer. The printer supports FDM, SLA, DLP, and PolyJet technologies.

C. 3D Scanning

3D scanning is a quick and accurate method that enables the digitization of physical objects. It captures the object's x, y, and z coordinates to create a digital replica. Reverse engineering, which is utilized across various industries, aids in product development while reducing design and manufacturing costs and time. By using 3D scanners, the data collection process from physical objects can be significantly accelerated. [7]

3D scanning technologies: [8]

- Optical-based 3D scanning
- Laser-based 3D scanning
- Laser pulse-based 3D scanning
- Photogrammetry
- Structural illumination-based 3D scanning
- Contact-based 3D scanning

The optical scanning method used in orthodontics was used to compare the analysis results obtained from other scans. Photogrammetry and lidar scanning, considered more accessible and cost-effective methods, were also tested with different applications. It was observed that the lidar scan obtained with the KiriEngine yielded the best results.

Pre-Scan Preparation:

1_ Using a platform: It was used as a height because scanning very small objects was difficult.

2_ Attaching stickers to the platform: The influence of colorful stickers on the surface

As S. Rianmora and K. Nuamchit noted in their work, dark tones, especially green, were not detected by a 3D laser scanner in experiments using additive colors (RGB color model). While red stickers were detected quickly, they resulted in noise and unwanted areas. I applied this sticker-based scanning method suggested by the researchers in my own study and achieved better results, especially on larger objects. However, even with smaller objects, I achieved extremely clear and successful scanning results on the maxilla and mandible models I used in the project by positioning the stickers on the platform surface (because it was an organic model, attaching stickers to the surfaces was difficult).

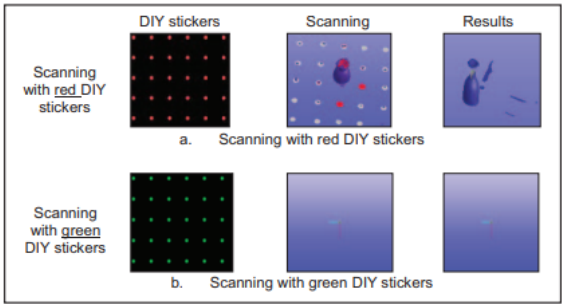


Figure 1: compared red (a.) and green (b.) DIY sticker that used in experiment

3_ Covering the surface with baby powder to prevent the object from shining.

4_ Proper lighting

Scanning Process:

Throughout the scanning process, scanning methods such as Optical-Based 3D Scanning, Laser Pulse-Based 3D Scanning, and Photogrammetry were used. When comparing photogrammetry and lidar results obtained with various apstore applications, lidar was observed to yield better results, so we continued with it. İstinye Dent Hospital resources were utilized for optical scanning.

Post-Scan Editing:

To prepare the scanned object for analysis Autodesk Meshmixer used .Some operations applied are cropping, resizing and smoothing



Figure 2: To avoid problems arising from the object's small size during scanning, the added platform was removed from the image.

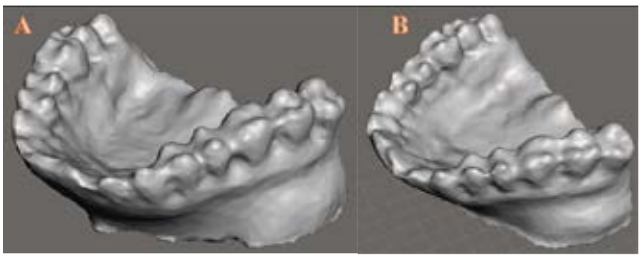


Figure 3: The scanned image was solidified, and sharp corners were smoothed. A: Light scanned image of the part coated with baby powder, edited in Meshmixer., B: Scanned image with dot stickers, edited in Meshmixer.

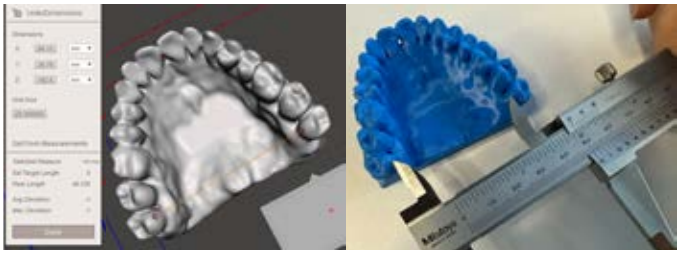


Figure 4: scanned image were measured using Meshmixer and compared with a caliper

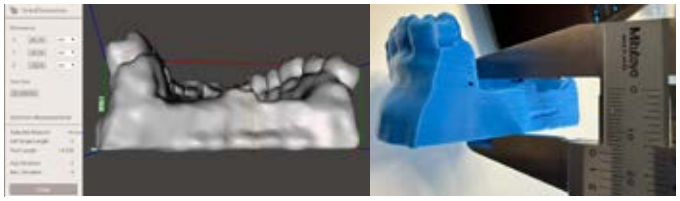


Figure 5: To verify the accuracy of the 3D printed and scanned images, the lengths of certain points in the scanned image were measured using Meshmixer and compared with a caliper.

D. Deviaation Analysis

Comparing the 3D model with the scanned model will be carried out using CatiaV5R21, a method similar to that used by ALPACA to characterize the evolutionary and ecological diversity of species. As shown in Figure a, two different 3D structures, blue and red, will be superimposed by matching their x, y, and z coordinates. The parts we define in different colors will reveal the differences between the structures after registration. These differences will provide us with sensitivity values. [10]

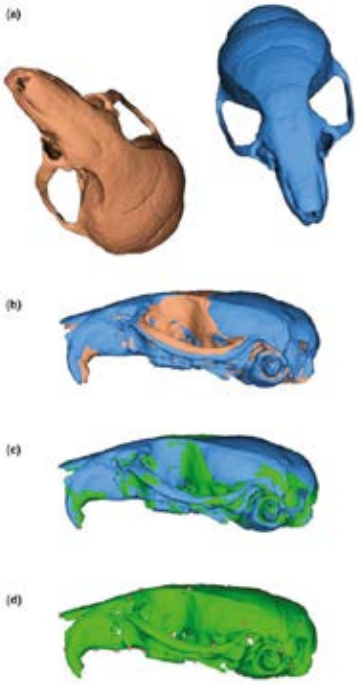


Figure 6: Image of two three-dimensional models mapping using ALPACA

Deviation Analysis Steps with Catia:
1. Start > Shape > Digitized Shape Editor
2. Import > Select to Files (Since the scanned image was 10^3 times smaller for the size control, a 10^3 scale was selected when uploading the scanned image to Catia, while the digital twin image was selected as 1.)
3. Mapping (performed by aligning the objects on top of each other).
4. Deviation Analysis > Choose Reference Part and Part to be Measured > Apply



Figure 7: deviation analysis steps

IV. Result And Discussion

Material	Scannıng method	Uygulama/ Maximum Sapma	Eksi yöndeki maksimum sapma (mm)	Pozitif yöndeki maksimum sapma (mm)
PLA	LIDAR SCANNİNG	Mandibula	-3,23	3,20
		Maxilla	**-4,27**	**3,05**
	OPTICAL SCANNING	Mandibula	-3,33	1,38
		Maxilla	**0,79**	**2,65**

In the comparison made for the maxilla, the deviation analyses resulted in the following order: optical scanning of the PLA model < lidar scanning of the PLA model.

When evaluated using the Lidar scanning method, the PLA model showed higher deviation values compared to optical scanning. This indicates that optical scanning provided more accurate results for the PLA model, especially for the maxilla, with lower maximum deviation values in both negative and positive directions.

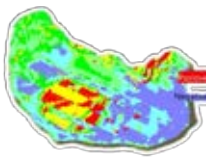


Figure 8: Pla based lidarl scanning results of maxilla model.

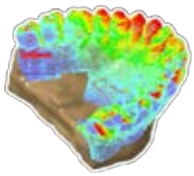


Figure 9: Pla based optical scanning results of maxilla model.

V. Conclusion

In conclusion, this study validated the theoretical accuracy range of orthodontic optical scanning methods under clinical-like conditions and benchmarked them against more affordable and readily available alternatives. Findings indicated that, when applied to a PLA model, the optical scanner achieved superior performance compared to competing systems.

Acknowledgment

I would like to thank İstinye University XR Labs for providing access to their facilities and İstinye Dent Hospital for granting the use of the optical scanner. I am also grateful to the team behind the KIRI Engine application for their guidance during the LiDAR scanning process, particularly for suggesting the application of baby powder on the model surfaces to improve scan quality. We would like to thank Asst. Prof. İpek Necla Güldiken for providing access to the scanning device used in this study, and Asst. Prof. Emine Nur Kahraman for sharing the CT images of the model.

This study has been supported by TÜBİTAK (The Scientific and Technological Research Council of Turkey) under project number 1919B012325675 -TUBITAK 2209A.

REFERENCES

[1] Etemad-Shahidi Y, Qallandar OB, Evenden J, Alifui-Segbaya F, Ahmed KE. Accuracy of 3-Dimensionally Printed Full-Arch Dental Models: A Systematic Review. J Clin Med. 2020 Oct 20;9(10):3357. doi: 10.3390/jcm9103357. PMID: 33092047; PMCID: PMC7589154.

[2] 1: Erdoğuş, H. B. (2019). CERRAHİ UYGULAMALARDA MEDİKAL GÖRÜNTÜLEMENİN ÖNEMİ VE ÜÇ BOYUTLU ANATOMİK MODEL KULLANIMI. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, 24(3), 491-502. https://doi.org/10.17482/uumfd.415607

[3] Lin, HH., Chang, HW. & Lo, LJ. Development of customized positioning guides using computer-aided design and manufacturing technology for orthognathic surgery. Int J CARS **10**, 2021–2033 (2015). https://doi.org/10.1007/s11548-015-1223-0

[4] Narita M, Takaki T, Shibahara T, Iwamoto M, Yakushiji T, Kamio T. Utilization of desktop 3D printer-fabricated "Cost-Effective" 3D models in orthognathic surgery. Maxillofac Plast Reconstr Surg. 2020 Aug 1;42(1):24. doi: 10.1186/s40902-020-00269-0. PMID: 32821740; PMCID: PMC7395922.

[5] 3d printer nedir, nasıl çalışır, neler yapar https://maker.robotistan.com/3d-yazici-printer/

[6] what are the different types of 3d printing? https://www.futurelearn.com/info/courses/getting-started-withdigital-manufacturing/0/steps/184102

[7] Sarmah, M., Neelima, A. & Singh, H.R. Survey of methods and principles in three-dimensional reconstruction from two-dimensional medical images. Vis. Comput. Ind. Biomed. Art **6**, 15 (2023). https://doi.org/10.1186/s42492-023-00142-7

[8] Exploring the potential of 3D scanning in Industry 4.0: An overview Abid Haleem, Mohd Javaid, RaviPratap Singh, Shanay Rab, Rajiv Suman, Lal it Kumar , Ibrahim,,Haleem Khan https://www.sciencedirect.com/science/article/pii/S2666307422000171

[9] S. Rianmora, K. Nuamchit, N. Vanasrivilai, P. Tantipiched and A. Rammbuth, "Applying scanning techniques to create 3D model," IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Yokohama, Japan, 2015, pp. 005217-005222, doi: 10.1109/IECON.2015.7392920. keywords: {Three-dimensional displays;Surface treatment;Solid modeling;Surface reconstruction;Data acquisition;Cameras;Design automation;reverse engineering;data acquisition;automated selective acquisition system;3D laser scanner},

[10] ALPACA: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures https://besjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/2041-210X.13689

A Unified Pipeline for Consistent Multi-Label Mask Generation in Pediatric Cardiac Segmentation

Omer Faruk Sahin
Bioengineering
Marmara University
Istanbul, Turkey
umarfaruksahin@outlook.com

Mohamad Humam Alzaeim
Biomedical
Engineering & Bioinformatics
Istanbul Medipol University
Istanbul, Turkey
mohamad.Alzaeim@std.medipol.edu.tr

Beyza Nur Yazici
Computer Engineering
Istinye University
Istanbul, Turkey
beyzanur.yazici@stu.istinye.edu.tr

Sakip Onder
Biomedical Engineering
Yildiz Technical University
Istanbul Turkey
sonder@yildiz.edu.tr

Doruk Gunveren
Cemberlitas Anatolian High School
Istanbul, Turkey
gunverendoruk@gmail.com

Senol Piskin*
Mechanical
Engineering Department
Istinye University
Istanbul, Turkey
*Corresponding author:
senol.Piskin@istinye.edu.tr

Fatma Zeynep Gungoren
Department of Radiology
The Anadolu Medical Center
Istanbul, Turkey
fzgungoren@yahoo.com

Ruqaiyah Mirza
Software Engineering
Istinye University
Istanbul, Turkey
ruqaiyah.mirza@stu.istinye.edu.tr

Irem Yilmaz
Biomedical Engineering
Yildiz Technical University
Istanbul, Turkey
irem.yilmaz@std.yildiz.edu.tr

Ibrahim Faress
Software Engineering
Istinye University
Istanbul, Turkey
ibrahim.faress@stu.istinye.edu.tr

Kevser Banu Kose
Biomedical
Engineering Department
Istanbul Medipol University
Istanbul Turkey
kbkose@medipol.edu.tr

JS Femilda Josephin
Computer Engineering Department
Istinye University
Istanbul, Turkey
femilda.bai@istinye.edu.tr

Musab Ahmed Khan Umair
Artificial Intelligence
Istinye University
Istanbul, Turkey
musabkhan2003@gmail.com

Fatmanur Aydogdi
Zuhtu Kurtulmus Anatolian High School
Istanbul, Turkey
nurraydogdi21@gmail.com

***Abstract*— Artificial Intelligence based cardiac image segmentation requires consistent labels, harmonized volumes, and computationally efficient dataset. The present study proposes a reproducible pipeline that (i) merges segmented cardiac structures into a single multi-label mask (0: background, 1: aorta, 2: chambers, 3: pulmonary structures), (ii) resamples masks to perfectly match scan dimensions using 3D Slicer's "Resample Image (BRAINS)" module (nearest-neighbor interpolation), and (iii) performs automatic region-of-interest (ROI) cropping by adding 10-pixel margin around mask-derived bounds. On our internal dataset, the workflow standardizes inputs, reduces data volume substantially, and preserves full cardiac anatomy. The resulting data enabled stable multi-class and task-specific training and contribution to a better Dice score in downstream models. The ROI script operate on a simple folder structure and can be executed batch-wise, supporting anonymity, interoperability, and rapid experimentation.**

Keywords*— *multi-label segmentation; deep learning; convolutional neural networks (CNNs); image resampling; ROI cropping; data preprocessing; medical image segmentation; pediatric cardiac image; segmentation AI

I. INTRODUCTION

For reliable supervised segmentation in pediatric and complex cardiac Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), the process heavily relies on high-quality and spatially aligned labels. However, model accuracy is often degraded while training complexity often increases because of variations in label encoding, inconsistent mask dimensions, and the presence of irrelevant background voxels. Hence, significant mismatches between image volumes and segmentation masks (e.g., 129-slice scans paired with 646-slice masks), heterogeneous labeling conventions, and extraneous anatomical structures beyond the cardiac region. Now, to address these challenges, a reproducible three-stage preprocessing helps by: (1) generating standardized multi-label masks; (2) harmonizing spatial dimensions via isotropic resampling; and (3) automatically cropping the region of interest (ROI) using margin-expanded mask bounds. This paper details the design, implementation, and benefits of this preprocessing workflow, especially highlighting its impact on data uniformity and segmentation performance across cardiac dataset.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

II. Related Work

Classical pipelines often rely on manual re-labeling and ad hoc cropping, which introduce variability and risk reproducibility [1]. Even though automated multi-label strategies exist, many do not enforce strict spatial alignment or incorporate margin-controlled cropping critical for consistent downstream analysis for 3D cardiac data [2]. In this work, established tools such as 3D Slicer and BRAINS Resample [3] were utilized to present a unified, scriptable pipeline that integrates label standardization, precise volumetric alignment, and reproducible cropping into a single automated workflow.

III. Datasets

Original masks and corresponding scans were generated through in-house segmentations.

IV. Methods

A. Multi-Label Mask Generation

Initially, existing segmentations were imported into 3D Slicer in a defined order to ensure consistent label assignment and exported as binary labelmaps [3]. A unified mask with fixed integer codes was adopted: 0 (background), 1 (aorta), 2 (chambers), 3 (pulmonaries) as shown in Fig. 1.

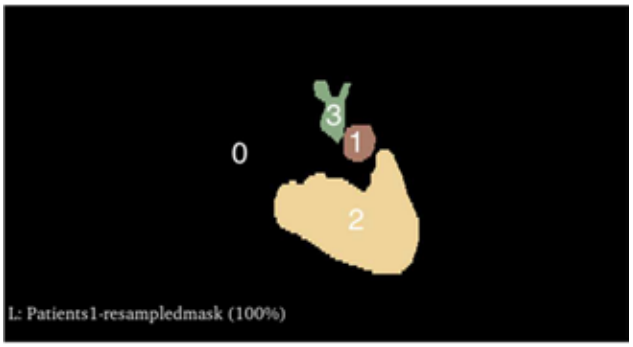


Fig. 1. Label Assignment; 0 = background, 1 = aorta, 2 = cardiac chambers (atria + ventricles), and 3 = pulmonary arteries.

B. Size Harmonization via Resampling

Mismatches in slice count, spacing, and origin between scans and masks are illustrated in Fig. 2. The scan (top) is 512 × 512 × 129 voxels at 0.37 × 0.37 × 1 mm spacing with origin (58, −100, −280 mm) and a flipped IJK→RAS direction, whereas the label volume (bottom) is 465 × 462 × 646 voxels at 0.127 mm isotropic spacing, origin (−78, −226, −261 mm), and a different orientation. These mismatches in geometry must be harmonized to avoid misalignment and were corrected using the "Resample Image (BRAINS)" module within 3D Slicer [3]. Nearest-neighbor interpolation and uchar pixel type were applied to retain integer labels [4]. Masks were aligned to the reference scan geometry. An example as shown in Fig. 3, both scan (top) and label (bottom) now share identical image dimensions, isotropic spacing, origin, and orientation, confirming successful resampling and alignment. This step is essential for correct model training, as misaligned tensors can silently corrupt losses and gradients.

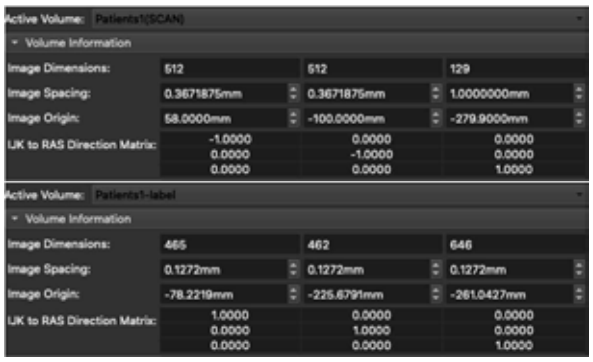


Fig. 2. Mismatch in volume geometry metadata between the scan and its label before preprocessing

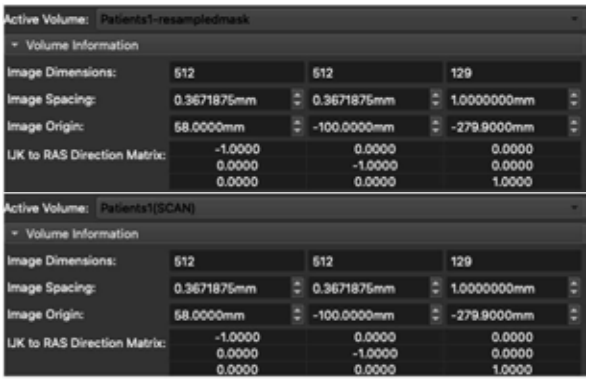


Fig. 3. Geometry metadata after preprocessing.

C. ROI Cropping with Margin Preservation

To reduce computational load and focus learning on cardiac anatomy in segmentation [5], min–max coordinates were extracted from masks to define a tight 3D bounding box, then added margin by 10 pixels on all sides to avoid cutting critical structures [6,7]. Both scan-mask pairs were cropped identically. A Pyhton Jupyter Notebook script executes batch cropping based on a simple folder structure (input/, output/, scripts/), enabling reproducible processing for each patient. Fig. 4 illustrates an example scan output, Top: original (non-cropped) scan with extensive background. Bottom: cropped volume centered on the cardiac anatomy with a fixed safety margin, preserving spacing/orientation while removing extraneous voxels to reduce memory use and accelerate training. Similarly, Fig. 5 demonstrates the corresponding ground truth mask, where the cropped volume is focused on the main structures (aorta, chambers, pulmonary arteries), allowing faster learning and keeping input dimensions input consistent among data.

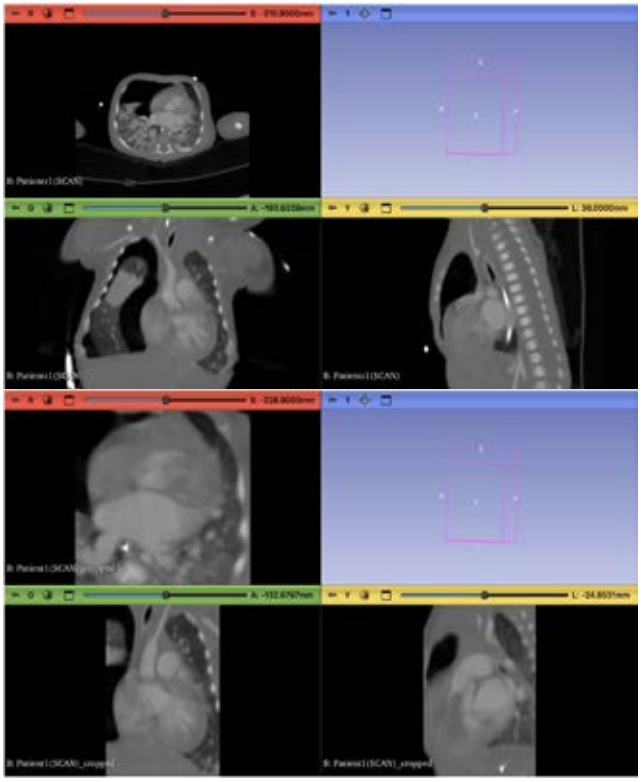


Fig. 4. Field-of-view comparison for compute-aware scan cropping.

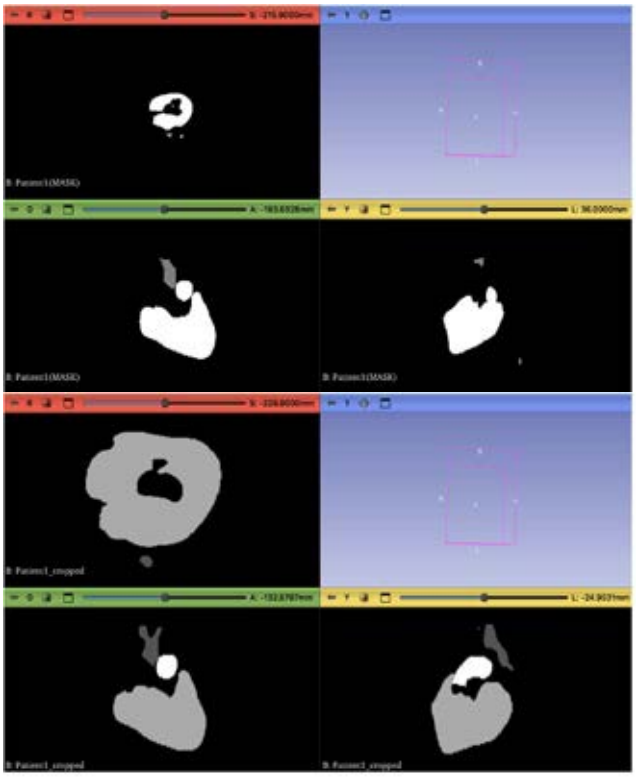


Fig. 5. Effect of cropping on label volumes

V. Implementation Details

Using pseudonymous IDs (e.g., Patient1, Patient2), patient files were organized. Each folder contained a single raw scan and the multi-label mask paired it. The proposed pipeline loops through each folder, preserves the original scan and outputs harmonized and cropped volumes that are prepared for analysis in order to guarantee uniform preparation. The widely utilized scientific packages NumPy to facilitate efficient array manipulation was employed in the implementation, which had been carried out in Python 3.8+.

VI. Experiments & Result

A. *Qualitative Verification*

Middle-slice overlays confirmed accurate alignment of post-resampling and cropping. No spatial drift was noted between labels and intensity volumes.

B. *Quantitative Effects*

Data volume was significantly reduced while retaining all cardiac anatomy, leading to faster training and inference. Although exact voxel count reductions vary by case, the approach consistently trimmed irrelevant body regions.

VII. Discussion

The proposed pipeline closes three common gaps in medical image segmentation preparation: unified label encoding, geometric consistency, and compute-aware cropping. By standardizing these steps, minimized human error and improved repeatability. While effective, the method assumes initial segmentations are anatomically valid; future work could integrate QA metrics (e.g., histogram-based sanity checks) directly into the pipeline. Additionally, explicit logging of voxel-count reductions and time savings would strengthen the cost–benefit analysis.

VIII. Conclusion

The overall methodology integrates multi-label mask generation, precise nearest neighbor resampling in 3D Slicer ("Resample Image (BRAINS)", unsigned char (uchar) labels), and ROI cropping with a fixed 10 pixel safety margin into a single, aligned, and reproducible preprocessing pipeline for cardiac MRI segmentation. Across internal scan–mask pairs, the workflow ensured exact label–image alignment, curtailed non-cardiac volume to reduce data size and time, and produced standardized inputs that supported stable, strong performance in both multi-class and task-specific models. These results suggest straightforward transferability to other multi-structure organs; and support adoption as a reproducible preprocessing standard. The method assumes anatomically valid initial segmentations; next steps include integrating fully automated quality assurance checks, systematically logging and profiling voxel count reductions and time/memory usage, and evaluating adaptive margin sizing.

IX. ACKNOWLEDGMENT

This work was funded by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under the 1001 Grant Program (Project No. 122E488).

X. AFFLIATIONS

This study was conducted in collaboration with Istinye University, Istanbul Medipol University, and Anadolu Medical Center.

REFERENCES

[1] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, "Variability and reproducibility in deep learning for medical image segmentation," Scientific Reports, vol. 10, no. 1, p. 13724, Aug. 2020, doi: https://doi.org/10.1038/s41598-020-69920-0.

[2] F. Maleki et al., "RIDGE: Reproducibility, Integrity, Dependability, Generalizability, and Efficiency Assessment of Medical Image Segmentation Models," Deleted Journal, Nov. 2024, doi: https://doi.org/10.1007/s10278-024-01282-9.

[3] A. Fedorov et al., "3D slicer as an image computing platform for the Quantitative Imaging Network," Magn. Reson. Imaging, vol. 30, no. 9, pp. 1323–1341, Nov. 2012. doi: 10.1016/j.mri.2012.05.001.

[4] W. Kim, Y. Kang, S. Lee, H.-J. Lee, and Y. Nam, "Label-Preserving Data Augmentation for Robust Segmentation of Thin Structure in MRI," Investigative Magnetic Resonance Imaging, vol. 28, no. 3, p. 107, 2024, doi: https://doi.org/10.13104/imri.2024.0014.

[5] R.-R. Galea, L. Diosan, A. Andreica, L. Popa, S. Manole, and Z. Bálint, "Region-of-Interest-Based Cardiac Image Segmentation with Deep Learning," Applied Sciences, vol. 11, no. 4, p. 1965, Feb. 2021, doi: https://doi.org/10.3390/app11041965.

[6] I. S. Isler, D. Mohaisen, C. Lisle, D. Turgut, and U. Bagci, "Uncertainty-Guided Coarse-to-Fine Tumor Segmentation with Anatomy-Aware Post-Processing," arXiv.org, 2025. https://arxiv.org/abs/2504.12215

[7] T. Zhou, L. Li, G. Bredell, J. Li, and Ender Konukoglu, "Quality-Aware Memory Network for Interactive Volumetric Image Segmentation," arXiv (Cornell University), Jan. 2021, doi: https://doi.org/10.48550/arxiv.2106.10686.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Volume Control in a Hybrid Mock Circuit of the Cardiovascular System using Command Filtered Backstepping Approach

Ahmed Alhajyounis*, Yaren Kuzucu†, Fatih Adıgüzel*, Kamuran Kadıpaşaoğlu†
*Control and Automation Engineering Department.
Yildiz Technical University, İstanbul, Türkiye
ahmed.alhajyounis@std.yildiz.edu.tr, fadiguzel@yildiz.edu.tr
†Biomedical Engineering Department.
Yildiz Technical University, İstanbul, Türkiye
yaren.kuzucu@std.yildiz.edu.tr, kamuran@yildiz.edu.tr

***Abstract*—Cardiovascular diseases often necessitate implantable therapies such as mechanical circulatory support devices, which are evaluated using mock circulatory systems. This study presents two nonlinear control strategies—classical backstepping and command-filtered backstepping for regulating the hydraulic dynamics of a hybrid pneumo-hydraulic mock circuit. After introducing the design steps of the two controllers, the closed-loop stability analyses are presented in the context of Lyapunov theory. To show the effectiveness of the two controllers, some simulations are performed by also evaluating in terms of tracking error and robustness to measurement noise, and command-filtered backstepping provides the best balance of performance and efficiency for precise, safe device testing.**

***Keywords*—Nonlinear Control, Backstepping, Mock Circulatory Loop.**

I. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide [1]. While certain conditions can be managed through lifestyle modification, pharmacotherapy, or surgical intervention, advanced pathologies—such as malignant arrhythmias, valvular and arterial lesions, end-stage heart failure, and peripheral arterial disease—frequently necessitate implantable solutions, including vascular prostheses, mechanical circulatory support devices, grafts, and stents [2]. The development of these biomedical devices typically follows a structured progression: virtual design and simulation (in silico), experimental evaluation in physical test platforms (in vitro), preclinical validation in live animals (in vivo), and, ultimately, clinical trials in humans (in clinico). If deficiencies are identified at any stage, the design undergoes revision, and the testing cycle is repeated [3]. Consequently, early identification of potential limitations is essential to streamline development, reduce costs, and minimize resource expenditure prior to market approval.

To enhance physiological fidelity during physical testing, mock circuits are employed to replicate cardiovascular behavior under both normal and pathological conditions. Cardiovascular system hybrid mock circuits (MCs) combine numerical and physical analogues in a hardware-in-the-loop configuration, enabling the controlled emulation of hemodynamic variables such as flow, pressure, and volume, as well as parameters including resistance and compliance [4]. In this framework, the numerical model generates reference pressure signals (e.g., left ventricular and aortic pressures), which are transmitted to the physical circuit via an electro-hydraulic interface, where they are tracked and regulated. Flow control ensures reservoir volume stability by adjusting pump flow rates, thereby preventing overflow or depletion and maintaining consistent experimental conditions [5].

Various MC architectures have been developed employing hydraulic, mechanical, or pneumatic actuation. Mechanical and hydraulic actuators provide simplicity in modeling and control but are hindered by inertia-related limitations, which result in sluggish responses and reduced accuracy [6] [7]. Pneumatic actuators, in contrast, exhibit faster response times and improved reliability. However, their compressibility introduces nonlinear dynamics that complicate both system modeling and controller design [8].

In order to solve these, a pneumo-hydraulic hybrid MC in which the hydraulic and pneumatic subsystems were controlled using integrator backstepping and Lyapunov-based error linearization algorithms are proposed, respectively, in [9]. This design successfully reproduced a range of physiological and pathological states, including normal, failing, and ventricular assist-device-supported conditions, with high tracking fidelity. On the other hand, the classical backstepping control is insufficient in terms of attenuating measurement noise, so the command-filtered backstepping controller delivers superior accuracy with reduced computational demand in noisy environments.

This study extends the paper in [9] by applying two nonlinear control strategies—classical backstepping and command-filtered backstepping—to regulate the hydraulic compartment of the hybrid MC. The objective of this paper is to compare their tracking accuracy, robustness, and computational efficiency under varying operating conditions.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

II. CARDIOVASCULAR SYSTEM MOCK CIRCUIT

The mock circuit (MC) employed in this study integrates a seven-node cardiovascular numerical model that was coupled to a physical pneumo-hydraulic rig via a hardware-in-the-loop interface. Pressure–volume relationships in the left ventricle (LV) and aorta (AO) defined the dynamics, and the model supplied reference LV/AO pressures to regulate the hardware. The setup comprised two pneumo-hydraulic reservoirs (LV and AO) reproducing LVAD preload/afterload, a continuous-flow LVAD pump, and a volume-displacement gear pump between reservoirs for circulation and level balancing. Solenoid valves modulated air inflow/outflow to track the references, while pressure transducers and electromagnetic flow probes measured reservoir pressures and pump flows in real time. Sensor data were acquired and processed in MATLAB/Simulink; further details on dynamics, signal processing, and control appear in [9].

This study focuses on the hydraulic subsystem dynamics and introduces the conceptual design of two backstepping method based controllers implemented to regulate the fluid level difference between the reservoirs. All governing equations and control expressions are derived in MATLAB and implemented in Simulink to simulate system behavior and evaluate controller performance.

III. SYSTEM DYNAMICS

Combining the volume difference dynamics given in [9] and the gear pump (the actuator) dynamics given in [5] the system dynamics can be expressed as

$$\Delta\dot{V}_h = 2Q_{vad} + 2R_{bf}\Delta P - 2\alpha_\omega\omega \tag{1a}$$

$$\dot{\omega} = \frac{1}{J}\left(-\tau_1 \text{sgn}(\omega) - \alpha_P\Delta P - \tau_2\omega + K_T I\right) \tag{1b}$$

$$\dot{I} = \frac{1}{L}\left(-K_{emf}\omega - R_{gp}I + V_{in}\right). \tag{1c}$$

where ΔV_h is the fluid level difference, Q_{lvad} is the flow rate generated by the lvad, α_ω is the gear pump stroke volume, R_{bf} is back flow hydraulic resistance, J is the motor inertia, ω is the motor rotation speed, τ_1 is the dry friction constant, τ_2 is the viscous friction constant, K_T is the torque constant, α_P is the load pressure constant, L is the motor inductance, I is the motor current, K_{emf} is the back-electromotive force voltage constant, R_{gp} is the resistance of the motor winding, and V_{in} is the motor input voltage.

For notational simplicity, the system is parameterized as follows: $x_1 = \Delta V$, $x_2 = \omega$, $x_3 = I$, $u = V_{in}, f_1 = 2Q_{lvad} - 2R_{bf}\Delta P$, $\theta_1 = \frac{1}{J}$, and $\theta_2 = \frac{1}{L}, f_2 = -\tau_1\,\text{sgn}(\omega) - \alpha_s\Delta P$, $g_1 = -2\alpha_s\omega, g_2 = -\tau_2$, $g_3 = K_T$, $g_4 = -K_{\text{emf}}$, and $g_5 = -R_{\text{gp}}$.

These assignments yield the simplified dynamical model:

$$\dot{x}_1 = f_1 + g_1 x_2 \tag{2a}$$

$$\dot{x}_2 = \theta_1 f_2 + \theta_1 g_2 x_2 + \theta_1 g_3 x_3 \tag{2b}$$

$$\dot{x}_3 = \theta_2 g_4 x_2 + \theta_2 g_5 x_3 + \theta_2 u. \tag{2c}$$

To design the control system, the tracking error (z_i) definitions are introduced as $z_i = \phi_i - x_i$ where ϕ_1 is the user-defined reference value for x_1, usually equal to 0, and ϕ_i (for $i = 2, 3$) is the virtual control signal for the corresponding states. Substituting Equations (2) into the time derivative of tracking error yields the tracking error dynamics:

$$\dot{z}_1 = \dot{\phi}_1 - f_1 - g_1(\phi_2 - z_2) \tag{3a}$$

$$\dot{z}_2 = \dot{\phi}_2 - \theta_1 f_2 - \theta_1 g_2(\phi_2 - z_2) - \theta_1 g_3(\phi_3 - z_3) \tag{3b}$$

$$\dot{z}_3 = \dot{\phi}_3 - \theta_2 g_4(\phi_2 - z_2) - \theta_2 g_5(\phi_3 - z_3) - \theta_2 u. \tag{3c}$$

A. Classical Backstepping

The first control method employs classical backstepping, in which a Lyapunov function is selected to establish stability and ensure convergence of the tracking errors. We begin by choosing a positive definite candidate Lyapunov function as

$$V = \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2 + \frac{1}{2}z_3^2. \tag{4}$$

Combining Equations (3) and the time derivative of Equation (4) yields

$$\begin{aligned}\dot{V} = {} & z_1\left[\dot{\phi}_1 - f_1 - g_1(\phi_2 - z_2)\right] \\ & + z_2\left[\dot{\phi}_2 - \theta_1 f_2 - \theta_1 g_2(\phi_2 - z_2) - \theta_1 g_3(\phi_3 - z_3)\right] \\ & + z_3\left[\dot{\phi}_3 - \theta_2 g_4(\phi_2 - z_2) - \theta_2 g_5(\phi_3 - z_3) - \theta_2 u\right].\end{aligned} \tag{5}$$

The controller signals are designed as follows:

$$\phi_2 = \frac{1}{g_1}(\dot{\phi}_1 - f_1 + k_1 z_1) \tag{6a}$$

$$\phi_3 = \frac{1}{\theta_1 g_3}(\dot{\phi}_2 + g_1 z_1 - \theta_1 f_2 - \theta_1 g_2\phi_2 + \theta_1 g_2 z_2 + k_2 z_2) \tag{6b}$$

$$\begin{aligned}u = {} & \frac{1}{\theta_2}(\dot{\phi}_3 + \theta_1 g_3 z_2 - \theta_2 g_4\phi_2 + \theta_2 g_4 z_2 - \theta_2 g_5\phi_3 \\ & + \theta_2 g_5 z_3 + k_3 z_3)\end{aligned} \tag{6c}$$

where $k_i > 0$ for $i = 1, 2, 3$ are the control gains and $\dot{\phi}_2$ and $\dot{\phi}_3$ are defined as the derivative of Equations (6a) and (6b) respectively. Substituting Equations (6) into Equation (5) yields $\dot{V} = -k_1 z_1^2 - k_2 z_2^2 - k_3 z_3^2$. In this condition since $\dot{V}(0) = 0$ and $\dot{V}(z) < 0\ \forall\ z \neq 0$, $\dot{V}(z)$ is a negative defininte function. Therefore, it can be seen that the closed-loop system is globally asymptotically stable in the sense of Lyapunov.

B. Command Filtered Backstepping

This section incorporates a command filtering-based backstepping controller to improve tracking performance in the presence of measurement noise. The filtered tracking error compensates the influence of noise on the system, and a control law is derived to guarantee stability [10]. Because the initial steps of command-filtered backstepping mirror those of classical backstepping, we proceed from Equations (6). Define the filtered tracking error and the compensated tracking error, respectively, as

$$\tilde{z}_i = \phi_{ic} - x_1 \tag{7}$$

$$v_i = \tilde{z}_i - \xi_i \tag{8}$$

where ϕ_{ic} is the filtered command signal and ξ_i is the compensated virtual control signal for the corresponding state for $i = 1, 2, 3$. The pseudocontrol signals can be defined as

$$\bar{\phi}_2 = \frac{1}{g_1}(\dot{\phi}_{1c} - f_1 + k_1\tilde{z}_1) \tag{9a}$$

$$\bar{\phi}_3 = \frac{1}{\theta_1 g_3}(\dot{\phi}_{2c} + g_1 v_1 - \theta_1 f_2 - \theta_1 g_2\bar{\phi}_2 + \theta_1 g_2 v_2 + k_2\tilde{z}_2) \tag{9b}$$

$$\bar{u} = \frac{1}{\theta_2}(\dot{\phi}_{3c} + \theta_1 g_3 v_2 - \theta_2 g_4\bar{\phi}_2 + \theta_2 g_4 v_2 - \theta_2 g_5\bar{\phi}_3 + \theta_2 g_5 v_3 + k_3\tilde{z}_3). \tag{9c}$$

Solving Equations (9) for the command derivatives yields

$$\dot{\phi}_{1c} = g_1\bar{\phi}_2 + f_1 - k_1\tilde{z}_1 \tag{10a}$$

$$\dot{\phi}_{2c} = \theta_1 g_3\bar{\phi}_3 - g_1 v_1 + \theta_1 f_2 + \theta_2 g_4 v_2 + \theta_1 g_2\bar{\phi}_2 - \theta_1 g_2 v_2 - k_2\tilde{z}_2 \tag{10b}$$

$$\dot{\phi}_{3c} = \theta_2\bar{u} - \theta_2 g_3 v_2 + \theta_2 g_4\bar{\phi}_2 - \theta_2 g_4 v_2 + \theta_2 g_5\bar{\phi}_3 - \theta_2 g_5 v_3 - k_3\tilde{z}_3. \tag{10c}$$

The command filter is implemented as a low-pass filter in state-space form to attenuate noise:

$$\dot{X} = \begin{bmatrix} 0 & \omega_n \\ -\omega_n & -2\zeta\omega_n \end{bmatrix} X + \begin{bmatrix} \omega_n \\ 0 \end{bmatrix} U \tag{11a}$$

$$Y = \begin{bmatrix} 1 & 0 \\ 0 & \omega_n \end{bmatrix} X \tag{11b}$$

where ω_n is the filter natural frequency, ζ is the filter damping ratio, X is the filter state vector, $U = \bar{\phi}_i$ and $Y = \begin{bmatrix} \phi_{ic} & \dot{\phi}_{ic} \end{bmatrix}^T$. Thus, the filter receives the pseudocontrol signals as input and outputs the filtered command and its derivative. Substituting Equations (10) and (8) into the time derivative of Equation (7) gives the filtered-error dynamics:

$$\dot{\tilde{z}}_1 = -k_1\tilde{z}_1 + g_1(\bar{\phi}_2 - \phi_{2c} + \xi_2 + v_2) \tag{12a}$$

$$\dot{\tilde{z}}_2 = -k_2\tilde{z}_2 + \theta_1 g_2(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_1 g_3(\bar{\phi}_3 - \phi_{3c} + \xi_3 + v_3) - g_1 v_1 \tag{12b}$$

$$\dot{\tilde{z}}_3 = -k_3\tilde{z}_3 + \theta_2 g_4(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_2 g_5(\bar{\phi}_3 - \phi_{3c} + \xi_3) - g_3 v_2 + \theta_2(\bar{u} - u). \tag{12c}$$

Substituting Equations (12) in the time derivative of Equation (8) yields the compensated tracking error dynamics as

$$\dot{v}_1 = -k_1\tilde{z}_1 + g_1(\bar{\phi}_2 - \phi_{2c} + \xi_2 + v_2) - \dot{\xi}_1 \tag{13a}$$

$$\dot{v}_2 = -k_2\tilde{z}_2 + \theta_1 g_2(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_1 g_3(\bar{\phi}_3 - \phi_{3c} + \xi_3 + v_3) - g_1 v_1 - \dot{\xi}_2 \tag{13b}$$

$$\dot{v}_3 = -k_3\tilde{z}_3 + \theta_2 g_4(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_2 g_5(\bar{\phi}_3 - \phi_{3c} + \xi_3) - g_3 v_2 + \theta_2(\bar{u} - u) - \dot{\xi}_3. \tag{13c}$$

Define a candidate Lyapunov function for the compensated tracking errors as

$$V_f = \frac{1}{2}v_1^2 + \frac{1}{2}v_2^2 + \frac{1}{2}v_3^2. \tag{14}$$

In this condition, the candidate Lyapunov function is radially unbounded and positive definite. Combining Equations (13) and the time derivative of Equation (14) yields

$$\begin{aligned}\dot{V}_f =& v_1\Big[-k_1\tilde{z}_1 + g_1(\bar{\phi}_2 - \phi_{2c} + \xi_2 + v_2) - \dot{\xi}_1\Big] \\ &+ v_2\Big[-k_2\tilde{z}_2 + \theta_1 g_2(\bar{\phi}_2 - \phi_{2c} + \xi_2) \\ &+ \theta_1 g_3(\bar{\phi}_3 - \phi_{3c} + \xi_3 + v_3) - g_1 v_1 - \dot{\xi}_2\Big] \\ &+ v_3\Big[-k_3\tilde{z}_3 + \theta_2 g_4(\bar{\phi}_2 - \phi_{2c} + \xi_2) - g_3 v_2 \\ &+ \theta_2 g_5(\bar{\phi}_3 - \phi_{3c} + \xi_3) + \theta_2(\bar{u} - u) - \dot{\xi}_3\Big].\end{aligned} \tag{15}$$

The filtered controller signals are devised as follows:

$$\dot{\xi}_1 = -k_1\xi_1 + g_1(\bar{\phi}_2 - \phi_{2c} + \xi_2) \tag{16a}$$

$$\dot{\xi}_2 = -k_2\xi_2 + \theta_1 g_2(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_1 g_3(\bar{\phi}_3 - \phi_{3c}) \tag{16b}$$

$$\dot{\xi}_3 = -k_3\xi_3 + g_3 v_2 + \theta_2 g_4(\bar{\phi}_2 - \phi_{2c} + \xi_2) + \theta_2 g_5(\bar{\phi}_3 - \phi_{3c} + \xi_3) - g_3 v_2 + \theta_2(\bar{u} - u) \tag{16c}$$

where $k_i > 0$ for $i = 1, 2, 3$ are the control gains. Substituting Equations (16) into Equation (15) yields $\dot{V}_f = -k_1 v_1^2 - k_2 v_2^2 - k_3 v_3^2$. In this condition, since $\dot{V}_f(z)$ is a negative definite function. Therefore, the compensated system is also globally asymptotically stable in the sense of Lyapunov. Finally, substituting Equations (7) and (8) into Equations (9) yields the pseudocontrol derivatives in their implementable form:

$$\dot{\phi}_{1c} = g_1\bar{\phi}_2 + f_1 - k_1(\phi_{1c} - x_1) \tag{17a}$$

$$\dot{\phi}_{2c} = \theta_1 g_3\bar{\phi}_3 - g_1(\phi_{1c} - \xi_1 - x_1) + \theta_1 f_2 + \theta_1 g_2\bar{\phi}_2 - \theta_1 g_2(\phi_{2c} - \xi_2 - x_2) - k_2(\phi_{2c} - x_2) \tag{17b}$$

$$\dot{\phi}_{3c} = \theta_2\bar{u} - \theta_2 g_3(\phi_{2c} - \xi_2 - x_2) + \theta_2 g_4\bar{\phi}_2 + \theta_2 g_5\phi_3 - \theta_2 g_4(\phi_{2c} - \xi_2 - x_2) - \theta_2 g_5(\phi_{3c} - \xi_3 - x_3) - k_3(\phi_{3c} - x_3). \tag{17c}$$

Note that ϕ_{ic} is obtained by filtering the pseudocontrol signals and ξ_i is obtained by integrating Equations (16). Consequently, all variables in Equations (17) are computable from the system states and known parameters.

IV. SIMULATION WORKS

TThe system and controller dynamics were implemented separately in Simulink to test the proposed strategies. First, the nominal (noise-free) plant was simulated. For the Classical Backstepping controller, additive noise was injected into measured signals to emulate real-world disturbances and benchmark performance. The Command-Filtered Backstepping controller was then tested under identical noise conditions, enabling a direct robustness comparison.

All controllers produced input commands within 0–24 V. Figure 1 shows the input signals (top) and actuator responses (bottom). While steady-state behavior was similar, noise in Classical Backstepping caused sharp peaks and dips due to

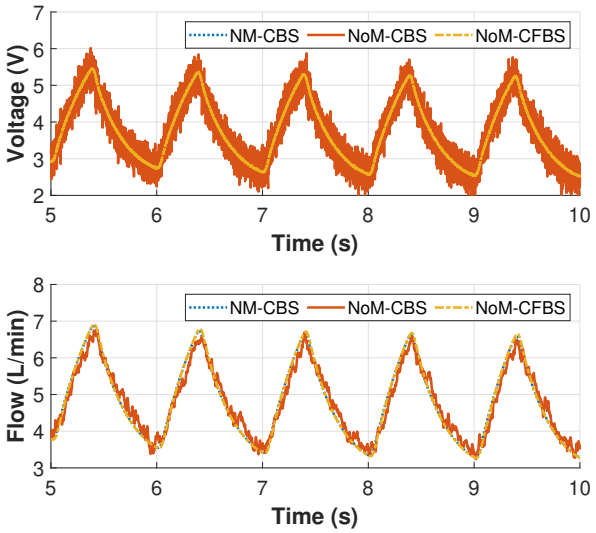


Fig. 1: Control signals over time and the corresponding actuator responses. Here, NM represents the nominal model, NoM the model with signal noise, CBS the classical backstepping, and CFBS the command-filtered backstepping.

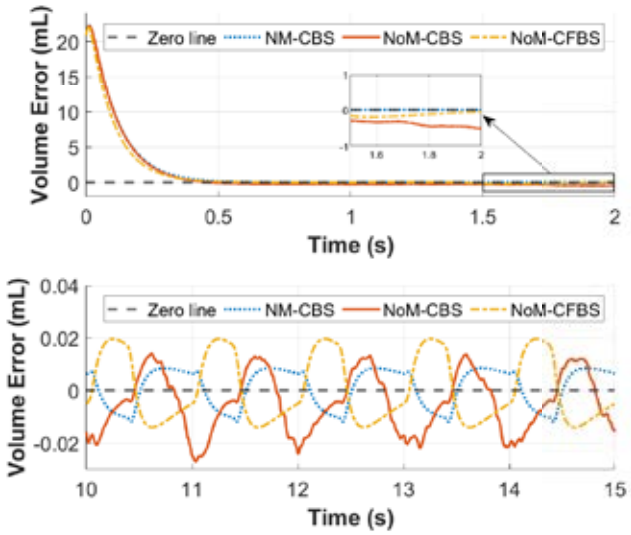


Fig. 2: Transient and Steady-state tracking error.

differentiation of noisy signals, leading to actuator chatter and potential instability. In contrast, Command-Filtered Backstepping smoothed inputs by filtering states before control computation, yielding more stable actuator responses. Figure 2 presents tracking-error signals for both transient (top) and steady-state (bottom) phases.

Across all cases, the inter-reservoir volume difference stayed within limits. However, under noise with Classical Backstepping, oscillations and steady-state error appeared. Command-Filtered Backstepping improved tracking accuracy, with only a slight phase shift. Specifically, Classical Backstepping achieved an average error of 1.1e-3 mL in the nominal case but degraded to –4.4e-3 mL with noise, while Command-Filtered Backstepping reduced the error to 6.2e-4 mL under noise. These results confirm that the command-filtered approach suppresses noise-induced oscillations and steady-state error.

V. CONCLUSIONS

Classical and command-filtered backstepping controllers were evaluated under identical conditions in the hydraulic subsystem of the hybrid pneumo-hydraulic mock circuit. For both control structures, the closed-loop stability is established. Each approach generated similar control inputs within the 0–24 V range and maintained tracking within acceptable limits. When measurement noises were introduced, the command-filtered backstepping controller achieved performance comparable to that of the classical backstepping controller operating in an ideal, noise-free setting. These findings demonstrate the effectiveness of the command-filtered backstepping architecture for systems subject to sensor noise. In future work, we plan to subject this approach to benchtop tests to assess its real-time performance and robustness.

ACKNOWLEDGMENT

The Cardiovascular System Mock Circuit used in this study was designed within the scope of the TUBITAK 1003 project (No. 118S098) and the Yildiz Technical University (YTU) Scientific Research Projects Coordination Unit (BAP) project (No. FBA-2024-6491) at the Physiological Control Laboratory of the Biomedical Engineering Department, YTU.

REFERENCES

[1] National Center for Health Statistics, "Multiple cause of death 2018–2021 on cdc wonder database," https://wonder.cdc.gov/mcd.html?utm=0pt, 2018.

[2] C. Tsao, A. Aday, Z. Almarzooq *et al.*, "Heart disease and stroke statistics—2023 update: A report from the american heart association," *Circulation*, vol. 147, pp. e93–e162, 2023.

[3] A. Alshammari, F. Alabdah, W. Wang, and G. Cooper, "Virtual design of 3d-printed bone tissue engineered scaffold shape using mechanobiological modeling: Relationship of scaffold pore architecture to bone tissue formation," *Polymers*, vol. 15, 2023.

[4] F. Cappon, T. Wu, T. Papaioannou, X. Du, P. L. Hsu, and A. W. Khir, "Mock circulatory loops used for testing cardiac assist devices: A review of computational and experimental models," *Int. J. Artif. Organs*, vol. 44, no. 11, pp. 793–806, 2021.

[5] B. Misgeld, D. Rüschen, S. Schwandtner, S. Heinke, M. Walter, and S. Leonhardt, "Robust decentralised control of a hydrodynamic human circulatory system simulator," *Biomed. Signal Process. Control*, vol. 20, pp. 35–44, 2015.

[6] J. Zhang, L. Li, X. Zhang, T. Zhang, and Y. Yuan, "Delay analysis and the control of electro-hydrostatic actuators," *Appl. Sci.*, vol. 12, no. 6, p. 3089, 2022.

[7] J. Li, Y. Lu, F. He, G. Zhou, and L. Miao, "Analysis and compensation control of engine valve response delay based on the electro-hydraulic variable valve actuator," *Machines*, vol. 10, no. 8, p. 701, 2022.

[8] D. Telyshev, A. Pugovkin, S. Selishchev, D. Ruschen, and S. Leonhardt, "Hybrid mock circulatory loop for training and study purposes," in *Proc. Ural Symp. Biomed. Eng., Radioelectron. and Inf. Tech. (USBEREIT)*. IEEE, 2018, pp. 29–32.

[9] A. Alhajyounis *et al.*, "Nonlinear control of a hybrid pneumo-hydraulic mock circuit of the cardiovascular system," *Int. J. Robust Nonlinear Control*, 2024.

[10] J. A. Farrell, M. Polycarpou, M. Sharma, and W. Dong, "Command filtered backstepping," in *2008 American Control Conference*, Seattle, WA, USA, 2008, pp. 1923–1928.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Design and Virtual Performance Testing of the Accessory Equipment for a Left Ventricular Assist Device

Sinemis Z. ÇELİK
Dept. of Electrical and Electronics Eng.,
Dogus University, Istanbul, Türkiye
scelik@dogus.edu.tr

Kamuran KADIPAŞAOĞLU
Dept. of Biomedical Engineering
Yildiz Technical University, Istanbul, Türkiye
kamuran@yildiz.edu.tr

***Abstract*—Cardiovascular diseases are the leading cause of global mortality, and heart failure remains a major consequence. While heart transplants are the gold standard for end-stage heart failure, donor shortages necessitate alternatives like left ventricular assist devices (LVADs). This study aimed to design and optimize the inlet and outlet cannulae and housing for an LVAD to minimize thrombosis risk and enhance device efficiency. Key design parameters included the diameters, lengths, and bend angles of the cannulae, ensuring smooth blood flow and reduced turbulence. Through iterative design modifications and computational fluid dynamics analysis, optimal configurations were identified. The results demonstrated that specific configurations significantly improved flow dynamics, reducing stagnant zones and wall shear stress. This study contributes to the development of a safer and more effective LVAD, addressing the critical need for mechanical circulatory support in patients with severe heart failure.**

***Keywords*—Left Ventricular Assist Device, computational fluid dynamics, cannulae design, thrombosis risk, design optimization, hydraulic performance.**

I. Introduction

Cardiovascular diseases, involving the heart and blood vessels, are the leading cause of global mortality, accounting for 32% of deaths in 2019 (17.9 million) [1]. Heart transplants are the best treatment for end-stage heart failure but are limited by donor shortages; only 2,000 transplants occur annually in the US [2]. In Turkey, in 2017, only 75 heart donors were available for 521 waiting patients, with 214 dying while awaiting transplantation [3]. Mechanical circulatory support (MCS), including ventricular assist devices (VADs) and total artificial hearts (TAHs), is a critical therapy for advanced heart failure. VADs, such as left ventricular assist devices (LVADs) and right ventricular assist devices (RVADs), initially served as a "bridge-to-transplant [4]." Now, they also provide long-term "destination therapy" and lifelong support for patients who are not suitable for transplants [5]. LVADs comprise a turbine, an inlet cannula (IC), an outlet cannula (OC), a driveline, and housing. The inlet cannula connects the left ventricle (LV) to the pump, while the outlet cannula links the pump to the aorta. Proper cannulae design is vital for steady blood flow. The driveline connects the pump to an external mechanism, and the housing contains the pump. Thrombosis is a primary complication of VADs, requiring patients to undergo anti-coagulation therapy [6]. Factors affecting VAD performance include cannulae geometry [7], inlet cannula insertion angle [8] and length [9], post-operative malposition [10], and pressure from the outflow graft on the right ventricle (RV) [11]. Optimizing LVAD's accessories involves iterative computational fluid dynamics (CFD) simulations to predict blood flow. These simulations solve fluid dynamics equations numerically, providing insights into velocity fields, pressure distributions, and shear stresses.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

II. Materials & Methods

A. Cannulae Design Parameters

The concept design for the turbine housing and the inlet and outlet cannulae was initiated. The main design parameters for the cannulae were inner diameter, bend angles, and tangent lengths (Fig. 1, a and c). For the inlet cannula, a step diameter reduction angle was considered to allow the larger diameter of the housing to meet the smaller diameter at the left ventricular entry site (LV apex) without disturbing the flow. The bend angle of the inflow cannula ranged from 55° to 65°, while the bend angle of the outlet cannula ranged from 100° to 110°. These variations influenced the formation of stagnant zones. For the housing design, the only parameter selected was the distance between the cannulae and the turbine (Fig. 1, b). The turbine length was fixed at 82.5 mm [12] and the housing length varied depending on this distance, ranging from 14 mm to 26 mm in addition to the turbine length.

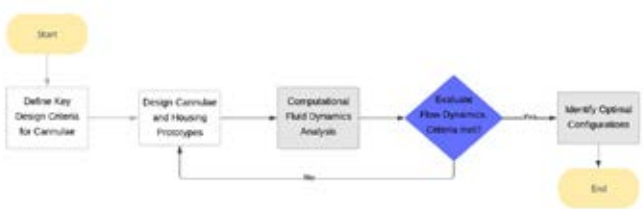
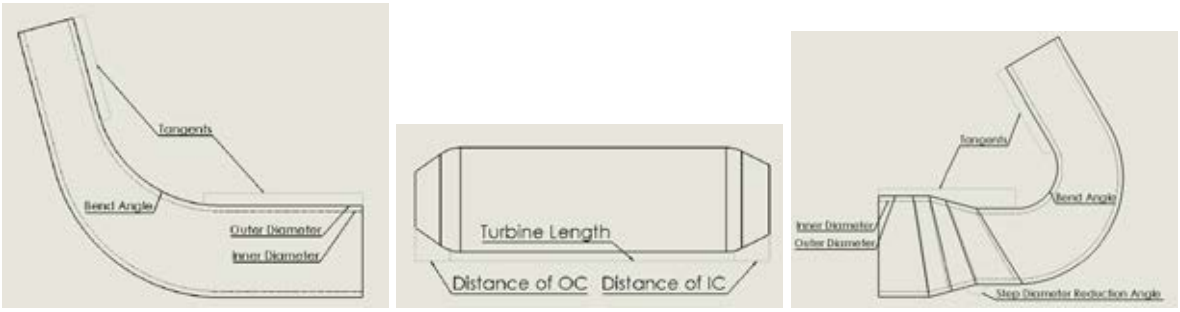


Fig. 1: Schematic diagram of the methodology workflow.



(a) (b) (c)

Fig. 2: Design Parameters of the Device: (a) Outlet cannula, (b) Housing, (c) Inlet cannula.

B. Cannulae Performance Analysis

Stagnant zones, where the velocity is below 5 mm/s, were identified as potential sites for thrombosis formation due to low shear rates [13]. Wall shear stress (WSS) was analyzed on the walls of both the inlet and outlet cannulae. WSS values below 0.3 Pa were associated with an increased risk of thrombosis formation, values between 0.3 Pa and 9 Pa were considered physiological and ideal for preventing both thrombosis and blood cell damage, and values above 9 Pa could trigger Von Willebrand factor activation, potentially causing blood cell damage [13].

C. Computational Fluid Dynamics Analysis

The CAD-designed models were used as the base geometry for CFD analysis. The meshing process started using two methods to improve mesh quality. The inflation mesh generation method, used to create a fine mesh near walls to observe boundary layer effects, was applied to three critical regions: the outer housing, the inducer blades, and the diffuser blades. In areas where inflation was not applied, the sizing mesh generation method was used, applied to all surfaces except the outer housing. The viscous k-epsilon model with enhanced wall treatment was used. Blood was defined as a fluid with a density of 1050 kg/m³ and viscosity of 0.0035 kg/ms. Boundary conditions included setting the inlet pressure as left ventricular pressure and the outlet pressure as aortic pressure. Rotor walls were defined as rotational moving walls to accurately simulate motion within the fluid. The SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) algorithm was used for pressure–velocity coupling, ensuring mass conservation and determining pressure fields.

D. Post-Processing Analysis

Stagnant zones (velocity < 5 mm/s) were identified in specific cross-sections of the inlet and outlet cannulae. For the inlet cannula, due to the diameter change through step diameter reduction, two areas were selected. IC Area-1 was the region just before the cannula meets the housing, where the diameter is larger. IC Area-2 was the region before the diameter change occurred, where the cannula's diameter is smaller than in IC Area-1. OC Area was the region located after the housing, at the beginning of the outlet cannula. Wall shear stress was analyzed along the walls of the cannulae. Hydraulic efficiency is calculated to evaluate the performance of the combined system of the turbine and cannulae.

$$\eta_h = \frac{Q \cdot \Delta P}{T \cdot \omega} \tag{1}$$

where Q is the flow rate $(\mathrm{L/min})$, ΔP is the pressure rise across the inlet and outlet (mmHg), T is the torque $(\mathrm{mN \cdot m})$, and ω is the rotational speed $(\mathrm{rad/s})$.

III. Results

Cannulae and housing geometric iterations were performed under the same physiological condition, just before aortic valve opening, which corresponds to LV pressure of 0 mmHg and aortic pressure to 84 mmHg, with an angular speed set to 550 rad/s. As the initial model, the housing had an inner diameter of 24.5 mm and a length of 82.5 mm, which was also the length of the turbine. The housing diameter was reduced to match the inlet and outlet cannulae diameters with a 10 mm fillet. The inlet cannula had an inner diameter of 20 mm, bend tangents of 25.51 mm (upper) and 20 mm (lower), a distance from the turbine of 7 mm, and a bend angle of 60°. A single-step diameter decrease was used between the housing and the

TABLE I: Housing Configurations and Performance Results

NO	Geometry			Performance				Stagnant Zone (mm^2)			WSS (Pa)	
	Dist. to IC (mm)	Dist. to OC (mm)	Length (mm)	Q (L/min)	P rise (mmHg)	Torque (mN.m)	Efficiency (%)	IC Area-1	IC Area-2	OC Area	IC Walls	OC Walls
v1	7	10	99.5	3.3242	84.6595	5.2330	21.7315	2.1578	0.9611	0.2183	3.6338	5.5535
v2	7	13	102.5	3.1527	84.2169	5.1920	20.6643	292.4740	143.4360	0.2107	15.4729	-8.4249
v5	7	7	96.5	3.1939	84.3069	5.2388	20.7695	2.2425	0.9660	0.2140	3.6179	5.4532
v12	10	10	102.5	3.4420	84.9895	5.3340	22.1616	0.6819	0.7482	0.2107	4.1214	6.0010

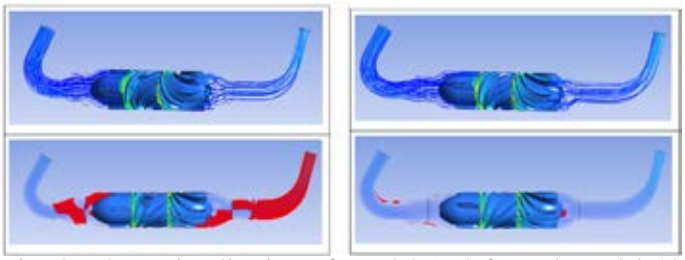


Fig. 3: Flow Visualization of Model 1 (left) and Model 12 (right): Streamlines (top) and Stagnant Zones (bottom). Red regions in the stagnant zone figures indicate flow velocities below 5 mm/s.

bend, with a reduction angle of 155°. The outlet cannula had an inner diameter of 14 mm, bend tangents of 25 mm (lower) and 24.77 mm (upper), a distance from the turbine of 10 mm, and a bend angle of 105°.

A. Housing

The length of the housing was adjusted based on the distance between the cannulae and the turbine. Outlet cannula distances of 7 mm, 10 mm, and 13 mm were tested. The inlet cannula with a 13 mm distance was not tested because preliminary analysis of the outlet cannula showed a significant increase in stagnant flow area. Inlet cannula distances of 7 mm and 10 mm were iterated. Based on these cannula distances, the housing length was determined to be 102.5 mm.

B. Outlet Cannula

The inner diameter was initially set at 14 mm. Any change in this diameter by ±1 mm resulted in a decrease in efficiency. Therefore, an inner diameter of 14 mm was selected. The tangent before the bend (lower) was initially set at 20 mm but was later reduced to 15 mm without a significant improvement in efficiency. The bend angle was tested at 100 and 110°, with 105° determined to be optimal.

C. Inlet Cannula

The initial step diameter reduction angle at the housing entrance was modified to a smooth transition, which reduced efficiency and flow rate. A double-step reduction also produced low efficiency. For single-step reduction, the initial angle of 155° varied to 170 and 165°. The most uniform flow was observed at 165°. The initial inner diameter was 20 mm and was reduced in increments to 18 mm and 16 mm. An inner diameter of 18 mm was selected for subsequent tests. The tangent lengths before (upper) and after (lower) the bend were adjusted and the bend angles of 50° and 60° were tested.

IV. Discussion & Conclusion

The housing geometry was found to improve both hydraulic performance and flow. This optimization revealed the importance of the distance between the turbine and cannulae in preventing flow separation. As can be seen in Table 1, the optimal design was observed when the distance between the turbine and the cannulae was 10 mm (Model 12). This configuration achieved the highest flow rate (3.4420 L/min) among all the other models, while also producing minimal stagnant zones in IC Area-1 and IC Area-2 (0.6819 mm² and 0.7482 mm², respectively). When the OC-turbine distance was increased to 13 mm (Model 2), it did not cause significant decreases in flow rate and efficiency, but the stagnant zone in IC Area-1 increased by 400 times compared to the optimal configuration. This result indicates that increasing the distance between the turbine and the cannulae may lead to the formation of low-velocity recirculation zones that can cause thrombosis. The WSS analysis also supported these results. In Model 2, high shear stress on the IC walls (15.4729 Pa) and negative stress on the OC walls (-8.4249 Pa) indicate reverse flows and flow separation. In contrast, the WSS values in Model 12 (4.1214 Pa on the IC walls and 6.0010 Pa on the OC walls), which is the optimal housing configuration, indicate more stable flow patterns.

It was observed that the performance of the outlet cannula varies depending on the inner diameter and bend angle. The optimal configuration, Model 17, featuring a 14 mm diameter, 25 mm lower tangent, 26.48 mm upper tangent, and a 100° bend angle, achieved the highest efficiency (21.4751%), low stagnant zones (0.2261 mm²), and moderate WSS (5.7229 Pa). Increasing the diameter to 15 mm (Model 3) slightly decreased efficiency (20.6625%) but increased stagnant zones (0.3733 mm). This increased diameter accelerated flow out of the housing, suggesting that this acceleration could cause flow separation near the wall. Reducing the diameter to 13 mm (Model 4) reduced stagnant zones (0.1599 mm²), a decrease of about 57% compared to the Model 3. However, it did not increase flow rate, suggesting a potential trade-off in adjusting the diameter. Bend angle was another influencing parameter. Models closer to 105° (Models 14 and 17) provided higher

TABLE II: Outlet Cannula Configurations and Performance Results

NO	Geometry				Performance				Stagnant Zone (mm^2)	WSS (Pa)
	Inner Diameter (mm)	Upper Tangent (mm)	Lower Tangent (mm)	Bend Angle (deg)	Q (L/min)	P rise (mmHg)	Torque (mN.m)	Efficiency (%)	OC Area	OC Walls
v3	15	25	24.77	105	3.1818	84.2844	5.2446	20.6625	0.3733	6.9969
v4	13	25	24.77	105	3.2106	84.3519	5.1778	21.1356	0.1599	5.8998
v14	14	15	27.36	105	3.3095	79.9941	5.3036	20.1712	0.2201	6.1817
v17	14	25	26.48	100	3.3037	84.6070	5.2596	21.4751	0.2261	5.7229
v18	14	25	23.27	110	3.2718	84.5094	5.2195	21.4058	0.2066	5.6539

flow rates (> 3.30 L/min), and models deviating from this angle showed a decrease in efficiency and an increase in WSS. Model 18, with a 110° bend angle, exhibited high efficiency, but also showed an increase in stagnant zones compared to Model 17. This indicated that higher bend angles could produce localized recirculation patterns. Overall, the findings indicate that bend angles between 100 and 105° yield more favorable results.

Finally, the inlet cannula geometry was optimized. A single-step diameter reduction angle of 165° yielded the most favorable performance. Model 16, with this reduction angle and a 50° bend angle, achieved the highest flow rate (3.3574 L/min) and highest efficiency (21.7353%). While the stagnant zones (IC Area-1 and IC Area-2) in this model were not the lowest among the other models (0.6057 mm² and 0.7920 mm², respectively), they still remained at moderate levels, indicating the overall performance of this model. In Model 10, a slight decrease in efficiency and flow rate was observed when the inner diameter was reduced from 18 mm (Model 16) to 16 mm. Changes to the bend angle showed that reducing the angle from 60° to 50° (Model 16) increased the flow rate without changing the WSS. Initially, smooth connections between the inlet and outlet of the IC were tested, but this resulted in low efficiency and reverse flow. Therefore, diameter reduction angle were implemented to solve this problem. The smooth and double-step reduction models (Models 6 and 8) produced the lowest flow rates and efficiencies (9.1774% and 8.5007%, respectively), which is less than half of the efficiency achieved in the optimal model (21.7353%). WSS was generally stable (4 Pa), except in Models 6 and 8 which can be due to flow jet at the housing entrance.

This study aimed to optimize the inlet and outlet cannulae of an LVAD accesories using iterative CAD and CFD analyses. This optimization reduced stagnant flow regions and improved flow stability and hydraulic efficiency. These design improvements have the potential to reduce thrombosis formation and anticoagulation requirements in patients and improve long-term device durability. This approach aims to develop a safer and more reliable LVAD for clinical use. However, the results of this study were limited to computational simulation without *in vitro* validation. Future work includes prototyping the optimized models, performing benchtop tests with the LVAD turbine model used in this study and performing PIV-based flow visualization in a circulatory mock circuit to compare and converge the simulation results.

ACKNOWLEDGMENT

The LVAD turbine model used in this study was designed within the scope of the TUBITAK 1003 project (No. 118S098) and the Yildiz Technical University Scientific Research Projects Coordination Unit (BAP) project (No. FBA-2024-6491) at the Physiological Control Laboratory of the Department of Biomedical Engineering, Yildiz Technical University (YTU).

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)." Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29 (Accessed: 13 Aug. 2025).
[2] M. J. Everly, "Cardiac transplantation in the United States: an analysis of the UNOS registry," *Clin. Transplants*, pp. 35–43, 2008.
[3] European Committee on Organ Transplantation, *Newsletter Transplant*, 2017.
[4] R. C. Starling *et al.*, "Results of the post-U.S. Food and Drug Administration-approval study with a continuous flow left ventricular assist device as a bridge to heart transplantation: a prospective study using the INTERMACS," *J. Am. Coll. Cardiol.*, vol. 57, no. 19, pp. 1890–1898, 2011.
[5] Cleveland Clinic, "Left ventricular assist devices (LVADs)." Available at: https://my.clevelandclinic.org/health/treatments/17192-left-ventricular-assist-devices-mechanical-circulatory-support-mcs (Accessed: 13 Aug. 2025).
[6] Y. Yaacoub *et al.*, "Do you need antiplatelets and anticoagulation for HeartMate 3," *J. Am. Coll. Cardiol.*, vol. 81, no. 8 (Suppl.), p. 2982, 2023.
[7] S. Liao *et al.*, "Numerical prediction of thrombus risk in an anatomically dilated left ventricle: the effect of inflow cannula designs," *Biomed. Eng. Online*, vol. 15, Suppl. 2, p. 136, 2016.
[8] V. K. Chivukula *et al.*, "Left ventricular assist device inflow cannula angle and thrombosis risk," *Circ. Heart Fail.*, vol. 11, no. 4, p. e004325, 2018.
[9] K. May-Newman *et al.*, "Reducing regional flow stasis and improving intraventricular hemodynamics with a tipless inflow cannula design: an in vitro flow visualization study using the EVAHEART LVAD," *Artif. Organs*, vol. 43, no. 9, pp. 834–848, 2019.
[10] J. Sacks *et al.*, "Utility of cardiac computed tomography for inflow cannula patency assessment and prediction of clinical outcome in patients with the HeartMate II left ventricular assist device," *Interact. Cardiovasc. Thorac. Surg.*, vol. 21, no. 5, pp. 590–593, 2015.
[11] R. M. Adamson *et al.*, "Principles of HeartMate II implantation to avoid pump malposition and migration," *J. Card. Surg.*, vol. 30, no. 3, pp. 296–299, 2015.
[12] M. Yigit, "Design, production and performance testing of a novel axial-flow, axial-flux left ventricular assist device (LVAD) with feedback-controlled physiologic flow and two-way wireless communication," Heart Failure 2025, 2025.
[13] M. Ghodrati *et al.*, "The influence of left ventricular assist device inflow cannula position on thrombosis risk," *Artif. Organs*, vol. 44, no. 9, pp. 939–946, 2020.

TABLE III: Inlet Cannula Configurations and Performance Results

NO	Geometry					Performance				Stagnant Zone (mm^2)		WSS (Pa)
	Inner Diameter (mm)	Lower Tangent (mm)	Upper Tangent (mm)	Diameter Reduction Angle (deg)	Bend Angle (deg)	Q (L/min)	P rise (mmHg)	Torque (mN.m)	Efficiency (%)	IC Area-1	IC Area-2	IC Walls
v6	20	19.45	20	Smooth	60	2.1462	81.9742	7.7464	9.1774	0.1778	0.1161	13.4400
v8	20	25.79	20	Double 150	60	2.1553	79.9941	8.1958	8.5007	0.0900	0.1834	9.6312
v10	16	23.60	20	165	60	3.1955	84.3144	5.2450	20.7571	0.2943	0.2521	4.5011
v16	18	27.33	20	165	50	3.3574	84.7495	5.2899	21.7353	0.6057	0.7920	3.9552

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Portable Smartphone-Integrated Incubator for Real-Time Cell Culture Monitoring

Ali Avci*, Bilge Can*†, Meryem Beyza Avci*‡, Arif E. Cetin*†
*Izmir Biomedicine and Genome Center, Izmir 35340, Turkiye
†Department of Biophysics, Faculty of Medicine, Dokuz Eylul University, Balcova, Izmir 35340, Turkiye
‡Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Balcova, Izmir 35340, Turkiye

***Abstract*—Incubators play a crucial role in maintaining living conditions of healthy cell cultures by adjusting the temperature, humidity and gas composition. During conventional workflows, samples are usually removed for imaging purposes, leading them to expose inappropriate conditions that may compromise experimental reliability. Therefore, we present a compact, low-cost smartphone-based incubator capable of real-time cellular monitoring. The necessity for complex computer-integrated monitoring systems is eliminated by the implementation of smartphone for both imaging purposes and control interface. The enclosed chamber preserves optimal growth conditions, ensuring cell viability throughout observation, while its portable design supports precise environmental regulation and uniform culture media distribution. With this approach, an affordable, portable and smartphone integrated solution is achieved, enabling continuous cell tracking with a wide accessibility for real-time characterization in various laboratories and field environments.**

***Keywords*—cell analysis, confluency, incubator, image processing.**

I. Introduction

For cell culture experiments, a controlled environment is required and often maintained by incubators, operating at 37°C, 5% CO_2 and high humidity [1], [2]. Changes in environment, such as temperature, gas concentration and humidity, can occur due to removal of cell cultures for imaging, which can affect the results of the experiment [3]. Since conventional setups consist of incubators and imaging systems separately, this issue is crucial for long-term or real-time monitoring. Under these circumstances, data reliability and cell health can be affected in a negative way by unnecessary exposure to external conditions, the possibility of contamination and delays in returning the samples to incubator. To solve this problem, existing commercial devices either place a microscope directly inside an incubator or use incubation chambers that enclose the microscope to integrate imaging with incubation. Although these solutions provide real-time imaging, they can be costly, increase risk of contamination, be incompatible with different sizes of incubators, and damage electronics due to high humidity CO_2, which all can be considered as downsides. Researchers tried to develop low-cost prototypes [4], [5], [6], [8]; however, their portability and accessibility are constrained by customized optics, motorized stages, and computerized image acquisition. Smartphones are strong candidates for micro-imaging [9], [10] and integrating them eliminates the need for complex optical systems or specialized computers, enabling real-time imaging. Nevertheless, there isn't an existing system, consisting of an integrated smartphone for imaging and a small-sized incubator for continuous cell monitoring. In this work, a low-cost portable platform with an integrated incubator and a smartphone for imaging is introduced for cell culture, proliferation and continuous monitoring, as demonstrated in Figure 1 [11]. The aim of design is sustaining environmental stability, high-quality image capturing, transferring outcomes directly with a smartphone application, removing the requirement of computerized bulky microscopes [12]. Its portability and compatibility with existing incubators enable sample imaging in an appropriate environment, eliminating the possibility of contamination, and increasing the reliability of experimental outcomes. Owing to low-cost production, incubator-based imaging can be adopted by a wider range, ensuring reliable experiments across various laboratory systems.

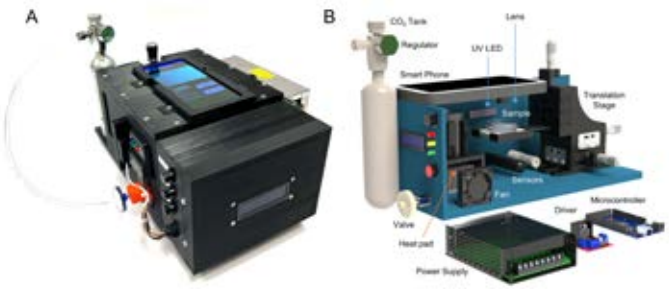


Fig. 1: Portable, low-cost incubator system: (A) photograph and (B) schematic diagram.

II. Methodology

A. System Architecture

A portable incubator has been developed by coupling a smartphone-based imaging module with an environmental control system for real-time cell observation and monitoring without bulky and computer-based microscopes. The incubator contains a small chamber that maintains intact cell growth under defined stable conditions. Additionally, it can ensure environmental homogeneity since the chamber has a small internal volume. Samples are illuminated by an LED array

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

beneath the chamber, and images are captured through a low-cost focusing lens mounted in front of the smartphone camera. A micrometric XYZ stage adjusts sample–lens distance for optimal focus. Environmental regulation consists of an Arduino control unit that controls the LED light source, heat pad, Thermal–Humidity (TH) and CO_2 sensors, and safety valve solenoid. The control unit energizes a heat pad to maintain the chamber at 37°C. The temperature is continuously monitored by the TH sensor. Humidity is maintained (¿80%) generated from a water container above the heat pad, with the TH sensor sending alerts for refill downstream of 81% relative humidity. CO_2 concentration is maintained at 5% from a tank with a regulator and a sensor–valve system with feedback to limit overshoot. All environmental parameters are monitored during the run and reported in real time by an internal LCD.

B. Control Electronics

CO_2 levels were monitored using a nondispersive infrared (NDIR) sensor (0–100,000 ppm, UART output), while temperature and humidity were measured with a waterproof I²C sensor (–40 to 125 °C, 0–100% RH). To minimize random fluctuations, each reading was averaged across three consecutive measurements. Two independent relay channels delivered 12 V to the heating pad and the CO_2 valve. Optocoupler isolation was included to protect the microcontroller against current surges. An I²C LCD module displayed the measured values in real time. The chamber illumination consisted of parallel LEDs that incorporated an array of 1 W white LEDs. These LEDs were powered directly by the 3 V output on the Arduino. The L298N driver regulated current to the valve and heating element circuits. The primary power supply was a 12 V switched-mode power supply. A common cooling fan that also operated on 12 V was added and placed to homogenize the heat, humidity, and CO_2. Near-UV SMD LED UV-C (265 to 275 nm) was also mounted inside the chamber, and was periodically turned on as a microbial sterilization design feature.

C. PID-Based Regulation

PID feedback is obtained via the TH sensor with a setpoint of 38°C to regulate temperature control, accomplishing a consistent 37°C. The Peltier element power is switched between 0-190 PWM by Arduino for prevention of overheating. The reversion of current flow is triggered by surpassing of the target, initiating the controlled cooling. To avoid continual depletion cycles, CO_2 regulation utilizes a familiar PID logic, operating only in the incrementing direction. The valve opens for a maximum of 2 s per cycle, regulated by CO_2 sensor readings, to restore the 5% target concentration. Based on the CO_2 sensor readings, the valve opens (for up to 2 seconds per cycle) to restore the target concentration of 5%.

D. Image Processing Workflow

Cell confluency is calculated with an OpenCV-based pipeline to analyze images taken in BGR format. After denoising the image with a median filter, the image is converted to a grayscale image and then a binary mask is created using adaptive Gaussian thresholding. Morphological opening and closing are used to remove artifacts and topologically fill small holes. Cell edges are measured using the Laplacian operator. This approach allows for providing accurate spatial mapping of cell coverage and for determining a quantitative parameter for assessing culture (Figure 2).

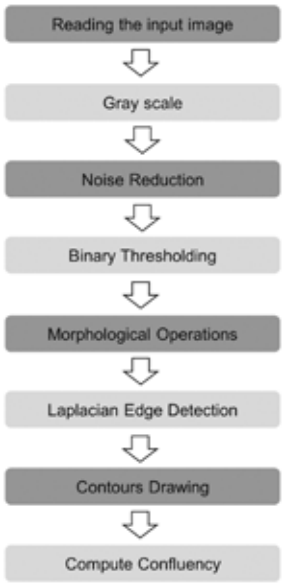


Fig. 2: Workflow diagram of the image processing method used for confluency estimation.

III. Results

A. Incubator Performance

Environmental stability was accomplished by implementation of a microcontroller-based relay control system with sensor input. The regulation of CO_2 was held at 5.0% with ± 0.1% variation by triggering valves for each 0.2s when the concentration was below 4.7% (94% of the set point). The loop maintains the setpoint without overshoot while counteracting the gradual return toward ambient conditions. Temperature control held the chamber at 37°C within ±0.2°C, switching heaters continuously when below 90% of the target or in 0.5 s pulses when near the setpoint. The humidity was maintained at 82% in 8 to 10 minutes. When it goes down below 81%, minor manual refilling was required (Figure 3). Optocoupler isolation was utilized to protect separate relay channels, powering the heating pad and valve.

B. Confluency Analysis

The cell culturation of MCF-7 breast cancer cells seeded at 50,000 cells/mL took 2 and 3 days. The outputs of the image processing algorithm showed confluency levels of 54% and 67% respectively, demonstrating the ability of distinguishment of growth stages. Cluster regions were quantitively marked by images processed by the algorithm, indicating the method offers spatial mapping alongside quantitative assessment.

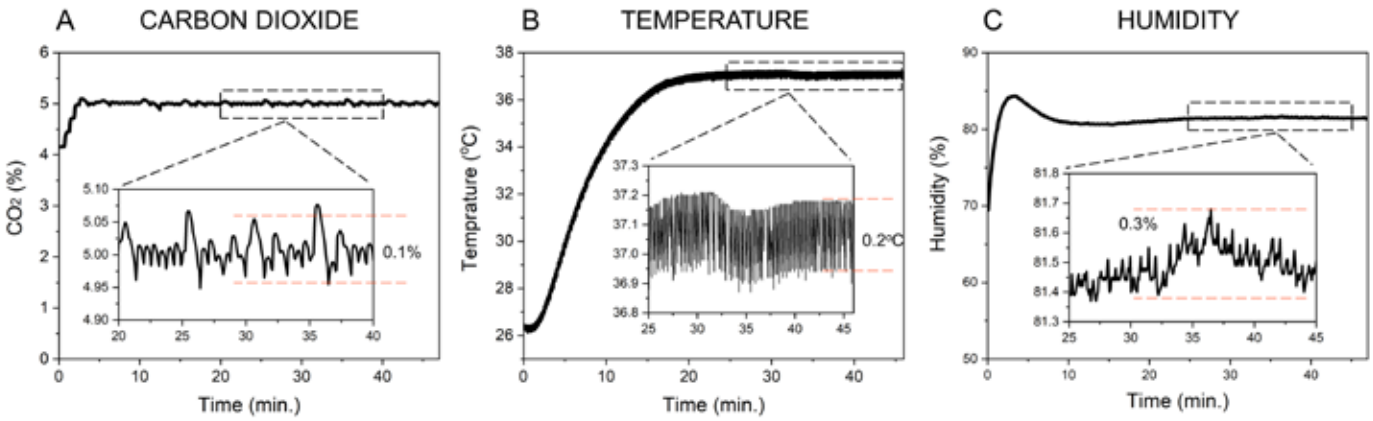


Fig. 3: Performance of the incubator in sustaining CO_2 at 5%, temperature at 37 °C, and humidity for cell culture. (A) CO_2 regulation via sensor–valve feedback. (B) Temperature control using sensor–heating pad feedback. (C) Humidity stabilization with the heating pad.

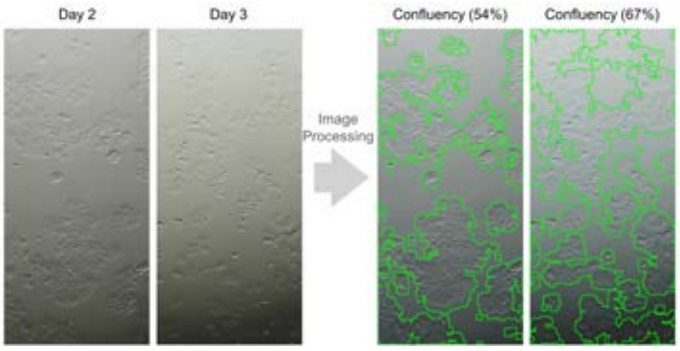


Fig. 4: MCF7 cell images captured on the second and third day of incubation.

C. Mobile Application

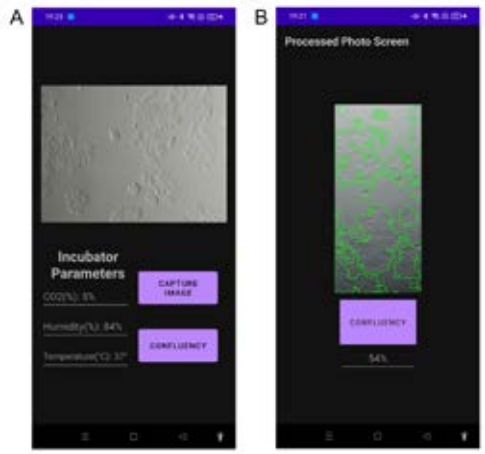


Fig. 5: Mobile application.

A Java-based application is developed with Android Studio for real-time monitoring of the incubator's CO_2, temperature and humidity sensor via USB connection. Instantaneous analysis is achieved through integration of image processing algorithm to device, eliminating the requirement of additional computing, enabling a portable workflow for real-time cell imaging.

IV. Conclusion

To conclude, we introduced a compact smartphone-based solution with incubator system for live cell tracking. A lens assembly is integrated into smartphone to capture petri dish images, while the CO_2 and temperature sensors are controlled by a microcontroller. The system is enclosed in a 3D-printed case, merging environmental control and imaging in a portable, cost-effective platform as a practical solution for real-time cell imaging.

References

[1] Y. K. Jawale, U. Rapol, and C. A. Athale, "3D-printed focussing mechanism for cellphone-based cellular microscopy," *J. Microsc.*, vol. 273, pp. 105–114, 2019.

[2] V. Müller, J. M. Sousa, H. C. Koydemir, M. Veli, D. Tseng, L. Cerqueira, A. Ozcan, N. F. Azevedo, and F. Westerlund, "Identification of pathogenic bacteria in complex samples using a smartphone based fluorescence microscope," *RSC Adv.*, vol. 8, pp. 36493–36502, 2018.

[3] A. Orth, E. R. Wilson, J. G. Thompson, and B. C. Gibson, "A dual-mode mobile phone microscope using the onboard camera flash and ambient light," *Sci. Rep.*, vol. 8, p. 3298, 2018.

[4] N. A. Switz, M. V. D'Ambrosio, and D. A. Fletcher, "Low-cost mobile phone microscopy with a reversed mobile phone camera lens," *PLoS One*, vol. 9, p. e95330, 2014.

[5] H. Zhu, I. Sencan, J. Wong, S. Dimitrov, D. Tseng, K. Nagashima, and A. Ozcan, "Cost-effective and rapid blood analysis on a cell-phone," *Lab Chip*, vol. 13, pp. 1282–1288, 2013.

[6] M. P. Walzik, V. Vollmar, T. Lachnit, H. Dietz, S. Haug, H. Bachmann, M. Fath, D. Aschenbrenner, S. A. Mofrad, O. Friedrich, and D. F. Gilbert, "A portable low-cost long-term live cell imaging platform for biomedical research and education," *Biosens. Bioelectron.*, vol. 64, pp. 639–649, 2015.

[7] Q. Wei, H. Qi, W. Luo, D. Tseng, S. J. Ki, Z. Wan, Z. Göröcs, L. A. Bentolila, T. Wu, R. Sun, and A. Ozcan, "Fluorescent imaging of single nanoparticles and viruses on a smart phone," *ACS Nano*, vol. 7, pp. 9147–9155, 2013.

[8] D. K. Rajan, J. Kreutzer, H. Välimäki, M. Pekkanen-Mattila, A. Ahola, A. Skogberg, K. Aalto-Setälä, H. Ihalainen, P. Kallio, and J. Lekkala, "A portable live-cell imaging system with an invert-upright-convertible architecture and a mini-bioreactor for long-term simultaneous cell imaging, chemical sensing and electrophysiological recording," *IEEE Access*, vol. 6, pp. 11063–11075, 2018.

[9] K. Su, Y. Pan, Z. Wan, L. Zhong, J. Fang, Q. Zou, and H. Li, "Smartphone-based portable biosensing system using cell viability biosensor for okadaic acid detection," *Sens. Actuators B Chem.*, vol. 251, pp. 134–143, 2017.

[10] K. Yang, J. Wu, H. Peretz-Soroka, L. Zhu, Z. Li, Y. Sang, J. Hipolito, M. Zhang, S. Santos, C. Hillier, R. L. de Faria, Y. Liu, and F. Lin, "Mkit: A cell migration assay based on microfluidic device and smartphone," *Biosens. Bioelectron.*, vol. 99, pp. 259–267, 2018.

[11] A. E. Cetin, "A portable and low-cost incubator system enabling real-time cell imaging based on a smartphone," *Turkish Journal of Physics*, vol. 47, no. 5, pp. 279–296, 2023.

[12] M. B. Avci, S. D. Yasar, and A. E. Cetin, "An optofluidic platform for cell-counting applications," *Analytical Methods*, vol. 15, no. 18, pp. 2244–2252, 2023.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Hemodynamic Effects on Acute Mesenteric Ischemia of Superior Mesenteric Artery

Hazal AZAZI
Department of Electrical and Electronics Engineering
Istanbul Bilgi University
Istanbul, Türkiye

Burak KAMBUROGLU
Department of General Surgery
Sakarya University
Sakarya, Türkiye

Enis DIKICIER
Department of General Surgery
Sakarya University
Sakarya, Türkiye

Ibrahim Basar AKA
Department of Mechatronics Engineering
Istanbul Bilgi University
Istanbul, Türkiye

***Abstract*— This study evaluates the role of superior mesenteric artery (SMA) geometry and hemodynamic parameters in diagnosing acute mesenteric ischemia (AMI) using patient-specific computational fluid dynamics (CFD) simulations. SMA models reconstructed from CT scans of five AMI patients and five controls were analyzed to quantify aortomesenteric (AOM) angle and distance, while transient pulsatile CFD simulations provided wall shear stress, oscillatory shear index, relative residence time (RRT), endothelial cell activation potential, and energy loss values. Results showed that AMI patients exhibited larger mean AOM angles and distances than controls, and reduced angles/distances correlated with increased energy loss. At end systole, RRT values were significantly higher in the AMI group, indicating its potential as a key discriminative marker. These findings highlight the promise of a non-invasive computational approach to complement imaging techniques for improved AMI diagnosis.**

Keywords— *Superior Mesenteric Artery, Ischemic Vascular Disease, Hemodynamics, Acute SMA Ischemia*

I. INTRODUCTION

Acute ischemia of the superior mesenteric artery (SMA) is a life-threatening condition that can lead to devastating outcomes, including bowel necrosis, systemic complications, and significant mortality if not promptly diagnosed and treated [1]. The SMA is a vital vessel responsible for supplying oxygenated blood to a large portion of the intestines, and its occlusion disrupts intestinal perfusion, leading to ischemia and necrosis. The condition arises from various causes, including embolism, thrombosis, and non-occlusive mesenteric ischemia (NOMI), each presenting distinct clinical and therapeutic challenges [2], [3], [4], [5]. Untreated acute mesenteric ischemia (AMI) is associated with extremely high mortality, which underscores the urgency of early diagnosis and intervention.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

Beyond clinical management, hemodynamic factors play a critical role in the progression of ischemia. Parameters such as wall shear stress (WSS), oscillatory shear index (OSI), and relative residence time (RRT) are known to influence endothelial function and thrombus formation, which are key contributors to ischemic pathophysiology [6], [7]. These parameters help in understanding the complex interactions between blood flow dynamics and vascular health, providing insights that can guide clinical interventions and improve patient outcomes [6]. While these hemodynamic markers have been extensively studied in coronary artery disease, their application to SMA ischemia is relatively novel.

This study aims to address the knowledge gap in SMA ischemia by integrating advanced CFD simulations with clinical imaging data. Our findings aim to identify key markers that could aid in the early diagnosis and management of this challenging condition, offering a novel computational framework to complement traditional diagnostic methods.

II. MATERIALS AND METHODS

A. Data acquisition and patient-specific geometries

The patients diagnosed with AMI are considered a patient group labeled as P1, P2, P3, P4, and P5, and the control group C1, C2, C3, C4, C5.

CTA data containing one hundred or more slices were transferred to 3D Slicer 5.4.0 software in DICOM format. The region of interest (ROI) was separated by segmentation, and 3D geometries were exported in the Standard Triangle Language (STL) form for patient-specific CFD analyses. To reduce surface roughness caused by imaging artifacts, 3D models were edited using Autodesk Meshmixer (San Rafael, CA, USA) software.

B. Calculating the geometric relations of the SMA

After the segmentation, the centerline data of the models is obtained by the 3D Slicer Vascular Modeling Toolkit (VMTK) Centerline Extraction extension. Then, its coordinates are exported in Comma-Separated Values (CSV) format and used for further calculations.

The AOM angles were calculated using the angle calculation formula between two vectors using MATLAB (R2023a, MathWorks, Inc.). The vector u was formed by subtracting the last coordinates of the SMA from its initial coordinates, while the vector v was similarly constructed for the aorta. The cross and dot products of vectors u and v were then provided as parameters to MATLAB' "atan2" function for the relevant calculation. The AOM distance was found with the distance formula between two points, measured 1 cm below the bifurcation point of the SMA using SolidWorks 2020 (Waltham, MA, USA).

C. CFD method and boundary conditions

Patient-specific CFD simulations were run individually to understand whether the geometric relations are related to the WSS-based hemodynamic indicators and to find a descriptor showing a significant difference between the patient and control groups.

Firstly, the boundaries of the reconstructed 3D geometries obtained from segmentation were extended between 30 and 50 mm using SolidWorks 2020 (Waltham, MA, USA) to avoid divergence and backflow.

Then, the 3D geometries were divided into volumes with poly-hexcore elements using ANSYS Fluent (2023 R1, ANSYS, Inc., Pittsburgh, PA, USA) with eight boundary layers to capture the velocity gradients correctly. The maximum element size was 8 x 10^{-4} m. A no-slip boundary condition was used on the rigid walls. Each model resulted in the same fineness; the final meshes contained 0.7 – 2.1 million cells in total.

The blood was treated as Newtonian and incompressible with a 1050 kg/m^3 density and 0.0035 Pa.s dynamic viscosity.

The CFD simulations presented in this research were performed on Fluent software. The governing equations (1)(2) were solved numerically using the finite volume method. A second-order upwind scheme was used for spatial discretization and the SIMPLE algorithm (Semi-Implicit Method for Pressure Linked Equations) for pressure–velocity coupling. The k–ω SST turbulence model was employed in the simulations.

$$\rho \frac{dv}{dt} = -\nabla p + \mu \Delta v \quad (1)$$

$$\nabla . V = 0 \quad (2)$$

A time-variant parabolic velocity profile was imposed at the inlet, and pulsatile pressure was applied to the distal abdominal aorta and SMA outlets. The reference pressure, P_0, was set at 80 mmHg. The outlet pressure was assigned using the simplest outlet lumped model, which includes volumetric flow rate and downstream resistance (3) in each time step.

$$P = P_0 + RQ - P_d \quad (3)$$

P_0 is a reference pressure, R is the downstream resistance, Q is the volumetric flow rate, and Pd is the dynamic pressure.

Resistance (R) values were calculated patient-specifically using the Hagen-Poiseuille Equation (4), and User-Defined Functions (UDF) were compiled separately.

$$R = \frac{8\mu L}{\pi r^4} \quad (4)$$

Where μ is the blood viscosity, L is the vessel's length, and r is the vessel's radius.

Transient simulations were run over three complete cardiac cycles, assuming a per-cycle duration of 0.8 seconds, to ensure thorough flow development and a periodic solution. The constant time step was set as 1 × 10^{-3} s, and the maximum residual for convergence was 10^{-5}.

Data extraction and visualization were performed with Ansys's post-processing tool, CFD-Post (2023 R1, ANSYS, Inc.). Numerical results taken from the last cardiac cycle' early systole (t = 1.6 s), peak systole (t = 1.8 s), end systole (t = 1.95 s), and end cardiac cycle (t = 2.45 s) results with maximum and mean values of each parameter investigated were visualized using MATLAB (R2023a, MathWorks, Inc.) to observe the effects of the hemodynamic factors.

The hemodynamic parameters analyzed in this study have direct clinical implications. WSS and OSI are linked to endothelial dysfunction, thrombus formation, and vascular remodeling, whereas RRT is increasingly recognized as a sensitive marker of ischemic risk. By quantifying these parameters in patient-specific SMA models, CFD provides functional information that conventional CT angiography alone cannot capture. This integration of anatomical and hemodynamic assessment may enhance early diagnosis of AMI and help identify patients at higher risk, supporting clinical decision-making.

III. RESULTS

The geometric relationships of the patient's SMAs, energy loss, and stress hemodynamic factors on the SMA wall were investigated in detail for ten cases in total, consisting of 5 control groups and 5 patient groups. Energy loss calculations, maximum and average values of TAWSS, OSI, maximum values of ECAP, and log(RRT), were reported from four different time steps of the last cardiac cycle.

Although no definitive distinction was observed, the patient group (48.26 degree for AOM angle and 16.6 mm for AOM distance) had higher mean values for aortomesenteric angle and distance compared to the control group (34.33 degree for AOM angle and 13.22 mm for AOM distance). The association between the geometric relations of the SMA and energy loss revealed that Control 1, with a 22.95 degree angle and 8.6 mm distance, representing the smallest aortomesenteric angle and distance, experienced the highest energy loss during peak systole. Conversely, the lowest energy loss results were observed in Patient 3, with a 78.55 degree

angle and 30 mm distance, representing the highest AOM angle and distance.

All findings presented in this paragraph were observed during the end systole. Control 3, with a max TAWSS value of 4.18 Pa, and Patient 2, with an avg TAWSS value of 0.65, had the smallest TAWSS values. Patient 5, whose max OSI value was closest to 0.5, and Patient 2, whose avg OSI value was 0.27, had the highest OSI values. Analysis of the maximum ECAP values showed that Control 3 exhibited the highest value at 0.120, whereas Patient 4 had the lowest at 0.030. Finally, according to the calculations made from the maximum values, Patient 5 had the highest RRT value with 25.91, and Control 2 had the lowest RRT value with 13.49. Particularly, due to the minimal disparity observed in max OSI values, log (RRT) emerged as the primary discriminatory factor. Within the scope of these results, max and avg TAWSS, avg OSI and max ECAP values did not significantly differ between the patient and control groups. However, a notable dissimilarity was identified between the patient and control groups during the end systole in terms of log (RRT) values.

Table I presents the hemodynamic parameters that demonstrated meaningful differences between the AMI and control groups. For improved data visualization, Figure 1 presents the OSI and RRT values of both the patient and control groups.

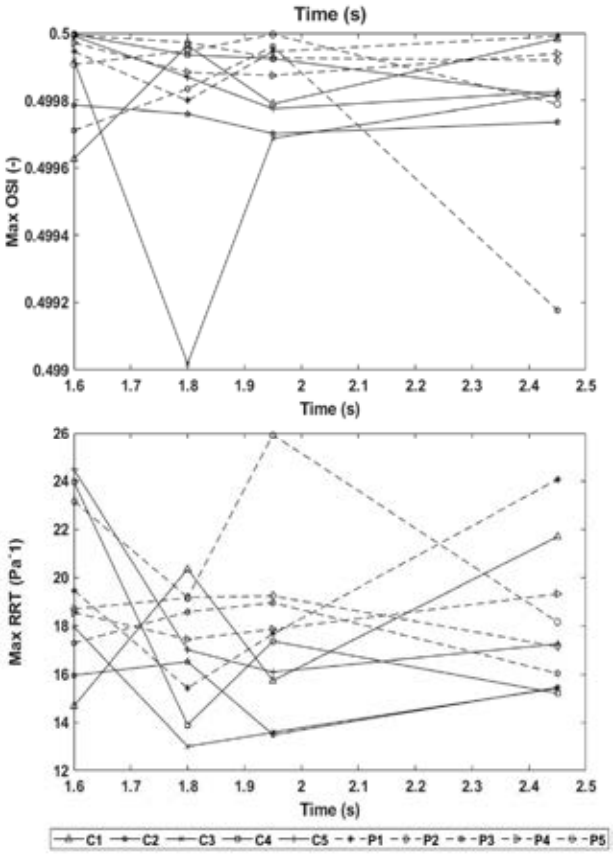


Fig. 1. The numerical results' plots of maximum OSI (top), and log (RRT) (bottom) for patient and control groups.

TABLE I. PATIENT-SPECIFIC ANATOMICAL AND HEMODYNAMIC CHARACTERISTICS

Patient #	SMA Characteristics		Time (s)	CFD Results		
	AOM Angle (°)	AOM Distance (mm)		Energy Loss (mW)	OSI (-) Avg (10^{-3})	RRT (Pa^{-1}) log
C1	22.95	8.6	t = 1.6	0.31	89.13	14.67
			t = 1.8	28.75	87.05	20.34
			t = 1.95	10.46	88.53	15.73
			t = 2.45	0.19	89.55	21.68
C2	31.03	11.4	t = 1.6	0.41	72.23	15.95
			t = 1.8	11.45	68.00	16.51
			t = 1.95	2.63	71.57	13.49
			t = 2.45	0.44	72.63	15.44
C3	42.07	12.6	t = 1.6	0.19	85.00	17.92
			t = 1.8	13.00	82.76	13.01
			t = 1.95	3.71	84.32	13.59
			t = 2.45	0.24	85.30	15.41
C4	44.11	25.71	t = 1.6	3.30	79.75	23.95
			t = 1.8	22.24	77.29	13.89
			t = 1.95	1.15	80.45	17.35
			t = 2.45	3.23	82.68	15.2
C5	31.48	7.80	t = 1.6	0.65	79.79	24.47
			t = 1.8	10.28	77.40	16.99
			t = 1.95	1.59	79.55	16.09
			t = 2.45	0.67	79.82	17.24
P1	31.20	11.90	t = 1.6	0.03	62.38	19.47
			t = 1.8	24.36	59.24	15.42
			t = 1.95	8.10	61.58	17.65
			t = 2.45	0.06	62.29	24.05
P2	74.36	9.60	t = 1.6	0.61	274.23	23.16
			t = 1.8	15.86	275.79	19.18
			t = 1.95	3.62	273.86	19.23
			t = 2.45	0.67	274.56	17.13
P3	78.55	30.00	t = 1.6	17.40	90.88	17.28
			t = 1.8	9.90	86.18	18.57
			t = 1.95	1.30	90.59	18.95
			t = 2.45	18.90	90.96	16.03
P4	32.55	15.80	t = 1.6	0.03	109.48	18.54
			t = 1.8	10.21	108.78	17.44
			t = 1.95	3.26	110.01	17.85
			t = 2.45	0.03	109.95	19.32
P5	24.64	15.70	t = 1.6	1.35	162.41	18.69
			t = 1.8	14.64	163.44	19.15
			t = 1.95	1.05	162.31	25.91
			t = 2.45	1.36	162.67	18.15

IV. DISCUSSION

At peak systole, the inverse relationship between AOM angles and distances and energy loss was notable. While the largest angle and distance appeared to reduce energy loss, it does not necessarily indicate hemodynamic efficiency. Instead, these extreme geometries may disrupt flow, leading to low WSS areas or elevated RRT, both of which are associated with increased thrombotic and ischemic risk. The research presented by Wu et al. [8] demonstrates that variations in the SMA-distal aorta angle did not significantly affect flow patterns, but larger angles induced regionally high OSI on the superior convex of the SMA. These findings highlight the complex interplay between vascular geometry and hemodynamic performance, underscoring the importance of considering additional parameters in clinical risk assessment.

While previous studies have highlighted the importance of RRT, its role in SMA-related ischemia remains underexplored. RRT holds significant clinical importance as a distinguishing factor between groups compared to the other hemodynamic parameters [9], [10]. Elevated RRT values in the AMI group are consistent with previous studies that associate high RRT with increased thrombotic risk and vascular remodeling [11], [12], [13].

V. CONCLUSION

As a summary, the AMI group exhibited larger mean AOM angles (48.26° vs. 34.33°) and distances (16.6 mm vs. 13.22 mm) compared to controls. At peak systole (1.8 s), reduced AOM angles and distances correlated with elevated energy loss, with Control 1 experiencing the highest energy loss (5.18×10^{-2} J). Conversely, Patient 3, with the largest AOM angle (78.55°) and distance (30 mm), showed the lowest energy loss (1.78×10^{-2} J). At end systole, a significant difference in RRT was observed, with the AMI group showing higher mean values (19.92 Pa^{-1}) compared to controls (15.25 Pa^{-1}).

This study has limitations, including the small sample size and modeling assumptions. Future work should validate RRT in larger populations, incorporate patient-specific inflow profiles, and explore advanced modeling to improve accuracy.

CONFLICT OF INTEREST

All authors declare that they have no conflict of interest.

ETHICS APPROVAL

This study was approved by the Ethics Committee of Sakarya University School of Medicine (No: 333; Date: 26/06/2020). CT scans used for patient-specific reconstructions were acquired between January 2015 and December 2019.

REFERENCES

[1] A. Bouassria *et al.*, "Superior Mesenteric Arterial Embolism Associated with an Acute Limb Ischemia: A Case Report and Literature Review," *Open J Gastroenterol*, vol. 04, no. 04, 2014, doi: 10.4236/ojgas.2014.44027.

[2] T. Ichiba *et al.*, "Baseline disease is a more important predictor of intestinal necrosis than CT findings in patients with acute mesenteric ischemia," *American Journal of Emergency Medicine*, vol. 34, no. 12, 2016, doi: 10.1016/j.ajem.2016.08.016.

[3] K. Molyneux, J. Beck-Esmay, A. Koyfman, and B. Long, "High risk and low prevalence diseases: Mesenteric ischemia," 2023. doi: 10.1016/j.ajem.2023.01.001.

[4] A. Nuzzo, Y. Castier, and O. Corcos, "Superior mesenteric artery dissection does not necessarily mean acute mesenteric ischemia," 2016. doi: 10.1016/j.ajem.2016.09.031.

[5] L. Garzelli *et al.*, "Insights into acute mesenteric ischaemia: an up-to-date, evidence-based review from a mesenteric stroke centre unit," 2023. doi: 10.1259/bjr.20230232.

[6] Y. Zhao *et al.*, "Evaluating Short-Term and Long-Term Risks Associated with Renal Artery Stenosis Position and Severity: A Hemodynamic Study," *Bioengineering*, vol. 10, no. 9, 2023, doi: 10.3390/bioengineering10091002.

[7] X. Zhang *et al.*, "Localized Elevation of Wall Shear Stress Is Linked to Recent Symptoms in Patients with Carotid Stenosis," *Cerebrovascular Diseases*, vol. 52, no. 3, 2023, doi: 10.1159/000526872.

[8] Z. Wu *et al.*, "The Significance of the Angle between Superior Mesenteric Artery and Aorta in Spontaneous Isolated Superior Mesenteric Artery Dissection," *Ann Vasc Surg*, vol. 45, 2017, doi: 10.1016/j.avsg.2017.06.156.

[9] G. Wang *et al.*, "The hemodynamic effect of eccentricity in visceral branched aneurysms with multilayer stents," *Proc Inst Mech Eng H*, vol. 236, no. 8, 2022, doi: 10.1177/09544119221106829.

[10] C. Trenti, M. Ziegler, N. Bjarnegård, T. Ebbers, M. Lindenberger, and P. Dyverfeldt, "Wall shear stress and relative residence time as potential risk factors for abdominal aortic aneurysms in males: a 4D flow cardiovascular magnetic resonance case–control study," *Journal of Cardiovascular Magnetic Resonance*, vol. 24, no. 1, 2022, doi: 10.1186/s12968-022-00848-2.

[11] J. Hashemi, B. Patel, Y. S. Chatzizisis, and G. S. Kassab, "Study of Coronary Atherosclerosis Using Blood Residence Time," *Front Physiol*, vol. 12, 2021, doi: 10.3389/fphys.2021.625420.

[12] Z. Chen *et al.*, "Vascular Remodelling Relates to an Elevated Oscillatory Shear Index and Relative Residence Time in Spontaneously Hypertensive Rats," *Sci Rep*, vol. 7, no. 1, 2017, doi: 10.1038/s41598-017-01906-x.

[13] N. R. Pillalamarri, S. Piskin, S. S. Patnaik, S. Murali, and E. A. Finol, "Patient-Specific Computational Analysis of Hemodynamics in Adult Pulmonary Hypertension," *Ann Biomed Eng*, vol. 49, no. 12, 2021, doi: 10.1007/s10439-021-02884-y.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

From Arousal to Acceptance: A Review of EDA/GSR and BVP Applications in Food Consumer Research

Riza Sadikzade
Dept. of Gastronomy and Culinary Arts
Izmir University of Economics
Izmir, Turkey
riza.sadikzade@ieu.edu.tr

K. Nazan Turhan
Dept. of Gastronomy and Culinary Arts
Izmir University of Economics
Izmir, Turkey
nazan.turhan@ieu.edu.tr

***Abstract*— Gastronomic experiences extend beyond taste and presentation, engaging multiple senses, emotions, and physiological responses. While traditional methods such as surveys and Interviews rely on subjective self-reports, recent research has demonstrated the value of psychophysiological measures—particularly electrodermal activity (EDA/GSR) and blood volume pulse (BVP/PPG)—as objective, time-resolved indicators of consumer arousal and engagement. This paper reviews the application of GSR in food consumer science, highlighting its role in assessing emotions and motivations, product and sensory evaluations, consumer behavior, and food packaging design. Evidence from recent studies indicates that GSR effectively captures implicit affective states in various contexts, including taste perception, beverage testing, pricing strategies, branding, and packaging, often revealing subconscious responses that diverge from explicit self-reports. These findings highlight the methodological advantages of integrating GSR with complementary tools, such as EEG, heart rate, and facial electromyography, to gain a comprehensive understanding of the consumer experience. Despite challenges related to signal interpretation and valence detection, standardized protocols and multimodal approaches strengthen the validity and predictive power of GSR in consumer research. Overall, the literature supports GSR as a robust tool for advancing sensory science and marketing, providing actionable insights into how emotions and physiological arousal shape consumer decision-making in food-related contexts.**

Keywords:—Gastronomy; GSR; EDA; Neuromarketing; Consumer Behaviour; Food Pacaking; Consumer Emotion

I. INTRODUCTION

Gastronomy is no longer just about taste and presentation—it is a multisensory experience that engages emotions, memory, and physiological reactions [1, 2]. Traditional methods like surveys and interviews rely on subjective self-reports, which can be biased or imprecise [2]. In contrast, physiological measures such as Galvanic Skin Response (GSR) and Blood Volume Pulse (BVP/PPG) offer time-resolved, objective indicators of arousal and affect that can be aligned with dining events [3-6] and have begun to show feasibility in food and beverage contexts [7].

Galvanic Skin Response (GSR), also known as electrodermal activity (EDA), is a physiological measure of emotional and sympathetic nervous system arousal [3, 8]. It detects changes in the skin's electrical conductivity caused by sweat gland activity, which is innervated by the autonomic nervous system. When a person experiences emotional stimulation—such as tasting an unexpectedly flavorful dish or reacting to a displeasing texture—their sweat glands secrete minute amounts of moisture, altering skin conductance. This makes GSR a powerful tool for objectively quantifying emotional engagement in gastronomy, thereby bypassing the biases inherent in self-reported surveys [1, 3].

GSR Data Processing GSR signals are categorized into two components:

1. SCL (Skin Conductance Level) or Tonic – Represents the slow-changing, background level of skin conductance influenced by long-term factors like stress, ambient temperature, or an individual's general arousal state. In a dining context, tonic levels might reflect a guest's overall comfort (e.g., relaxed vs. stressed) before eating.

2. SCR (Skin Conductance Response) or Phasic – Rapid, short-term spikes triggered by immediate stimuli, such as the first bite of a dish, a surprising flavor combination, or an aromatic garnish. These transient responses are key to identifying specific moments of pleasure, surprise, or aversion during a meal [3, 8].

The tonic Skin Conductance Level (SCL) was obtained by low-pass filtering the SC signal at ~0.05 Hz (very slow trends). The phasic Skin Conductance Response (SCR) was computed as the fast component of SC, either as the residual SCR(t)=SC(t)−SCL(t) or via a band-pass filter (≈0.05–1.0 Hz) that preserves stimulus-locked fluctuations [664, 9]. SCL change (ΔSCL) was quantified as the mean SCL in the response window minus the baseline SCL averaged over a pre-stimulus interval (e.g., 5 s immediately before the cue) [8]. In event-related SCR analysis, a post-stimulus window—typically 1–10 s after the cue (longer for olfaction)—is examined. Peaks in the phasic signal are detected with a refractory period of at least 1 s and a minimum-prominence threshold.

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

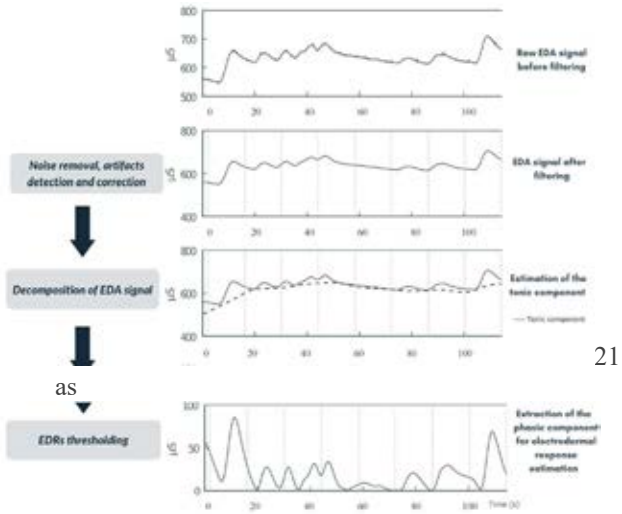


Fig 1. Schematic representation of EDA signal processing. Adapted from Zangróniz et al. (2017).

Amplitude is defined as the peak height above the preceding trough (µS), and latency as the time from stimulus to peak [8,10]. Researchers can use a range of GSR processing software—Ledalab (a MATLAB-based toolbox) among the most common.

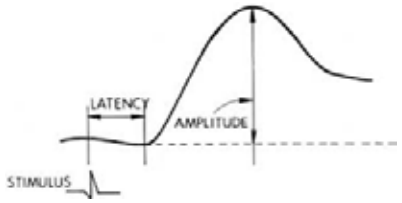


Fig. 2. Example of a typical SCR. Following stimulus onset. Adopted from Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011)

Blood Volume Pulse (BVP)

BVP, often measured via photoplethysmography (PPG), tracks changes in blood flow near the skin's surface, reflecting heart rate and vascular responses. A sensor (usually on the finger or earlobe) emits light into the skin and detects reflected light to measure blood volume changes. Increased heart rate or vasoconstriction/vasodilation can indicate stress, excitement, or relaxation. By segmenting the BVP signal into epochs aligned to the stimulus (e.g., ±5 s around the ad), researchers can directly compare ad-period BVP to pre-stimulus baseline. For noise reduction of the BVP signal, a 0.5–5 Hz band-pass filter is applied; this range is chosen to retain heart-rate frequencies [5, 6].

II. Applications of EDA in Food Consumer Science Research

Electrodermal activity (EDA) measurements in consumer research require careful interpretation, as skin conductance fluctuations may reflect not only emotional arousal but also responses to environmental stimuli (e.g., ambient noise, odors) or cognitive states (e.g., task concentration, spontaneous thoughts) [11]. Standardized protocols are essential to control confounding variables and ensure valid inferences from EDA data [12]. In food consumer science, EDA has emerged as a valuable tool to complement self-reports, capturing implicit emotional and motivational responses that traditional methods may miss, particularly non-conscious affective states [13]. EDA has been used in conjunction with explicit consumer emotions and motivations, product/experience evaluation, consumer behavior and food packaging research.

A. EDA in Assessing Consumer Emotions and Motivations

Caruelle et al. [13] review the use of electrodermal activity (EDA) in consumer research, highlighting its value as a psychophysiological measure of emotional arousal. They emphasize that EDA captures only arousal, not valence, and note methodological issues such as inconsistent reporting, latency, and attentional influences. The authors call for greater transparency and standardization, suggesting EDA is most effective when combined with complementary measures of emotion. Spinelli et al. [14] examine skin-conductance responses (SCR) to oral food stimuli, illuminating how taste quality and intensity, as well as individual personality traits, modulate emotional engagement during consumption. Their findings indicate that SCR reliably reflect arousal triggered by sensory characteristics such as novelty, perceived quality, and flavor intensity, underscoring that these physiological responses originate from multiple emotional drivers. This study highlights the value of assessing consumer reactions to food at the implicit, autonomic level, suggesting that SCR can uncover deeper layers of emotional experience—especially those not captured by traditional self-report—thus offering a more nuanced understanding of consumer sensory and affective experience. Larrañaga-Ayastuy et al. [7] explore the interplay between electrodermal activity (EDA) and conscious liking in the context of beer tasting, revealing critical insights into how implicit emotional arousal aligns—or diverges—from explicit evaluation. Their findings demonstrate that EDA, as a marker of autonomic emotional responses, does not always mirror self-reported liking, suggesting that consumers' immediate physiological reactions during tasting can be distinct from their reflective judgments. This divergence underscores the value of integrating implicit measures like EDA alongside explicit assessments in sensory research to capture the full spectrum of consumer experience. The study positions EDA as a powerful tool for uncovering underlying emotional engagement that may inform product development and marketing strategies in the beverage sector. Rita et al [15] investigate the autonomic emotional responses elicited by private label versus national food brands using electrodermal activity (EDA) measurements. Their experimental study reveals that national brands consistently generate stronger physiological arousal responses compared to private label alternatives, suggesting deeper emotional engagement with established brand identities. Notably, these implicit physiological reactions often diverged from participants' explicit brand attitude reports, revealing subconscious preferences undetectable through traditional surveys. The study demonstrates how autonomic nervous system responses provide unique insights into brand perception, particularly for low-involvement food products where consumers may lack strong conscious brand preferences. These findings advance our understanding of how different brand types unconsciously influence consumer emotions, with important implications for brand positioning strategies and private label marketing approaches in the food sector.

B. Product/Experience Evaluation

GSR has been employed in studies that measure consumers' perceived quality and perceived value, providing implicit insights into their underlying emotional responses. Álvarez-Pato et al. [16] developed an innovative multimodal framework for predicting food acceptance by integrating facial emotion recognition (FER), galvanic skin response (GSR), cardiac pulse measurements, and traditional self-reported liking scores during evaluations of both pleasant and unpleasant food samples. Their results revealed important limitations and opportunities in sensory evaluation methods: while FER alone proved insufficient for reliable acceptance prediction, the combination of physiological measures - particularly GSR as a robust indicator of emotional arousal - significantly improved predictive accuracy when analyzed through machine learning models (neural networks and decision trees). The study demonstrated that GSR measurements captured crucial aspects of consumer engagement that were not evident in either facial expressions or self-reports alone, with cardiac pulse providing additional, though less pronounced, predictive value. These findings challenge conventional reliance on subjective ratings in sensory science, instead advocating for a data fusion approach that combines multiple physiological signals with traditional measures. The research establishes GSR as particularly valuable for objectively assessing emotional responses to food stimuli, while also highlighting the potential of machine learning to automate and enhance acceptance predictions in both research and product development settings. This multimodal framework offers food scientists and marketers a more comprehensive, objective tool for understanding consumer preferences that accounts for both conscious evaluations and subconscious physiological responses. Martínez-Levy et al. [17] conducted a multimodal study examining consumer responses to chocolate pudding with varying sucrose concentrations (38, 83, 119, and 233 g/kg), combining both explicit self-reports (liking, sweetness, bitterness, and astringency ratings) and implicit physiological measures (heart rate/HR and galvanic skin response/GSR). Their findings revealed that increasing sucrose concentration enhanced self-reported liking and sweetness perception while reducing bitterness and astringency. Notably, HR - but not GSR - showed significant variation across sucrose levels, with the intermediate concentration (119 g/kg) eliciting peak HR responses that positively correlated with hedonic ratings, suggesting an optimal "sweet spot" for consumer preference. The study also found that adult participants exhibited stronger physiological responses than younger adults, and that HR patterns at lower sucrose levels reflected bitterness/astringency perception. These results demonstrate the value of integrating physiological and self-report measures in sensory science, providing actionable insights for product optimization (particularly suggesting ~120 g/kg sucrose as an ideal formulation target) while highlighting age-related differences in consumer responses. The research underscores how implicit physiological measures can reveal important consumer preferences that may not be captured through traditional self-report methods alone.

C. Consumer Behavior

Multiple studies demonstrate significant correlations between electrodermal activity (EDA) and key consumer behavior metrics, including purchase intention, willingness to pay premiums and brand loyalty formation suggesting EDA's predictive value in consumer research. Somervuori and Ravaja [18] employ psychophysiological measures (facial EMG, skin conductance) to reveal that price levels differentially modulate emotional and behavioral responses: lower-priced brands elicit stronger positive affect and approach tendencies (SC increase: 1.2±0.3μS), while premium pricing triggers negative affect despite higher perceived quality ratings. This dissociation between conscious evaluations (self-reported) and implicit reactions (physiological) demonstrates that pricing strategies activate distinct affective processes, with physiological measures uncovering latent emotional conflicts (e.g., desire for quality vs. price aversion) that traditional surveys fail to capture. Smith et al. [19] demonstrate how branded food packaging physiologically engages children through skin conductance response (SCR) measurements. Their experimental study with school-aged children found significantly stronger physiological arousal when children viewed their favorite branded food and beverage products compared to unbranded alternatives, even for products with identical nutritional content. These elevated SCR patterns correlated with both increased self-reported desire to consume and improved brand recognition, suggesting that food branding triggers subconscious physiological responses that may influence consumption behavior. The effect was particularly notable for energy-dense, nutrient-poor products, highlighting how marketing may unconsciously promote less healthy food choices. By combining psychophysiological measures with behavioral and self-report data, this study provides compelling evidence that children's physiological responses to food branding precede and potentially shape their conscious preferences, with important implications for policies regulating food marketing to children.

D. Food Packaging

Moya et al [20] examine how neuromarketing methods such as EEG, galvanic skin response (GSR), and eye-tracking have been applied to the study of food packaging, with the aim of understanding consumer attention, emotional arousal, and memory. Their review highlights that while these techniques are increasingly used in both academic and commercial contexts, concerns remain about the validity, reliability, and standardization of measures in industry practice. The authors argue that neuromarketing can meaningfully complement traditional self-report approaches, offering deeper insights into unconscious and automatic consumer reactions to packaging design. At the same time, they stress the need for a stronger scientific framework to ensure methodological rigor and to translate physiological data into actionable strategies for food marketers. Vila-López et al [21] investigated consumer responses to low-fat food packaging by comparing reactions in virtual and physical shopping environments. Using both physiological measures (electrodermal activity, heart rate, and facial muscle activity) and self-report evaluations, they tested six packaging designs that varied in color and messaging. Their findings showed that low-order emotional responses captured through physiological indicators were consistent across both contexts and strongly predicted actual product choice, whereas high-order cognitive evaluations did not align with purchase behavior. This suggests that implicit physiological measures provide a more reliable reflection of

consumer decision-making processes in packaging research than conscious, self-reported judgments.

III. Conclusion

Overall, the reviewed literature demonstrates that electrodermal activity (EDA/GSR) and related physiological measures provide a valuable complement to traditional self-report methods in food consumer science, offering time-sensitive and objective insights into emotional arousal and motivation. Across applications ranging from taste evaluation [14, 17] and beverage testing [7] to brand perception [15, 19], pricing strategies [18], and packaging design [20-22], GSR consistently captures implicit reactions that often diverge from explicit evaluations. These divergences highlight its capacity to uncover subconscious affective processes—whether positive arousal in response to branding, negative tension in relation to price, or nuanced reactions to flavor intensity—that self-reports alone may overlook. At the same time, methodological challenges such as signal variability, latency, and the inability to capture valence underscore the need to combine GSR with complementary measures such as EEG, heart rate, or facial EMG, and to employ standardized protocols for data collection and analysis. Taken together, this body of work positions GSR as a robust tool for advancing sensory and consumer research, enabling a richer understanding of how emotions, motivations, and physiological arousal interact to shape food-related experiences and decision-making.

References

[1] Spence, C. (2017). *Gastrophysics: The new science of eating*. Viking.
[2] Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology, 22*(3), 332–351.
[3] Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer.
[4] Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Journal of Neuroscience Methods, 190*(1), 80–91.
[5] Tamura, T., Maeda, Y., Sekine, M., & Yoshida, M. (2014). Wearable photoplethysmographic sensors — Past and present. *Electronics*, 3(2), 282–302.
[6] Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement, 28*(3), R1–R39.
[7] Park, J., Kim, S., Lee, K., & Kim, H. (2022). Photoplethysmogram analysis and applications: A review. *Sensors, 22*(17), 6587.
[8] Larrañaga-Ayastuy, E., Soroa, G., Larrayoz, I., Arana, M., & Iriarte, Á. (2023). Electrodermal response and its relationship with explicit response in controlled and real contexts: A case study with different beer styles. *Journal of Sensory Studies, 38*(6), e12799.
[9] Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed., pp. 217–243). Cambridge University Press.
[10] Zangróniz, R., Martínez-Rodrigo, A., López, M. T., Pastor, J. M., & Fernández-Caballero, A. (2017). Electrodermal activity sensor for classification of calm/distress condition. *Sensors, 17*(3), 1–13.
[11] Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011). The skin conductance response, anticipation, and decision-making. *Journal of Neuroscience, 31*(21), 7812–7818.
[12] Shimmer. (2018). *Shimmer3 GSR+ user guide*. Shimmer Sensing.
[13] Quigley, K. S., Lindquist, M. A., & Barrett, L. F. (2014). Inducing and measuring emotion and affect: Tips, tricks, and secrets. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 220–252). Cambridge Unv Press.
[14] Caruelle, D., Gustafsson, A., Shams, P., & Lervik-Olsen, L. (2019). The use of electrodermal activity (EDA) measurement to understand consumer emotions – A literature review and a call for action. *Journal of Business Research, 104*, 146–160.
[15] Spinelli, S., Ares, G., Antúnez, L., & Monteleone, E. (2023). Skin conductance responses to tastes: Effects of quality, intensity and novelty. *Food Quality and Preference, 106*, 104807.
[16] Rita, P., Guerreiro, J., & Omarji, A. (2021). Unveiling consumer emotions towards private labels: An electrodermal activity perspective. *Journal of Consumer Behaviour, 20*(6), 1419–1434.
[17] Álvarez-Pato, V. M., Sánchez, C. N., Domínguez-Soberanes, J., Méndoza-Pérez, D. E., & Velázquez, R. (2020). A multisensor data fusion approach for predicting consumer acceptance of food products. *Foods, 9*(6), 774.
[18] Martínez-Levy, A. C., Moneta, E., Rossi, D., Trettel, A., Peparaio, M., Civitelli, E. S., di Flumeri, G., Cherubino, P., Babiloni, F., & Sinesio, F. (2021). Taste responses to chocolate pudding with different sucrose concentrations through physiological and explicit self-reported measures. *Foods, 10*(7), 1527.
[19] Somervuori, O., & Ravaja, N. (2013). Purchase behavior and psychophysiological responses to different price levels. *Psychology & Marketing, 30*(6), 479–489.
[20] Smith, R., Kelly, B., Yeatman, H., & Boyland, E. (2019). Food marketing influences children's attitudes, preferences and consumption: A systematic critical review. *International Journal of Environmental Research and Public Health, 16*(11), 2059.
[21] Moya, I., García-Madariaga, J., & Blasco, M. F. (2020). What can neuromarketing tell us about food packaging? *Foods, 9*(12), 1856.
[22] Vila-López, N., Kuster-Boluda, I., & Alacreu-Crespo, A. (2021). Designing a low-fat food packaging: Comparing consumers' responses in virtual and physical shopping environments. *Foods, 10*(2), 211.
[23] Liao, L. X., Corsi, A. M., Chrysochou, P., & Lockshin, L. (2015). Emotional responses towards food packaging: A joint application of self-report and physiological measures of emotion. *Food Quality and Preference, 42*, 48–55.

Mahremiyet Odaklı Mobil Acil Müdahale Sistemi Tasarımı

Privacy-Aware Mobile Emergency Response System Design

Melahat KIZIL
Sağlık Hizmetleri Meslek Yüksekokulu
İzmir Ekonomi Üniversitesi
İzmir, Türkiye
melahat.kizil@gmail.com

Zeynep ŞİŞLİ
Hukuk Fakültesi, Hukuk Bölümü
İzmir Ekonomi Üniversitesi
İzmir, Türkiye
zeynep.sisli@izmirekonomi.edu.tr

Çınar GEDİZLİOĞLU ve Kutluhan EROL
Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü
İzmir Ekonomi Üniversitesi
İzmir, Türkiye
cinar.gedizlioglu@izmirekonomi.edu.tr, kutluhan.erol@gmail.com

***Özetçe*—Gelişen mobil sağlık teknolojileri, acil müdahale süreçlerinde hız ve doğruluk sağlarken, hasta mahremiyetinin korunmasında etik kuralların ihlali ve hukuki sorunlara sebep olma tehlikesi doğurmaktadır. Bu çalışmada, "Mahremiyet Kalkanı" adı verilen, yapay zekâ destekli anonimleştirme ve katmanlı veri erişim tabanlı denetlenebilirlik özelliklerini bir arada sunan entegre bir mobil acil müdahale sistemi tasarımı sunulmaktadır. Sistem, müdahale süresini etkilemeden kişisel verilerin korunmasını hedeflemekte; paramedik, doktor ve diğer sağlık personelinin görev tanımlarına uygun erişim yetkisi ile çalışmaktadır. Tasarım, literatür taraması, Acil Sağlık Hizmetlerinde çalışan sağlık personeli ve paramedik öğrencilerinden elde edilen geri bildirimlerle şekillendirilmiştir. Sistemin acil sağlık hizmetlerinde hem etkin müdahale hem de güçlü mahremiyet koruması sağlayarak, mevcut teknolojilere kıyasla önemli etik ve operasyonel avantajlar sunacağı düşünülmektedir.**

Anahtar Kelimeler — Mobil sağlık sistemleri, acil sağlık hizmetleri, veri mahremiyeti, yapay zekâ, katmanlı erişim kontrolü

***Abstract* — Although mobile health technologies enhance the speed and accuracy of emergency response, they also risk violating ethical standards and raising legal issues concerning patient privacy. This study introduces** Privacy Shield**, an integrated mobile emergency response system that incorporates AI-based anonymization, layered data access, and traceability features. The system is designed to safeguard personal data without compromising response time, granting access rights based on the specific roles of paramedics, physicians, and other healthcare professionals. Its development was informed by a literature review as well as feedback from Emergency Medical Services staff and paramedic students. By combining ethical safeguards with operational efficiency, Privacy Shield is expected to offer clear advantages over existing technologies, ensuring both effective emergency care and robust privacy protection.**

Keywords — Mobile health systems, emergency health services, data privacy, artificial intelligence, layered access control.

I. GİRİŞ

Teknolojik gelişmeler tıbbın her alanı gibi acil sağlık hizmetleri kapsamındaki hasta yönetimi ve müdahale süreçlerini dönüştürmektedir. Özellikle hastane öncesi acil müdahalede kullanılan mobil sağlık uygulamaları, giyilebilir sensörler ve tele-tıp sistemleri, sağlık profesyonellerinin karar alma süreçlerini destekleyerek hayat kurtarıcı sonuçlar

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

doğurmaktadır. Ancak bu teknolojilerin yoğun kullanımı, hasta verilerinin toplanması, işlenmesi ve iletilmesi sırasında hasta mahremiyeti ve veri güvenliği konularında gerekli tedbirler alınmadığı takdirde ciddi etik ihlallere yol açma ve hukuki sorunlara sebep olma riski taşımaktadır.

Avrupa Birliği'nin Genel Veri Koruma Yönetmeliği (GDPR), Biyoloji ve Tıbbın Uygulanması Bakımından İnsan Hakları ve İnsan Haysiyetinin Korunması Sözleşmesi: İnsan Hakları ve Biyotıp Sözleşmesi, Türkiye Cumhuriyet Anayasası'nın "Özel Hayatın Gizliliği" başlıklı 20. maddesi, Kişisel Verileri Koruma Kanunu (KVKK), Hasta Hakları Yönetmeliği ve diğer ilgili mevzuat kapsamında, kişinin mahremiyet hakkı çerçevesinde "hassas kişisel veri" niteliğindeki sağlık verilerinin korunmasına ilişkin düzenlemelerin, acil müdahale ortamında uygulanmasını kolaylaştıracak sistemlere ihtiyaç duyulmaktadır.

Sağlık çalışanlarının, görüntüleme sistemleri ve anlık veri iletim araçlarını kullanırken hastanın kimliğini, fiziksel mahremiyetini ve kişisel verilerini korumaları her zaman mümkün olmamaktadır. Acil durumların dinamik ve karmaşık yapısı, sağlık çalışanlarının hasta gizliliği ile müdahale gereklilikleri arasında hassas bir denge kurmasını zorlaştırmaktadır. Bu bağlamda, hem müdahale hızını düşürmeyecek hem de hasta haklarını ihlal etmeyecek teknolojik tasarımların geliştirilmesi gerekmektedir.

Bu makale, söz konusu etik ve hukuki ikilemleri ele alarak, "Mahremiyet Kalkanı" adlı entegre bir sistemin tasarımını sunmaktadır. Bu sistem, yapay zekâ destekli anonimleştirme, katmanlı erişim mekanizmaları tabanlı denetlenebilirlik gibi yenilikçi bileşenlerle, hem etkin müdahale hızını korumayı hem de hasta haklarını güçlendirmeyi amaçlamaktadır. Tasarımımız, paramedik ve sağlık profesyonellerinin saha operasyonlarında etik kaygılardan arınmış, hesap verebilir ve güvenli bir şekilde çalışmalarına olanak tanıyacak bir çözüm sunmayı amaçlamaktadır

II. LİTERATÜR TARAMASI

Sağlık hizmetlerinde gelişen otomasyon ile birlikte giderek artan yapay zeka uygulamaları ile ilgili olarak bu teknolojinin doğası gereği sağlık hizmetlerine uygunluğu tartışılmakta, güvenlik ve verimlilik açısından gizlilik, etik sorunlar ve sorumluluk endişeleriyle bağlantılı hukuki çerçevenin yeterliliği sorgulanmaktadır[1]. Yapay zeka teknolojisi kullanımının, benzer şekilde üzerinde durulması gereken hukuki sorumluluk, mahremiyet ve gizlilik gibi etik sorunlar, çalışanların iş kaybı ve eğitim ihtiyaçları gibi zorlukları bulunmakla birlikte, tanı ve tedavi sırasında klinisyenlerin karar alma süreçlerine destek olma ve sağlık hizmetlerinin yönetiminde operasyonel etkinlik ve maliyetleri düşürme gibi gibi bir çok açıdan umut verici bir potansiyel gösterdiği bir çok araştırma ve gerçek hayat deneyimlerine dayalı olarak ifade edilmektedir [2]. Genel olarak tıbbi otomasyon, özel olarak artan yapay zeka uygulamaları ile, hastanın temel kişilik haklarının güvence altına alınmasını amaçlayan tıbbi etik ve hukuki sorumluluk açısından uygulamada karşılaşılabilecek sorunların her biri hakkında, sağlık hizmetinin insan yaşamı ile yakından ilgili niteliği gereği düşünülmesi ve çözüm yollarının araştırılması gerekmektedir.

Mahremiyet ve veri güvenliği etik ikilemler kapsamında tartışılan konu başlıklarından ikisidir[3]. Yunanca "kişilik, karakter" anlamına gelen "ethos" sözcüğünden türeyen "etik" sözcüğünün amacı, hizmet verilen kişilerin temel haklarını güvence altına almaktır. Etik/ahlak felsefesi, doğru davranışlarda bulunmak, iyi bir insan olmak ve insani değerler hakkında düşünme pratiğidir. Tıp alanında geçerli biyoetik ise; uygulayıcılar (sağlık hizmeti sağlayıcılar), hastalar ve toplum arasındaki ilişkilerle ilgilenir, karşılıklı hak ve sorumlulukları insan haklarına uygun düzenlemeyi amaçlar. Biyoetik ile hukuk arasında sıkı bir ilişki bulunmaktadır[4]. Sağlık hukukunda "hasta hakları", biyoetiğin "hak" formatında ifadesidir denebilir. Hukukta sağlık hizmeti sağlayıcıların sorumluluklarının temelini biyoetik oluşturur. Bu durum, hukuk ve biyoetik tarafından "hasta hakları" ve "ölüm hakkı" gibi ifadelerde olduğu gibi "haklar" temelli ortak terimlerin kullanımından anlaşılabilir[5].

Otomasyon ile toplanan kişisel sağlık verilerinin işlenmesi, saklanması ve korunması, kişinin yaşam hakkı kapsamında özerkliğine saygı temel ilkesinin ve bu bağlamda mahremiyet hakkının gereği olmakla hasta güvenliğinin önemli bir parçasıdır. Avrupa Birliği'nin GDPR ve Türkiye'nin KVKK düzenlemeleri, sağlık verilerinin korunmasına yönelik çerçeveler sunsa da, olay yerindeki acil karar anlarında bu düzenlemelerin uygulanması sınırlı kalmaktadır. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü Kişisel Sağlık Verileri sayfasında yer alan mahkeme karar özetlerinde görüldüğü üzere sağlık hizmetleri ile ilgili birçok mahremiyet ihlali hukuki uyuşmazlık konusu olmuş[6], bu konuda Türkiye Cumhuriyeti Anayasası 20.maddesi ile herkesin özel hayatına ve aile hayatına saygı gösterilmesini isteme hakkı, Hasta Hakları Yönetmeliği 21.maddesi ile hastanın mahremiyetine saygı gösterilmesi esas ve zorunluluğu, Yataklı Sağlık Tesislerinde Acil Servis Hizmetlerinin Uygulama Usul ve Esasları 2.maddesi ile her türlü tıbbi işlem sırasında hasta mahremiyeti, tıbbi etik kurallar ve hasta hakları ilkelerine uygun davranılması kuralı ve diğer düzenlemelere atıfla Sağlık Bakanlığı tarafından 2016/10 sayılı Genelgesi ve ilgili diğer mevzuat çıkarılmıştır[7].

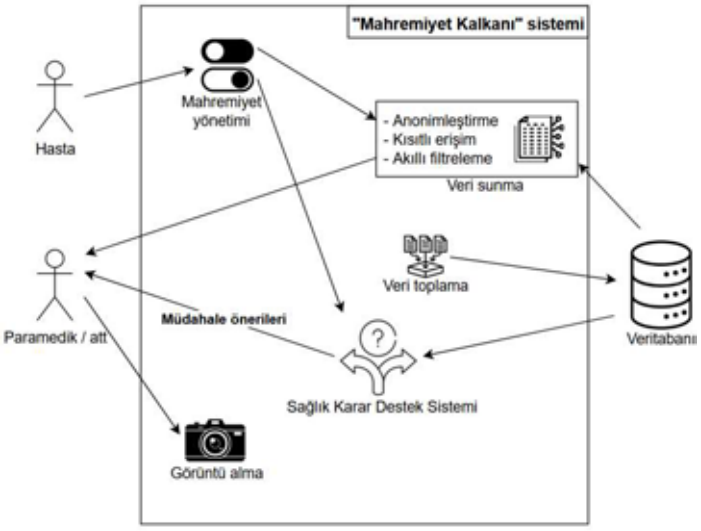
Hasta güvenliği kapsamında sağlık verilerinin gizliliği ve mahremiyet hakkı, genel olarak sağlık çalışanları tarafından bilinmekte ve kabul edilmekte, ancak uygulamada hayata geçirilmesi açısından gerek Türkiye, gerek Finlandiya'da yapılan çalışmaların gösterdiği gibi farklı bakış açıları ve öncelikli haklar konusunda farklılıklar söz konusu olabilmektedir [8][9][10]. Bu çalışmalar ile varılan ortak nokta eğitim ve araştırma gerekliliği olmakla birlikte, bu çalışmada kurulması önerildiği üzere Özata ve Özer'in çalışması sonuç kısmında "sağlık sistemi üzerinde kayıtlı veriler için mahremiyeti sağlayacak filtreleme işlemlerinin hayata geçirilmesi" nin gereğine vurgu yapıldığı dikkat çekmektedir[10].

III. YÖNTEM

Bu çalışma, teknoloji tasarımı ve etik denge yaklaşımıyla karma yöntemli bir yaklaşıma sahiptir. Literatürdeki hasta mahremiyet sorunları incelenmiş, 112 Acil Sağlık Hizmetlerinde staj yapan paramedik öğrencilerinin geri bildirimlerine başvurulmuş, aynı zamanda burada çalışan

paramedik ve acil tıp teknisyenlerinden fikir alınmış ve görüşler doğrultusunda prototip sistemin bileşenleri geliştirilmiştir.

Bu prototip sistem bir mobil uygulama çerçevesinde düşünülmekle beraber, bu sistem etrafında tasarlanmış bir donanım da gelecekte düşündüğümüz projelerdendir. Sistemin ve bileşenlerinin çalışma prensipleri aşağıdaki şekilde görülebilir:



Şekil. 1. Mahremiyet Kalkanı sistemi tasarımı

IV. "MAHREMİYET KALKANI" SİSTEM TASARIMI

A. *Akıllı Filtreleme*

Tasarladığımız Mobil Acil Müdahale Sistemi, öncelikle yapay zekâ destekli görüntü işleme teknolojileri ile donatılmıştır. Bu teknoloji sayesinde, acil müdahale sırasında kaydedilen görüntülerdeki hasta yüzü ve cinsel bölgeler otomatik olarak bulanıklaştırılmaktadır. Aynı zamanda kritik olarak öngörülen alanların netliğinin artırılması da yine sistemin işlevlerindendir. Böylece, hastanın kimliği ve mahremiyeti korunurken, sağlık personelinin müdahale için ihtiyaç duyduğu kritik bilgiler net şekilde görünür kılınmaktadır. Bu sayede hem etik kurallara uyum sağlanmakta hem de hasta mahremiyeti üst düzeyde korunmaktadır.

B. *Katmanlı Veri Erişim Sistemi*

Sistemin en önemli bileşenlerinden biri olan katmanlı veri erişim sistemi, farklı sağlık personeli rollerine göre değişen erişim yetkileri tanımlamaktadır. Paramedikler, acil durumlarda yalnızca hastanın yaşamsal belirtilerine ulaşabilirken, uzman doktorlar hasta onayı ile birlikte daha detaylı tıbbi verilere erişebilmektedir. Bu erişim süreçleri şifrelenmiş bir şekilde kayıt altına alınmakta ve bu sayede veri erişim geçmişi tam anlamıyla denetlenebilir hale gelmektedir. Böylece, yetkisiz erişimlerin önüne geçilirken, sağlık çalışanlarının hesap verebilirliği de sağlanmaktadır. Sistem kayıtlı verilere erişimi, erişenin rolüne, görevine, amacına uygun şekilde redakte ederek sağlamaktadır.

C. *Anlık Mahremiyet Modu*

Bu mod, hastanın bilinç durumuna göre aktif hale getirilen bir özelliktir. Eğer hasta bilinci açıksa, sağlık çalışanları cihaz ekranı üzerinden doğrudan hastanın "mahremiyet onayı"nı alabilmektedir. Ayrıca, hastanın önceden e-Nabız sistemi veya TC kimlik kartına entegre edilmiş mahremiyet tercihleri de sisteme tanımlanabilir. Bu mekanizma, hastanın kişisel veri paylaşımı üzerinde daha fazla kontrol sahibi olmasını ve tercih ettiği gizlilik seviyesine uygun müdahalenin yapılmasını mümkün kılmaktadır.

D. *Zaman Kısıtlı Görüntü ve Veri Paylaşımı*

Veri paylaşım sürecinde, sistem tarafından belirlenen 30 dakikalık şifreli erişim penceresi bulunmakta olup, bu sürenin sonunda paylaşılan görüntü ve veriler otomatik olarak anonimleştirilmekte veya tamamen silinmektedir. Bu zaman kısıtlaması, verilerin gereksiz yere uzun süre erişilebilir kalmasının önüne geçerken, hasta verilerinin güvenliğinin sürekli olarak korunmasına olanak tanımaktadır.

E. *Yapay Zekâ Destekli Karar Modülü*

Son olarak, yapay zekâ destekli karar modülü, hastanın kimlik bilgilerine erişmeden, semptomlar üzerinden analiz yapabilmekte ve sağlık personeline hızlı, doğru ve hasta mahremiyetine saygılı öneriler sunmaktadır. Bu modül, acil müdahalelerde hem etkinliği artırmakta hem de hasta bilgilerini koruma noktasında kritik bir rol üstlenmektedir.

V. TARTIŞMA

Bu çalışma, mobil sağlık teknolojilerinin sunduğu imkanlarla birlikte ortaya çıkan etik ve hukuki zorluklara çözüm üretme çabası çerçevesinde değerlendirilebilir. Geliştirilen sistem, acil sağlık müdahaleleri gibi zamanla yarışılan kritik süreçlerde, mahremiyetin ihlal edilmeden de etkin bir müdahalenin mümkün olduğunu göstermektedir. Böylece uzun süredir karşı karşıya gelmiş iki önemli değer olan hızlı müdahale ve veri gizliliği, teknolojik bir entegrasyon sayesinde birlikte ele alınabilmiştir.

Acil sağlık hizmetlerinin doğası gereği müdahale süreleri hayati önem taşımaktadır. Bu bağlamda, sistemin tüm bileşenlerinin müdahale hızını olumsuz yönde etkilemeden çalışması, hasta güvenliği ile birlikte zaman yönetimi açısından da önemli bir kazanımdır. Bu bağlamda, yapay zekâ destekli modüllerin zaman yönetimi açısından da ek kısıtları olması öngörülmektedir.

"Mahremiyet Kalkanı" mahremiyet ile müdahale arasındaki dengeyi yeniden tanımlayan bir uygulama olarak konumlanmaktadır. Bununla birlikte, sistemin uygulanabilirliği ve kabulü açısından hukuki altyapı ve eğitim desteği hayati öneme sahiptir. Paramediklerin teknolojiyi etik çerçevede kullanabilmeleri için hizmet içi eğitimlerle desteklenmesi gerekir.

Mahremiyet teknolojileri ne kadar gelişmiş olursa olsun, sistemin çalışmasını izleyen, denetleyen ve gerektiğinde müdahale eden etik gözetim mekanizmalarının kurulması önemlidir. Bu yapılar, kullanıcıların sisteme ne zaman, nasıl, hangi amaçla eriştiğini denetlemeli ve olası suistimalleri önlemek için şeffaf raporlama sistemleriyle desteklenmelidir.

Etik gözetim, yalnızca teknolojik denetimi değil, aynı zamanda kurumsal sorumluluğu da kapsamalıdır.

Bu tür teknolojik sistemlerin, özellikle önerdiğimiz sistem gibi prototip aşamasında bulunan sistemlerin, gerçek yaşam koşullarında nasıl işlediğini ve farklı acil müdahale senaryolarında ne ölçüde etkili olduğunu değerlendirebilmek için kapsamlı saha çalışmaları büyük önem taşımaktadır. Farklı kullanıcı profilleri, çevresel koşullar ve müdahale dinamikleri dikkate alınarak yapılacak uygulamalı testler, sistemin işlevselliği ve sürdürülebilirliği konusunda daha sağlam veriler sunacaktır.

VI. SONUÇ VE ÖNERİLER

"Mahremiyet Kalkanı" hasta mahremiyetini ihlal etmeden acil müdahale sürecinde kullanılabilecek yenilikçi bir teknolojidir. Bu sistem sayesinde sağlık çalışanları daha güvenli, hesap verebilir ve etik bir müdahale süreci yürütebilir. Öncü Mahremiyet Kalkanı sistemi, acil sağlık hizmetlerinde mahremiyet ve müdahale etkinliği arasındaki geleneksel ikilemi çözmeye yönelik bütüncül bir yaklaşım sunmaktadır. Bu sistemin en önemli katkısı, hasta mahremiyetini pasif bir yasal gereklilik olmaktan çıkarıp, müdahale sürecinin aktif ve teknolojik olarak desteklenen bir parçası haline getirmesidir. Yapay zekâ destekli bulanıklaştırma modülü, sağlık profesyonellerinin müdahale için kritik bilgilere erişimini sağlarken, hastanın kişisel tanımlayıcı bilgilerini otomatik olarak gizleyerek etik ihlal riskini en aza indirmektedir.

Katmanlı veri erişim sistemi, "en az ayrıcalık ilkesi" (least privilege principle) doğrultusunda, her profesyonele yalnızca görevi için zorunlu olan verilere erişim yetkisi tanıyarak veri güvenliğini artırmaktadır. Bu yapı, gereksiz veri paylaşımını ve kötüye kullanımını önlerken, aynı zamanda her erişimin şeffaf ve değiştirilemez bir dijital iz bırakmasını sağlamaktadır. Bu özellik, olası bir etik veya hukuki problem çıkması durumunda hesap verebilirliği en üst düzeye çıkarmaktadır.

"Mahremiyet Kalkanı", her ne kadar bu yazıda mobil uygulama çerçevesinde ele alınsa da, gelecekte özel tasarlanmış bir donanım yardımıyla işlevlenecektir. Bu donanımın görüntü işleme, kontrollü veri erişimi ve yapay zeka karar modülü gibi unsurları içinde barındırması, dolayısıyla önerdiğimiz sistemi tam haliyle yansıtması öngörülmektedir. Yapacağımız bu çalışma sayesinde otomasyonu arttıracak, insan hatasını en aza indirecek ve bir yandan bahsetmiş olduğumuz etik ikilemlere de çözümler sunacağız.

Aynı zamanda, bu sistemin kullanım rahatlığını ve verimliliğini artırmak adına bir sanal yardımcı eklemeyi planlıyoruz. Bu sanal yardımcı, günümüzde insanla birebir etkileşimde oldukça başarılı olduğu bilinen doğal dil modellerinin sisteme uygun entegrasyonuyla gerçekleştirilecektir. Yazı ve ses tabanlı çalışan versiyonları mevcut olacaktır.

Sistemin uygulanabilirliği ve başarısı, teknik tasarımın ötesinde, birtakım kritik faktörlere bağlıdır. Öncelikle, sağlık profesyonellerinin bu yeni teknolojiye olan kabulü ve entegrasyonu büyük önem taşımaktadır. Paramediklerin bu sistemi doğru ve etik bir şekilde kullanabilmeleri için, yoğun ve sürekli hizmet içi eğitim programlarına ihtiyaç duyulacaktır. İkincil olarak, sistemin yasal altyapı ile tam uyumluluğu sağlanmalıdır. GDPR ve KVKK gibi düzenlemelerin acil durum senaryolarına özel olarak nasıl yorumlanacağı ve uygulanacağı konusunda net yönergeler oluşturulması gerekmektedir. Son olarak, sistemin donanım ve yazılım maliyetleri ile operasyonel sürdürülebilirliği, kamu ve özel sağlık kuruluşları tarafından dikkate alınmalıdır. Gelecekte, sistemin tele-tıp, afet yönetimi ve evde bakım gibi alanlara genişletilmesi, sağlık hizmetlerinin dijital dönüşümünde önemli bir referans modeli oluşturabilir.

KAYNAKLAR

[1] Pashkov VM, Harkusha AO, Harkusha YO. ARTIFICIAL INTELLIGENCE IN MEDICAL PRACTICE: REGULATIVE ISSUES AND PERSPECTIVES. Wiad Lek. 2020;73(12 cz 2):2722-2727. PMID: 33611272.

[2] Lee D, Yoon SN. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. Int J Environ Res Public Health. 2021 Jan 1;18(1):271. doi: 10.3390/ijerph18010271. PMID: 33401373; PMCID: PMC7795119.)

[3] Farhud DD, Zokaei S. Ethical Issues of Artificial Intelligence in Medicine and Healthcare. Iran J Public Health. 2021 Nov;50(11):i-v. doi: 10.18502/ijph.v50i11.7600. PMID: 35223619; PMCID: PMC8826344.

[4] (Iserson K.V. Bioethics. In: Marx J.A., Hockberger R.S., Walls R.M., editors. Rosen's Emergency Medicine: Concepts and Clinical Practice. 5th edition. C.V. Mosby; St. Louis (MO): 2002. pp. 2725–2733.).

[5] Iserson KV. Ethical principles--emergency medicine. Emerg Med Clin North Am. 2006 Aug;24(3):513-45. doi: 10.1016/j.emc.2006.05.007. PMID: 16877128; PMCID: PMC7127201.

[6] Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü Kişisel Sağlık Verileri, Karar Özetleri.https://kisiselveri.saglik.gov.tr/TR-55834/karar-ozetleri.html , erişim tarihi 08.08.2025.

[7] Sağlık Bakanlığı Hasta Mahremiyetine Saygı Gösterilmesi Genelgesi, 2016/10.https://dosyamerkez.saglik.gov.tr/Eklenti/4283/0/hasta-mahremiyetine-saygi-gosterilmesi-2016-10pdf.pdf?_tag1=96E870A967D2F4DBCBA0A86DF8607E4E533E467F, erişim tarihi 08.08.2025.

[8] Koskimies EM, Koskenniemi J, Leino-Kilpi H. Patient's informational privacy in prehospital emergency care: Paramedics' perspective. Nurs Ethics. 2020 Feb;27(1):53-66. doi: 10.1177/0969733019834977. Epub 2019 Apr 10. PMID: 30971182.

[9] Bilir O. Acil Servis Çalışanlarının Hasta Haklarına Bakışı. Konuralp Medical Journal. April 2015;7(1):28-33. doi:10.18521/ktd.22718

[10] Özata M, Özer K Sağlık Çalışanlarının Hasta Mahremiyeti Konusundaki Tutumlarının İncelenmesi.Hacettepe Sağlık İdaresi Dergisi, 2017; 20(1):1-21.Rudnicky, A. I., Polifroni, Thayer, E H., and Brennan, R. A. "Interactive problem solving with speech", *J. Acoust. Soc. Amer., Vol. 84, 1988, p S213(A).*

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

AI Assisted Polyp Detection and Segmentation in Colonoscopy

Değer Ayata, PhD
Artificial Intelligence and Computer Vision Department
AIATUS AI
Istanbul, Türkiye
deger.ayata@aiatus.com

***Abstract*—This study presents a two-stage deep learning pipeline for real-time polyp localization and segmentation in endoscopic imagery. The system integrates a YOLO detector for rapid, frame-level candidate identification and a U-Net segmenter with a ResNet-50 encoder for pixel-accurate mask generation. Evaluation on a held-out test set demonstrated strong performance: the detection model achieved 92.3% precision, 86.0% recall, and a mean Average Precision (mAP@0.5) of 94.8%, while the segmenter reached a Dice score of 0.936 and IoU of 0.672. Model selection during training was guided by background-ignored IoU@0.5, contributing to a decoupling between Dice and IoU metrics due to their differing sensitivity and aggregation methods. Both components individually satisfied real-time throughput requirements on the target GPU.**

Keywords—Polpy detection; image segmentation; colonoscopy; medical image processing;plyp detection; object detection

I. Introduction

Colorectal cancer remains a major global health concern, with nearly 1.9 million new cases reported each year [1]. Since the majority of CRC cases originate from adenomatous polyps, early detection and removal during colonoscopy are key to reducing both incidence and mortality. However, studies have shown that even experienced clinicians may miss 17–28% of polyps during routine procedures [2,3]. Missing polyps—especially small or flat ones—can be a serious issue, as they may later turn into cancer, reducing the effectiveness of screening efforts [4]. This makes it clear that there's an urgent need for supportive technologies that can help doctors detect polyps more consistently and accurately.

In recent years, AI has become a valuable tool for helping doctors detect polyps during colonoscopies [1]. Modern computer-aided detection systems can now analyze colonoscopy footage in real time, flagging polyps that might be missed by the human eye [2,4]. Deep learning model and computer vision techniques have shown impressive accuracy in medical imaging tasks, including both polyp detection and segmentation [5]. For instance, real-time object detectors such as those in the YOLO family have proven effective at spotting polyps with high precision during live procedures. AI's potential here is significant—it can help reduce miss rates and improve consistency across different clinicians, ultimately making colonoscopy screenings safer and more reliable. Still, building accurate detection models is no easy task. Polyps vary greatly in appearance—some are small, flat, and hard to distinguish from healthy tissue [4]. Challenging visual conditions like folds, glare, or fluid in the colon further complicate detection [4]. To be practical in real-world settings, models also need to be computationally efficient, ideally running at 25 frames per second or more to keep up with live video.

To address this challenge, real-time AI-based computer vision systems are being developed to assist physicians during colonoscopies. These systems can process live colonoscopy camera feeds and provide consistent detection support without fatigue or variability.

II. Literature Review

Early polyp detection approaches applied conventional computer vision and older machine learning techniques (e.g. hand-crafted features, classic CNNs) [7]. In recent years, the field has shifted almost entirely to deep learning object detectors for this task. Two main families of object detection models have been explored: (1) two-stage detectors like Faster R-CNN (first generate region proposals, then classify) and (2) one-stage detectors like YOLO and SSD (which directly predict boxes in one pass) [1]. Given the real-time requirements of colonoscopy, one-stage models (particularly the YOLO series) have become the dominant choice due to their speed advantage, while still achieving high accuracy.

Recent advances in computer-aided detection have focused on fast, single-stage detectors that maintain the frame rates needed for clinical use while enhancing the ability to detect small lesions. Reviews and benchmarking studies highlight YOLO (You Only Look Once) style models as practical solutions well-suited for integration into endoscopy procedures [4,6,7]. For example, variant of YOLO with improvements for small target detection and attention mechanisms—has shown increased average precision on public datasets without sacrificing efficiency required for real-time deployment. Besides YOLO, Hyb-SSDNet offers an updated version of the SSD (Single Shot Detector) framework, incorporating lightweight backbone networks and multi-scale feature fusion to deliver competitive performance when speed is essential [5]. Traditional end-to-end approaches have also proven that high frame-per-second (FPS)

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

processing pipelines can successfully handle entire colonoscopy videos [3].

Another key task alongside polyp detection is polyp segmentation, which involves labeling every pixel in an image as either part of the polyp or the surrounding tissue, resulting in a clear binary mask highlighting the polyp area. Accurate segmentation is especially important because it helps determine the size of the polyp, defines precise margins for removal, and can assist in identifying the polyp type based on its shape and texture. However, segmenting polyps is a complex challenge due to their diverse appearances and often indistinct or fuzzy edges. These unclear boundaries arise from factors such as uneven lighting and the smooth transition between the polyp and the surrounding mucosa. On the segmentation front, UNet variants achieve high Dice scores for polyp mask accuracy [8, 9].

III. Scope and Contribution

In this study a pragmatic, two-module setup was targeted for single-workstation clinical use 1) computer-aided detection (CAD) and computer-aided segmentation (CAS). CAD module is a fast detector localizes candidates every frame, and a precise segmenter (CAS) refines boundaries on demand. Public, reproducible datasets and mainstream architectures (YOLO variants, U-Net variants) were emphasized due to adoptability and prior success in colonoscopy. The contributions consist of a compact real-time inference pipeline with measured latencies, a clear training recipe with reproducible open dataset splits, and a concise, project-relevant literature recap positioning the design.

IV. Methods

A. Colonoscopy Data Source and Dataset

Colonoscopy images are collected as part of standard gastrointestinal diagnostic procedures. During the exam, a thin, flexible instrument known as a colonoscope—which includes a high-definition camera—is guided through the rectum to visually examine the colon's interior. As the device records live video, clinicians can pause the feed to capture individual frames that show suspicious regions, such as potential polyps. These selected images are then annotated by specialists, who outline abnormal areas for use in medical research. This expert-labeled image data plays a vital role in training and validating AI systems for tasks like polyp detection and segmentation.

For all experiments in this study, the Kvasir-SEG dataset was used exclusively. This dataset includes 1,000 colonoscopy images, each paired with a manually verified binary mask that outlines the polyp region [9]. These images were collected during routine diagnostic colonoscopy procedures, where a flexible endoscope equipped with a high-resolution video camera captures real-time footage of the colon's interior. From this video stream, still frames containing visible polyps are extracted and annotated by medical experts to serve as training data for computer vision models. The images in Kvasir-SEG vary in resolution, ranging from approximately 332×487 up to 1920×1072, and each image comes with exactly one corresponding mask, making it well-suited for supervised segmentation tasks. To structure the dataset for training and evaluation, a deterministic 80/10/10 split was applied, resulting in 800 images for training, 100 for validation, and 100 for testing.

B. Polyp Detection

Polyp detection refers to the task of identifying and localizing polyps within endoscopic images, typically by drawing bounding boxes around the regions of interest. This process plays a critical role in assisting clinicians by automatically highlighting potential polyps that may require closer examination. In recent years, object detection models—particularly those in the YOLO (You Only Look Once) family—have become widely used for this task due to their balance of speed and accuracy. YOLO models are designed for real-time detection, making them highly suitable for live colonoscopy procedures. They work by dividing an image into grids and predicting both the bounding boxes and class probabilities for each region in a single forward pass. In this study, various YOLO variants were employed for polyp detection, leveraging their proven efficiency in medical imaging applications.

The effectiveness of YOLO models in polyp detection is typically assessed using several well-established evaluation metrics. Precision reflects the proportion of correctly identified polyps among all predicted instances, indicating how often the model is right when it makes a detection. Recall, on the other hand, measures the model's ability to find all actual polyps in the dataset, highlighting its sensitivity. A commonly used metric, mAP@0.5 (mean Average Precision at an Intersection over Union threshold of 0.5), provides an overall accuracy score based on how well the predicted bounding boxes align with the ground truth. For a more comprehensive assessment, mAP@0.5:0.95 averages the model's performance across a range of stricter IoU thresholds—from 0.5 to 0.95—offering deeper insight into its consistency and robustness.

To ensure consistent supervision across both detection and segmentation tasks, bounding boxes were automatically generated from the existing segmentation masks. This approach allowed both branches of the model to learn from a unified annotation source, reducing discrepancies between tasks. All input images were resized to a fixed resolution of 640×640 using letterboxing, a technique that maintains the original aspect ratio by adding padding where necessary. The detection model was trained for 100 epochs using a cosine learning rate schedule, starting from an initial learning rate of 0.01. This schedule gradually reduced the learning rate over time in a smooth, wave-like manner to promote stable convergence. The model checkpoint for final evaluation was selected based on its performance on the validation set, specifically using IoU@0.5 as the selection criterion.

During inference, the model operated using a single-scale input, matching the resolution used during training to ensure consistent performance. Figure 2 provides a visual example of the detector's output, showing both the predicted bounding box and confidence score on a sample frame. This visualization underscores how detection complements segmentation by identifying object locations, while segmentation further refines the boundaries.

C. Polyp Segmentation

In medical image analysis, segmentation refers to the task of assigning each pixel in an image to a specific anatomical or pathological class—such as identifying tumors, lesions, or in this case, polyps. Accurate segmentation is essential for diagnosis, treatment planning, and monitoring, especially in endoscopic procedures where polyps must be precisely localized.

This study employed U-Net architecture variants for segmentation, using a supervised learning approach with binary masks where each pixel was labeled as either "polyp" or "background." This binary formulation simplifies the task while directing the model's capacity toward clinically relevant regions.

The input images were resized to a fixed resolution of 512×608 pixels to standardize dimensions across the dataset. Prior to feeding into the model, images were normalized using ImageNet mean and standard deviation, which aligns data distributions with pretrained backbones. To enhance the model's robustness against positional variance, horizontal flip augmentation was applied during training.

Training was performed over 100 epochs using the U-Net architecture, optimized with the Adam optimizer (learning rate: 1e-4, batch size: 2). To address the significant class imbalance—where polyp regions are often small and sparsely distributed compared to the background—the Dice loss function was employed. Dice loss is particularly effective in medical segmentation tasks, as it directly optimizes for spatial overlap between predicted and ground truth masks, emphasizing accuracy in the regions of interest.

Fig. 1 illustrates the overall data pipeline, from preprocessing to model selection, highlighting the parallel flow of detection and segmentation. The U-Net is trained with binary masks and evaluated using Dice and IoU metrics, while detection differs only in its target outputs—bounding boxes instead of pixel-wise masks.

Model selection was based on IoU@0.5, calculated while excluding background pixels, ensuring that the evaluation focused solely on the model's ability to segment polyp regions accurately. To provide a more comprehensive view of performance, two complementary evaluation strategies were employed: Dice score was computed directly from the raw (non-thresholded) probability maps and averaged across the entire dataset. This approach captures overall segmentation quality in a smooth, probabilistic manner. IoU, on the other hand, was calculated after applying a 0.5 threshold to binarize the model outputs. These values were then averaged on a per-image basis, offering insight into prediction consistency at the individual image level.

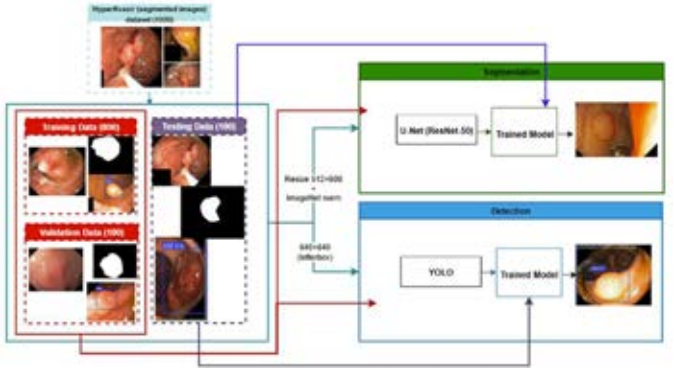


Fig. 1. Training and data workflow for the models.

Fig. 2 presents a sample inference result, where segmentation masks and their contours are overlaid on an original frame, showcasing boundary refinement around localized polyps.

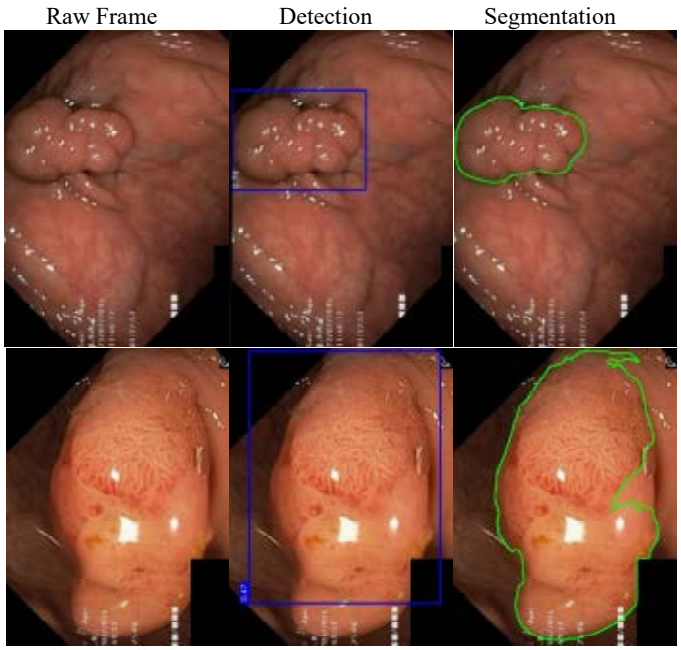


Fig. 2. Qualitative example on a representative colonoscopy frame.

Fig. 3 shows the real-time deployment pipeline, beginning with frame capture and shared preprocessing, followed by independent detection and segmentation stages, and concluding with overlay composition.

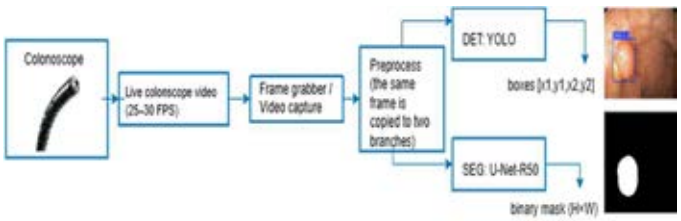


Fig. 3. Real-time inference pipeline

V. Results and Discussion

The YOLOv8-m object detection model demonstrated strong performance in localizing polyp regions within the held-out test set. As shown in Table 1, the model achieved high scores across key evaluation metrics—indicating both accurate and consistent detection behavior.

Specifically, the model reached a precision of 0.923, meaning the majority of its predicted bounding boxes corresponded to true polyp locations, with minimal false positives. It also attained a recall of 0.860, reflecting its ability to detect most of the true polyp instances in the test images. These values combine to produce an F1 score of 0.890, which balances precision and recall into a single metric. Finally, the model achieved a mean Average Precision at IoU 0.5 (mAP@0.5) of 0.948, indicating high overall detection quality across varying confidence thresholds.

Table 1: YOLO variants detection performance

Method	Precision	Recall	F1	mAP@0.5
YOLOv8-m	0.923	0.860	0.890	0.948

Following detection, the U-Net segmentation model—based on a ResNet-50 backbone—was evaluated on the same test split. It demonstrated a high degree of pixel-level alignment between predicted masks and ground truth, confirming its effectiveness in detailed polyp boundary delineation. Performance metrics for segmentation are summarized in Table 2.

Table 2: U-Net segmentation performance

Method	Dice	IoU
U-Net (ResNet-50)	0.936	0.672

VI. Conclusion

The proposed system combined a fast object detector and a precise segmentation model to achieve both reliable localization and accurate boundary refinement. YOLOv8-m effectively identified polyp candidates on a frame-by-frame basis, while the U-Net segmenter refined these regions into high-resolution, pixel-level masks. Both models, when run independently, met real-time performance requirements on the target GPU.

Despite promising results, the current pipeline has limitations. Training was conducted on a single dataset, which restricts generalizability to other clinical environments. Additionally, the system has not been stress-tested in continuous video streams, where temporal noise and performance drift may emerge. Future improvements should prioritize cross-dataset validation, integration of lightweight temporal smoothing to reduce frame-level inconsistencies, and latency-aware inference strategies that maintain a balance between responsiveness and accuracy in real-world use.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[2] S. Zhao, S. Wang, P. Pan, T. Xia, X. Chang, X. Yang, L. Guo, Q. Meng, F. Yang, W. Qian, Z. Xu, Y. Wang, Z. Wang, L. Gu, R. Wang, F. Jia, J. Yao, Z. Li, and Y. Bai, "Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-Analysis," Gastroenterology, vol. 156, no. 6, pp. 1661–1674.e11, 2019, doi: https://doi.org/10.1053/j.gastro.2019.01.260.

[3] D. Jha, S. Ali, N. K. Tomar, H. D. Johansen, D. Johansen, J. Rittscher, M. A. Riegler, and P. Halvorsen, "Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning," IEEE Access, vol. 9, pp. 40496–40510, Mar. 2021, doi: https://doi.org/10.1109/ACCESS.2021.3063716.

[4] Y. Yu, K. Lin, J. Hong, R.-G. Tsai, and Y. Huang, "PD-YOLO: Colon Polyp Detection Model Based on Enhanced Small-Target Feature Extraction," Computers, Materials & Continua, vol. 82, no. 1, pp. 913–928, 2025, doi: https://doi.org/10.32604/cmc.2024.058467.

[5] Souaidi M, El Ansari M. Multi-Scale Hybrid Network for Polyp Detection in Wireless Capsule Endoscopy and Colonoscopy Images (Hyb-SSDNet). Diagnostics. 2022;12:2030. https://www.mdpi.com/2075-4418/12/8/2030 MDPI

[6] J. Wan, W. Zhu, B. Chen, L. Wang, K. Chang, and X. Meng, "CRH-YOLO for precise and efficient detection of gastrointestinal polyps," Scientific Reports, vol. 14, no. 1, p. 30033, Dec. 2024, doi: https://doi.org/10.1038/s41598-024-81842-9.

[7] S. Gupta and J. Singla, "YOLO-LAN: Precise Polyp Detection via Optimized Loss, Augmentations and Negatives," arXiv preprint arXiv:2509.19166 [cs.CV], 2025. doi: 10.48550/arXiv.2509.19166

[8] S. Safarov and T. K. Whangbo, "A-DenseUNet: Adaptive Densely Connected UNet for Polyp Segmentation in Colonoscopy Images with Atrous Convolution," Sensors, vol. 21, no. 4, p. 1441, 2021, doi: https://doi.org/10.3390/s21041441.

[9] Jha, D., Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, (2020). Kvasir-SEG: A Segmented Polyp Dataset. In: Ro, Y., et al. MultiMedia Modeling. MMM 2020. Lecture Notes in Computer Science(), vol 11962. Springer, Cham. https://doi.org/10.1007/978-3-030-37734-2_37

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Evaluation of No-Reference Image Quality Measurements for Medical Imaging

Dilek Özer
Institute of Biomedical Engineering
Boğaziçi University
Istanbul, TÜRKİYE

Prof. Dr. Albert Güveniş
Institute of Biomedical Engineering
Boğaziçi University
Istanbul, TÜRKİYE

***Abstract*— Medical imaging technologies are essential for making clinical diagnosis and treatments, but issues like blurriness, noise, and loss of contrast can impair diagnostic precision. This research study assesses the effectiveness of three commonly used No-Reference Image Quality Assessment(NR-IQA) models – PIQE, BRISQUE, and NIQE-on X-ray computed Tomography(CT) images. A collection of 50 images, consisting of both ideal and artificially distorted CT scans, was evaluated subjectively by three doctors and objectively using NR-IQA models. The results reveal that PIQE demonstrates strong agreement (82%) with human assessment, while BRISQUE and NIQE show a poor alignment (30% and 28%, respectively). These findings suggest that PIQE is more appropriate for applications in medical imaging, recommending its incorporation into automated quality control systems to enhance diagnostic results.**

Keywords— medical imaging, image distortion, image quality assessment(IQA), subjective IQA, no-reference IQA, clinical decision support.

I. INTRODUCTION

The advent of medical imaging technologies has transformed diagnostics, allowing for precise and prompt medical interventions[1]. Nevertheless, the quality of these images is essential for facilitating accurate clinical decision-making[2]. No-Reference Image Quality Assessment(NR-IQA) models are crucial in this area, as they gauge image quality without the need for reference images[3].

Despite their common application, NR-IQA models encounter challenges within the medical field due to the specific imaging demands. Medical images frequently exhibit localized artifacts that differ from those typically seen in natural images, where conventional NR-IQA models have mainly been tested[4]. Prior research highlights the challenges of applying IQA models that were initially developed for natural scenes to clinical datasets, emphasizing the need for specialized solutions that consider the structural intricacies of medical data[5],[8]. This reinforces the significance of thoroughly examining how effectively NR-IQA models align with expert evaluations in clinical settings.

Among the frequently used NR-IQA models, BRISQUE[1] and NIQE[2] are often employed; however, they have shown varying effectiveness in the context of medical imaging[8]. On the other hand, the Perception-based Image Quality Evaluator(PIQE) has been developed as a block-wise, unsupervised method that focuses on localized distortions and does not depend on Natural Scene Statistics(NSS)[3],[9]. These features render PIQE especially pertinent for clinical imaging, where issues such as blur, loss of contrast, and noise usually appear in specific areas with clear diagnostic significance[5].

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

A. *Quality of Medical Images*

Medical imaging, particularly Computed Tomography(CT), plays a vital role in clinical diagnostics by offering a non-invasive view of anatomical structures[5]. The reliability of CT images in diagnosis is heavily dependent on their quality, as issues such as blur, noise, and contrast loss can hinder physicians from interpreting images correctly. Consequently, Image Quality Assessment(IQA) is essential to guarantee that only diagnostic-quality images are used in clinical settings[6].

B. *No-Reference Image Quality Assessment(IQA)*

Conventional Image Quality Assessment(IQA) techniques rely on reference images[7], which may frequently be unavailable in numerous clinical scenarios. As a result, No-Reference Image Quality Assessment(NR-IQA) methods have gained traction for automated quality control. Among the most commonly used NR-IQA methods are BRISQUE[1], which utilizes Natural Scene Statistics(NSS) along with opinion-aware training, and NIQE [2], an opinion-unaware approach that assesses statistical deviations from naturalness. However, both methods were mainly created for natural images, which restricts their applicability in medical settings[8].

C. *PIQE and Considerations Relevant to Specific Domains*

Recently, the Perception-based Image Quality Evaluator (PIQE) has been introduced as a block-wise, unsupervised approach that focuses on localized artifacts[3]. In contrast to BRISQUE and NIQE, PIQE operates independently of NSS assumptions, which may make it a better fit for medical images where distortions are often localized and have clinical relevance[9]. To explore this theory, a dataset of CT scans was created using images from NIH DeepLesion repository[4], featuring both pristine and synthetically distorted images.

D. *Motivation and Contribution*

Prior research indicates the challenges of utilizing natural-image IQA models for medical datasets[10],[11]. To tackle this issue, we assess PIQE, BRISQUE, and NIQE on CT images, using subjective ratings from physicians as the benchmark. The contributions include:

- Creation of a CT dataset incorporating clinically relevant synthetic distortions.
- Comparative analysis of NR-IQA models(BRISQUE ,NIQE, and PIQE) against the ratings provided by physicians.
- Determining the model that aligns most closely with clinical requirements.

Table 1: Essential Features of NR-IQA Models.

Model	Evaluation Method	Key Features	Assumptions
PIQE	Block-wise evaluation	Localized artifact detection	None
BRISQUE	Statistical measures	Uses NSS for quality assessment	NSS-dependent
NIQE	Statistical measures	Works without training on datasets	NSS-dependent

Table 1 summarizes PIQE has become a strong contender because of its block-wise assessment approach, which does not depend on NSS, thus making it more suitable for medical images that frequently exhibit localized distortions[3].

II. MATERIALS AND METHODS

The assessment of NR-IQA models centers on their effectiveness and precision in medical imaging, specifically CT scans. This research presents a systematic approach to evaluate the PIQE, BRISQUE, and NIQE models, highlighting their clinical significance. The study has four phases, Dataset creation, Subjective and Objective assessment, Comparative Analysis. As shown in Table 2, the procedure initiated with dataset creation consisting of 20 pristine and 30 distorted images, which was then followed by subjective evaluations in a clinical environment and objective assessments using BRISQUE, NIQE, and PIQE. The outcomes were subsequently analyzed and compared to delivering an extensive evaluation of the quality of CT images.

A. Dataset

A collection of 20 unmodified CT scans was obtained from the NIH DeepLesion repository[4]. To mimic actual degradations, synthetic distortions such as Gaussian blur, noise, and contrast were introduced[12], leading to a final count of 50 images. These images constituted the evaluation dataset(FYPDataset). Table 2 provides an overview of the evaluation dataset.

Table 2. Composition of the Evaluation Dataset.

Image Type	Source	Number of Images
Pristine	NIH DeepLesion repository	20
Artificially Distorted	Synthetic distortions applied	30

Figure 1 displays processed images sourced from the identical original CT scan (a). The samples feature Gaussian blur (b, σ=2.25), grayscale contrast alteration (c, range [0.2–0.4]),and salt-and-pepper noise (d, density=0.02). These examples illustrate the various distortions implemented in the FYPDataset.

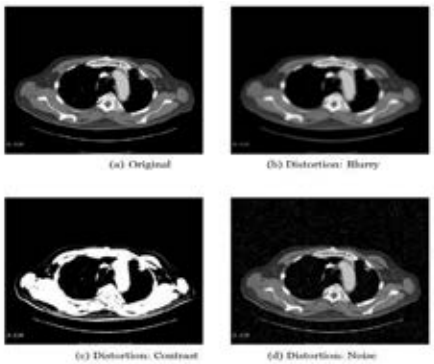


Fig. 1. A set of processed images as a sample.

B. Subjective Assessment

For reliable evaluation, subjective assessments were conducted by three physicians, which included two clinicians and one radiologist, and assessed the dataset with the help of a 5-point Likert scale. The subjective evaluation utilized the Single Stimulus Method, where each image was assessed on its own without any comparison to a reference image[21]. Each image was rated according to four structured diagnostic quality inquiries, concentrating on aspects such as clarity, diagnostic utility, and noticeable artifacts[13]. The consensus and average ratings were considered the subjective gold-standard baseline.

Table 3. Subjective IQA Observers.

Observer Index	Position	Experience(years)	Time Cost(min)	Data Consumption
1	Clinician	13+	31	Nominative
2	Clinician	3	40	Testing
3	Radiologist	13+	23	Nominative

Table 3 shows information about the observers, including observer index, position, experience, time cost and data usage. A preliminary assessment questionnaire was created to collect the viewpoints of doctors, with the last question encouraging involvement in Subjective IQA. From the responses, three doctors were chosen as observers (Table 3) although the selection was constrained by their type and level of expertise.

C. Objective NR-IQA Models

All three NR-IQA models(BRISQUE, NIQE, and PIQE) produce numerical ratings from 0 to 100, with scores close to 0 indicating optimal image quality and scores close to 100 representing the lowest quality. Three models for No-Reference Image Quality Assessment(NR-IQA) were developed using Python: BRISQUE[1], NIQE[2], and PIQE[3]. Each model was utilized on the dataset to generate objective scores, which were subsequently compared to the ratings given by physicians to evaluate their results. In particular, BRISQUE[1] operates as an opinion-aware model that has been trained using natural scene statistics (NSS) features, NIQE[2] functions as an opinion-unaware model that relies on statistical naturalness, while PIQE[3] serves as a block-based unsupervised evaluator that focuses on localized distortions and does not rely on NSS[9], rendering it more appropriates for applications in medical imaging.

D. Agreement Calculation

Agreement between subjective and objective evaluations was quantitatively assessed as the ratio of instances where subjective and objective rankings aligned. For each image group gi, denote Rs(gi) as the subjective ranking and Ro(gi) as the objective ranking. The indicator function can be expressed as equation (1):

$$I(gi) = \{1, \text{if } Rs(gi) = Ro(gi); 0, \text{otherwise}\} \quad (1)$$

The overall agreement is calculated as equation (2):

$$A = (1/N) * \Sigma I(gi) \times 100\% \quad (2)$$

N denotes the total count of assessed images. This metric was chosen to offer an intuitive indication of how often model-generated rankings align with expert assessments, serving as a complement to the analysis. According to this calculation, PIQE achieved 41 out of 50 matches(82%), while BRISQUE and NIQE had significantly lower rates of 30% and 28%, respectively.

III. RESULTS

The correlation in Table 4 indicates that PIQE aligns most closely with physician evaluations, where lower objective scores are typically associated with higher subjective ratings(Pearson r = -0.16, Spearman ρ = -0.18). BRISQUE shows a more moderate yet inconsistent correlation, which is evidenced by its minimal correlation coefficients(Pearson r = 0.03, Spearman ρ = 0.07). In terms of performance, NIQE ranks the lowest, characterized by poorly clustered scatter points and weak positive correlation(Pearson r = 0.11, Spearman ρ = 0.09), which goes against the anticipated pattern. In summary, PIQE stands out as the most dependable NR-IQA model when compared to the other two models.

Table 4. Correlation of NR-IQA Models with Subjective Scores.

	Model	Pearson r	Spearman ρ
1	PIQE	-0.163655602	-0.178549706
2	BRISQUE	0.029057302	0.070579188
3	NIQE	0.106123323	0.089692466

To confirm previous results and eliminate potential mistakes or coincidences, further tests were carried out with a new set of 20 medical images, including 8 originals, 7 blurred images, 4 with contrast issues, and 1 containing noise. The subjective Image Quality Assessment (IQA) relied on ratings from the secondary observer (index 002) in Table 3, while the Objective IQA utilized the same three models, placing particular focus on PIQE. The analysis comparing Subjective and Objective findings produced results matching with the initial study, confirming that the previous experiments did not contain major errors or discrepancies, thus enhancing the reliability of the results across primary, clinical, and computational assessments.

The degree of consensus between NR-IQA models and expert results was evaluated. Among these models, PIQE showed the highest result with subjective evaluations at 82%, whereas BRISQUE and NIQE scored 30% and 28%, respectively. This indicates that PIQE is a more reliable option for evaluating CT images in clinical settings.

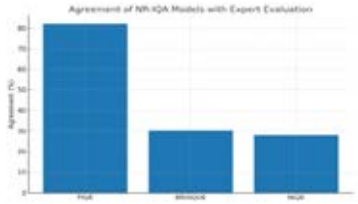


Fig. 2. Proportions of agreement between PIQE, BRISQUE, and NIQE and evaluations from experts.

Figure 2 illustrates the agreement rates between three NR-IQA models and expert evaluations. PIQE exhibits the highest performance at roughly 82%, whereas BRISQUE and NIQE show much lower agreement rates of about 30% and 28%, respectively. This indicates that piqe aligns more closely with physicians' evaluations of CT image quality than the other two models.

Table 5. Examples of Match Case.

	Scores	
	Subjective IQA	Objective IQA
Original	53	5
Blurry	52	4
Contrast	46	3

Table 6. Examples of Mismatch Case.

	Scores	
	Subjective IQA	Objective IQA
Original	46	5
Blurry	52	4
Contrast	53	3

Table 5 and Table 6 are directly linked to Equation(1) and (2). Instances labeled as "Match" receive a value of 1 in the indicator function, while "Mismatch" instances are given a value of 0. These tables illustrate cases where subjective and objective evaluations either align or differ, providing clarity on the methodology used to determine agreement percentages. A match indicates that subjective and objective rankings align for a particular group of images, whereas a mismatch shows a lack of agreement. Collectively, these examples offer tangible insight into how the reported agreement rates-82% for PIQE, 30%for BRISQUE, and 28% for NIQE- were calculated.

Table 7. Comparison of IQA Models.

Criterion	PIQE	NIQE	BRISQUE
Advantage	Fairly effective; appropriate for medical imaging [3],]9].	Excels in processing natural images[2].	Commonly utilized[1].
Disadvantage	Restrictions in different areas[3],[9].	Inappropriate for medical images[2],[8].	Insufficiently effective for medical images[1],[5].
Performance Evaluation	High perf [3],[14].	Low perf [2],[15].	Medium perf[1],[15].
Performance with X-ray CT Scans	Superior perf[3],[5].	Inconsistent[2][5]	Inconsistent[1],[5].
Applicability to Medical Imaging	Fairly efficient[3],[5].	Limited, similar to BRISQUE[2],[5].	Restricted, difficulties with texture/contrast[1],[5].
Dependence on NSS	Independent[3][9]	High[2],[8].	High[1],[8].
Sensitivity to Local Distortions	High [3],[9],[16].	Low[2],[15].	Low[1],[5].
Correlation with Subjective Evaluations	Positive (~82%)[3],[5],[14]	Poor (~28%)[2],[5],[14]	Poor (~30%)[1],[5],[14]
Requirement for Training Data	Minimal[3]	Uses training data[2]	Significant[1]
Performance in This Study	Good [5]	Poor [5]	Poor [5]

Table 7 summarizes the comparison of different IQA models. The findings indicate that the BRISQUE model exhibited a low relation with expert reviews, attaining only around %30. Likewise, the NIQE model also fell short, achieving a result of nearly 28%. On the other hand, the PIQE model exhibited significantly better performance in this study, reaching 82% when compared to expert evaluations. This enhanced result with human assessments implies that PIQE may be more adept at accurately measuring image quality concerning perceptual distortions.

IV. DISCUSSION

A. Findings

Despite their widespread use, NR-IQA models are widely used, they face difficulties in the medical domain because of the unique imaging requirements. Medical images often contain localized artifacts that are distinct from those found in natural images, where traditional NR-IQA models have primarily been evaluated[4]. Earlier research underscores the obstacles in adapting IQA models designed for natural images to clinical datasets, stressing the necessity for customized solutions. Thus, understanding how effectively these models correspond with clinical evaluations is crucial for enhancing their applicability in medical imaging.

The results show that PIQE significantly performs both BRISQUE and NIQE in evaluating the quality of CT images[14]. Unlike other approaches, PIQE does not rely on Natural Scene Statistics(NSS), which often fail to capture the complex structural characteristics found in medical imaging[9]. This represents a significant advantage for PIQE, as it can detect distortions that directly impact the diagnostic quality of CT images. The poor performance of BRISQUE and NIQE highlights the statistical differences between NSS and the structural features of medical images, indicating that these models- originally designed for natural images- are ineffective at recognizing distortions critical for clinical use[15]. In contrast, PIQE's block-wise, unsupervised method effectively identifies blur and localized noise[16], making it more capable of detecting degradations that influence physicians' evaluations of diagnostic value. Instances of match and mismatch

evaluations can be found in Table 5 and Table 6, which further clarifies how agreement is calculated.

B. Comparison with Previous Research

The performance of BRISQUE and NIQE underscores the statistical disparities between NSS and the structural characteristics of medical images, which aligns with earlier research reporting comparable shortcomings of NSS-based techniques. Conversely, the effectiveness of PIQE reinforces prior observations that highlight the benefits of localized, block-wise, and unsupervised approaches over models constructed on natural image statistics[9],[15],[16].

Previous research[3],[5] has also indicated that PIQE outperforms other NR-IQA models in analyzing X-ray CT scans; however, those findings were primarily derived from broad evaluations or review analysis. In contrast, this study builds on that evidence by presenting a direct comparison of PIQE with BRISQUE and NIQE using a specially curated CT dataset, while also validating the outcomes against assessments by physicians. This approach provides stronger clinical relevance and empirical validation, illustrating that PIQE not only corresponds more closely with expert evaluations but also offers improved practicality for clinical use. This signifies important progress toward dependable automated quality control in CT imaging.

C. Potential Clinical Applications

The proven capability of PIQE to identify distortions that impact diagnostic quality has real-world implications for medical imaging processes. Given its greater alignment with physicians' subjective evaluations, PIQE could be incorporated into automated quality assurance systems. This integration would facilitate the prompt identification of low-quality CT images prior to their application in clinical diagnosis, consequently decreasing the chances of diagnostic errors and improving reliability in healthcare decision-making.

D. Limitations of the Current Research

The findings of this study should be interpreted with certain limitations. Firstly, the dataset comprised only 50 CT images, which could limit its broader applicability; larger datasets are necessary to account for more extensive variability[5]. In addition, the application of synthetic distortions may not fully capture the intricacies of actual clinical artifacts, underscoring the importance of validation with authentic medical data[14]. This research was also confined to CT images and assessments from only three clinicians, which may lead to potential bias[5],[11]. Lastly, the research study did not investigate real-time processing, which is vital for clinical settings and should be considered in future research. [16].

E. Future Work

Future research should focus on improving NR-IQA models and exploring hybrid approaches that incorporate deep learning techniques to advance the field of medical image quality assessment[11]. The effect of the extent of distortions may also be investigated. Such advancements could help bridge the gap between algorithmic evaluations and the more nuanced assessments made by human experts, facilitating the integration of technology into healthcare systems. Broadening research to encompass larger datasets and various imaging modalities will enhance the validation of PIQE's versatility across different clinical scenarios, while also promoting the advancement of deep learning-driven IQA models specifically designed for medical imaging[17]-[20].

V. CONCLUSION

The evaluation of NR-IQA models indicates that PIQE significantly outperforms both BRISQUE and NIQE in assessing the quality of CT images, which is crucial for clinical purposes. Integrating PIQE into imaging workflows could enhance automated quality control, support the early detection of poor scans, and boost diagnostic reliability.

REFERENCES

[1] A. Mittal, A. Moorthy, and A. Bovik, "BRISQUE: Blind/Referenceless Image Spatial Quality Evaluator," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 3449–3461, 2012.

[2] A. Mittal, R. Soundararajan, and A. Bovik, "NIQE: A Natural Scene Statistic-based Quality Evaluator," IEEE Signal Processing Letters, vol. 20, no. 11, pp. 154–158, 2013.

[3] S. Venkatanath, R. Praneeth, B. Chandrasekhar, et al., "PIQE: Perception Based Image Quality Evaluator," in *Proc. IEEE National Conf. of Communications (NCC)*, 2015, pp. 1–6.

[4] K. Yan et al., "DeepLesion: Automated Mining of Large-Scale Lesion Annotations and Universal Lesion Detection with Deep Learning," *Journal of Medical Imaging*, vol. 5, no. 3, 036501, 2018.

[5] R. J. Ferrari et al., "Medical image quality assessment: A review," *Journal of Digital Imaging*, vol. 30, pp. 449–459, 2017.

[6] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in Proc. *IEEE Int. Conf. on Pattern Recognition(ICPR),* 2010, pp. 2366–2369.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity*," IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[8] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[9] P. Le Callet, S. Möller, and A. Perkis, "Qualinet white paper on subjective methods for image quality assessment," *COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services*, 2012.

[10] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010.

[11] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[12] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in Proc. *IEEE Int. Conf. on Image Processing(ICIP)*, 2017.

[13] X. Liu, J. van de Weijer, and A. D. Bagdanov, "RankIQA: Learning from rankings for no-reference image quality assessment," in Proc. *IEEE Int. Conf. on Computer Vision(ICCV)*, 2017, pp. 1040–1049.

[14] Y. Zhang, J. Gao, and F. Zhu, "Saliency-inspired blind image quality assessment using a naturalness model," *Neurocomputing*, vol. 172, pp. 392–402, 2016.

[15] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[16] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in Proc. *IEEE Computer Vision and Pattern Recognition(CVPR),* 2017.

[17] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain*," IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[18] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1317–1329, 2006.

[19] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[20] S. Winkler and R. Subramanian, "Overview of eye tracking datasets," in Proc. *Int. Workshop on Quality of Multimedia Experience (QoMEX)*, 2013.

[21] International Telecommunication Union, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R BT.500-13, 2012.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Altered Salience-Network Functional Connectivity in Depression

Mehmet Akif ÖZÇOBAN Gaziantep University
Naci Topçuoğlu Vocational School,
Department of Electronic and Automation
Gaziantep, Türkiye
mehmetakifozcoban@gmail.com

***Abstract*— This study assesses salience-network (SN) connectivity changes in major depressive disorder (MDD) using EEG-based, frequency-resolved wavelet coherence (WC). It examines the single-feature classification performance of these measures. Resting-state, eyes-closed EEG data were collected from 60 MDD patients and 60 healthy controls. WC values were calculated within canonical frequency bands, and group differences were evaluated with independent-samples t-tests. MDD patients showed significantly higher wavelet coherence across multiple frequency bands. (delta, theta, alpha, beta, gamma) in frontal and fronto-central connections. In classification, several individual connection–band features achieved moderate to good discrimination, with the highest performance observed for F3–F4 theta (77.8% accuracy, AUC = 0.84) and Fp1–F4 delta (79% accuracy, AUC = 0.81). These results suggest SN-related hyper-synchrony in MDD, consistent with fMRI reports of disrupted network switching. EEG-based WC shows promise as a biomarker, with its performance likely to improve through the integration of multiple features.**

I. INTRODUCTION

Functional neural networks are temporally organized patterns of coupling among widely distributed ensembles of brain regions that exhibit correlated activity, reflecting the functional interdependence of structures within large-scale circuits [1]. Data-driven analyses of intrinsic functional magnetic resonance (fMRI) signals have delineated a set of prevalent macroscale networks in the human brain [2]. Among these, the default mode network (DMN), the frontoparietal control network (FPN, also referred to as the central executive network, CEN), and the salience network (SN) are particularly significant for cognition and affect. These networks are robustly expressed at rest and are therefore considered canonical resting-state networks. Broadly, the DMN supports internally oriented mentation (e.g., self-referential thought), while the FPN underpins goal-directed control and working memory. The SN detects salient internal/external events and helps coordinate switching between the DMN and the CEN. Coherence between different brain regions reflects the efficiency and strength of information transfer among these regions. Higher values can be interpreted as similar neural activation of two brain regions. Whereas decreased coherence refers to poor functional connectivity between these areas. Most of the previous studies used fMRI to evaluate the strength of functional connectivity between the brain networks in major depressive disorder (MDD) [3]. The disease is a leading cause of disability worldwide and is increasingly conceptualised as a systems-level brain disorder characterised by disturbed large-scale network dynamics [4]. Although substantial progress has been made in understanding and treating MDD, its pathophysiology remains incompletely understood [5]. In addition to regional abnormalities, accumulating evidence reveals dysfunctional interactions among the DMN, CEN, and SN—a triad essential for switching between internally oriented thinking and goal-directed control. In MDD, fMRI studies consistently report dysconnectivity within and between the DMN, CEN, and SN networks [6]. Motivated by the millisecond dynamics of network coordination, Electroencephalography (EEG) is used in the study [7]. fMRI offers excellent spatial resolution; however, several factors limit its routine clinical application. These factors include high costs and inconvenience, sensitivity to motion and physiological noise, and poor temporal resolution. Since BOLD signals reflect hemodynamic responses, estimates of functional connectivity are primarily restricted to fluctuations below 1 Hz. On the other hand, EEG allows for millisecond-level sampling and provides frequency-resolved functional connectivity across a range of 1 to 50 Hz (and beyond). This provides direct insights into neural synchronization among distributed networks. Previous surface-EEG studies have identified scalp proxies for well-known resting-state networks, such as the SN, and have specified corresponding electrode pairs. In particular, frontal and fronto-midline connections (e.g., Fp1/Fp2–F3/F4 and C3/C4–Fz) have been highlighted as relevant to the SN. Building on this framework, the present study investigates frequency-specific synchronization within these connections using wavelet coherence (WC). This framework addresses the multiple-testing inflation that occurs in localized spectra by requiring that significant regions exceed a scale-dependent critical patch size derived from the reproducing kernel of the wavelet transform [8]. This approach effectively removes most of the spurious, short-lived patches that can arise even in noisy data. By utilizing WC, estimated in the time-frequency domain with

979-8-3315-5565-8/25/$31.00 ©2025 IEEE

methodologies like continuous wavelets, we gain access to band-limited, high-temporal-resolution coupling without the need for stringent stationarity assumptions.

To our knowledge, this is the first study that investigates salience network connectivity in MDD using an electrophysiological connectivity method. Our aim is to contribute to the literature and provide this information.

II. Material and Methods

A. Subjects

Researchers collected EEG data from 60 patients diagnosed with MDD and 60 age-matched healthy controls (HC) at the prestigious Üsküdar University Neuropsychiatry Health Practice and Research Centre in Istanbul, Turkey. MDD diagnoses were rigorously confirmed through the Structured Clinical Interview for DSM-5, ensuring the precise selection of participants. All patients were medication-free for at least two weeks prior to EEG recording and had not received psychiatric therapy, making the data pure and reliable. The healthy controls were thoroughly screened to exclude any history of stroke, neurological, or neuropsychiatric disorders and were not on any medication, ensuring a robust comparison group. Informed consent was obtained from all participants in accordance with the Declaration of Helsinki. The study protocol was reviewed and approved by the university's ethics committee, ensuring compliance with the highest ethical research standards. To assess symptom severity, participants completed the well-established Beck Depression Inventory (BDI) and Beck Anxiety Inventory (BAI), both 21-item self-report measures that provide comprehensive insights into depressive and anxiety symptoms. This meticulous methodology underscores the study's commitment to scientific rigour and the pursuit of impactful, trustworthy findings.

B. EEG Acquisition and Preprocessing

Resting-state EEG was recorded for 7 minutes with participants' eyes closed at a sampling rate of 512 Hz using a 19-channel international 10–20 montage, including electrodes placed at Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2, Fz, Cz, and Pz. Electrode impedances were maintained below 5 kΩ. A band-pass filter (0.5–70 Hz) and a 50 Hz notch filter were applied using Scan Edit 4.3, and subsequent spectral analyses and asymmetry computations were conducted in MATLAB. To suppress slow drifts while preserving phase, filtering was implemented in zero-phase form using a Butterworth high-pass filter with a cutoff frequency of 0.5 Hz, within the 0.5–70 Hz passband. An experienced neurophysiology expert initially identified and removed ocular and muscle artefacts (including EOG) using an amplitude threshold of 50 µV peak-to-peak, along with visual inspection. Residual artefacts were further minimised using independent component analysis (ICA).

C. Classical Conventional Method

The magnitude-squared coherence between two wide-sense stationary signals, *x(t)* and *y(t),* at frequency *f* is defined as the normalized cross-spectral density. It is calculated as the squared magnitude of the cross-spectrum divided by the product of the auto-spectra (power spectral densities).

$$C_{xy}(f) = \frac{|W_{xy}(f)|^2}{W_x(f)\, W_y(f)} \quad (1)$$

$W_{xy}(f)$ is denoted the cross-spectral density.

$$W_{(xy)}(f) = X(f)Y(f)^* \quad (2)$$

X(f) and *Y(f)* are the products of the Fast Fourier Transform (FFT) of *x* and *y*, respectively. The asterisk sign refers to the complex conjugate operator. The conventional magnitude-squared coherence, computed using two signals, is defined. The frequency components were obtained for estimating the power spectral density and cross-spectral density.

D. Wavelet Coherence Method

EEG is inherently nonstationary, and its spectral content varies over time[9, 10]. To effectively capture these time-varying spectral features, we employed the continuous wavelet transform (CWT), enabling a signal to be represented simultaneously in both time and frequency domains. The CWT of a signal x(t) is defined as follows:

$$\mathrm{CWT_x}(\tau, a) = \int_{-\infty}^{+\infty} x(t)\ \Psi^*_{\tau,a}(t)\, dt \quad (3)$$

$\Psi^*_{\tau,a}(t)$ is the scaled and translated copy of the mother wavelet Ψ_0 obtained by dilation by a and translation by τ.

The CWT of a signal depends on time (τ) and scale (a), with scale inversely related to the central frequency. Thus, each scale corresponds to a unique pseudo-frequency. Among various wavelet families, the Morlet wavelet is widely used for EEG analysis due to its Gaussian-windowed sinusoidal form, which balances time and frequency localization. The Morlet mother wavelet is defined as:

$$\psi_0(t) = \pi^{-1/4}\, e^{j2\pi f\, t}\, e^{-t^2/2} \quad (4)$$

The CWT in (3) is obtained by correlating the signal *x(t)* with time-shifted and dilated replicas of the mother wavelet defined in (4). The wavelet power was given in equation 5.

$$WP_x(t,f) = |CWT_x(t,f)|^2 \quad (5)$$

Hence, the CWC method provides a time-frequency representation *Wx(t,f)* that depends on time *t* and the wavelet's centre frequency *f*. The WC method between two signals *x* and *y*, which highlights regions of high common power, is defined as equation 6.

$$W_{xy}(t,f) = CWT_x(t,f)\,CWT_y(t,f)^* \quad (6)$$

Having defined the smoothing operators, the wavelet coherence is [8].

$$\mathrm{WC}(t,f) = \frac{\left|S\left((1/a)\,XWT_{xy}(t,f)\right)\right|^2}{S\left((1/a)\,WP_x(t,f)\right)S\left((1/a)\,WP_y(t,f)\right)} \quad (7)$$

The Schwartz inequality ensures that *WCxy* has values between 0 and 1 [11, 12].

III. RESULTS

MDD exhibited higher coherence than controls across multiple frontal and fronto-central links, especially in, most consistently in delta–theta: C3–Fz: Delta (p = .000) and theta (p = .037), higher in MDD. F4–Fz: Delta (p = .016) higher in MDD. C4–Fz: Delta (p = .000) and theta (p = .006) higher in MDD. Fp1–F3: Delta, theta, alpha, and gamma (all p = .000) are higher in MDD. Fp1–F4: Delta (p = .000) and theta (p = .001) higher in MDD. F3–F4: Delta (p = .000), theta (p = .000), alpha (p = .003) higher in MDD. Fp2–F3: Delta, theta, alpha (all p = .000) are higher in MDD. Fp2–F4: Delta (p = .001), theta (p = .001), alpha (p = .000), beta (p = .002) higher in MDD. Group differences were given in Table 1.

TABLE I:WAVELET COHERENCE ANALYSIS RESULTS

		MDD group		HC group		
		Mean	SD	Mean	SD	*p*
C3-Fz	Delta	0.668	0.115	0.587	0.117	0.000
	Theta	0.653	0.086	0.615	0.108	0.037
F4-Fz	Delta	0.747	0.151	0.688	0.153	0.016
C4-Fz	Delta	0.680	0.116	0.543	0.117	0.000
Fp1-F3	Delta	0.666	0.110	0.582	0.162	0.000
	Theta	0.739	0.092	0.634	0.185	0.000
	Alpha	0.805	0.785	0.792	0.157	0.000
	Gamma	0.636	0.640	0.537	0.141	0.000
Fp1-F4	Delta	0.587	0.119	0.465	0.136	0.000
	Theta	0.625	0.120	0.543	0.163	0.001
F3-F4	Delta	0.666	0.152	0.526	0.115	0.000
	Theta	0.695	0.127	0.588	0.123	0.000
	Alpha	0.747	0.119	0.681	0.148	0.003
Fp2-F3	Delta	0.571	0.133	0.439	0.114	0.000
	Theta	0.613	0.126	0.517	0.144	0.000
	Alpha	0.699	0.116	0.612	0.173	0.000
Fp2-F4	Delta	0.682	0.122	0.601	0.170	0.001
	Theta	0.755	0.086	0.672	0.185	0.001
	Alpha	0.824	0.067	0.727	0.186	0.000
	Beta	0.725	0.104	0.655	0.160	0.002

Classification using WC features (Table 2). With a bagged-trees ensemble (100 learners, max 20 splits) and 5-fold CV, single connection/band models achieved moderate–good performance on several pairs:

- C3–Fz (delta): 75.00%, AUC = 0.68.
- Fp1–F4 (delta): 79.17% accuracy, AUC = 0.81.
- F3–F4 (theta): 77.08% accuracy, AUC = 0.84 (highest AUC).
- F3–F4 (alpha): 79.17%, AUC = 0.81.
- Fp2–F3 (delta): 75.00%, AUC = 0.76; high recall (78.57%).

Some significant group differences showed only modest discriminability (e.g., C4–Fz delta, AUC = 0.42), indicating that statistical significance does not always imply strong single-feature classification. The results were given in Table 2.

TABLE II: CLASSIFICATION RESULTS

		Accuracy	AUC	Precision	Recall	F1
C3-Fz	Delta	75.00	0.68	60.00	42.86	50.00
	Theta	52.08	0.50	9.09	7.14	8.00
F4-Fz	Delta	66.67	0.65	41.67	35.71	38.46
C4-Fz	Delta	68.75	0.42	40.00	14.29	21.05
Fp1-F3	Delta	60.42	0.61	27.27	21.43	24.00
	Theta	62.50	0.71	37.50	42.86	40.00
	Alpha	60.42	0.65	33.33	35.71	34.48
	Gamma	66.67	0.60	42.86	42.86	42.86
Fp1-F4	Delta	79.17	0.81	70.00	50.00	58.33
	Theta	68.75	0.76	47.37	64.29	54.55
F3-F4	Delta	68.75	0.81	47.37	64.29	54.55
	Theta	77.08	0.84	66.67	42.86	52.17
	Alpha	79.17	0.81	70.00	50.00	58.33
Fp2-F3	Delta	75.00	0.76	55.00	78.57	64.71
	Theta	66.67	0.77	42.86	42.86	42.86
	Alpha	64.58	0.67	36.36	28.57	32.00
Fp2-F4	Delta	60.42	0.63	22.22	14.29	17.39
	Theta	58.33	0.57	36.36	57.14	44.44
	Alpha	77.08	0.73	61.54	57.14	59.26
	Beta	75.00	0.69	62.50	35.71	45.45

IV. DISCUSSION

The findings of the study show increased slow wave coherence in MDD across insular and ACC regions (frontal, fronto-central). These results are consistent with neuroimaging studies of the salience network in neurological and neuropsychiatric diseases. This pattern aligns with existing literature that highlights the dysfunction of the SN central to the pathophysiology of depression. In a study involving rTMS treatment, connectivity within the SN decreased after therapy, and greater reductions in SN connectivity were linked to better

responses to antidepressants. These findings indicate that depression is characterized by pathological hyper-coupling of the SN at rest and inefficient switching between the DMN and executive systems. Treatment appears to reduce SN coupling and enhance interactions between the SN and DMN. [13]. This pattern aligns with findings from network-level fMRI studies that show enhanced connectivity within the SN and stronger coupling between the SN and the DMN. At the same time, there is weaker coupling between the SN and the FPN. This configuration is thought to bias attention toward internal or salient signals while impairing top-down control. Although these findings were reported in the context of another dysregulated behavior—problematic smartphone use—they converge on a dysregulation pattern centered around the SN, which likely generalizes to affective disorders such as MDD [14]. Additionally, precision functional mapping shows a near twofold cortical expansion of the frontostriatal salience network in depression due to network-border shifts; this topology is trait-like (stable across mood state, detectable pre-onset), while within-person frontostriatal connectivity fluctuations track and predict anhedonia over time. Taken in conjunction with our slow-wave coherence increases over the insula/ACC, these results reinforce an SN-centred account, in which an enlarged SN territory and inefficient SN-gated switching bias DMN–FPN interactions are implicated [15]. In contrast, studies on late-life depression, particularly those with high levels of apathy, indicate decreased connectivity within the SN and altered coupling between the SN and the dorsolateral prefrontal cortex (DLPFC). This suggests a specific signature of SN hypoconnectivity that varies by age and phenotype. The differences observed may be attributed to variations in cohort characteristics (such as elderly individuals with high apathy compared to a mixed-age group), symptom profiles, and methods used (like resting-state fMRI focusing on the right anterior insula/DLPFC versus our approach using EEG to examine slow-wave activity in the insula, anterior cingulate cortex, and fronto-parietal coherence) [16]. These findings highlight the importance of stratifying research by age and motivational phenotype and emphasize the need to map the source-level nodes of the salience network in future studies [17].

V. CONCLUSION

Abnormal connectivity within the salience network (SN) might contribute to cognitive deficits and rumination in MDD. Our single-edge models demonstrated moderate ability to distinguish cases, with a maximum AUC of 0.84. Future research will involve comparing additional synchronization metrics such as PLV, PLI, wPLI, imaginary coherence, and PSI, as well as combining source-level EEG with fMRI to validate dysregulation in the SN, DMN, and FPN. This multimodal strategy could improve the clinical relevance of EEG findings in MDD.

REFERENCES

[1] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in cognitive sciences,* vol. 14, no. 6, pp. 277-290, 2010.

[2] L. Q. Uddin, B. Yeo, and R. N. Spreng, "Towards a universal taxonomy of macro-scale functional human brain networks," *Brain topography,* vol. 32, no. 6, pp. 926-942, 2019.

[3] W. Marx *et al.*, "Major depressive disorder," *Nature Reviews Disease Primers,* vol. 9, no. 1, p. 44, 2023.

[4] M. Fava and K. S. Kendler, "Major depressive disorder," *Neuron,* vol. 28, no. 2, pp. 335-341, 2000.

[5] R. H. Kaiser, J. R. Andrews-Hanna, T. D. Wager, and D. A. Pizzagalli, "Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity," *JAMA psychiatry,* vol. 72, no. 6, pp. 603-611, 2015.

[6] Z. Fang, E. Lynn, V. J. Knott, and N. Jaworska, "Functional connectivity profiles in remitted depression and their relation to ruminative thinking," *NeuroImage: Clinical,* vol. 45, p. 103716, 2025.

[7] D. A. Pizzagalli, "Frontocingulate dysfunction in depression: toward biomarkers of treatment response," *Neuropsychopharmacology,* vol. 36, no. 1, pp. 183-206, 2011.

[8] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear processes in geophysics,* vol. 11, no. 5/6, pp. 561-566, 2004.

[9] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE transactions on biomedical engineering,* vol. 53, no. 2, pp. 266-272, 2006.

[10] H. Osterhage, F. Mormann, T. Wagner, and K. Lehnertz, "Measuring the directionality of coupling: phase versus state space dynamics and application to EEG time series," *International journal of neural systems,* vol. 17, no. 03, pp. 139-148, 2007.

[11] C. Torrence and P. J. Webster, "Interdecadal changes in the ENSO–monsoon system," *Journal of climate,* vol. 12, no. 8, pp. 2679-2690, 1999.

[12] J.-P. Lachaux *et al.*, "Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence," *Neurophysiologie Clinique/Clinical Neurophysiology,* vol. 32, no. 3, pp. 157-174, 2002.

[13] K. E. Godfrey, S. D. Muthukumaraswamy, C. M. Stinear, and N. Hoeh, "Decreased salience network fMRI functional connectivity following a course of rTMS for treatment-resistant depression," *Journal of Affective Disorders,* vol. 300, pp. 235-242, 2022.

[14] J. Ahn, D. Lee, K. Namkoong, and Y.-C. Jung, "Altered functional connectivity of the salience network in problematic smartphone users," *Frontiers in Psychiatry,* vol. 12, p. 636730, 2021.

[15] C. J. Lynch *et al.*, "Frontostriatal salience network expansion in individuals in depression," *Nature,* vol. 633, no. 8030, pp. 624-633, 2024.

[16] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "Alzheimer's disease: models of computation and analysis of EEGs," *Clinical EEG and Neuroscience,* vol. 36, no. 3, pp. 131-140, 2005.

[17] G. S. Yuen *et al.*, "The salience network in the apathy of late-life depression," *International journal of geriatric psychiatry,* vol. 29, no. 11, pp. 1116-1124, 2014.

TIPTEKNO'25 MEDICAL TECHNOLOGIES CONGRESS
26-28 October 2025
Kaya Artemis Hotel, Famagusta, TRNC

Modeling of Optimal Load Capacity Based on Surface Electromyography and the Borg Scale

Ayse Sabire Kesgin ve Alpaslan Koc
Biomedical Engineering
Samsun University Faculty of Engineering and Natural Sciences Ballıca Campus (Samsun, Türkiye)
aysekesginn@gmail.com, alpaslan.koc@samsun.edu.tr

***Abstract*— Prolonged and intensive use of the musculoskeletal system can lead to muscle fatigue, physiological strain within muscle tissue, and the development of various musculoskeletal disorders. Consequently, there is a need for systems that can prevent individuals from overexerting their muscles and sustaining potential injuries during daily activities or physical tasks. This study aims to develop a muscle-specific load-lifting model by assessing both subjective and objective measures of individual load-carrying capacity. Data were collected from the biceps brachii muscle of a healthy 20-year-old participant, comprising 120 samples recorded using the BIOPAC MP36 system. Subjective assessments were obtained using the Borg CR 10 scale, while objective measurements included the applied muscle force and the average muscle electrical potential recorded within a 2-second interval. Relationships between muscle force and Borg score, muscle force and normalized EMG (sEMG) signals, and Borg score and sEMG signals were analyzed to model the applicable muscle load using sEMG signals. The results demonstrated strong positive correlations across all three pairings. Furthermore, the time-domain integrated sEMG signal was modeled with respect to the Borg score, and at a Borg score of 7, the individual's optimal muscle force capacity was estimated as 16.29 kg. The proposed model offers a valuable alternative for defining safe physical load thresholds and provides a basis for ergonomic load planning.**

Keywords*— *muscle strength; load-lifting modeling; ergonomic load planning; sEMG; Borg CR-10 scale

I. INTRODUCTION

The worldwide prevalence of musculoskeletal disorders reaches 1.71 billion people [1]. People must handle different amounts of weight, changing from light shopping bags to heavy equipment, based on their daily activities, jobs, and sports training [2,3]. The musculoskeletal system experiences physiological strain and metabolic changes in muscle tissue when people use it extensively during long periods [4]. The system that carries physical loads experiences direct effects from this process, leading to the development of muscle fatigue. The symptoms of muscle fatigue include decreased muscle performance, weakened neuromuscular connections, and changes in muscle metabolism [5-7]. The changes in muscle function and tissue stress from these alterations can lead to short-term performance declines and small-scale tissue damage, which may develop into long-term musculoskeletal diseases. The impact of the negative conditions can vary according to inter-individual structural and physiological differences. Therefore, modeling individual optimal load limits is critical for preventing weakness and degeneration of the skeletal muscles.

In the literature, muscle fatigue evaluation has generally relied on either the temporal or frequency characteristics of sEMG signals or purely subjective measurement methods. Troiano et al. investigated the relationship between force and fatigue in the trapezius muscle, combining subjective and objective indicators [8]. They claimed that these indicators can provide information on endurance time and force. Wu et al. focused on detection of fatigue integrating the time and frequency domain features of sEMG with a hybrid TCN and CNN-based model [9]. While these studies successfully validate the utility of sEMG and the Borg scale, their focus remains on general relationships and endurance time prediction; they do not provide a specific predictive model for determining the instantaneous, individual limit value of the maximum load a person can safely lift without overstraining the muscle. Therefore, the developed model enables general muscle endurance evaluation and, more importantly, creates a unique, quantitative method that supports individualized load management based on personal subjective perception.

II. MATERIALS AND METHODS

The study was conducted on a healthy 20-year-old female participant, randomly selected from the Department of Biomedical Engineering students at Samsun University, per ethical approval and voluntary participation principles. Surface EMG (sEMG) signals were recorded from the biceps brachii (m. biceps brachii) muscle using a high-precision BIOPAC MP36 data acquisition system (BIOPAC Systems Inc., Goleta, CA, USA). The signals were acquired in real time through the BIOPAC Student Lab 4.1 software integrated into the system. Fig. 1 shows the data acquisition position.

The participant received an isopropyl alcohol wipe cleaning on their forearm to reduce skin impedance before electrode placement. The forearm was positioned supine, with the elbow slightly flexed downward or at a right angle. Electrodes were placed 3–4 cm apart between the active and reference electrodes, while the ground electrode was positioned over the tendon [10]. At the beginning of the experiment, the participant's maximum average force applied by the dominant arm was measured. Subsequently, the range between zero and

979-8-3315-5565-8/25/$31.00 ©2025 IEEE



Figure 1. The data acquiring position

the maximum average force was divided into four equal intervals to determine four distinct load levels in ascending order. For each of these four threshold force levels, the participant's sEMG signal was recorded for two seconds. The procedure was performed thirty times while the subject rested for five minutes between each attempt. One of the raw EMG signals obtained during these measurements is presented in *Fig. 2*. *Fig. 3* shows force signals corresponding to the four threshold levels exerted by the participant measured using an SS25LB dynamometer. And *Fig. 4* indicates the time-dependent integrated EMG signals. In addition, the perception level of the subject experienced was assessed and recorded using the Borg scale provided in *Table I*.

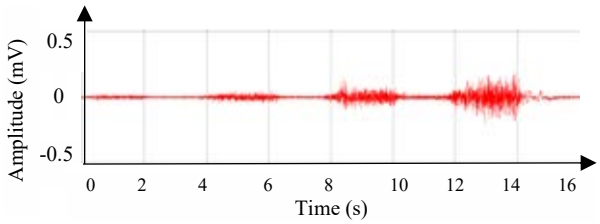


Figure 2. Raw EMG signal

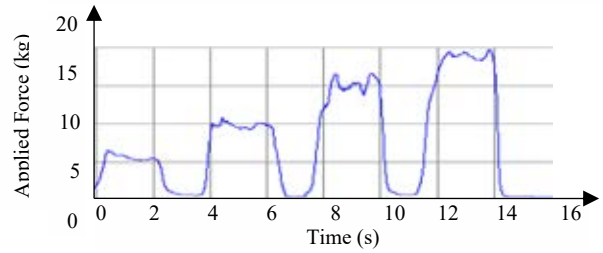


Figure 3. Forces applied during the experiment in 2-second intervals.

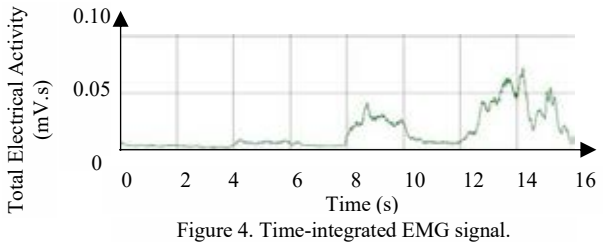


Figure 4. Time-integrated EMG signal.

TABLE I. BORG CR-10 SCALE [11]

Borg Scale of Perceived Exertion	
Score	***Effort Level***
0	Nothing at all (no exertion/pain)
0.5	Very, very easy
1	Very easy
2	Easy
3	Moderate
4	Comfortable
5	Somewhat difficult
6	Difficult
7	Hard
8	Extremely hard
9	Near maximal exertion
10	Maximal exertion / exhaustion

Lower scores on the Borg CR-10 scale indicate that the activity was perceived as very light and caused little to no exertion. In contrast, higher scores represent extremely demanding activities that require maximal effort. The participant was informed to indicate each exercise's perceived difficulty level by selecting the appropriate value from the numerical range based on personal perception. As the scale is subjective, there are no right or wrong responses; rather, it reflects the individual's experienced level of exertion. The sEMG, Borg CR-10, and muscle force data obtained from the participant were statistically analyzed using SPSS software (version 31.0).

III. RESULTS

We analyzed the subject muscle force against their reported perceived exertion levels. By considering the average muscle force as the independent variable and the Borg CR-10 scale scores as the dependent variable, a statistically significant and strongly positive correlation was revealed. The correlation coefficient ($R^2 = 0.824$) and p-value ($p < 0.001$) demonstrate a strong positive relationship between the variables which proves that muscle force averages directly affect perceived exertion levels. The corresponding graph and regression model derived from the data provides a visual illustration of this relationship as shown in *Fig. 5*. The two variables show a strong relationship because of the biological processes which control muscle strength generation and the sense of work intensity. The amount of muscle force people use determines their energy expenditure and metabolic needs which leads to increased feelings of exertion.

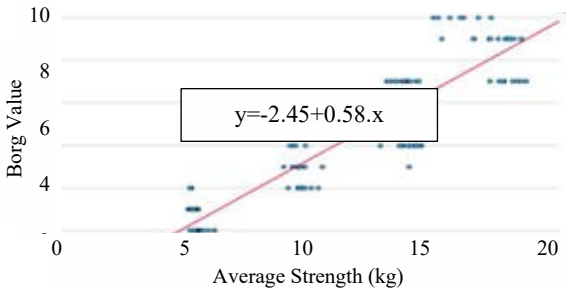


Figure 5. The linear relationship between average muscle strength and Borg CR-10 score

The regression model developed through statistical analysis reveals essential information about the biological processes which link muscle force to electrical activity. By examining the *Fig. 6* and the regression model, it is evident that there is a strong positive correlation between the average applied muscle force and the total electrical activity of the muscle. The muscle activation level rises when the applied force strength increases according to these results.

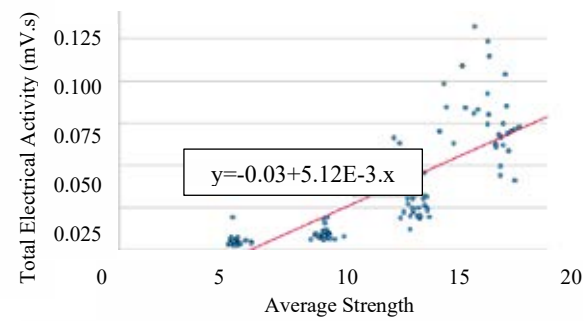


Figure 6. Linear relationship between the applied average muscle force and total electrical activity

The relationship between perceived exertion and muscle electrical activity is a complex phenomenon that has been explored through the utilization of the Borg CR-10 scale and sEMG signals. We used the Borg scale to measure subjective exertion while sEMG signal integration serves as an objective measure of electrical activity to establish their quantitative relationship. As shown in *Fig. 7,* perceived exertion directly correlates with muscle electrical activity because muscle electrical activity increases when perceived exertion increases ($R^2 = 0.722$, $p < 0.001$). The findings of this analysis have significant implications for our understanding of physiological responses to exercise and physical activity. Research that measures both perceived exertion against force and muscle electrical activity against force enables us to correlate the perception of the subject and sEMG signal.

This robust correlation establishes a connection that allows for load management independent of direct force measurement. Specifically, if we can model the sEMG signal based on the individual's perceived exertion scale, we can effectively manage the lifting load by monitoring sEMG values. For example, a Borg score of 7 [11], which represents challenging yet acceptable exertion, corresponds to a calculated muscle force of 16.29 kg and an sEMG value of 0.0534 mVs. This ability to link a subjective, acceptable threshold (Borg score) to an objective, measurable physiological output (sEMG value) offers a powerful tool for personalized load guidance.

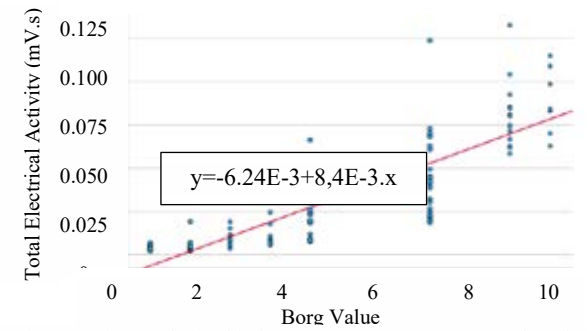


Figure 7 Linear relationship between Borg CR-10 score and total electrical activity

IV. DISCUSSION AND CONCLUSION

The evaluation of muscle fatigue has been a topic of interest in literature, with various studies relying on either the temporal or frequency characteristics of surface electromyography (sEMG) signals or purely subjective measurement methods. Troiano et al. examined the relationship between force and fatigue in the trapezius muscle, combining subjective and objective indicators [8]. The researchers established that these indicators enable scientists to evaluate both endurance duration and force production which are vital for muscle fatigue research. Wu et al. conducted research which presented an opposing perspective to the study. focused on the detection of fatigue by integrating the time and frequency domain features of sEMG with a hybrid Temporal Convolutional Network (TCN) and Convolutional Neural Network (CNN)-based model [9]. The researchers proved that sEMG technology and the Borg scale work effectively as tools for measuring physical exertion. While their approach demonstrated the potential of machine learning models in predicting muscle fatigue, it still relied on general relationships between sEMG features and fatigue. The absence of individualized load management predictive models requires a quantitative system which incorporates personal subjective perception data.

The research conducted by Troiano et al. and Wu et al. have contributed significantly to the understanding of muscle fatigue, they have certain limitations that highlight the need for a more comprehensive approach. A predictive model which unites quantitative and qualitative assessment techniques and considers personal variations in subjective perception will create a customized quantitative system for muscle fatigue assessment. In our study, the strong linear relationship identified between the Borg CR-10 scale and the sEMG electrical activity (as detailed in *Fig. 7*) allowed us to develop a personalized predictive model for load management. This model leverages the known relationship to derive a quantitative guide for the safe limits of physical exertion, which is a unique contribution distinguishing our work from general correlational studies [8] or classification models [9].

Specifically, the established linear regression equations (derived from the data in *Fig. 5, Fig. 6, and Fig. 7*) were used to generate *Table II*, which presents the individual-specific safe limits for the tested muscle and the subject. The Borg CR-10 scores directly match the Estimated Force (kg) values and limited sEMG (mV.s) values which establish the needed force to achieve particular perceived exertion levels. The Borg score of 7 which indicates 'challenging yet acceptable' exertion level corresponds to 16.29 kg as the maximum safe load and 0.0534 mV.s as the measurable sEMG activity threshold for the subject tested.

The *Table II* operates as a system which provides instant decision support to protect the person with the maximum sEMG values that prevent harmful perceived exertion levels (Borg 8 or higher). Real-time sEMG monitoring enables clinicians and coaches to track individual lifting loads which help them enforce

weight limits that protect users from overexertion and musculoskeletal strain.

TABLE II. TABLE OBTAINED THROUGH MODELING

Borg CR-10	Estimated Force (kg)	Limited yEMG (mV.s)
0	4,22	-0,0084
0,5	5,09	-0,004
1	5,95	0,0005
2	7,67	0,0093
3	9,4	0,0181
4	11,12	0,0269
5	12,84	0,0358
6	14,57	0,0446
7	16,29	0,0534
8	18,02	0,0622
9	19,74	0,0711
10	21,46	0,0799

In conclusion, this study demonstrates that the combined assessment of psychophysical and electrophysiological data can define individualized safe loading limits, thereby supporting personalized programs in health, safety, ergonomics, rehabilitation, and sports training.

REFERANCES

[1] "Musculoskeletal health". Accessed: 1 August 2025. [Online]. Access address:https://www.who.int/newsroom/factsheets/detail/musculoskeletal-conditions.

[2] B. D. Fox, L. W. Judge, D. C. Dickin, and H. Wang, "Biomechanics of Military Load Carriage and Resulting Musculoskeletal Injury: A Review", J. Orthop. Orthop. Surg., c. 1, sy 1, ss. 6-11, Nis. 2020.

[3] K. H. D. Tang, "The Prevalence, Causes and Prevention of Occupational Musculoskeletal Disorders", Glob. Acad. J. Med. Sci., c. 4, sy 2, ss. 56-68, Mar. 2022.

[4] T. K. Gill vd., "Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021", Lancet Rheumatol., c. 5, sy 11, ss. e670-e682, Kas. 2023.

[5] D. E. Lidstone, J. A. Stewart, R. Gurchiek, A. R. Needle, H. van Werkhoven, and J. M. McBride, "Physiological and biomechanical responses to prolonged heavy load carriage during level treadmill walking in females," Journal of Applied Biomechanics, 2017.

[6] A. Benwali, Q. R. Ferguson, S. Farahat, R. H. Sandler, E. Hill, and H. A. Mansy, "Non-invasive Evaluation of Muscle Fatigue Using Mechanomyography and Surface Electromyography - A Pilot Study", içinde 2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA: IEEE, Ara. 2022, ss. 1-4.

[7] D. Constantin-Teodosiu and D. Constantin, "Molecular Mechanisms of Muscle Fatigue", Int. J. Mol. Sci., c. 22, sy 21, s. 11587, Eki. 2021.

[8] A. Troiano, F. Naddeo, E. Sosso, G. Camarota, R. Merletti, and L. Mesin, "Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface EMG signal and perceived exertion scale," *Gait & Posture*, vol. 28, no. 2, pp. 179–186, Aug. 2008.

[9] W.-T. Wu, S.-J. Ruan, Y.-W. Tu, Y.-L. Huang, and H.-J. Guo, "A Hybrid Model for Muscle Fatigue Detection: Integrating Time and Frequency Domain Features of sEMG Signals With RPE Scale During Isometric Contraction," IEEE Transactions on Instrumentation and Measurement, vol. 74, 4004411, pp. 1–14, Mar. 2025.

[10] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures", J. Electromyogr. Kinesiol., c. 10, sy 5, ss. 361-374, Eki. 2000.

[11] A. S. Jackson, G. Borg, J. J. Zhang, K. R. Laughery, and J. Chen, "Role of physical work capacity and load weight on psychophysical lift ratings", Int. J. Ind. Ergon., c. 20, sy 3, ss. 181-190, Eyl. 1997.

TIPTEKNO'25

MEDICAL TECHNOLOGIES CONGRESS

26-28 October 2025

Kaya Artemis Hotel, Famagusta, TRNC

Biyomedikal Ve Klinik Mühendisliği Derneği

Biyomedikal ve Klinik Müh. Derneği

Babıali Cad. Ziya Gün Is Hanı No:28 D:12 Cagaloglu - Fatih / Istanbul

www.biyoklinikder.org / e-mail: info@tiptekno13.org

Organizasyon Sekreteryası

BROS

Congress, Incentive and Events

Bilge Yüksel

Esentepe Mah. Sağlam Fikir Sok. Esen Palas Apt. A Blok No:2 D:9 Esentepe / Şişli / İstanbul

Tel : +90 (212) 296 66 70 / **Fax:** +90 (212) 296 66 71

E-mail: bilge.yuksel@brosgroup.net