REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

FUSION OF DYNAMIC AND STATIC FEATURES IN SIGNATURE VERIFICATION

Mustafa Semih SADAK

DOCTOR OF PHILOSOPHY THESIS
Department of Electronics and Communication Engineering
Program of Telecommunications

Supervisor
Assoc. Prof. Dr. Nihan KAHRAMAN

Co-supervisor
Dr. Umut ULUDAĞ

November, 2022

REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

FUSION OF DYNAMIC AND STATIC FEATURES IN SIGNATURE VERIFICATION

A thesis submitted by Mustafa Semih SADAK in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 24.11.2022 in Department of Electronics and Communication Engineering, Program of Telecommunications.

Assoc. Prof. Dr. Nihan KAHRAMAN
Yildiz Technical University
Supervisor

Dr. Umut ULUDAĞ
TÜBİTAK
Co-supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. Nihan KAHRAMAN, Supervisor
Yildiz Technical University _____

Prof. Dr. Ece Olcay GÜNEŞ, Member
İstanbul Technical University _____

Assoc. Prof. Dr. Ahmet SERBES, Member
Yildiz Technical University _____

Assoc. Prof. Dr. Sadiye Nergis TURAL POLAT, Member
Yildiz Technical University _____

Assist. Prof. Dr. Gökalp TULUM, Member
İstanbul Nişantaşı University _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Fusion of Dynamic and Static Features in Signature Verification supervised by my supervisor, Assoc. Prof. Dr. Nihan KAHRAMAN and my co-supervisor, Dr. Umut ULUDAĞ. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Mustafa Semih SADAK

Signature

Dedicated to my family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Nihan KAHRAMAN, for her guidance during this research. Many thanks to Prof. Dr. Ece Olcay GÜNEŞ and Assoc. Prof. Dr. Ahmet SERBES, the thesis monitoring committee members, for their support and encouragement throughout the process. I thank my co-supervisor, Dr. Umut ULUDAĞ, for his mentoring and valuable research ideas. To conclude, my deepest, warm thanks to my family for their belief in me and unconditional support.

Mustafa Semih SADAK

TABLE OF CONTENTS

LIST OF SYMBOLS vii

LIST OF ABBREVIATIONS viii

LIST OF FIGURES x

ABSTRACT xii

ÖZET xiv

1 INTRODUCTION 1

1.1 Literature Review . . . 3

1.1.1 Literature Review on Sound of Signature . . . 4

1.1.2 Literature Review on Offline (Static) Signature Image . . . 11

1.2 Objective of the Thesis . . . 17

1.3 Hypothesis . . . 17

1.4 Contribution . . . 17

1.5 Outline . . . 18

2 DATA EXTRACTION 20

2.1 Pen Types . . . 21

2.2 Paper Types . . . 22

2.3 Phone Models . . . 23

2.4 Data Collection Procedure . . . 24

3 PREPROCESSING 26

3.1 Signature Sound Data Preprocessing . . . 26

3.1.1 Audio-based sound data preprocessing . . . 26

3.1.2 Image-based sound data preprocessing . . . 28

3.2 Signature Image Data Preprocessing . . . 30

4 FEATURE EXTRACTION 31

4.1 Feature Extraction for Audio Data . . . 31

4.2 Feature Extraction for Image Data . . . 33

4.2.1 Shallow (non-deep) learning-based feature extraction 34
4.2.2 Deep learning-based feature extraction 35

5 CLASSIFICATION **38**

6 SHALLOW (NON-DEEP) LEARNING BASED APPROACH **42**
6.1 Test Results 44
6.1.1 Results of the Tests Using Only Static Offline Signature Image Data 44
6.1.2 Results of the Tests Using Only Dynamic Signature Sound Data 46
6.1.3 Results of the Tests Using Fusion of Offline (Static) Signature Data and Signature Sound (Dynamic) Data 50
6.1.4 Verification Results and Analysis when Query and Reference Signature Pen-Paper-Phone Combinations are Different 52

7 DEEP LEARNING BASED APPROACH **56**
7.1 Test Results 58
7.1.1 Results of the Tests for the Deep Learning-Based Approach Using Only Offline (Static) Signature Image Data 59
7.1.2 Results of the Tests Using Only Dynamic Signature Sound Data 60
7.1.3 Results of the Tests Using Fusion of Offline (Static) Signature Data and Signature Sound (Dynamic) Data 64

8 RESULTS AND DISCUSSION **68**
8.1 Future Work 71

REFERENCES **72**

PUBLICATIONS FROM THE THESIS **77**

LIST OF SYMBOLS

Ω	Omega
E	Energy
ε_t	Gaussian White Noise
IR	Impulse Response
Q	Query Signature
R	Reference Signature
σ	Sigma
θ	Theta
δ	Delta
$C_{l,r}$	Curvelet Coefficient
T	Time
N	Window Size
h	Hop Size

LIST OF ABBREVIATIONS

ROC	Receiver Operating Characteristics
FAR	False Accept Rate
FRR	False Reject Rate
GAR	Genuine Accept Rate
EER	Equal Error Rate
AER	Average Error Rate
CNN	Convolutional Neural Network
DTW	Dynamic Time Warping
SFOSE	Spectral Flux Onset Strength Envelope
SC	Spectral Centroid
WD	Writer Dependent
WI	Writer Independent
AR	Autoregressive
SINR	Signal to Interference plus Noise Ratio
STD	Trend-Seasonal Decomposition
SSIM	Structural Similarity Index Measure
PSNR	Peak Signal-to-Noise Ratio
MSE	Mean Squared Error
LR	Logistic Regression
NB	Naive Bayes
RF	Random Forest
SVM	Support Vector Machine
DTW	Dynamic Time Warping

I	In-phase
Q	Quadrature
LOS	Line of Sight
ZC	Zadoff–Chu Sequence
STFT	Short Term Fourier Transform
HOG	Histogram of Oriented Gradients
LBP	Local Binary Patterns
ULBP	Uniform Local Binary Patterns
NN	Nearest Neighbour
DRT	Discrete Radon Transform
PNN	Probabilistic Neural Network
FV	Fisher Vector
SIFT	Scale Invariant Feature Transform Descriptors
PCA	Principal Component Analysis
GLCM	Gray Level Co-occurrences Matrix
HPFI	High Priority Index Feature
SKcPCA	Skewness-Curtosis controlled PCA
OC-SVM	One-Class Support Vector Machine
RBF	Radial Basis Function
LSTM	Long-Short Term Memory
BLSTM	Bidirectional Long–Short Term Memory (BLSTM)
RNN	Recurrent Neural Network
OSS	Onset Strength Signal
FFT	Fast Fourier Transform
OCC	One-Class Classification
CIR	Channel Impulse Response

LIST OF FIGURES

Figure 1.1 CNN Architecture [13] 11
Figure 2.1 Two mobile phones are displayed in an experimental setup with a BIC Cristal ballpoint pen and thin paper with the auto-copy feature (To the right is a rollerball fine-point pen) 20
Figure 2.2 BIC 0.5mm extra fine point rollerball pen 21
Figure 2.3 BIC 1mm Cristal disposable ballpoint pen 21
Figure 2.4 Scanned A4 plain paper (80 g/m2 - 24 lb.) 22
Figure 2.5 Scanned A5 thin paper with auto copy feature (55 g/m2 - 15 lb.) . 23
Figure 3.1 Signature Sound Data Preprocessing 27
Figure 3.2 Handwritten signature 28
Figure 3.3 Handwritten signature of a participant as raw sound signal data . . 28
Figure 3.4 Image graphs of the audio signal's onset strength envelope and spectral centroid produced from the signature: a) Image of original spectral-flux onset strength envelope graph b) Image of preprocessed spectral-flux onset strength envelope graph c) Image of original spectral centroid graph d) Image of preprocessed spectral centroid graph 29
Figure 3.5 Preprocessing phases for static handwritten signature a) Original image b) Grayscale image c) Preprocessed image d) Inverted image 30
Figure 4.1 Raw signal 32
Figure 4.2 Onset strength envelop of signal 33
Figure 4.3 Spectral centroid of signal 33
Figure 4.4 Operators for different radius values and point counts: a) Radius=1, Number of points=8. b) Radius=2, Number of points=8. c) Radius=2, Number of points=16. 34
Figure 4.5 One of the architectures employed in SigNet [24], A series of transformations using convolutional layers, max-pooling layers, and fully-connected layers are applied to the input image. 36
Figure 5.1 Separation of two different classes by hyper planes 38
Figure 5.2 Block diagram of the proposed shallow learning-based approach for multimodal signature verification (sound and image) 39

Figure 5.3 Block diagram of the proposed deep learning-based approach for multimodal signature verification (sound and image) 40
Figure 6.1 Flowchart for the proposed methodology of sound and multimodal signature verification 43
Figure 6.2 ROC curve for static offline signature verification according to averaged error rates obtained from all pen-paper combinations. . . 46
Figure 6.3 ROC curves for spectral flux onset envelope of audio data, spectral centroid of audio data and fusion of spectral flux onset envelope and spectral centroid of audio data 48
Figure 6.4 ROC Curves for signature sound (dynamic) data only, offline (static) signature data only, and fusion of signature sound (dynamic) data and offline (static) signature data 51
Figure 7.1 Flowchart for deep learning-based approach 57
Figure 7.2 ROC curve for static offline signature verification according to averaged error rates obtained from all pen-paper combinations. . . 60
Figure 7.3 ROC curves for spectral flux onset envelope of audio data, spectral centroid of audio data and fusion of spectral flux onset envelope and spectral centroid of audio data 62
Figure 7.4 ROC Curves for signature sound (dynamic) data only, offline (static) signature data only, and fusion of signature sound (dynamic) data with offline (static) signature data 65
Figure 8.1 Comparison of genuine signature image with skilled forgery image: a) Genuine signature image b) Forged signature image 70
Figure 8.2 Comparison of graphic images obtained from signature sound data a) Image of Onset Strength Envelope of genuine signature sound b) Image of Onset Strength Envelope of forged signature sound c) Image of Spectral Centroid of genuine signature sound d) Image of Spectral Centroid of forged signature sound 70

ABSTRACT

Fusion of Dynamic and Static Features in Signature Verification

Mustafa Semih SADAK

Department of Electronics and Communication Engineering
Doctor of Philosophy Thesis

Supervisor: Assoc. Prof. Dr. Nihan KAHRAMAN
Co-supervisor: Dr. Umut ULUDAĞ

In biometrics, accurately verifying individuals with handwritten signatures is one of the most challenging problems due to high intra-class and low inter-class variability. In this thesis, to help overcome this difficulty, the sound produced by the friction of paper and pen during the signing process is evaluated as separate biometric data. Some datasets in the literature, like the GPDS and MCYT, which are accessible to the general public, only contain static signature images. However, the signature sound data required for this research is not available in any publicly accessible dataset. Therefore, in this study, a new dataset consisting of genuine and forged signatures is built from scratch by collecting samples from 93 participants. Each participant is asked to sign on two different paper types using two different pen types. Four samples are taken for each paper-pen combination. The sound emerging from each signature is recorded with the internal microphones of two particular mobile phone models including different operating systems. As a result, a dataset consisting of signature sounds and corresponding signature images is constructed. These two data types (signature sound and signature image) are evaluated for biometric signature verification, both together and separately. For the feature extraction stage, spectral flux onset envelope and spectral centroid graphs of the sound data are plotted, and these graphs are converted to image files. Afterward, feature vectors representing dynamic sound data are obtained from these image files employing local binary patterns (LBP) and scale-invariant feature transform (SIFT) algorithms. Feature vectors of static signature images are also obtained by performing LBP and SIFT algorithms. As a classifier, the

one-class support vector machine (OC-SVM) is trained with only genuine signatures obtained from each user. Forgeries are used for testing only. In the second approach (deep learning-based) proposed as an alternative in this study, feature extraction is conducted with a convolutional neural network (CNN) based deep learning algorithm instead of LBP and SIFT. The results are compared with the shallow (non-deep) learning-based approach. Signature verification is performed with only dynamic signature sound data, only static signature image data, and the fusion of both sound and image data. According to different pen-paper-phone model combinations used, the equal error rates (EER) obtained using only sound data are in the 1.08-5.38% (Average: 3.02%) range for the shallow learning-based approach and the 1.94-5.59% (Average: 3.48%) range for the deep learning-based approach. It is also observed that the fusion of sound and image further increased the verification success to EER of 0.00-1.08% (Average: 0.29%) interval for the shallow learning-based approach, similarly EER of 0.00-2.15% (Average: 0.67%) interval for the deep learning-based approach.

Keywords: score-level fusion, feature fusion, onset detection, signature verification, local binary patterns, scale-invariant feature transform, image processing, audio signal processing, support vector machines, convolutional neural networks

ÖZET

İmza Doğrulamada Dinamik ve Statik Özelliklerin Birleştirilmesi

Mustafa Semih SADAK

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Nihan KAHRAMAN
Eş-Danışman: Dr. Umut ULUDAĞ

Biyometri alanında, bireylerin el yazısı imzalarıyla başarılı bir şekilde doğrulanması, sınıf içi yüksek ve sınıflar arası düşük değişkenliğin olması nedeniyle en zorlu yöntemlerden biridir. Bu tezde, bu zorluğun üstesinden gelinmesine yardımcı olmak için, imzalama sırasında kağıt ve kalemin sürtünmesiyle ortaya çıkan ses, ayrı bir biyometrik veri olarak değerlendirilmiştir. Literatürde, GPDS ve MCYT gibi bazı erişime açık veri tabanları yalnızca statik imza görüntüleri içerir. Ancak bu çalışma için gerekli olan ses verilerini içeren erişime açık bir veri tabanı bulunmamaktadır. Bu nedenle, bu çalışmada 93 katılımcıdan örnekler toplanarak orijinal ve sahte imzalardan oluşan yeni bir veri tabanı sıfırdan oluşturulmuştur. Her katılımcıdan iki farklı kalem türü kullanarak iki farklı kağıt türüne imza atması istenmiştir. Her bir kağıt-kalem kombinasyonu için dört imza alınmıştır. Her imzadan çıkan ses, ayrı işletim sistemlerine sahip iki farklı cep telefonu modelinin dahili mikrofonları ile kayıt altına alınmıştır. Sonuçta imza sesleri ve bu seslere karşılık gelen imza görüntülerinden oluşan bir veri seti üretilmiştir. Bu iki veri tipi (imza sesi ve imza görüntüsü) hem birlikte hem de ayrı ayrı olarak biyometrik imza doğrulama için değerlendirilmiştir. Öznitelik çıkarma aşaması için ses verilerinin spektral akı başlangıç zarfı ve spektral merkez grafikleri çizilmiş ve bu grafikler görüntü (image) formatına dönüştürülmüştür. Bu görüntü dosyalarından yerel ikili desenler (LBP) ve ölçek değişmez özellik dönüşümü (SIFT) algoritmaları ile dinamik ses verilerini temsil eden özellik vektörleri elde edilmiştir. Statik imza görüntülerinin öznitelik vektörleri de LBP ve SIFT algoritmaları ile elde edilmiştir. Sınıflandırıcı olarak, tek-sınıf destek

vektör makinesi (OC-SVM), her kullanıcıdan alınan orijinal imzalarla eğitilmiştir. Sahte imza örnekleri sadece test için kullanılmıştır. Bu çalışmada alternatif olarak önerilen ikinci yaklaşımda (derin öğrenme tabanlı), LBP ve SIFT yerine evrişimsel sinir ağları (CNN) tabanlı bir derin öğrenme algoritması ile öznitelik çıkarımının gerçekleştirildiği bir imza doğrulama sistemi geliştirilmiş ve derin öğrenme tabanlı olmayan (sığ) yaklaşımla karşılaştırılmıştır. İmza doğrulama yalnızca dinamik imza ses verileri, yalnızca statik imza görüntü verileri ve hem ses hem de görüntü verilerinin birleştirilmesiyle gerçekleştirilmiştir. Kullanılan farklı kalem-kağıt-telefon modeli kombinasyonlarına göre, yalnızca ses verileri kullanılarak elde edilen eşit hata oranları (EER), sığ öğrenme tabanlı yaklaşım için %1.08-5.38 (Ortalama: 3.02%) aralığında ve derin öğrenme tabanlı yaklaşım için %1.94-5.59 (Ortalama: 3.48%) aralığında olmuştur. Ayrıca, ses ve görüntünün birleştirilmesinin, sığ öğrenme tabanlı yaklaşımda %0.00-1.08 (Ortalama: 0.29%) EER aralığına ve benzer şekilde derin öğrenmeye dayalı yaklaşımda %0.00-2.15 (Ortalama: 0.67%) EER aralığına doğrulama başarısını yükselttiği gözlemlenmiştir.

Anahtar Kelimeler: skor düzeyinde füzyon, özellik füzyonu, başlangıç tespiti, imza doğrulama, yerel ikili örüntüler, ölçek-değişmez özellik dönüşümü, görüntü işleme, ses sinyali işleme, destek vektör makineleri, evrişimsel sinir ağları

1
INTRODUCTION

Biometrics refers to measurable, distinctive data, including iris, face, fingerprint, palm vein map, voice, handwritten signature, and so forth, that serves to identify or verify individuals. This data can be categorized into behavioral biometrics and physical biometrics. Physical biometrics, such as the iris, face, and fingerprint, is the unique data that a person has without revealing his/her will. In behavioral biometrics, characteristic features such as voices, gestures, and handwritten signatures arise from consent and appropriate actions of individuals [1].

A handwritten signature is a behavioral biometric that represents people's will and is used to validate and adopt texts in many areas such as finance, education, health, security, and law. Thus, the success of verification of this biometric data, which causes people to assume responsibilities, is of great importance. Handwritten signature biometric systems, which are frequently used in daily life, can be evaluated under two categories: signature identification and signature verification. The signature identification process is to determine whether an individual's signature biometric is included in any of the signature classes previously introduced to the system. Signature verification, on the other hand, is to determine whether the signature belongs to a particular person (with claimed identity) or not. In addition, according to the technical characteristics of the signature biometric systems, the cited signatures can be divided into offline signatures and online signatures [2]. Online signatures are taken with the help of digital surface devices such as pressure-sensitive tablets. Thus, while signing, pen movements are monitored with motion detectors. Dynamic features of the signature, like pen pressure, the number of strokes, speed, and duration are captured together with its static shape. An offline signature is a static image of a signature on a piece of paper or a non-digital similar object. In offline signatures, there is no dynamic data of the signature, only the static form of the signature.

In signature verification research, the impact of methods, algorithms, and designs used in preprocessing, feature extraction, and classification stages is crucial. Recent research has revealed two commonly used methods for classification or feature

extraction: writer-dependent (WD) and writer-independent (WI) methods. Studies have also suggested combining these two methodologies to develop hybrid techniques. A unique classifier is prepared for every single participant to train the model in the WD classification strategy. A global classifier training model for all users is prominent in the WI classification technique. In the WI technique, there is just one classifier, and it is suitable for all users. These two approaches (WD and WI) are valid not only for classification purposes but also for the training networks and algorithms used in the feature extraction stages. In this thesis, a writer-independent procedure is used in the feature extraction step, while a writer-dependent technique is used in the classification step. In addition to all these, the number of signature samples used for the training of the signature verification system and whether forgeries are used together with genuine signature samples in the training of the system or not are also substantial. Because in real-life applications, many signature samples belonging to one person may not be available [3]. Likewise, forgeries of a person's signature are often unavailable beforehand. As a result, factors such as training a signature verification system with as few samples as possible and not using forgeries during training make that system more applicable. Forgeries can be used to test the system. In this case, the type of forgeries used in the system plays a key role. Forgeries may be examined under three headings: a) Random Forgery, b) Simple Forgery, and c) Skilled Forgery. If the forger does not know the name of the person whose signature he/she is imitating and has never seen his/her signature, such signatures are random forgeries. If the forger knows the name of the person whose signature he/she is imitating and has never seen his/her signature, such signatures are simple forgeries. If the forger has seen the signature of the person he/she is imitating and has practiced repeatedly to imitate that signature, such signatures are skilled forgeries. Skilled forgery-type signatures are the most effective samples to test for evaluating the system because such forgeries are in the category that most closely resembles genuine signatures. To be appropriate in real-world applications, the classifier network in this study was trained using just four genuine signatures. Skilled forgeries, the most compelling of forgery types for verification accuracy, were used only for testing purposes in the proposed system. The False Rejection Rate (FRR) of genuine signatures, the False Acceptance Rate (FAR) of forgeries, and the Genuine Accept Rate (GAR) (GAR=1-FRR) are typically used to measure system performance. Additionally, it is usual practice to report other metrics such as the Average Error Rate (AER), which is the average of FAR and FRR, or the Equal Error Rate (EER), which is the error rate where both FAR and FRR are equal. Another indicator of verification findings is the Receiver Operator Characteristics (ROC) curve, which is a graphical representation linking GAR and FAR acquired at differing acceptance thresholds.

In business, commerce, legal, and other related fields, signing documents can subject people to significant financial and moral obligations. Since signatures are such a common form of verification, people who wish to harm may abuse them or use them in fraudulent transactions. The detection and verification of the validity or forgery of signatures, and biometric features of individuals, is thus an essential research topic. The biggest challenge that signature verification systems face is high intra-class variability and low inter-class variability. That is, the similarity between an individual's signatures may be low, and the similarity between his signature and a forgery provided by someone else may be high. Using more than one biometric data is one of the helpful approaches to overcome this difficulty. When two or more different biometric features are combined, it is called multimodal biometric or biometric fusion and brings high confidence to verification or identification success (i.e., iris and fingerprint). Biometric approaches that verify or recognize only based on a single biometric data are called unimodal biometric systems [4]. In this thesis, in addition to offline signatures, the sound generated during signing as dynamic data is recorded as another biometric data. Hence the data required for successful verification is enriched. So, a kind of multimodal biometric system that makes verification using these two particular biometric data (static and dynamic) is designed. The motivation behind employing sound data is a dynamic sound signal is more difficult for imitators to imitate than a static image of the signature. In addition, acquiring, copying, and recording audio data is more complex than image data. Also, it is easier to record audio containing dynamic data with the help of just a microphone rather than using tablets with a customized digital surface, as is the case with capturing online signatures [5]. Besides, signing on such digital surfaces can be a bit clumsy. Moreover, the cost of processing separate raw data such as pressure, velocity, time, and acceleration obtained with tablets or devices used in online signature verification systems is higher than the cost of processing audio data recorded with a microphone, which contains an associated summary form of these data in a single signal [6].

Outcomes from the combination of signature sound and signature image used in signature verification for 75 participants are published as an article [7]. Furthermore, in a conference paper [8], preliminary findings were provided in the scope of this study, utilizing just the sound of signatures to extract features.

1.1 Literature Review

Signature verification studies have progressed in various stages from the 1970s to the present. Studies on online signatures were included later in the process that initially started with offline signatures. As a result, many methods, algorithms, and techniques

have been applied and published to date for online and offline signature verification. However, there are other studies published elsewhere on verifying the signature by analyzing the sound generated from the movement of the pen on paper or evaluating this sound as an extra feature, which relatively is a new approach investigated in this study, although not many. So the literature review has been conducted under two categories; a) Literature Review on Signature Sound (Dynamic) and b) Literature Review on Offline (Static) Signature Image

1.1.1 Literature Review on Sound of Signature

Many types of research in which signature sound, handwriting sound, and even the sound of signing on the tablet are evaluated separately in the studies published to date in terms of their performance in data collection, preprocessing, feature extraction, and classification steps. The detailed information of these studies is summarized in Table 1.1.

Table 1.1 Summary of the Literature Review on Sound of Signature

Study	**Number of Samples**	**Preprocessing**	**Feature Extraction**	**Classification**	**Results**
Seniuk and Blostein (2009) [9] (handwritten word or letter recognition)	Samples provided by a single writer	Gaussian smoothing, Rescaling power signal, Segmentation, Length normalization	Peak points and the structures obtained from scale space representations	Classification algorithms based on template matching	70% (alphabet) recognition rate and 90% word recognition rate
Li (2004, 2010) [6] [5]	50 genuine, 50 forged from each of 5 participants	Band pass filter, Segmentation, Rescaling, Normalizing	Normalized Hilbert envelope of sounds	A straightforward multi-layer back propagation neural network	More than 75% correctness for different scenarios
Khazei et al. (2012) [10]	10 genuine signatures from each of 30 participants	Segmentation, Normalization	Autoregressive (AR) coefficients with Burg algorithm, cepstrum based features obtained with the log-domain bicepstral method	Distance-based classifiers: Euclidean, Manhattan (Cityblock) and Chessboard.	EER between 49% and 50.133%
Armiato et al. (2016) [11]	10 genuine, 2 forged from each of 55 participants	Normalizing signals amplitudes	Combination of wavelet-based features	Euclidean classifier and Modified correlation classifier	Above 80% accuracy

Continued on next page

Table 1.1 – continued from previous page

Study	Number of Samples	Preprocessing	Feature Extraction	Classification	Results
Du et al. (2018) [12] (handwritten word or letter recognition)	-	Reducing sampling rate, Noise removal, Enhancing the Signal to Interference plus Noise Ratio (SINR), Energy normalization, Removing the dc component	Deep features using CNN	CNN	81% word recognition accuracy rate
Ding et al. (2019) [13]	112 genuine and 60 forged from each of 14 participants	Down converting signal, Reducing sampling frequency, Noise removal with Seasonal-Trend Decomposition (STD)	A novel chord-based method, to estimate phase-related changes caused by small activities	Deep CNN	EER: 5.5% AUC: 98.7%
Chen et al. (2020) [14]	20 genuine, 20 forged from each of 35 participants	Noise removal, Performing cross-correlation function to obtain impulse response	The structural similarity index measure (SSIM), Peak signal-to-noise ratio (PSNR), Mean squared error (MSE), and Hausdorff distance.	Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM)	EER: 1.25% AUC: 98.2%
Sadak et al. (2020) [8]	16 genuine, 16 forged from each of 40 participants	Segmentation, Normalization, Reducing sampling frequency	Picked peaks from spectral flux onset strength envelope of signal	Distance based classifier: Dynamic Time Warping (DTW)	EER: between 8.14% and 15.29%
Wei et al. (2021) [15]	70 genuine signatures, 60 random , 60 skilled forgeries from each of 12 participants	Bandpass filter for noise removal, Adaptive thresholds for segmentation of sound and vibration data, Normalization	Zero Crossing Rate, Spectral Centroid, Spectral Spread, Sprectral Flux, Spectral Entropy, Spectral Rolloff	One-Class classifier based on CNN	EER: 5% AUC: 98.4%
Zhao et al. (2021) [16]	32 genuine, 28 forged from each of 40 participants used in training	Smoothing, Segmentation, Phase Unwrapping, Zero Padding	Spatio-Temporal Features from Channel Impulse Response (CIR)	CNN-based Multi-Modal Siamese Network	EER: 3.27%

1.1.1.1 Data Extraction

The public offline signature databases like CEDAR [17] and GPDS [18] do not currently have the sound signals of offline signatures that this new technique requires. Since there is no signature dataset consisting of audio data publicly available, researchers who are interested in the topic must build their databases from scratch.

It is important to position the microphone so that the sounds of the signatures are well recorded. Li (2004) [6] tested the quality of the sound data obtained in each position separately by placing the microphone under the paper, at a distance of 20-25 cm from the pen-nib, and finally on the pen by attaching it to the pen. As a result of the test, it was concluded that the most efficient sound recording was obtained with the 3rd option, that is, the microphone mounted on the pen. Li [5], [6] used 50 samples of the author's "This pen" calligraphy to build the dataset, apart from 50 genuine samples, 50 forgeries signed by 5 volunteers, 100 random signing sound samples and 200 random texts of the same length with "This pen" phrase were also included. Volunteers were given time to practice before asking them to forge signatures. Khazaei et al. [10] collected ten signatures per person for the dataset they built with 30 participants. Signatures were taken in an isolated environment using the C2 condenser microphone, electronic pen, and pad. They recorded signals with a sampling frequency of 44100 Hz. Armiato et al. [11] developed a special device in the form of a 32×22×10 cm box to record the acoustic emissions generated during the signing. The box's interior is lined with soundproof foam, and a microphone is affixed to it. They obtained a total of 550 genuine signature samples and 110 forgery samples from 55 people, including 10 genuine signatures and 2 forgeries from each participant. All participants used the same kind of pen and paper and were in the same data collection setup while collecting samples. The sampling rate for each audio data was 44100 Hz. Seniuk and Blostein [9] analyzed the sound made up of acoustic emissions caused by friction between the pen and the surface when writing some text by a writer. They tried to extract characters, words, and meanings from this audio data. They built the dataset using 60 lb HP laser printer paper, Bic Round Stic Grip (fine) pen, and Labtec PC Mic 333 microphone. The sampling rates of the collected samples were 16 kHz, with 8 bits. Samples were saved in wav extension format and written by one person (Seniuk). Ding et al. [13] collected data by recording audio with the mobile application they developed, which uses the built-in microphone of the Samsung Galaxy s6. First, they collected 112 genuine signatures per person from 14 graduate students. They then asked students to generate 12 skilled forgeries for each of the five random participants from 14 students. As a result, they obtained a dataset including 112 genuine signature samples and 60 skilled forgery samples for each of the 14 students. Chen et al. [14], took a different approach and asked the candidates to sign using

the iPad Pro and Apple pencil. They captured the signature sounds produced by the friction between the iPad Pro and the Apple pencil with the SAMSUNG galaxy note 8 and its built-in microphone. With a mobile app they developed called SilentSign, they collected 20 genuine signatures from each of the 35 participants of various age groups, nationalities, and genders. Using the iOS screen recording function, they recorded the screen of how the genuine signatures were signed. Each participant watched the videos of the five randomly selected participants to imitate. After enough practice, each participant provided a total of 20 skilled forgeries, 4 for each randomly selected participant. Additionally, they extracted 15 random forgeries for each participant by taking one signature from the genuine signatures of the other 15 participants. As a result, the dataset was created with a total of 700 genuine signatures, 700 skilled forgeries, and 525 random forgeries from 35 participants. Zhao et al. [16] collected 40 genuine signature sounds and 35 forged signature sounds from each of the 40 participants using HONOR Play3 including the Android 10 operating system. Almehmadi [19] took a different approach by comparing the information manually obtained from the signature sounds (number of straight lines, number of circles, number of angles, number of dots, etc.) with the information obtained from the image of the signature. He conducted his study in a silent place with 20 participants, ranging in age from 25 to 53.

1.1.1.2 Preprocessing

Audio files are obtained by using mobile applications in some studies, microphones, computers, and recording devices in others. Making these audio files efficient for feature extraction is very important as it affects the success of the classification. One of the basic steps taken for this purpose to segmentation by cropping the parts from the beginning and end of the sound file that does not contain the necessary information of the signing process, as in [5] [6]. Li [5] filtered the digitized signals with a 250 to 6000 Hz 4th order bandpass filter to reduce ambient noise. In the same study, sound signals varying in lengths between 2.3 seconds and 3.7 seconds were rescaled taking 100 evenly spaced samples according to the equation;

$$T = 1/f = \frac{t_2 - t_1}{100} \tag{1.1}$$

where t_1 is the starting and t_2 is the terminating time of the sound signal. After the segmentation and rescaling steps, the envelope of the sound signal was normalized to its total energy;

$$E[n] = \frac{e[n]}{\sqrt{\sum_{i=1}^{100} e^2[n]}} \tag{1.2}$$

Gaussian smoothing is applied to the absolute value (modulus) of the sound signal by Seniuk and Blostein [9]. Using amplitude normalization, they rescaled the power signal to reach the maximum amplitude of unity. As a segmentation practice, they semi-automatically segmented the signal into constituent letters or words. In the normalization phase, they normalized each segment to a standard time interval. They distorted the signals to be used in tests to evaluate the robustness of classification algorithms. The signals are distorted by applying a uniform perturbation function to the amplitudes of the sound signals. Khazaei et al. [10] performed only segmentation and normalization as preprocessing steps in their work. The average of all audio data was calculated by Arimato et al. [11], and this value was subtracted from each audio signal for normalization purposes. They also normalized the amplitudes of the signals to be between the range of -1 and 1. Ding et al. [13] down-converted the received sound signal by multiplying $cos2\pi ft$ and $-sin2\pi ft$ together with the low-pass filter to extract the In-phase (I) and the Quadrature (Q) components of the baseband signal and the resulting data is down-sampled by a factor of 300. They also reduced the sampling frequency of the signal from 48 kHz to 160 Hz. To remove the periodic ambient noise, they used a function-based method called Seasonal-Trend Decomposition (STD). Chen et al. [14] tried to determine the vertical distance locations of the pen tip to the microphone and aimed to derive features through these points. A special "Adaptive Energy-based Line of Sight (LOS) Detection" technique was developed for this purpose. They use the cross-correlation function to extract impulse responses after the recording has begun;

$$IR(t) = ZC_R^*(-t) * ZC_{1024bits}(t) \tag{1.3}$$

where $ZC_R^*(-t)$ is the conjugation of the incoming baseband signal. It was presumed that the residual noise power would resemble a Gaussian distribution and they adopted certain formulae for noise elimination. The approximate beginning of the LOS path was then determined using an adaptive energy-based technique in the next stage. Du et al. [12] have a study about acoustic-based handwriting recognition. They considered the handwriting sound that arises from the friction between paper and pen as the main feature. By segmenting the auditory stream, they sought to isolate and recognize the sounds of each letter. First, they reduced the sampling rate of 44100 Hz sounds to 4410 Hz, which is sufficient to extract handwriting signal features. In their work, an adaptive threshold according to ambient noise which varies significantly over time is set for peak detection. To deal with the negative effect of varying ambient noise, they used two sliding windows of different sizes. They tried to eliminate the noise in the sound by calculating and comparing the power of the noisy parts with one window, and the power of the meaningful parts with the

other window. Word and letter segmentation was made according to a certain time threshold T_{word} and T_{letter} between the peak values in the signals. They removed the DC component to amplify the dynamic component. Specifically, they calculated the average level of the sound signal and subtracted it from the sound signal. To improve the Signal-to-Interference-plus-Noise Ratio (SINR), the high-frequency component was removed using a low-pass filter. Finally, they used sound signal energy normalization. Sadak et al. [8] reduced sample rates of each audio signal, they converted signals to 'wav' extension format and performed normalization and segmentation for each signature sound data. Zhao et al. [16] employed smoothing, handwritten motion segmentation, phase unwrapping, and zero padding for Channel Impulse Respons (CIR) information extracted from acoustic signals.

1.1.1.3 Feature Extraction

In the studies published to date, for the feature extraction phase, various approaches have been developed. Signing sound's Normalized Hilbert envelope was employed by Li [5] [6] as a feature space. He obtained the envelope signal by the equation;

$$e[n] = |s[n] + jH\{s[n]\}| \tag{1.4}$$

where $e[n]$ is Hilbert transform envelope of signal, $s[n]$ is n^{th} sample of the writing sound and $H\{s[n]\}$ is the discrete Hilbert transform of $s[n]$. Peak points and structures derived from scale-space representations were regarded by Seniuk and Blostein [9] as feature vectors. Khazei et al. [10] extracted Autoregressive (AR) coefficients with the Burg algorithm and they obtained cepstrum-based features with the log-domain bicepstral method. The mathematical representation of AR is as follows:

$$x_t = \sum_{i=1}^{N} a_i x_{t-i} + \varepsilon_t \tag{1.5}$$

where N is the order of the filter, a_i is the i^{th} AR coefficient, x_t is the sound signal series, and ε_t is assumed as Gaussian white noise. Armiato et al. [11] designed a feature extraction approach based on the combination of wavelet-based features. They employed the wavelet-packet transforms, trying 48 different wavelet filters. To estimate phase-related changes caused by small activities, Ding et al. [13] adopted a chord-based method. They extracted frequency-domain features using a discrete cosine transform (DCT), considering the estimated phase-related changes. They obtained 8 pairs of $\Delta Chord$ and $Acceleration$ sequences for each sound signal sequence. Then DCT is used to get the frequency-domain characteristics for each $\Delta Chord$ or $Acceleration$ sequence scaled to a 0-1 range. Thus, they

achieved effective size reduction by adopting DCT. Chen et al. [14] calculated The Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Hausdorff Distance as features and they constructed a four-dimensional similarity feature vector from two Impulse Responses IR_A and IR_B. They called the feature vector genuine or forgery, depending on whether IR_A and IR_B are produced from the same genuine signature dataset. Du et al. [12] transformed the sound signal into a gray-scale image. They obtained the time-ordered frequency feature sequence by applying Short Term Fourier Transform (STFT) to the audio signal. In their study, Convolutional Neural Network (CNN) algorithm-based deep features are determined from the images, instead of artificial or handcrafted features. Sadak et al. [8] extracted time-ordered spectral flux onset points as features. Wei et al. [15] adopted Zero Crossing Rate, Spectral Centroid, Spectral Spread, Spectral Flux, Spectral Entropy, and Spectral Rolloff features. Zhao et al. [16] used Spatio-Temporal characteristics taken from the CIR data as features.

1.1.1.4 Classification

Many approaches in the studies published to date with the classification phase in which signatures are labeled as genuine or forgery. Li [5] [6] proposed a straightforward multi-layer back propagation neural network presented for classifying patterns of signature sounds. He used multi-layer network architecture with linear basis and sigmoid activation functions which are trained by a back-propagation algorithm. In his approach, the neural network includes 100 input neurons that receive input vectors containing 100 data points from the pre-processor. Moreover, there are two hidden layers with 40 and 5 neurons, respectively, and a single output neuron. In his studies, a single continuous node was used to facilitate the training process, and a predefined threshold was used for decision-making. Seniuk and Blostein [9] investigated three approaches based on template matching; Integrating the absolute difference between two signals is the first one. In this approach, training data is used directly as a set of templates. Then similarity of a query signal to a template is defined as the integral of the absolute value of the difference between the two timespan-normalized and amplitude-clipped signals. If the two signals are the same, the integral is equal to zero. Bigger numbers mean less similarity. The second one is to use the edit distance algorithm to compare the two signals' time-ordered peaks. The final one is comparing the structures produced from the signal representations in scale space. In this last approach, using structure and characteristics obtained from a scale space representation of the signal, they devised an algorithm independent of window size. Khazei et al. [10] compared three distance-based classifiers: Euclidean, Manhattan (Cityblock), and Chessboard. According to the results, the top-performing classifier

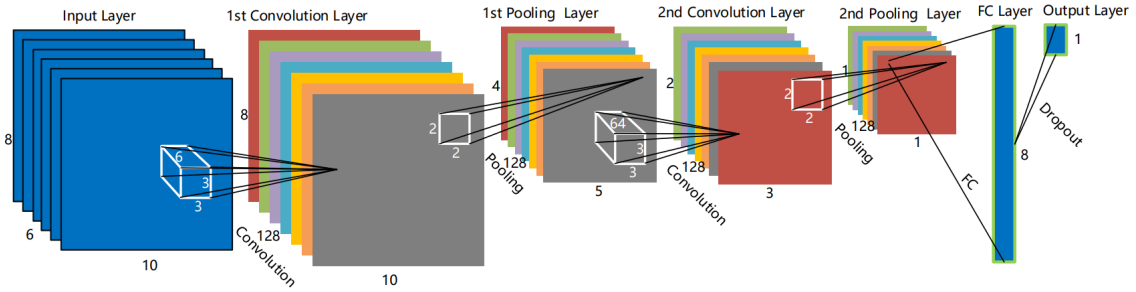


Figure 1.1 CNN Architecture [13]

was the one that used Euclidean distance and complex cepstrum features. Armiato et al. [11] proposed two separate classification approaches to compare the feature vectors. Euclidean classifier is their primary option because of its ease of use and prior pattern-matching findings. A modified correlation classifier, which is based on the notion of signal similarity, is their alternative option. This method combines two criteria to generate a similarity score using cross-correlation. According to their test results, when the feature vector is obtained with the Daub-4 wavelet transform filter; both classifiers have similar scores and high performance. Deep CNN was employed by Ding et al. [13] for robust binary classification. For training and prediction, they provided similarity distance matrices to the CNN model. Figure 1.1 gives an overview of their adopted CNN architecture. In order to improve verification performance, Chen et al. [14] sought to select the best classifier. Therefore, they looked at the four classification models: Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). Following their testing and comparisons, they concluded that the SVM model was the most effective classification model considering the features they employed. Using the Dynamic Time Warping (DTW) distance-based classifier, Sadak et al. [8] qualified the audio signal samples as genuine or forgery based on the values of the similarity ratios between the audio signals. A one-class classifier based on CNN was developed by Wei et al. [15]. Zhao et al. [16] utilized a CNN-based Multi-Modal Siamese Network as a classifier. They developed an application called SonarSign for their proposed signature verification system and designed the classification phase with 5-fold cross-validation. Each participant, as previously noted, provided 40 genuine signatures and 35 skilled forgeries. The training set for each 5-fold cross-validation consists of 32 genuine signatures and 28 skilled forgeries.

1.1.2 Literature Review on Offline (Static) Signature Image

The offline signature verification studies, summarized in Table 1.2, obtained more successful results, especially with competitive advances over public datasets (i.e.

GPDS, MCYT, CEDAR). In this competition, besides the success percentage, details such as how many samples are used during the training, whether forgeries are used or not, and what type of forgeries are used for the tests are decisive. Studies in Table 1.2 are detailed in the following sections.

Table 1.2 Summary of the Literature Review on Offline Signature

Study	# Samples per Participant	Preprocessing	Feature Extraction	Classification	Results
Guerbai et al (2015) [20]	CEDAR (12 sample) GPDS (12 samples)	Binarization. Noise removal with mean filter.	The energy of the curvelet coefficients.	One-class SVM classifier	AER (CEDAR): 5.60% AER (GPDS): 15.07%
Yılmaz and Yanıkoğlu (2016) [21]	GPDS-160 (12 samples)	Erased strokes that are far away from image centroid, The upper and lower contours of the signature are detected for eliminating variations in pen tip thickness	Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Scale Invariant Feature transform Descriptors (SIFT) features.	Combination of user-dependent SVM and global SVM classifiers	EER: 6.97%
Pal et al. (2016) [22]	GPDS-100 (8 samples)	Mean filter for removing noises, Determining minimum bounding box of the images	Local Binary Patterns (LBP), Uniform Local Binary Patterns (ULBP)	Nearest Neighbour (NN) classifier with an Euclidian distance measure	EER: 32.21%
Ooi et al. (2016) [23]	MCYT (10 samples)	Median filter for noise removal, Binarizing images	Discrete Radon Transform (DRT)	Probabilistic Neural Network (PNN)	EER: 9.87%
Hafemann et al. (2017) [24]	GPDS-960 MCYT-75 (10 samples) CEDAR (12 sample) Brazilian PUC-PR (30 samples)	Removed the background using OTSU's algorithm, Inverting and resizing image	CNN	SVM based classifiers	EER (GPDS-160) : 1.7% EER (MCYT) : 2.87% EER (CEDAR) : 4.63% EER (Brazilian PUC-PR) : 2.01%
Okawa (2017) [25]	MCYT-75	Moment-based normalization, Histogram normalization, Clipping strokes, Binarizing image	KAZE features based on the recent Fisher Vector (FV) encoding	SVM	EER: 5.9%
Alaei et al. (2017) [26]	GPDS-140 (12 samples)	Binarizing and cropping images, Mean filter for noise removal	LBP-based features	Fuzzy similarity distance based classifier	EER: 16.67%
Sharif et al. (2020) [27]	CEDAR (12 samples), MCYT(12 samples), GPDS (12 samples)	Binarizing images, Resizing binary image, Median filter	Local and global features which are utilized by best feature selection algorithm	SVM	AER (CEDAR): 4.67% AER (MCYT): 5% AER (GPDS): 5.42%

Continued on next page

Table 1.2 – continued from previous page					
Study	**# Samples per Participant**	**Preprocessing**	**Feature Extraction**	**Classification**	**Results**
Gosh (2021) [28]	GPDS-300 (12 samples), MCYT-75 (10 samples)	resized to a fixed size, skewness correction	Structural and directional features	RNN	EER (GPDS-300): 1.46% EER (MCYT-75): 0.34%

1.1.2.1 Data Extraction

In some of the research related to signature verification systems, those who design the biometric systems build an exclusive data set for themselves. In such cases, it becomes difficult to compare these studies with other studies in terms of feature extraction, classification, etc., because the datasets they work on are different. Thanks to the general availability of databases such as GPDS [18], CEDAR [17], MCYT [29], and Brazilian PUC-PR [30], studies published to date can compete more objectively by incorporating these datasets into their operations. Participants are generally required to sign a large number of genuine signatures and/or skilled forgeries on gridded sheets to produce these kinds of datasets. The pages are then scanned at resolutions of 300 dpi or 600 dpi to migrate data on the pages to the digital environment.

1.1.2.2 Preprocessing

The preprocessing step is an important factor since it directly affects classification success and computational performance. In this stage, several crucial tasks are completed, including noise reduction, background removal, alignments, size reduction, etc. Using OTSU's algorithm [31], Hafemann et al. [24] eliminated the background data and set background pixels to white. After that, the image was inverted (turned negative), and the image's size was adjusted using predetermined values that were supported by the neural network methods they were using. Using a distance threshold, Yanıkoğlu and Yılmaz [21] removed strokes that are distant from the picture centroid. To remove changes in pen tip thickness, they detected the upper and lower contours of the signature. They used the following alignment procedure for rotation, scaling, and fine translation.

$$argmin_{\sigma,\theta,\delta}\{||Q^i_{\sigma,\theta,\delta} - R^i||\} \tag{1.6}$$

where σ is scaling, θ is rotation, δ is fine translation Q is query signature, R is reference signature and $Q^i_{\sigma,\theta,\delta}$ is the transformed version of Q. With the optimum scaling, rotation, and translation parameters that minimize the distance between the query and reference signature, each query signature Q of a participant is aligned to

each reference signature R^i of that participant. To control the position and rotation and enhance low-contrast photos, Okawa [25] used moment-based normalization and histogram normalization procedures. Using a mask on the image to clip the strokes, he reduced background noises. Then, he obtained a binarized image from the original gray-level image using a thresholding approach based on discriminant analysis. To improve the strokes, he used a smoothing filter and dilation. In the preprocessing stage, Pal et al. [22] identified the minimal bounding box of the images and used a mean filter to eliminate noise from the signature images. Ooi et al. [23] used a median filter to remove noise and a straightforward thresholding approach to binarize the signature images. A histogram-based thresholding approach was used by Alaei et al. [26] to binarize greyscale signature images. They utilized a mean filter to reduce noise. Input images were under-sampled and cropped to determine the signature images' minimal bounding boxes. Dey et al. [32] fixed the sizes of all the signature images using bilinear interpolation. They converted the original images into negative images and then normalized each image by dividing the pixel values by the standard deviation of the pixel values across all the images in the dataset. To extract the binary image from the signature image, Sharif et al. [27] used Otsu segmentation. They downsized the binary picture to 256×256 pixel size, and on the resulting image, they performed certain morphological operations, including thinning (erosion) and closing (dilation + erosion). They utilized a median filter to reduce noise. Taking into account the largest signature, Each image of the signatures was scaled by Pinzon et al. [33] to a certain size. By turning the image grayscale and adjusting the contrast, they removed any stains or dirt and enhanced it. By applying Otsu's algorithm, Ruiz et al. [34] binarized signature images and used a 3×3 sized kernel to make signature strokes wider. They reduced the size of the signature images to the predetermined 128×128 pixel size. For neural network training, they normalized pixel data from (0, 255) to (0, 1). To eliminate differences in height and width amongst samples of the same signature, Gosh [28] scaled the images of all the signature samples to a fixed 128×128-pixel size. Then, the signature samples of the same person that are skewed and not horizontally aligned have been corrected by bringing them into horizontal orientation.

1.1.2.3 Feature Extraction

In general, there are two ways to extract the features obtained during the feature extraction phase: the first is by modifying feature extractors, and the second is by using feature learning methods. Numerous types of research published to date employ, examine and compare both techniques individually or together. From a different perspective, the features obtained during the feature extraction step may be split

into two groups: 1) Global Features, and 2) Local Features. Global features are those features that are obtained from the complete signature, such as length, height, aspect ratio, etc. Local features are those that can be extracted from each grid of the signature image. Cartesian and log-polar grids were utilized by Yılmaz and Yanıkoğlu [21] to extract characteristics from local zones. By combining the features of the Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Scale Invariant Feature Transform Descriptors (SIFT) in accordance with their success rates, they are obtained for the classification step. Convolutional Neural Networks were used by Hafemann et al. [24] to propose writer-independent feature extraction from signature images, which is intended to capture visual cues that discriminate between genuine signatures and forgeries, independent of the signer. Dey et al. [32] presented SigNet, a Siamese network that uses the Convolutional Siamese Network learning method to directly determine the characteristics to be utilized for classification from signature images, regardless of the signer. To get local features from significant areas of the signature images, Okawa [25] suggested KAZE features based on the current Fisher Vector (FV) encoding. The best feature selection algorithm makes use of the local and global characteristics that Sharif et al. [27] extracted. Aspect ratio, signature area, pure width, pure height, and normalized signature height are considered global characteristics in their study. Signature centroid, slope, angle, and distance are local characteristics. In order to obtain relevant characteristics for the best feature selection component, they used a genetic algorithm. The curvelet transform is used in the study of Guerbai et al. [20] to generate characteristics of the handwritten signatures. Their method takes advantage of the energy of the curvelet coefficient obtained from the whole image of the handwritten signature. To generate curvelet coefficients, they applied the curvelet transform on the signature image at different scales and different orientations, using the wrapping technique. Energy E of the curvelet coefficients is calculated as below:

$$E(l,r) = \sum_{i} \sum_{j} |C_{l,r}(i,j)| \tag{1.7}$$

where $C_{l,r}$ is the curvelet coefficient computed at the scale l and the orientation r. Ooi et al. [23] used the discrete radon transform (DRT) to extract features and principal component analysis (PCA) to reduce the dimensions. An efficient feature extraction method based on under-sampled bitmaps and LBP-based features was introduced by Alaei et al. [26]. Then, they retrieved LBP-based characteristics from the resulting under-sampled bitmap image. Twenty-two Gray Level Co-occurrences Matrix (GLCM) and eight geometric features were generated by Batool et al. [35], and they were merged using a method based on a high priority index feature (HPFI). They presented skewness-kurtosis controlled PCA (SKcPCA) to choose the best features for final categorization into forged and genuine signatures. Four structural and

direction-oriented features, including change of trajectory direction, trajectory slope, trajectory waviness, and center of mass, have been retrieved in differing quantities from each signature sample in the research Gosh [28] presented.

1.1.2.4 Classification

A One-Class Support Vector Machine (OC-SVM) based classification design was proposed by Guerbai et al. [20] using a writer-independent technique. To train the model, they employed genuine signatures and random forgeries made from genuine signatures from other participants. They evaluated the system using the CEDAR dataset and found that employing 4, 8, and 12 genuine signatures for model training, respectively, resulted in average error rates (AERs) of 8.70%, 7.83%, and 5.60%. When employing 4, 8, and 12 genuine signatures for training, they obtained AERs for the GPDS dataset of 16.92%, 15.95%, and 15.07%, respectively. Using 12 reference signatures, Hafemann et al. [24] developed classifiers for each participant based on writer-dependent Support Vector Machines (SVM). They used genuine signatures as well as random forgeries made out of genuine signatures from other participants. Both a linear formulation and the Radial Basis Function (RBF) kernel were used to train the SVM. They conducted their studies using the GPDS-960 [18], MCYT-75 [29], CEDAR [17], and Brazilian PUC-PR datasets [30]. They obtained an EER of 2.87% for the GPDS-160 dataset, 4.63% for the MCYT dataset, 2.01% for the CEDAR dataset, and 1.72% for the Brazilian PUC-PR dataset. Sharif et al. [27] adopted a writer-dependent SVM classifier, employing the GPDS Synthetic, CEDAR, and MCYT datasets. On the CEDAR, MCYT, and GPDS Synthetic datasets using 5, 10, and 12 genuine signatures. They attained minimal average error rates of 4.17%, 5.0%, and 5.42%, respectively. Yılmaz & Yanıkoğlu [21] used 5 or 12 reference signatures to perform writer-dependent and writer-independent SVM classifiers, and they trained both types of classifiers with RBF kernel. The combined performance of all classifiers has a state-of-the-art EER of 6.97%. They employed only genuine signatures in the model training phase of their suggested system in order to make their applications more realistic and relevant. Writer-independent offline signature verification using deep metric learning was proposed by Rantzsch et al. [36]. They used GPDS Synthetic, a portion of the Dutch Offline Signatures, and Japanese Offline Signatures datasets from the ICDAR SigWiComp2013 contest [37] for model training and assessment. It is claimed that the system outperforms the state-of-the-art for offline signature verification from the ICDAR SigWiComp 2013 competition. Based on a Siamese Neural Network architecture, Ruiz et al. [34] introduced a writer-independent signature verification system against random forgeries that can be applied to new participants without the need for extra training. Two kinds of

synthetic signatures, augmented signatures, genuine signatures, and combinations of all, were independently used to train the model. Better results were obtained using synthetic signatures than other datasets. However, combining all types of samples achieved the best results. Their method was evaluated on the GPDS Synthetic, MCYT, SigComp11 [38], and CEDAR datasets, and EER results of 6.51%, 3.93%, and 4.84%, respectively, were obtained. Long-Short Term Memory (LSTM) and Bidirectional Long–Short Term Memory (BLSTM), two specialized Recurrent Neural Network (RNN) classifier models, have been employed by Gosh [28] in their study on the verification of handwritten signatures. He also compared the system he developed using a CNN-based classification approach to an RNN-based classification approach, both of which are writer-dependent. The efficiency of the system was evaluated using six popular public signature databases. He claims that his experimental findings show that the proposed RNN-based signature verification system surpasses the CNN-based system and the current state-of-the-art results.

1.2 Objective of the Thesis

The objective of the thesis is to examine whether the friction sound between pen and paper, which occurs during the signing, has a biometric value or not for verification. It was also aimed to investigate the success rate of biometric verification of the signature sound alone and to determine whether the signature sounds increase the verification success when evaluated together (Fusion) with the corresponding signature images. In addition, it aimed to reveal the effect of the difference between the pen-paper and sound recorders (Mobile phones) on the verification success.

1.3 Hypothesis

The hypothesis of this thesis: "The sound caused by pen-paper friction during signing has a biometric value for verification, and when these signature sounds are fused with corresponding signature images, they verify with a higher success than the verification success with signature images only or sounds only."

1.4 Contribution

Our main contribution is the comprehensive consideration of the friction-induced sound between the pen and paper surface during the signing procedure. The factors contributing to the success of this system are listed below.

1) A new approach is proposed, which performs handwritten signature verification

with a very high degree of success, only taking into account the sound emanating from the friction between the pen and paper during the signing (signature sound).

2) Another approach is proposed, using a fusion of the signature sound and image to verify the signature. This method has a higher success rate than both signature verification using only signature image data and signature verification using only signature sound data. Statistical significance tests are used to validate these results.

3) A data set consisting of signature sounds and signature images from 93 participants is built from the ground up. This dataset contains a total of 2976 signature images and 5952 signature sounds, with 16 genuine signature images, 16 forged signature images, 32 genuine signature sounds, and 32 forged signature sounds per participant. Each participant is required to sign using two distinct types of paper and two distinct types of pens. For each paper-pencil combination, four samples are collected. The sound arising from each signature is recorded using the built-in microphones of two different phone models. The signers' ages ranged from 19 to 64, with 55 male and 38 female participants.

4) By analyzing the cases where the pen-paper-phone combinations of the reference and query signatures are different, it has been determined to what extent the differences in the paper, pen, and phone items affect the verification success.

5) A novel approach to feature extraction is provided, wherein sound-based features are extracted and transformed into images, and feature extraction is carried out using image processing techniques on the image. In the deep learning-based approach (Chapter 7) proposed in this study, the SigNet model [24] trained with only static signature images is used for the first time for feature extraction from sound-based data.

6) Two different signature verification approaches are proposed, in which feature extraction is carried out with deep and non-deep (shallow [39]) learning-based methodologies. These two approaches are compared in terms of their benefits and drawbacks.

1.5 Outline

Chapter 2: The methodology used to build the dataset, the total number of participants, the total number of samples collected, and the tools used to collect the samples—pen, paper, and sound recording equipment—are all described in this chapter.

Chapter 3: The operations performed in the preprocessing step for both signature sound data and signature image data are explained in detail.

Chapter 4: In this chapter, the processes performed for feature extraction are explained. Feature extraction is performed separately on the signature sound and the signature image.

Chapter 5: Detailed information is given on the classification methodology. Block diagrams, including classification phases of proposed approaches, are illustrated.

Chapter 6: In this chapter, the shallow learning-based approach is explained. Tables and graphs demonstrating the test results are provided.

Chapter 7: The deep learning-based proposed approach is explained. The tables and graphs providing test results are given. Comparative test results for the deep learning-based approach and the shallow learning-based approach are included in this chapter.

Chapter 8: The conclusions, summary, and recommendations for further research are discussed in this chapter.

2
DATA EXTRACTION

Offline signature datasets like GPDS [18], CEDAR [17], MCYT [29], and others that are publicly available to researchers lack the sound signal data that is specifically required for this study. As a result, a new dataset is built that includes sound signal data of the signing process as well as static images of the signatures like in the aforementioned public offline signature datasets. An illustration of the experimental setup is shown in Figure 2.1.



Figure 2.1 Two mobile phones are displayed in an experimental setup with a BIC Cristal ballpoint pen and thin paper with the auto-copy feature (To the right is a rollerball fine-point pen)

Additionally, the dataset of static signature images produced in this research and two public datasets (GPDS and MCYT) are compared using the proposed approach independently, and as a result, the coherence of the dataset built in this study is determined as provided in the following chapters (Chapters 6-7, Tables 6.1-7.2).

2.1 Pen Types

The types of pens used in the signing procedure are decided considering the most common usage. Additionally, taking into account that variations in pen nib thickness also change the noises made throughout the signing procedure, it is assumed that if pen nibs vary in thickness, verification would be more challenging. Thus, we subject our system to a more stringent test. The most widely used pens in the world, the BIC 1mm Cristal disposable ballpoint pen [40] and the BIC 0.5mm Extra Fine Point Rollerball Pen, are both selected. Figures 2.2 and 2.3 provide images of the selected pen types.



Figure 2.2 BIC 0.5mm extra fine point rollerball pen



Figure 2.3 BIC 1mm Cristal disposable ballpoint pen

2.2 Paper Types

It is aimed to choose the most frequently used paper types as well. There are two types of paper used in the study, A4 plain paper (80 g/m2 - 24 lb, size:210x297mm) and A5 thin paper with an auto-copy feature (55 g/m2 - 15 lb, size:210x148mm). Figures 2.4 and 2.5 provide representations of the selected paper types.



Figure 2.4 Scanned A4 plain paper (80 g/m2 - 24 lb.)

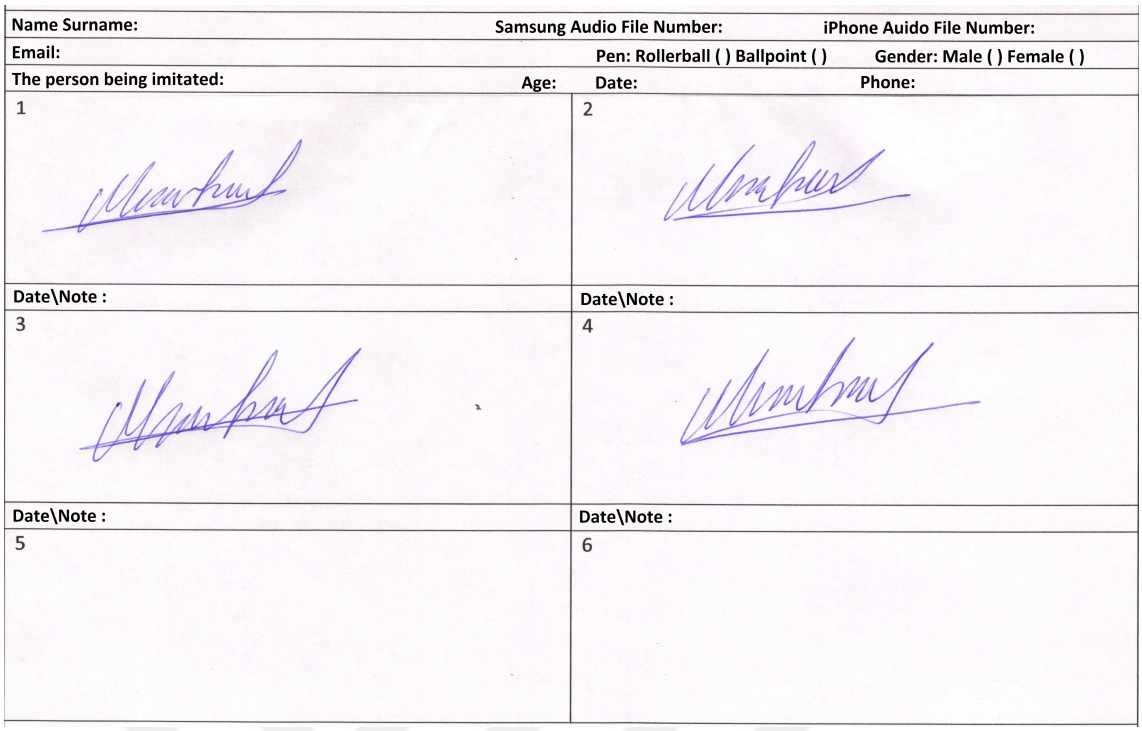


Figure 2.5 Scanned A5 thin paper with auto copy feature (55 g/m2 - 15 lb.)

The participant's name and surname, email address, phone number, date, gender, the kind of pen they use, and the ID numbers of the audio files that are acquired after recording are all indicated in the fields on the sheets (Due to privacy concerns, some of the data is not displayed for the samples in Figures 2.4-2.5.). There is not a section for paper types, though, because these can be determined on the papers themselves. There is a section for the name and last name of the person the participant is impersonating if they sign the skilled forgery of another participant. The participant leaves this area empty if he/she gives his/her genuine signature.

2.3 Phone Models

When building the dataset, it is crucial to keep in mind that the dataset should be interesting for research, useful for real-world applications, and cover as much usage area as possible. Two of the mobile phone models' internal microphones are used to record the sound signals that occur throughout the signing procedure. Android and iOS, the iPhone 7 Plus, and the Samsung Galaxy Note 3 were selected as the two mobile devices to compare, due to their broad use throughout the globe [41] and the fact that they support the two most popular mobile operating systems. The sound recording software utilized in the study is the built-in sound recorder for both mobile devices. The internal microphones of the phones are combined in the data collection setup so that the distance between the two microphones and the signature is aimed

to be equal (Figure 2.1).

2.4 Data Collection Procedure

One participant is required to sign four times (Since two paper types and two pen types are used, there are a total of four combinations) for each combination of pen and paper. The sound recording applications on both phone models are launched and begin recording the sound before the first signature is given. The recording lasts up to the conclusion of the fourth signature. As a result, each phone has one sound file that contains the sounds of four different signatures. As shown in Figures 2.4 and 2.5, there are additional fields on the signed paper where you can write the ID numbers of these sound files independently for each phone. 93 participants (55 male, 38 female) have provided signature samples, each of which includes four genuine signatures on plain paper with a rollerball pen, four genuine signatures on plain paper with a ballpoint pen, four genuine signatures on thin paper with a rollerball pen, and four genuine signatures on thin paper with a ballpoint pen. So, using a combination of two different pen types and two different paper types, a participant provides 16 genuine signatures. Due to the simultaneous recording of the sounds of these 16 genuine signatures by two phones, a total of 32 signature audio files—16 on each phone—are produced. 16 skilled forgeries and a total of 32 audio files associated with these forgeries are likewise gathered from one participant in the same manner and using the same combinations. In summary, each participant provides 16 genuine signature images, 16 skilled forgery images, 32 genuine signature audio files, and 32 skilled forgery audio files to build a signature data set of 93 participants (See Table 2.1). Before receiving the skilled forgeries, the person who supplies them is shown the name and signature of another participant to copy, and he/she is also allowed to do practice trials at least five times until they indicated they are ready. Thus, he/she could produce a signature that is close to the one they are trying to replicate; these forgeries are known as skilled forgeries. The forger does not watch how the signature to be forged is signed. If he/she had watched, it would be a great advantage in imitating the signature image but not as much of an advantage in imitating the signature sound. Because the sound is more abstract than the image, it is more difficult to remember and make the same sounds with a pen. In the usual workplace setting, there is an average noise level of 40 dB while the sounds of the signatures are being recorded. Participants are not instructed to sign with a louder or more pronounced motion. They are asked to sign just in a usual way. There is not any validation process in place for factors like signature length, volume, etc. Small (size) and quiet (with no distinguishable audio data) signatures from certain participants are accepted, and all of them are added to the dataset. The dataset is called SKU to make it easier to present in tables and results.

Table 2.1 Summary of dataset collected (SKU) from 93 participants (Age range is between 19 and 64).

#Participants	Pen Type	Paper Type	Phone Type	#samples
93 [55 male, 38 female]	Ballpoint Pen	Plain Paper	Samsung Galaxy Note 3	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Rollerball Pen	Plain Paper	Samsung Galaxy Note 3	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Ballpoint Pen	Thin Paper	Samsung Galaxy Note 3	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Rollerball Pen	Thin Paper	Samsung Galaxy Note 3	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Ballpoint Pen	Plain Paper	iPhone 7 Plus	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Rollerball Pen	Plain Paper	iPhone 7 Plus	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Ballpoint Pen	Thin Paper	iPhone 7 Plus	4 genuine signature, 4 skilled forgery
93 [55 male, 38 female]	Rollerball Pen	Thin Paper	iPhone 7 Plus	4 genuine signature, 4 skilled forgery

3
PREPROCESSING

In data processing, the preprocessing stage is crucial. Due to its direct impact on the feature extraction and classification phases, even the smallest improvement achieved at this step can result in substantial gains in verification performance. Processes performed with sound signal data and processes performed with signature images progress independently up to a point because of the nature of data collection phase in this study. In conclusion, there are two basic categories under which the preprocessing step is evaluated: 1) Sound data preprocessing, and 2) Image data preprocessing.

3.1 Signature Sound Data Preprocessing

The preprocessing of the signature sound is done in 2 stages. The first stage is audio-based preprocessing (Section 3.1.1), and the second is image-based preprocessing (Section 3.1.2). In the first stage, the audio signal data is preprocessed (i.e., segmentation, down-scaling, etc.). The preprocessed audio signal from the first stage is sent to the audio-based feature extraction phase in Chapter 4, Section 4.1. The output of this phase is two feature vectors, SFOSE and SC. These vectors are returned to the second stage, the Image-based sound data preprocessing (Section 3.1.2). Figure 3.1 shows the flowchart for preprocessing procedure of the signature sound.

3.1.1 Audio-based sound data preprocessing

Each participant had to provide at least four genuine signatures and four forgeries, as was specified in Chapter 2, according to each combination of pen and paper. Each mobile phone produces a sound signal that includes the sounds of these four signatures. Each signature sound's beginning and ending positions are identified. The Audacity tool [42] is used to segment each of those sound signals manually. To more easily distinguish the start and end points of the signature sound, the audio signal is segmented through the spectrogram view provided by this program. As a result, each sound file in the mobile phones is split into four independent signature sound files,

and the parts that did not include the dynamic data of the signing process are deleted. Sound files for signatures are converted from "m4a" to "wav" file formats. It is noticed that the energy density between 0 Hz and 22050 Hz offers adequate information about the signal, so the sampling rate is decreased to 22050 Hz to reduce the cost of data processing and make the utilized algorithms perform better.

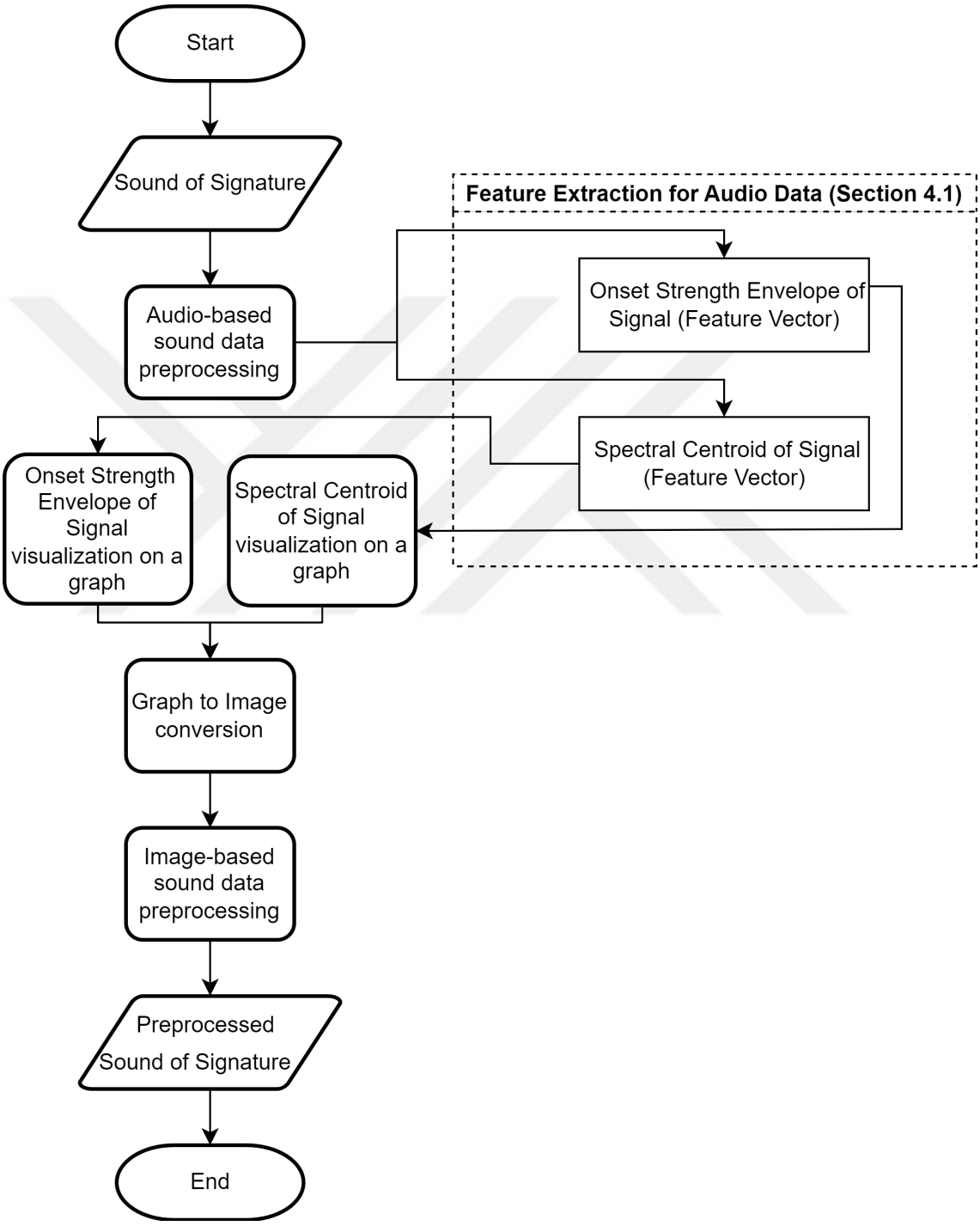


Figure 3.1 Signature Sound Data Preprocessing

3.1.2 Image-based sound data preprocessing

Following audio-based preprocessing, the raw audio signal data corresponding to a participant's static handwritten signature (Figure 3.2) is illustrated in Figure 3.3. The preprocessed audio signal is sent to the feature extraction stage (Section 4.1) to obtain feature vectors. These feature vectors are normalized (min-max) to reduce the impact of different pen-paper-phone combinations on verification success. Then, the plotted graphics of the feature vectors obtained in the feature extraction stage are converted into images and sent back to the preprocessing phase again. This time, an image-based preprocessing procedure is applied to these incoming image files; Axis lines and other axis-related information are removed (See Figure 3.4), the graphic image data of audio signals is scaled down by 50%, and it is transformed to gray-scale. The retrieved images are tagged with the sequence number, writer aliases, pen, and paper types, and signature class type (genuine or forgery).



Figure 3.2 Handwritten signature

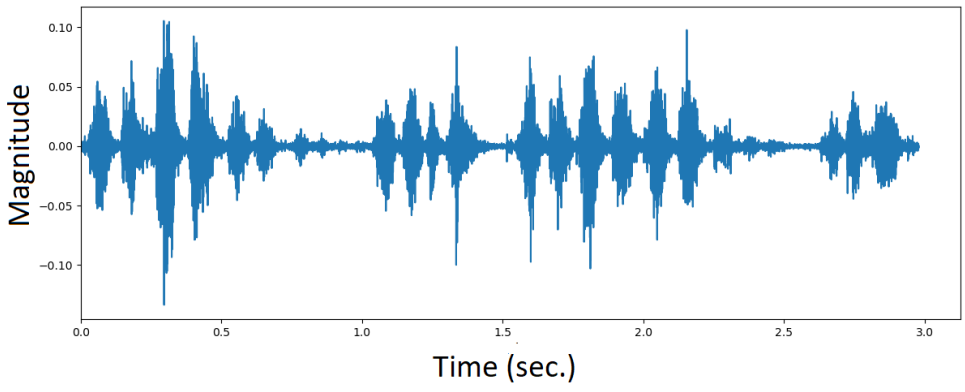


Figure 3.3 Handwritten signature of a participant as raw sound signal data

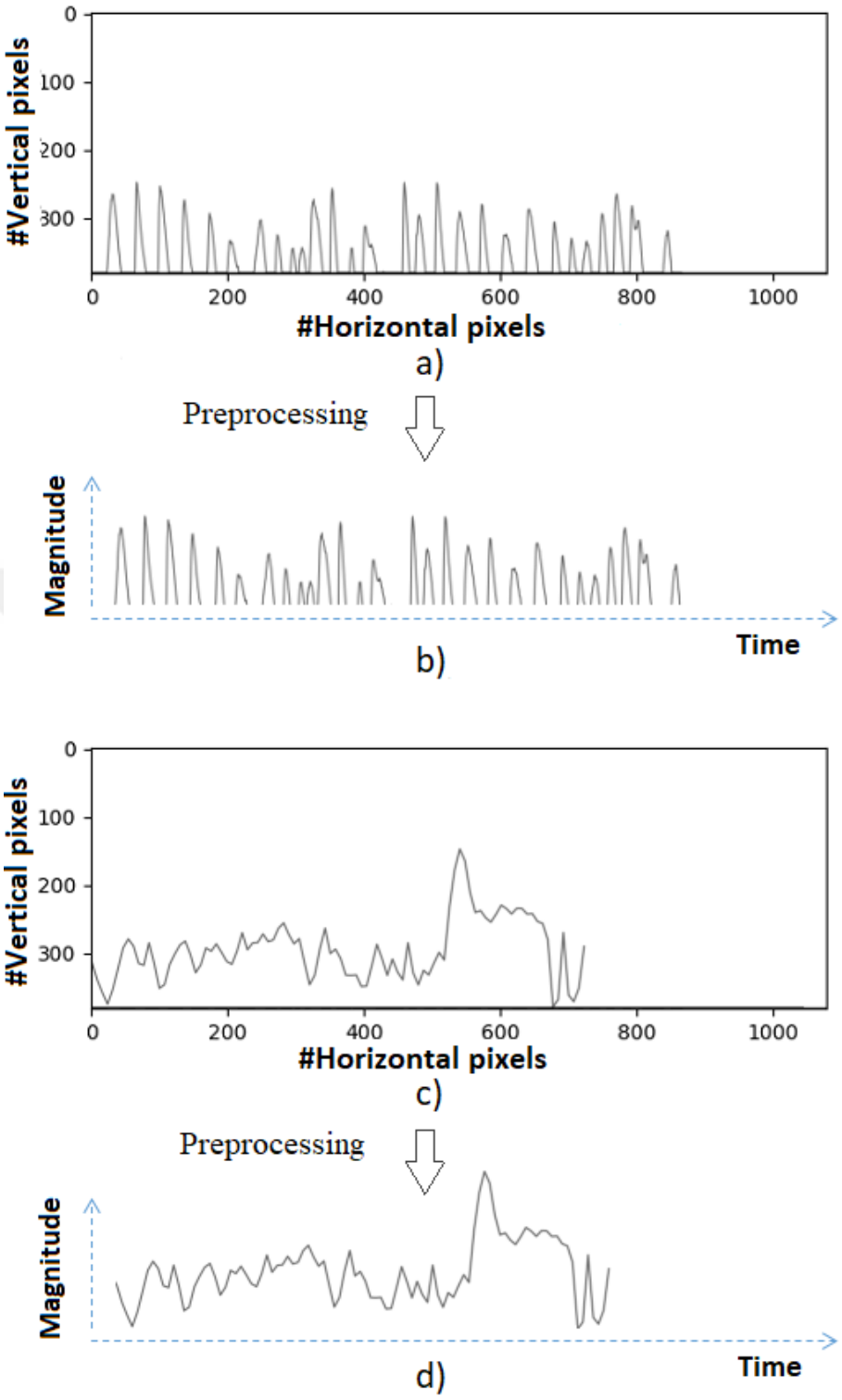


Figure 3.4 Image graphs of the audio signal's onset strength envelope and spectral centroid produced from the signature: a) Image of original spectral-flux onset strength envelope graph b) Image of preprocessed spectral-flux onset strength envelope graph c) Image of original spectral centroid graph d) Image of preprocessed spectral centroid graph

3.2 Signature Image Data Preprocessing

Each paper in the dataset holding image data of the static handwritten signatures is scanned at a resolution of 600 dpi in 24-bit color. An algorithm used to extract the individual images detects signature-containing rectangles and crops each signature to fit within these rectangles. Each cropped signature is named by the signer's aliases, the type of pen, the type of paper, the class of the signature (genuine or forgery), and the sequence number. The size of the image files is decreased by 50%. The images are converted to grayscale. For the shallow learning-based approach (Chapter 6): erosion and opening morphological procedures are performed to complete the gaps in the signature lines to increase the efficiency of the processes that would be carried out in the subsequent phases. Noise is reduced using Gaussian blur, and background noises and colors are eliminated using OTSU's thresholding algorithm [31]. The negative of the signature is produced by deducting each pixel from the image's maximum intensity value of 255. Images are then centered on a canvas. To sharpen signatures, a closing morphological operation is used. Once again, the image size is reduced. The center of the signature image is cropped to contain the signature. Finally, a non-local means denoising algorithm [43] is employed to eliminate noises. (see Fig. 3.5). For the deep learning-based approach (Chapter 7): only the inversion is applied to the signature images.

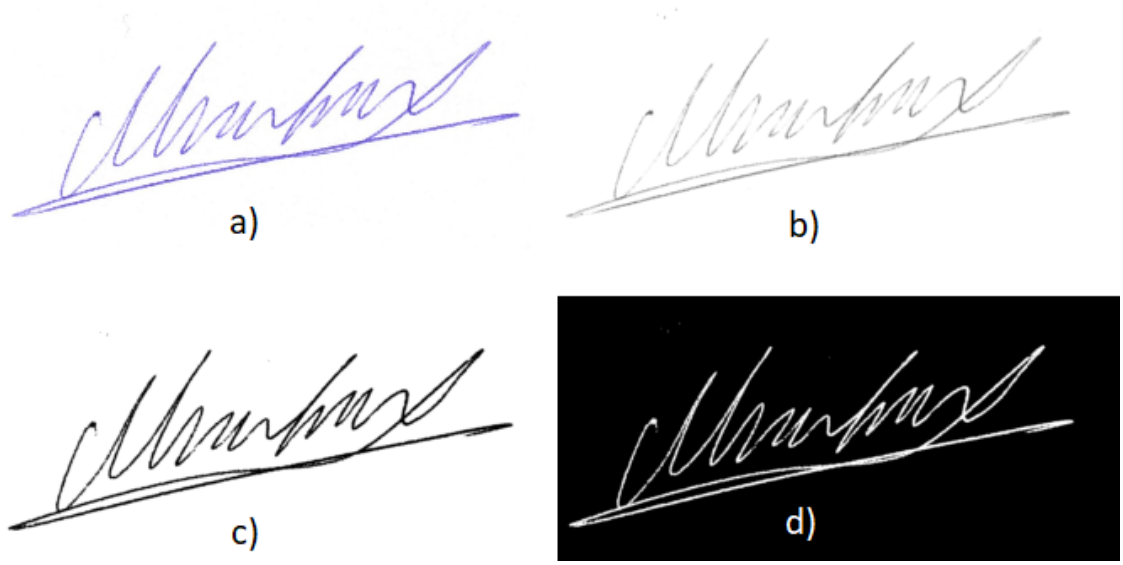


Figure 3.5 Preprocessing phases for static handwritten signature a) Original image b) Grayscale image c) Preprocessed image d) Inverted image

4
FEATURE EXTRACTION

The elements of the feature vectors acquired in this step are extracted using one of two methods in general: either hand-crafted feature extractors or deep feature learning algorithms. The implementer explicitly specifies features in the first case. The second case is an automated feature detection process that uses a deep learning algorithm to find the most useful features for classifying. The drawback of this deep learning-based feature extraction approach is that the implementer finds it very challenging to recognize, comprehend, or fine-tune these features on a human level. Despite this drawback, the detection of features using the deep learning-based approach is better at enhancing the classification's success. This study performs feature extraction by comparing both cases individually.

There are two distinct data types of a signature since the participant-provided signature data includes both audio signals and signature images. Distinct feature extraction strategies were used for these two different types of data, accordingly. The feature extraction procedures for both types of data are analyzed in the following sections.

4.1 Feature Extraction for Audio Data

Varying noises with different timbres are exposed throughout the signing process by utilizing various types of pens, papers, and mobile phones. Despite having distinct timbres, it is observed that when the sound signals of signatures are made by the same individual using various pen and paper combinations, they exhibit similar notary alterations. Spectral flux onset envelope curves are very convenient for comparing audio signals with different amplitude values because these curves stand similarly even if the amplitude values are different. So spectral flux onset envelope of signals provides an advantage in comparing audio signals with different amplitude values (timbres) obtained from different paper-pencil-phone combinations. The rate of the positive changes in the consecutive power spectrums of a signal over time is measured

by the concept of spectral flux. It is mathematically expressed as:

$$SF(n) = \sum_{i=1}^{i=\frac{N}{2}} H(|X(n,i)| - |X(n-1,i)|) \tag{4.1}$$

where $H(x) = \frac{x+|x|}{2}$ is half-wave rectifier function, n is frame number, N is window size, i is frequency bin index and $X(n,i)$ is the i_{th} frequency bin of the n_{th} frame. $X(n,i)$ is regarded as the short-time Fourier transform (STFT) of the signal $x(n)$. It is expressed as:

$$X(n,i) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(hn+m)w(m)e^{-\frac{2ijm\pi}{N}} \tag{4.2}$$

where $w(m)$ is a Hamming window, N is window size and h is hop size.

The process known as onset detection [44] is used to find the beginnings of all events connected to significant changes in an audio signal. The phrase "Detection Function" has also been used in studies published to date to refer to the Onset Strength Signal (OSS). A signal's envelope is the line that encircles it from the outside to encompass its oscillations. Another feature of a signature sound adopted for this study is the spectral centroid graph of the audio data. The spectral centroid—which reveals where the average of the spectrum weighted by amplitude is located—can be computed with the use of the Fast Fourier Transform (FFT). It may be formulated as:

$$SC = \frac{\sum_{i=0}^{i=N} iwA_i}{\sum_{i=0}^{i=N} A_i} \tag{4.3}$$

where w is the width of each spectral bin in Hz and A is the amplitude. Figures 4.1-4.2-4.3 illustrate the steps of a sound signal during the feature extraction phase.

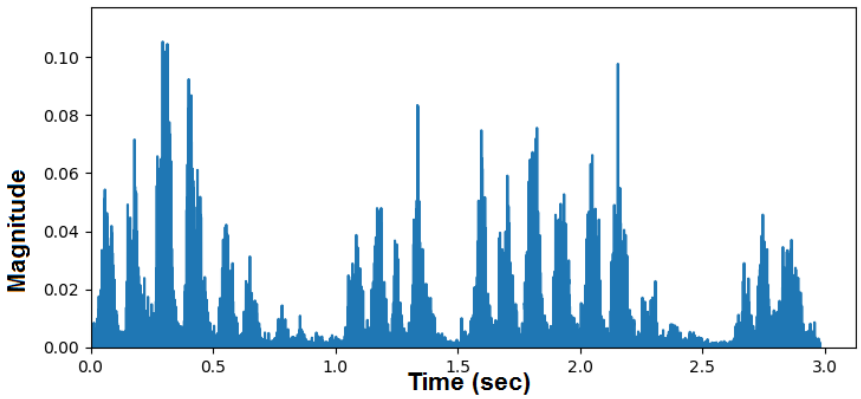


Figure 4.1 Raw signal

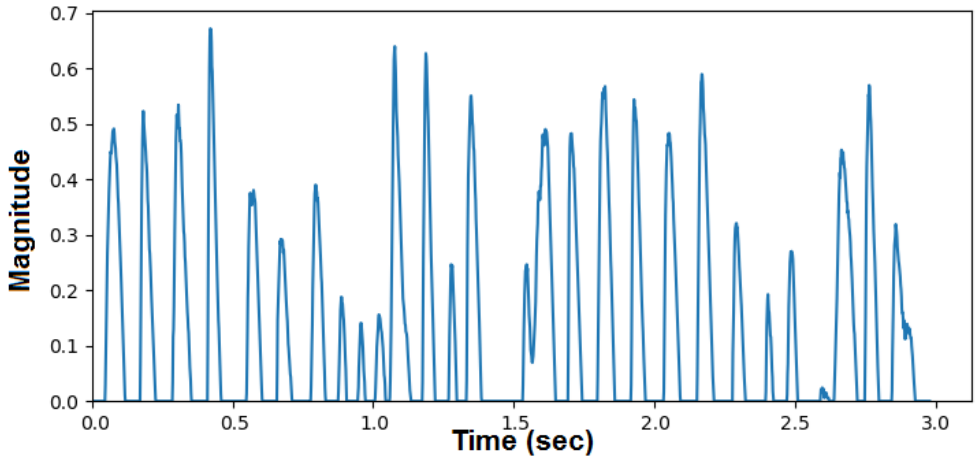


Figure 4.2 Onset strength envelop of signal

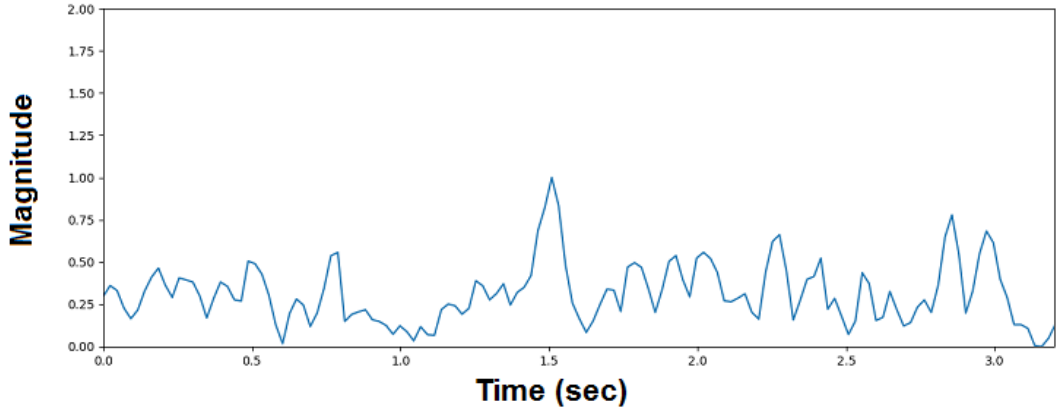


Figure 4.3 Spectral centroid of signal

4.2 Feature Extraction for Image Data

There are two types of images for feature extraction in this study. So, feature vectors are taken from both handwritten signature images and graphic images of signature sounds. The improved performance in the classification step is due to the combination of the feature vectors acquired from these two types of images.

The representations used in modern deep learning systems generally consist of tens or even hundreds of consecutive layers, and they are all automatically learned by exposure to training data. Other machine learning techniques, on the other hand, frequently concentrate on learning just one or two layers of data representations; as a result, they are also referred to as "Shallow Learning" techniques [39]. In this research, two separate approaches are proposed comparatively by performing feature extraction with two different methods, shallow (non-deep) learning-based feature extraction, and deep learning-based feature extraction.

4.2.1 Shallow (non-deep) learning-based feature extraction

At this level, Scale Invariant Feature Transform (SIFT) [45] and Local Binary Pattern (LBP) [46] algorithms were applied to each image, resulting in descriptors converted to histogram arrays and accepted as distinct feature vectors. The LBP feature extraction approach for gray-level independent image representation determines how closely related each pixel is to its neighbors. Applying the LBP operator to the image is simple and efficient. The operator's sole role is to evaluate each neighbor point in the radius distance using the specified radius parameter and compare it to the chosen center point on the image. A binary code is produced for each pixel in the image by thresholding the surrounding pixels in relation to the central pixel using an operator. LBP can be formulated as:

$$LBP(x_i, y_i) = \sum_{n=0}^{n-1} s(g_n - g_i)2^n \tag{4.4}$$

$$s(g_n - g_i) \begin{cases} 1, & (g_n - g_i) \geq 0 \\ 0, & (g_n - g_i) \leq 0 \end{cases} \tag{4.5}$$

where (x_i, y_i) is the center point, and n is the number of points in the radius distance. g_i is the gray value of the i_{th} center point and g_n is the gray value of the n_{th} neighboring pixel.

Example operators for various radius values and point counts are represented graphically in Figure 4.4. Interpolation is used to approximate the gray values of the circle's points that do not precisely match the pixel points.

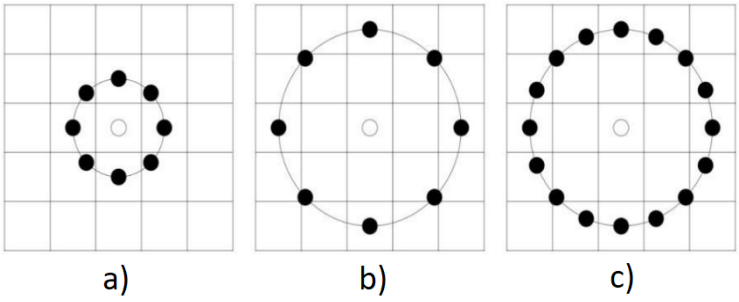


Figure 4.4 Operators for different radius values and point counts: a) Radius=1, Number of points=8. b) Radius=2, Number of points=8. c) Radius=2, Number of points=16.

The number of points is determined to be eight, and the radius is determined to be 2 in the proposed method. Depending on the size of each image file, a $m \times n$ dimensional matrix including LBP features is produced. The histogram vector of each LBP feature

matrix is computed. The histogram vector's size becomes $2^8 = 256$ since the number of points is adjusted to 8.

The SIFT algorithm, which finds and identifies regional features of an image that do not change with rotation and/or scaling, is used to construct the other feature vector. It is a systematic algorithm that is mostly composed of four parts; 1) Scale-space extrema detection: Identifying potential locations for features. 2) Keypoint localization: Accurately detecting the feature key points. 3) Orientation Assignment: This involves assigning directions to key points. 4) Keypoint Descriptor: A high-dimensional vector for identifying the key points. The perceptual hash of the image and a few contour features are also retrieved from each image file. These features, with the LBP and SIFT features, are included in the feature vectors. The aspect ratio of the image, the ratio of the contour area, the area of the bounding rectangle, and the ratio of the convex hull area to the bounding rectangle area are examples of these contour features. As previously stated, since the representation of the signature sound is an image file similar to that of a signature image, two-dimensional feature vectors (LBP and SIFT) are derived independently from both the sound of a signature and signature image. The size of these vectors is increased to 1×260 by including perceptual hashes of the images and other contour features. Two 260-dimensional feature vectors (LBP and SIFT) obtained from the sound of the signature, and two 260-dimensional feature vectors (LBP and SIFT) obtained from the signature image are sent to the classification phase for each signature in the dataset to combine the features obtained from these two types of data. The LBP and SIFT algorithms are used to accomplish writer-independent feature extraction in the shallow learning-based strategy outlined in Chapter 6.

4.2.2 Deep learning-based feature extraction

Deep learning algorithms, a subset of machine learning algorithms, are first popularised in 2006 [47]. Deep learning models have more complex hierarchical architectures than traditional data analytics. It has many hidden layers, so input data is modified numerous times before being utilized to generate the desired output. It also investigated whether deep learning-based features could successfully be used for the proposed verification system. Large amounts of data are necessary for the success of deep learning algorithms. The system is trained using this big data, which requires time, and decides which features to extract. Model output is generally produced in the initial training using programming tools to be used in the later stages. This approach saves time by avoiding repeating training for subsequent classifications. The size of the data set provided for this study is not sufficient to reach an efficient result by

training with deep learning and producing a model. So, using pre-trained models based on deep convolutional neural networks (CNN) [48] like SigNet [24] (It was produced using solely signature data.), ResNet [49], VGG-16 [50], and VGG-19 [50], feature extraction is carried out. Among these models, SigNet [24] is included in the proposed approach because it yields the most successful results (see Table 7.1 in Chapter 7) in terms of time and accuracy. The founders [24] of the SigNet model wanted to make their model learn features with a writer-independent approach by training the network with a non-supervised method. Being writer-independent has the benefit of being more suitable for applications in the real world because the network doesn't need to be retrained for learning features when a new participant is added to the system.

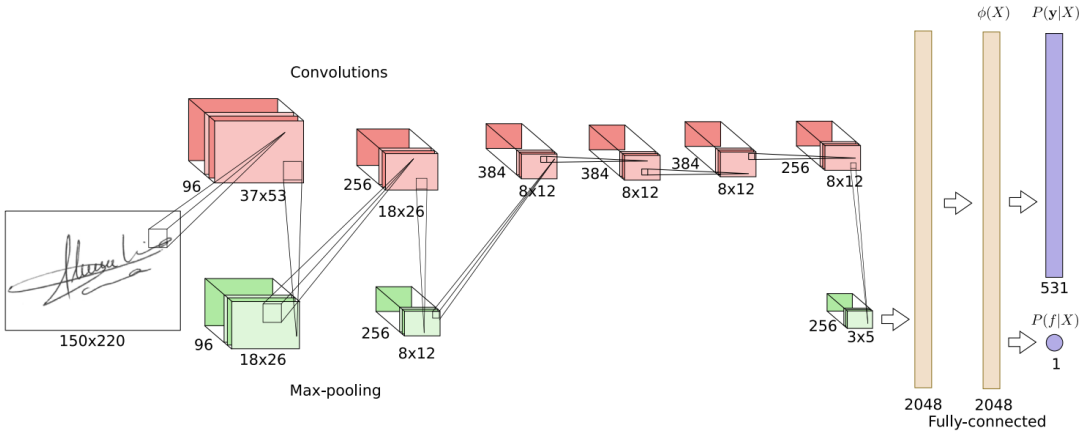


Figure 4.5 One of the architectures employed in SigNet [24], A series of transformations using convolutional layers, max-pooling layers, and fully-connected layers are applied to the input image.

Table 4.1 CNN Layers [24]

Layer	Size	Other Parameters
Input	$1 \times 150 \times 220$	
Convolution (C1)	$96 \times 11 \times 11$	Stride = 4, pad=0
Pooling	$96 \times 3 \times 3$	Stride = 2
Convolution (C2)	$256 \times 5 \times 5$	Stride = 1, pad=2
Pooling	$256 \times 3 \times 3$	Stride = 2
Convolution (C3)	$384 \times 3 \times 3$	Stride = 1, pad=1
Convolution (C4)	$384 \times 3 \times 3$	Stride = 1, pad=1
Convolution (C5)	$256 \times 3 \times 3$	Stride = 1, pad=1
Pooling	$256 \times 3 \times 3$	Stride = 2
Fully Connected (FC6)	2048	
Fully Connected (FC7)	2048	
Fully Connected + Softmax	M (#Users)	
Fully Connected + Sigmoid	1	

The input image is transformed using convolutional layers, max-pooling layers, and fully-connected layers in a series of steps. M units with a softmax activation, where M

is the total number of users in the data set, make up the neural network's final layer. Their proposed CNN architecture for M=531 users is shown in Figure 4.5. Layers are also described in Table 4.1. In the deep learning-based approach explained in Chapter 7, feature extraction was performed with the SigNet model.

5
CLASSIFICATION

Table 1.2 in Chapter 1 makes clear that the Support Vector Machines (SVM) technique is frequently employed for cutting-edge signature verification research. Based on statistical learning theory, SVM is a controlled classification method that was developed by V. Vapnik et al. [51]. It essentially uses a line to split data into two classes most effectively. This line is intended to be as close to both classes' extreme points and vectors as possible. Decision boundaries (hyperplanes) are established for this purpose as illustrated in Figure 5.1.

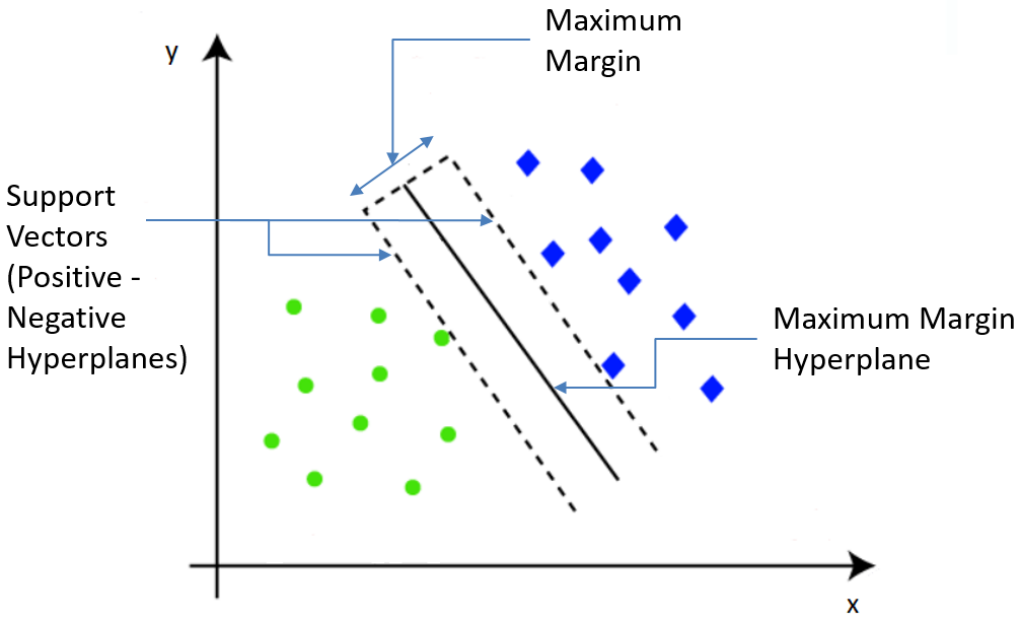


Figure 5.1 Separation of two different classes by hyper planes

The two types of SVM that are most frequently employed are Linear SVM and Non-linear SVM. A single straight line can be used to divide a dataset into two classes when the data is linearly separable. When a dataset cannot be categorized with a straight line, or if the features in the dataset are not distributed linearly, the non-linear SVM algorithm is utilized. It is chosen to classify only genuine signatures in this study so that it would be appropriate for real-life applications. To accomplish this objective, only genuine signatures are utilized in the training phase of the One-Class

SVM classification method, which is another variation of SVM. One Class Classification (OCC) uses single-class samples during training to distinguish between samples of one specific class. One-Class SVM classifies incoming data as being similar to or distinct from the samples used in training by employing a hypersphere rather than a hyperplane to divide two classes of instances. There is a competition to train classifiers with as few samples as possible and solely with genuine signatures (one class) to make signature verification acceptable for real-life applications in the studies published to date. In this aspect, OC-SVM is quite effective.

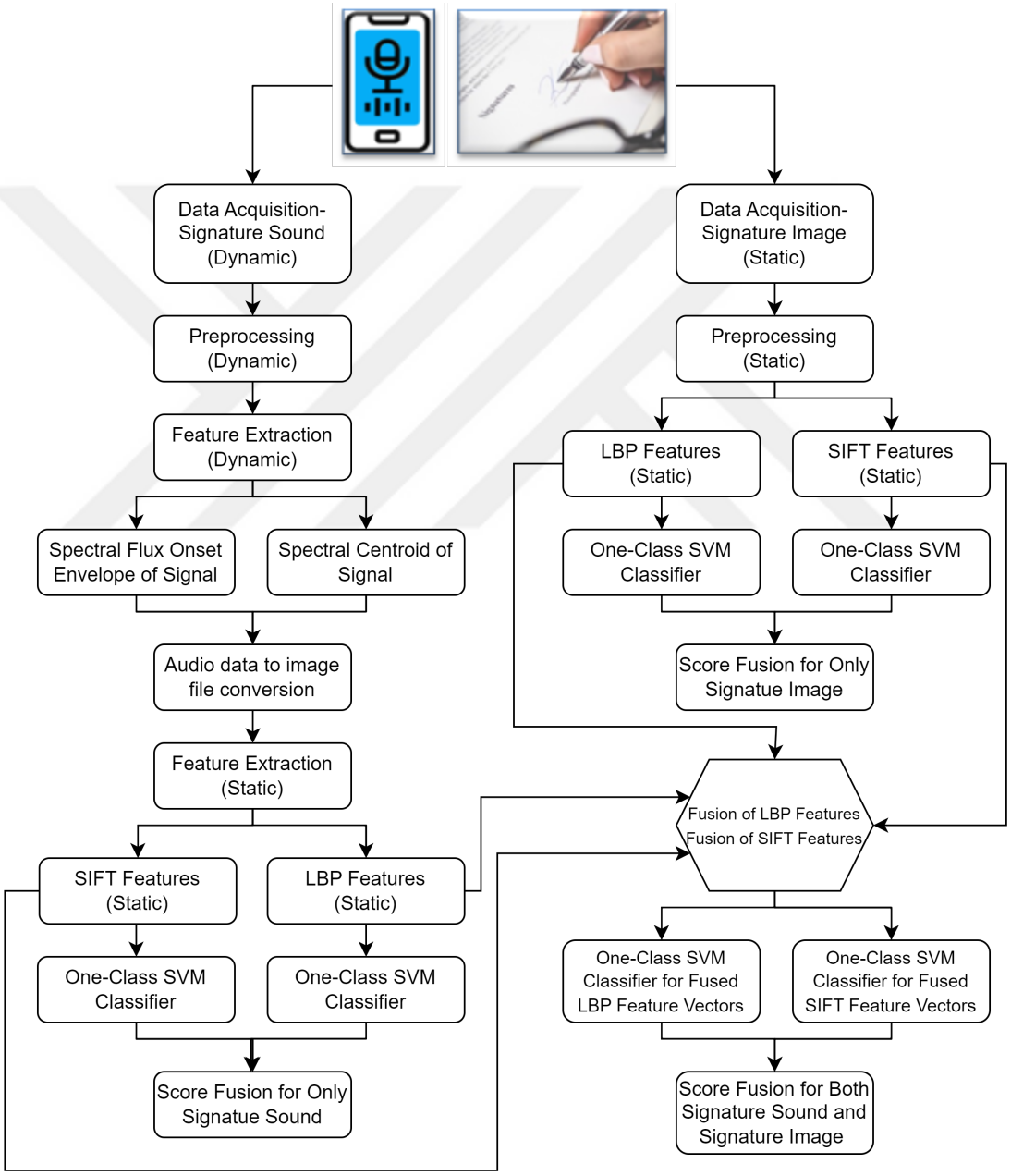


Figure 5.2 Block diagram of the proposed shallow learning-based approach for multimodal signature verification (sound and image)

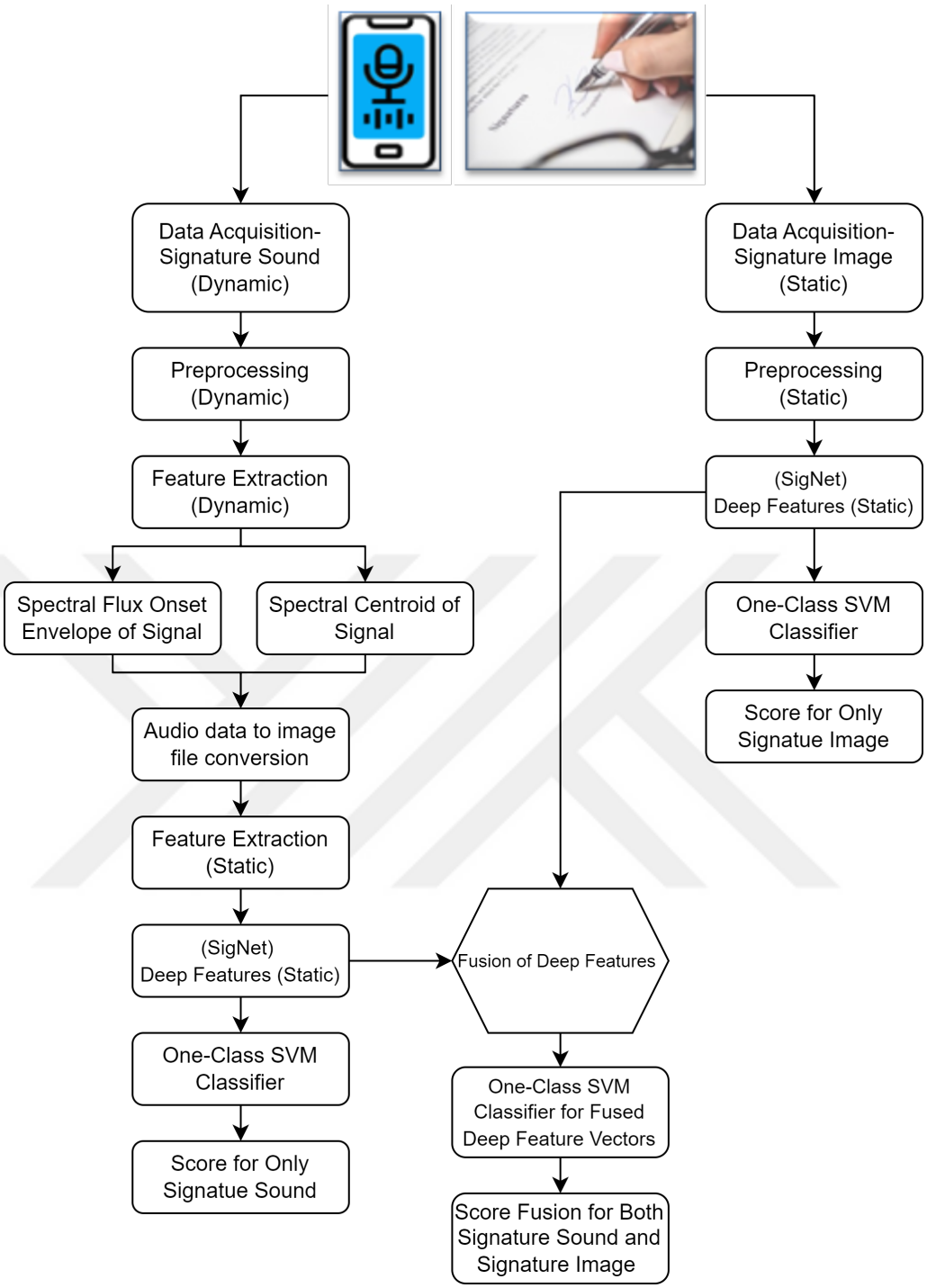


Figure 5.3 Block diagram of the proposed deep learning-based approach for multimodal signature verification (sound and image)

Figure 5.2 displays the block diagram for the proposed shallow learning-based approach (Chapter 6), and Figure 5.3 illustrates the block diagram for the proposed deep learning-based approach (Chapter 7). Both of them incorporate the classification phase.

For feature fusion based on audio data, feature vectors (LBP and SIFT for shallow

learning-based, SigNet for deep learning-based approach) extracted from each of the audio-based Spectral Flux Onset Strength Envelope (SFOSE) and Spectral Centroid (SC) images are combined to generate united feature vectors by eliminating the mean and scaling to unit variance feature vectors are standardized. For image-based signature verification, feature vectors (LBP and SIFT for shallow learning-based, SigNet for deep learning-based approach) are extracted only from the static image of the signature instead of SFOSE and SC graphical images. When audio and image data are fused to increase the verification accuracy, feature vectors from the audio data and feature vectors from the signature image data are combined. These vectors are employed independently in the OC-SVM classifier utilizing leave-one-out cross-validation. Four genuine signatures are used for training, and four forged signatures are used for testing. In the testing, each of the four forged signatures, and each left-out genuine signature during the leave-one-out cross-validation procedures, are sequentially utilized together. Since the shallow learning-based approach performs two feature extraction algorithms (LBP, SIFT), two distinct classification scores are calculated and normalized between 0 and 1 using min-max normalization. These scores are averaged (score-level fusion) in order to improve verification performance and get the final score. Since there is only one feature extraction algorithm (SigNet) in the deep learning-based approach, score-level fusion has not been performed. Only feature vectors extracted from SFOSE and SC images or static signature images are combined for verification.

In the following two chapters, shallow and deep learning-based approaches are detailed comparatively by providing experimental test results with tables and graphical illustrations.

6
SHALLOW (NON-DEEP) LEARNING BASED APPROACH

The audio data's extracted features (see Section 4.1) are transformed into graphics, which are then turned into an image file. The audio-derived image files and the signature images are both subjected to the use of LBP and SIFT feature extraction (see Section 4.2.1) methods.

To verify a signature using just audio data, two image files Spectral Flux Onset Strength Envelope of Signal (SFOSE) and Spectral Centroid (SC) of Signal are extracted from the audio data of the signature. From each of these images, two feature vectors (LBP and SIFT) are produced. 1×256-dimensional LBP and SIFT feature vectors are combined with 1x4 dimensional contour feature vectors to become 1×260-dimensional feature vectors. For feature fusion based on audio data, 1×260 size LBP+Contour feature vectors extracted from both audio-based images (SFOSE, SC) are combined to generate a 1×520 dimensional LBP+Contour feature vector, and 1×260 size SIFT+Contour feature vectors extracted from both audio-based images (SFOSE, SC) are combined to form a 1×520-dimensional SIFT+Contour feature vector. By eliminating the mean and scaling to unit variance, feature vectors are standardized. These two 1×520 size vectors are employed independently in the OC-SVM classifier (Chapter 5). For each vector, distinct classification scores are computed, and these values are normalized between 0 and 1 using min-max normalization. The average of two scores resulting from the classifier is calculated, indicating whether the signature represented by each vector is genuine or a forgery, according to a threshold value.

For verification using only the static signature image, a 1×260 LBP+Contour feature vector and a 1×260 SIFT+Contour feature vector are obtained for each signature sample. These vectors are used for training the classifier after the standard scaling step. Two scores corresponding to each vector (LBP, SIFT) resulting from testing the OC-SVM classifier are averaged after being subjected to the min-max normalization process, as previously mentioned. It is decided whether the tested signature is genuine or forgery based on whether the average score exceeds the threshold value.

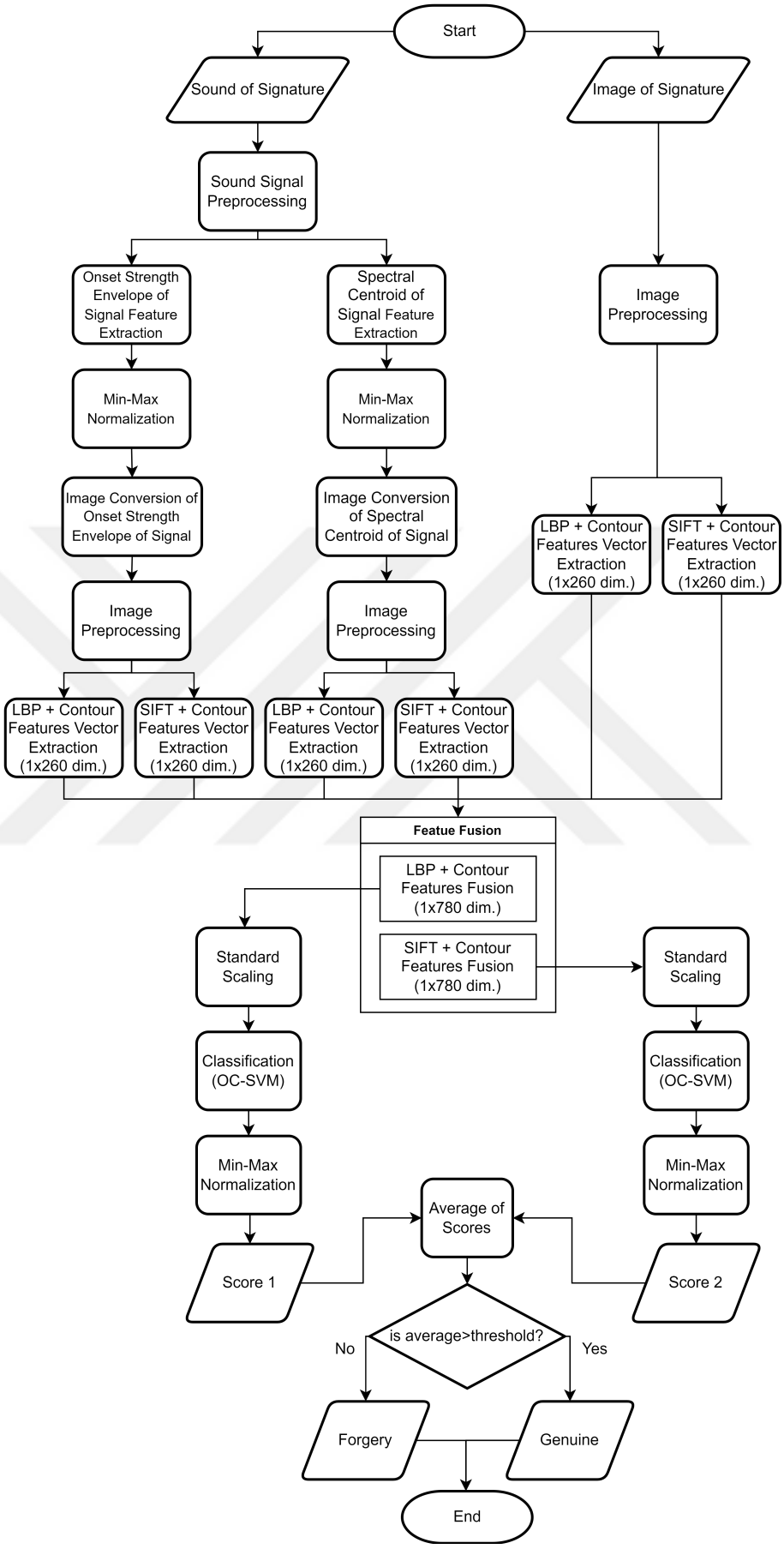


Figure 6.1 Flowchart for the proposed methodology of sound and multimodal signature verification

When audio and image data are fused to increase the verification success, two (SFOSE and SC) 1×260 size LBP+Contour feature vectors from the audio data and one 1×260 size LBP+Contour feature vector from the image data are combined to form 1×780-dimensional LBP+Contour feature vector. Likewise, two (SFOSE and SC) 1×260 SIFT+Contour feature vectors from the audio data and a 1×260 size SIFT+Contour feature vector from the image data are combined to generate a 1×780 size SIFT+ Contour feature vector. After the standard scaling phase, these vectors are utilized for training the OC-SVM classifier. Before the previously indicated min-max normalization procedure, two scores for each vector (LBP, SIFT) obtained from testing the classifier are averaged. Whether the average score surpasses the threshold value determines if the examined signature is genuine or a forgery. As a result, the fusion of a signature sound and an associated signature image has been achieved, increasing the success of classification.

Four genuine signatures are utilized for training by employing leave-one-out cross-validation, and four forgeries are used for testing. In the testing, each of the four forgeries, and each left-out genuine signature during the leave-one-out cross-validation procedures, are sequentially utilized together. Figure 6.1 shows the flowchart for the proposed shallow learning-based approach.

6.1 Test Results

For 93 participants consisting of 55 male and 38 female participants in the age range 19 to 64, the values of the False Acceptance Rate (FAR), Genuine Acceptance Rate (GAR), False Reject Rate (FRR), and Equal Error Rate (EER) are determined. The value at which the FRR and FAR values are equal is recognized as the EER. When the FRR and FAR values are not even, the closest FRR and FAR values are chosen, and the EER is computed across these values using linear interpolation. The verification outcomes are based on combinations of pen-paper types and phone models. For the convenience of presentation, the dataset developed for this study is given the name SKU (See Chapter 2).

6.1.1 Results of the Tests Using Only Static Offline Signature Image Data

The proposed method is applied to the static signature images in the GPDS-100 [18] and MCYT-75 [29] public offline signature datasets as well as the static signature images in the dataset collected from scratch (SKU) to get an idea about the reliability of the static signature image data collected within the context of this study. Table 6.1 presents results.

Table 6.1 Application of the proposed methodology to offline signature images in the dataset built from scratch (SKU) and to publicly available offline signature datasets (MCYT, GPDS).

Data Set	#Participants	#Samples	Results
MCYT-75 [29]	75	4	EER: 11.11%
GPDS-100 [18]	100	4	EER: 9.67%
SKU (ballpoint pen and plain paper)	**93**	**4**	**EER: 11.47%**
SKU (rollerball pen and plain paper)	**93**	**4**	**EER: 11.29%**
SKU (ballpoint pen and thin paper)	**93**	**4**	**EER: 5.02%**
SKU (rollerball pen and thin paper)	**93**	**4**	**EER: 10.22%**

Tables 6.2 and 6.3 present the results from the proposed method and a few state-of-the-art studies when they were applied to publicly available offline (static) signature datasets (MCYT, GPDS). In paper [52], which offers a comprehensive overview of recent works on offline signature verification, several kinds of research are discussed in detail.

Table 6.2 Based on the MCYT-75 dataset, a comparison between the proposed approach and some of the state-of-the-art publications.

Study	Feature Extraction	Classification	#Samples	Results
Masoudnia et al. [53]	CNN	SVM	10	EER: 5.85%
Ooi et al. [23]	DRT with PCA	PNN	5	EER: 9.87%
Zois et al. [54]	Poset Grid Features	SVM	5	EER: 6.02%
Maergner et al. [55]	Keypoint Graphs	GED, Bipartite	10	EER: 12.01%
Hafemann et al. [24]	CNN(SigNet)	SVM	10	EER: 2.87%
Proposed method for only signature image	**LBP and SIFT**	**OC-SVM**	**4**	**EER: 11.11%**

Table 6.3 Based on the GPDS dataset, a comparison between the proposed approach and some of the state-of-the-art publications.

Study	Feature Extraction	Classification	#Samples	Results
Hafemann et al. [56]	CNN	SVM	14	EER: 10.70%
Hafemann et al. [24]	CNN(SigNet)	SVM	5	EER: 2.41%
Xing et al.[57]	Convolutional Siamese	Cosine, Euclidean	54	EER: 10.37%
Narwade et al. [58]	Pixel Matching Features	SVM	12	EER:8.71%
Proposed method for only signature image	**LBP and SIFT**	**OC-SVM**	**4**	**EER: 9.67%**

Table 6.4 Equal error rate values for participants' genders (Male (M), Female (F)) and ages (<30, ≥30) when using just static data (signature image).

Combination (SKU dataset)	Gender		Age		Results (Gender)(EER)		Results (Age)(EER)	
	#M	#F	#(<30)	#(≥30)	M	F	<30	≥30
Ballpoint Pen-Plain Paper	55	38	63	30	7.27%	16.84%	12.06%	10.00%
Rollerball Pen-Plain Paper	55	38	63	30	11.82%	9.47%	9.52%	12.38%
Ballpoint Pen-Thin Paper	55	38	63	30	7.88%	1.32%	6.35%	2.22%
Rollerball Pen-Thin Paper	55	38	63	30	10.65%	9.47%	11.64%	6.67%

Only the offline signature (static) data from the dataset is used to derive the EER values of the verification successes, which are also computed specifically for gender and age (See Table 6.4).

Figure 6.2 shows Receiver Operating Characteristic (ROC) curve according to averaged results obtained from all pen-paper combinations in Table 6.1 based only on offline (static) signature data.

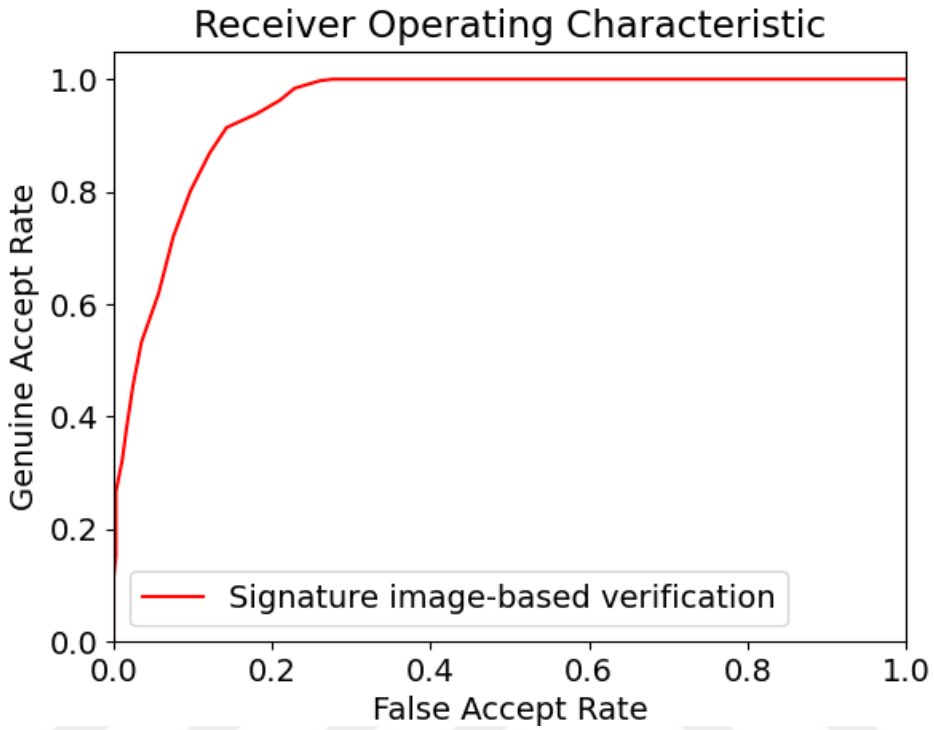


Figure 6.2 ROC curve for static offline signature verification according to averaged error rates obtained from all pen-paper combinations.

6.1.2 Results of the Tests Using Only Dynamic Signature Sound Data

The features extracted from images of the Spectral Flux Onset Strength Envelope (SFOSE) of the signature sound only, the features extracted from the Spectral Centroid (SC) image of the signature sound only, and the combined feature vectors (combined LBP+contour features and combined SIFT+contour features) extracted from these images (SFOSE and SC) produced for enhancing accuracy, are used separately for verification with audio data only. To make signature verification using sound data alone, four genuine and four forged signature sound samples from 93 participants are employed. The two image files (Image of SFOSE, Image of SC) are acquired from the audio data, and each of these image files is converted into two feature vectors (LBP, SIFT) separately (See Figure 6.1). OC-SVM is employed in the classification phase, along with leave-one-out cross-validation. A score-level fusion is performed by taking the average of the obtained scores resulting from LBP and SIFT features. Table 6.5 presents the findings.

Table 6.5 Verification results for audio data alone (Number of participants: 93)

Combination (SKU dataset)	Results (SFOSE [59])	Results (SC [60])	Results (SFOSE+SC)
Ballpoint Pen-Plain Paper-Galaxy Note 3	EER: 9.97%	EER: 10.04%	EER: 5.38%
Rollerball Pen-Plain Paper-Galaxy Note 3	EER: 1.61%	EER: 13.69%	EER: 1.08%
Ballpoint Pen-Thin Paper-Galaxy Note 3	EER: 5.38%	EER: 14.97%	EER: 4.30%
Rollerball Pen-Thin Paper-Galaxy Note 3	EER: 3.09%	EER: 15.16%	EER: 2.15%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	EER: 5.91%	EER: 16.62%	EER: 4.52%
Rollerball Pen-Plain Paper-iPhone 7 Plus	EER: 2.15%	EER: 9.95%	EER: 1.79%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	EER: 6.45%	EER: 9.14%	EER: 3.87%
Rollerball Pen-Thin Paper-iPhone 7 Plus	EER: 5.81%	EER: 6.99%	EER: 1.08%

According to gender and age, separate calculations are made for the Equal Error Rate values of the verification successes that resulted from the audio data. According to gender, EER values for verification using only audio data are shown in Table 6.6. Table 6.7 lists the Equal Error Rate values for verification using audio data exclusively by age ranges.

Table 6.6 Verification results for audio data alone according to gender (Number of male (M) participants: 55, number of female (F) participants: 38).

Combination (SKU dataset)	Results (SFOSE)(EER)		Results (SC)(EER)		Results (SFOSE+SC)(EER)	
	M	F	M	F	M	F
Ballpoint Pen-Plain Paper-Galaxy Note 3	8.73%	11.84%	8.36%	12.50%	3.38%	7.89%
Rollerball Pen-Plain Paper-Galaxy Note 3	0.91%	2.63%	10.65%	18.42%	0.00%	2.63%
Ballpoint Pen-Thin Paper-Galaxy Note 3	7.27%	2.63%	15.58%	11.84%	4.55%	3.95%
Rollerball Pen-Thin Paper-Galaxy Note 3	2.73%	3.95%	15.00%	13.16%	1.82%	2.63%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	5.09%	7.89%	11.95%	23.68%	3.38%	5.26%
Rollerball Pen-Plain Paper-iPhone 7 Plus	1.82%	2.63%	8.83%	11.58%	1.82%	1.97%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	4.24%	9.65%	12.73%	5.26%	6.06%	0.00%
Rollerball Pen-Thin Paper-iPhone 7 Plus	4.55%	7.89%	8.00%	4.93%	1.21%	2.63%

Table 6.7 Verification results for audio data alone according to age of participants (Number of participants younger than 30 (<30): 63, number of participants older than 30 (≥30): 30).

Combination (SKU dataset)	Results (SFOSE)(EER)		Results (SC)(EER)		Results (SFOSE+SC)(EER)	
	<30	≥30	<30	≥30	<30	≥30
Ballpoint Pen-Plain Paper-Galaxy Note 3	9.84%	10.00%	9.52%	10.83%	4.76%	6.67%
Rollerball Pen-Plain Paper-Galaxy Note 3	2.38%	0.00%	11.11%	17.50%	1.59%	0.00%
Ballpoint Pen-Thin Paper-Galaxy Note 3	6.35%	3.33%	12.02%	21.11%	5.82%	2.22%
Rollerball Pen-Thin Paper-Galaxy Note 3	3.70%	1.67%	13.85%	17.78%	3.17%	0.00%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	3.97%	10.00%	16.27%	16.67%	4.37%	3.33%
Rollerball Pen-Plain Paper-iPhone 7 Plus	3.17%	0.00%	8.25%	13.33%	1.67%	1.61%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	7.94%	3.33%	3.17%	23.33%	2.65%	6.67%
Rollerball Pen-Thin Paper-iPhone 7 Plus	6.98%	3.33%	5.56%	10.00%	1.59%	2.22%

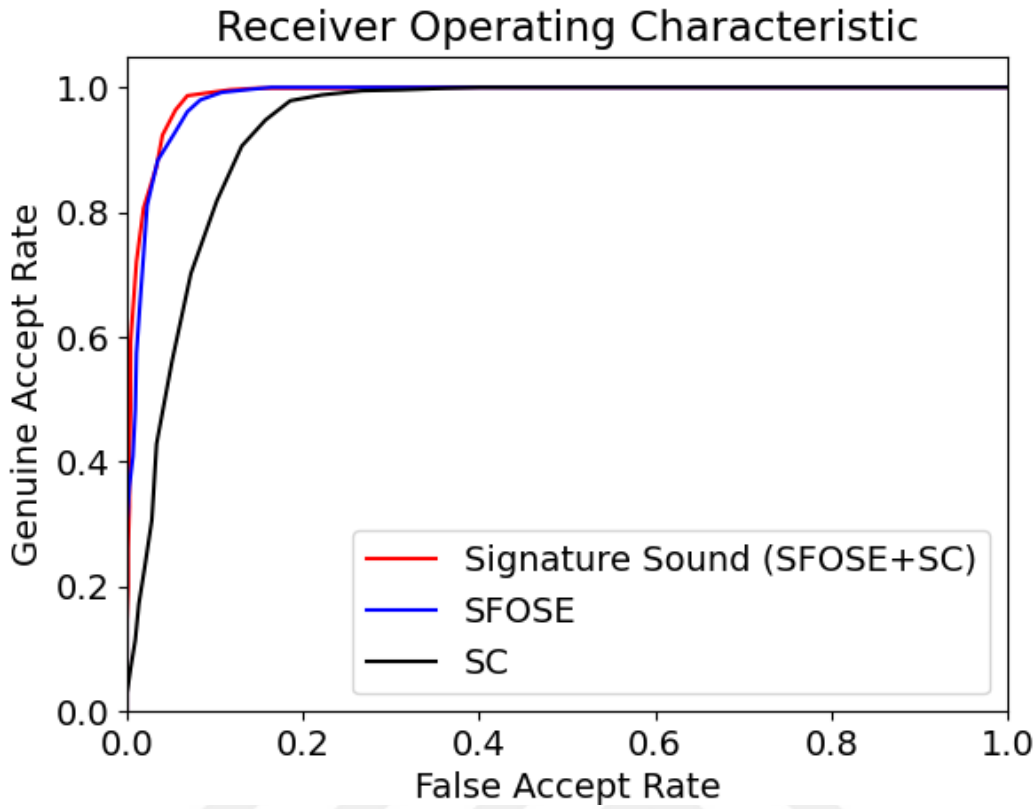


Figure 6.3 ROC curves for spectral flux onset envelope of audio data, spectral centroid of audio data and fusion of spectral flux onset envelope and spectral centroid of audio data

Figure 6.3 shows the ROC curves plotted from only sound-based signature verification, for all participants, according to features based on only the Spectral Flux Onset Envelope of signal, features based on only the Spectral Centroid of the audio signal, and features based on a fusion of Spectral Flux Onset Envelope and Spectral Centroid of the audio signal. Curves correspond to the averaged values obtained from pen-paper-phone type combinations in Table 6.5.

By fusing the two features (SFOSE and SC) obtained from sound, it is evident that the success of the verification increased. The t-test statistical significance test [61] is used to confirm this. P value is 0.0004808 as a consequence. The hypothesis that the combination of features enhances the success of verification can be considered true because this value is less than 0.05.

Numerous studies carried out signature verification with sound data (See Section 1.1.1). In Table 6.8, a comparison of these researches with the proposed approach based only on audio data can be seen.

Table 6.8 Comparison of the proposed approach on sound data only with some of the signature sound verification studies on their own dataset

Study	#Participants	#Samples	Feature Extraction	Classification	Results
Li (2004, 2010) [6] [5]	5	10 genuine, 10 forged	Normalized Hilbert envelope of sounds	multi-layer back propagation neural network	≥ 75% correctness for different scenarios
Khazei et al. (2012) [10]	30	10 genuine	AR coefficients, cepstrum based features	Euclidean, Manhattan, Chessboard	49%≤ EER≤ 50.133%
Armiato et al. (2016) [11]	55	10 genuine, 2 forged	Combination of wavelet-based features.	Euclidean classifier, Modified correlation classifier	≥ 80% accuracy
Ding et al. (2019) [13]	14	112 genuine, 60 forged	A chord-based method, to estimate phase-related changes caused by small activities.	Deep CNN	EER: 5.5% AUC: 98.7%
Chen et al. (2020) [14]	35	20 genuine, 20 forged	SSIM, PSNR, MSE, and Hausdorff distance.	LR,NB, RF, SVM	EER: 1.25% AUC: 98.2%
Wei et al. (2021) [15]	12	70 genuine, 60 forged	Zero Crossing Rate, Spectral Centroid, Spectral Spread, Sprectral Flux, Spectral Entropy, Spectral Rolloff.	One-Class classifier based on CNN	EER: 5% AUC: 98.4%
Zhao et al. (2021) [16]	40	32 genuine, 28 forged	Spatio-Temporal Features from Channel Impulse Response (CIR)	CNN-based Multi-Modal Siamese Network	EER: 3.27%
Proposed (Ballpoint Pen-Plain Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 5.38%**
Proposed (Rollerball Pen-Plain Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER:1.08%**
Proposed (Ballpoint Pen-Thin Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 4.30%**
Proposed (Rollerball Pen-Thin Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 2.15%**
Proposed (Ballpoint Pen-Plain Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 4.52%**

Continued on next page

Table 6.8 – continued from previous page

Study	#Participants	#Samples	Feature Extraction	Classification	Results
Proposed (Rollerball Pen-Plain Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 1.79%**
Proposed (Ballpoint Pen-Thin Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 3.87%**
Proposed (Rollerball Pen-Thin Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 1.08%**

6.1.3 Results of the Tests Using Fusion of Offline (Static) Signature Data and Signature Sound (Dynamic) Data

By using the signature image and the signature sound together, the proposed approach aims to enhance classification accuracy. This purpose is accomplished by combining the combined features (LBP and SIFT) of SFOSE and SC graphic images of audio data with the associated LBP and SIFT feature vectors extracted from the offline (static) signature image (See Figure 6.1). Then, signature verification results are generated for the LBP and SIFT vectors separately, and the results obtained for each of these vectors are averaged. Thus a kind of score-level fusion is performed. Table 6.9 provides EER values by gender and age for the verification of the fusion of offline (static) signature data with signature sound (dynamic) data.

Table 6.9 Results of verification for the fusion of offline (static) signature data with dynamic signature sound data, segmented by participant age (<30, ≥30) and gender (Male (M), Female (F)).

Combination (SKU dataset)	Gender		Age		EER (Gender)		EER (Age)	
	#M	#F	#(<30)	#(≥30)	M	F	<30	≥30
Ballpoint Pen-Plain Paper-Galaxy Note 3	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Rollerball Pen-Plain Paper-Galaxy Note 3	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Ballpoint Pen-Thin Paper-Galaxy Note 3	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Rollerball Pen-Thin Paper-Galaxy Note 3	55	38	63	30	1.82%	0.00%	1.59%	0.00%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	55	38	63	30	1.82%	0.00%	0.00%	3.33%
Rollerball Pen-Plain Paper-iPhone 7 Plus	55	38	63	30	0.00%	1.75%	1.06%	0.00%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	55	38	63	30	1.56%	0.00%	0.00%	2.67%
Rollerball Pen-Thin Paper-iPhone 7 Plus	55	38	63	30	0.00%	0.00%	0.00%	0.00%

Table 6.10 lists the verification findings obtained using signature sound (dynamic) data only (SFOSE+SC), offline (static) signature image data only, and a combination of both signature sound (dynamic) and offline (static) signature data. The statistical significance test revealed that the p-value is 3.80×10^{-5} to confirm the outcomes in

Table 6.10. It can be concluded from this outcome that the findings in this table are not a coincidence.

Table 6.10 Results of verification for the fusion of static (offline) signature data and dynamic signatures sound data using various types of used pens, papers, and mobile phone models

Combination (SKU dataset)	Sig. Sound	Sig. Image	Fusion of signature sound and signature image
B.Point Pen-Plain Paper-Galaxy Note 3	EER: 5.38%	EER: 11.47%	GAR: 100.00% FAR:0.00% FRR: 0.00% **EER: 0.00%**
R.Ball Pen-Plain Paper-Galaxy Note 3	EER: 1.08%	EER: 11.29%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
B.Point Pen-Thin Paper-Galaxy Note 3	EER: 4.30%	EER: 5.02%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
R.Ball Pen-Thin Paper-Galaxy Note 3	EER: 2.15%	EER: 10.22%	GAR: 98.92% FAR: 1.08% FRR: 1.08% **EER: 1.08%**
B.Point Pen-Plain Paper-iPhone 7 Plus	EER: 4.52%	EER: 11.47%	GAR: 98.92% FAR: 1.08% FRR: 1.08% **EER: 1.08%**
R.Ball Pen-Plain Paper-iPhone 7 Plus	EER: 1.79%	EER: 11.29%	GAR: 96.77% FAR: 0.00% FRR: 3.23% **EER: 0.08%**
B.Point Pen-Thin Paper-iPhone 7 Plus	EER: 3.87%	EER: 5.02%	GAR: 93.55% FAR: 0.00% FRR: 6.45% **EER: 0.09%**
R.Ball Pen-Thin Paper-iPhone 7 Plus	EER: 1.08%	EER: 10.22%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**

Figure 6.4 shows the ROC curves for verification findings obtained using signature sound (dynamic) data (SFOSE+SC), offline (static) signature data, and a fusion of signature sound (dynamic) data, and offline (static) signature data. Curves reflect the averaged results from the combinations in Tables 6.1, 6.5, and 6.10.

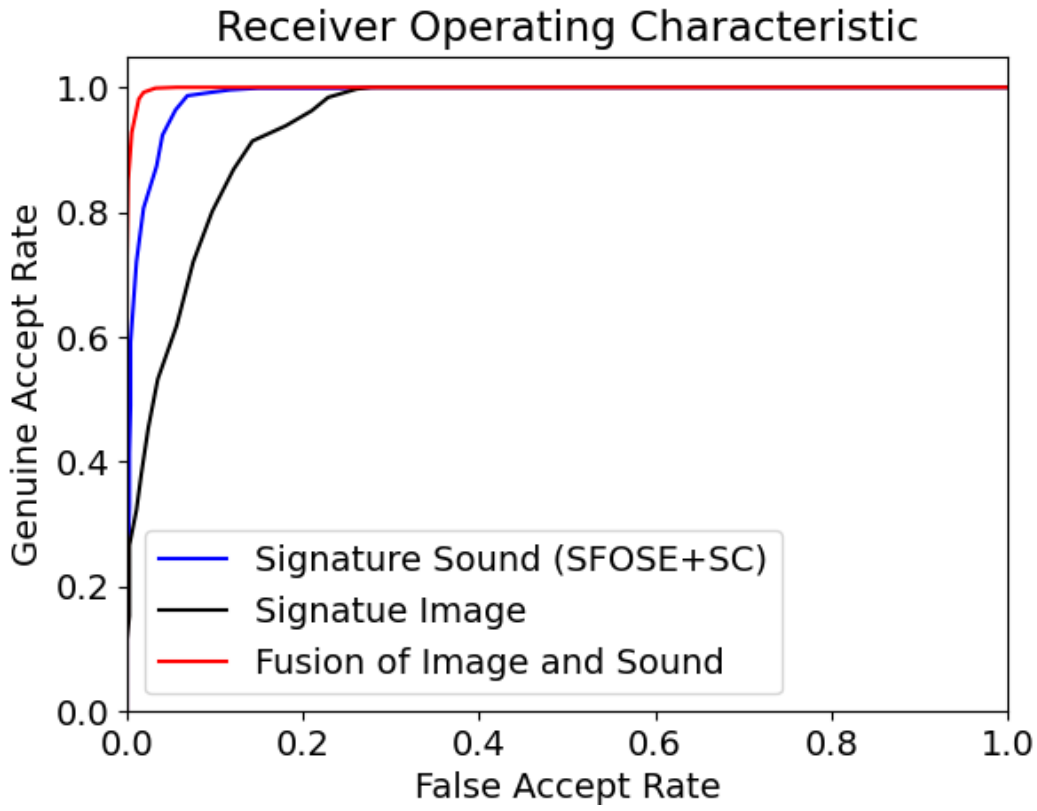


Figure 6.4 ROC Curves for signature sound (dynamic) data only, offline (static) signature data only, and fusion of signature sound (dynamic) data and offline (static) signature data

6.1.4 Verification Results and Analysis when Query and Reference Signature Pen-Paper-Phone Combinations are Different

It is analyzed how the cases where query signature data and reference signature data are obtained with different combinations of pen, paper, and mobile phones will affect the success of signature verification. Here, the reference signatures represent the four genuine signatures that the classifier is trained with leave-one-out cross-validation. Query signatures are signatures used to test the trained classifier. Query signatures can be genuine or forgery. To test the authenticity of signatures, query signatures using the different pen, paper, and phone combinations are cross-tested with reference signatures using different combinations.

In this test phase, forged query signatures and reference signatures had the same combination of paper and pen phone models. The genuine query signatures had different combinations than reference and forged query signatures. In other words, while we have four genuine (reference) signatures, the system is tested with four forgeries with the same combination (pen-paper-phone model) of these reference signatures. The system is also tested with four genuine signatures in different combinations (pen-paper-phone model) from the reference signatures. Thus, it is analyzed to what extent the system can detect genuine signatures when the reference and forgeries are in the same combination, while genuine signatures have different combinations (pen-paper-phone model). The results are in Table 6.11. For ease of display, the combination values in Table 6.11 are coded with the letter "B" instead of the ballpoint pen, the letter "R" instead of the rollerball pen, the letter "P" instead of plain paper, the letter "T" instead of thin paper, the letter "S" for Samsung Galaxy Note 3, and the letter "I" instead of iPhone 7 Plus.

Table 6.11 Verification results according to different genuine query signature combinations (pen-paper-phone model) versus the reference (same combination with forged query signatures) signature combinations

Reference Sign. Combination	Query Sign. Combination	Sound (SFOSE+SC)	Sound&Image Fusion
B-P-S	R-P-S	EER: 17.56%	EER: 12.44%
B-P-S	B-T-S	EER: 8.60%	EER: 7.17%
B-P-S	R-T-S	EER: 21.08%	EER: 13.99%
B-P-S	B-P-I	EER: 21.24%	EER: 7.53%
B-P-S	R-P-I	EER: 29.57%	EER: 20.30%
B-P-S	B-T-I	EER: 17.20%	EER: 15.41%
B-P-S	R-T-I	EER: 26.88%	EER: 19.51%
R-P-S	B-P-S	EER: 14.70%	EER: 12.26%
R-P-S	B-T-S	EER: 9.27%	EER: 8.24%
R-P-S	R-T-S	EER: 7.53%	EER: 5.91%
R-P-S	B-P-I	EER: 17.67%	EER: 15.41%
R-P-S	R-P-I	EER: 10.32%	EER: 1.08%
R-P-S	B-T-I	EER: 15.17%	EER: 9.46%
R-P-S	R-T-I	EER: 10.60%	EER: 6.67%

Continued on next page

Table 6.11 – continued from previous page

Reference Sign. Combination	Query Sign. Combination	Sound (SFOSE+SC)	Sound&Image Fusion
B-T-S	B-P-S	EER: 13.21%	EER: 9.52%
B-T-S	R-P-S	EER: 20.28%	EER: 15.86%
B-T-S	R-T-S	EER: 21.51%	EER: 16.13%
B-T-S	B-P-I	EER: 29.72%	EER: 21.15%
B-T-S	R-P-I	EER: 36.99%	EER: 25.81%
B-T-S	B-T-I	EER: 21.25%	EER: 2.69%
B-T-S	R-T-I	EER: 26.88%	EER: 21.15%
R-T-S	B-P-S	EER: 23.11%	EER: 19.83%
R-T-S	R-P-S	EER: 10.14%	EER: 8.60%
R-T-S	B-T-S	EER: 12.69%	EER: 14.11%
R-T-S	B-P-I	EER: 27.96%	EER: 27.31%
R-T-S	R-P-I	EER: 24.19%	EER: 18.28%
R-T-S	B-T-I	EER: 19.94%	EER: 20.26%
R-T-S	R-T-I	EER: 12.61%	EER: 1.61%
B-P-I	B-P-S	EER: 11.02%	EER: 0.00%
B-P-I	R-P-S	EER: 9.68%	EER: 6.45%
B-P-I	B-T-S	EER: 7.10%	EER: 5.38%
B-P-I	R-T-S	EER: 11.83%	EER: 9.06%
B-P-I	R-P-I	EER: 12.60%	EER: 9.68%
B-P-I	B-T-I	EER: 8.60%	EER: 5.65%
B-P-I	R-T-I	EER: 10.22%	EER: 7.05%
R-P-I	B-P-S	EER: 21.39%	EER: 16.13%
R-P-I	R-P-S	EER: 6.81%	EER: 2.15%
R-P-I	B-T-S	EER: 12.75%	EER: 10.75%
R-P-I	R-T-S	EER: 8.87%	EER: 7.89%
R-P-I	B-P-I	EER: 16.13%	EER: 12.66%
R-P-I	B-T-I	EER: 12.37%	EER: 9.68%
R-P-I	R-T-I	EER: 5.16%	EER: 5.02%
B-T-I	B-P-S	EER: 16.49%	EER: 16.13%
B-T-I	R-P-S	EER: 15.59%	EER: 11.18%
B-T-I	B-T-S	EER: 8.96%	EER: 2.15%
B-T-I	R-T-S	EER: 18.49%	EER: 10.75%
B-T-I	B-P-I	EER: 14.70%	EER: 13.44%
B-T-I	R-P-I	EER: 20.74%	EER: 13.86%
B-T-I	R-T-I	EER: 16.59%	EER: 9.68%
R-T-I	B-P-S	EER: 24.30%	EER: 21.51%
R-T-I	R-P-S	EER: 10.39%	EER: 9.68%
R-T-I	B-T-S	EER: 16.67%	EER: 18.49%
R-T-I	R-T-S	EER: 6.45%	EER: 0.07%
R-T-I	B-P-I	EER: 20.43%	EER: 18.28%
R-T-I	R-P-I	EER: 10.97%	EER: 11.67%
R-T-I	B-T-I	EER: 13.76%	EER: 16.13%

The average equal error rate (EER) of the proposed verification system is examined by analyzing Tables 6.10 and 6.11 to see how much it is impacted when the query signature's pen-paper-phone combinations are different from the reference signature's pen-paper-phone combinations. The impact of each pen, paper, and phone item on the average equal error rate is detected individually (See Table 6.12).

Table 6.12 The average rates of increasing the EER when the pen-paper-phone combination of the query signature is different from the pen-paper-phone combination of the reference signature

Differnt Item for Query Signature	Average Increase for Sound	Average Increase for Fusion of Sound and Image
Pen	EER: 12.67%	EER: 12.60%
Paper	EER: 6.84%	EER: 8.08%
Phone	EER: 9.31%	EER: 2.14%
Paper-Phone	EER: 12.55%	EER: 12.28%
Pen-Phone	EER: 17.02%	EER: 15.83%
Pen-Paper	EER: 14.17%	EER: 13.06%
Pen-Paper-Phone	EER: 18.41%	EER: 16.53%

If only the pen in the pen-paper-phone combination of the query signature is different from the pen-paper-phone combination of the reference signature, it has been calculated how the test (query) signature affects the EER. In this condition, the query signature increased EER by an average of 12.67% in the verification based on sound data only. In signature verification based on the fusion of signature sound and signature image, it increased the EER by an average of 12.60%.

If just the paper is changed in the paper-pen-phone combination for the query signature, it has been found that the test (query) signature raises the EER by an average of 6.84% in the verification based only on sound data. Additionally, based on the fusion of the signature sound and signature image, the query signature increased the EER by an average of 8.08%.

It has been determined that the query signature raised the EER by an average of 9.31% in the verification based solely on sound data if only the phone is different in the paper-pen-phone combination for the query signature. The same query signature increased the EER by an average of 2.14% in signature verification based on the fusion of signature sound and signature image. Since the sounds of the same signature (same signature image) are recorded by two different phones simultaneously, these values are relatively lower, and the verification success is higher. Therefore, it is necessary to test the system with query signatures using different combinations of phone-paper, phone-pen, and paper-pen pairs to determine the effect of the phone difference in the query signature, as detailed below.

If the paper and the phone are different in the paper-pen-phone combination for the query signature, it is calculated that the test (query) signature increases the EER by an average of 12.55% in the verification based on sound data only. In signature verification based on the fusion of signature sound and image, the query signature increased the EER by an average of 12.28%.

When the pen and the phone used for the query signature differ from the pen-paper-phone combination for the reference signature, it has been determined that the test (query) signature raises the EER by an average of 17.02% in the verification based solely on sound data. It also increased the EER by an average of 15.83% in signature verification based on the fusion of signature sound and image.

It has been found that the test (query) signature raised the EER by an average of 14.17% in the verification based only on sound data when the pen and the paper used for the query signature are different from the pen-paper-phone combination for the reference signature. Additionally, it decreased signature verification success based on the fusion of sound and image by an average of 13.06% EER.

When the pen-paper-phone combination of the query signature is completely different from the pen-paper-phone combination of the reference signature, the sound-only verification rate increases by an average of 18.41% EER. In verification based on the fusion of signature and voice, there is an average increase of 16.53% EER.

In light of these results, it can be deduced that the item that affects the signature verification performance the most is the changes in the pen type, then the phone model, and the least in the paper type.

Equal error rates due to the differences in the pen-paper combinations of the query and reference signatures used in the signature verification with only the signature image are given in Table 6.13. (Combinations in Table 6.13 do not include the phone item because only the signature image verifications are based, no sound data is used.)

Table 6.13 Verification results according to different genuine query **static signature** combinations (pen-paper) versus the reference (same combination with forged query signatures) **static signature** combinations

Reference Sign. Combination	Query Sign. Combination	Results (Static Signature Image)
B-P	R-P	EER: 21.29%
B-P	B-T	EER: 22.12%
B-P	R-T	EER: 22.89%
R-P	B-T	EER: 24.19%
R-P	R-T	EER: 13.73%
R-P	B-P	EER: 15.41%
B-T	B-P	EER: 21.15%
B-T	R-P	EER: 19.75%
B-T	R-T	EER: 20.07%

7
DEEP LEARNING BASED APPROACH

Deep Learning is used to solve challenges across many different sectors. Deep learning algorithms have been used for motion detection, face recognition, autonomous car systems, speaker recognition, data mining, and so forth. This study also examined the effectiveness of classification or verification utilizing features extracted via deep learning models. During the preprocessing stage, the size of the image files is cut in half. Grayscale conversion and inversion to the negative images are applied. With the help of pre-trained models built on the foundation of deep convolutional neural networks [62] [63], such as SigNet [24] (a model formed using just signature data), VGG-16 [50], VGG-19 [50], and ResNet-50 [49], features are extracted. The generated feature vectors are scaled to the 1×256 dimension. On each of the datasets we have available, these pre-trained models have been evaluated and contrasted (See Table 7.1). As in the preceding chapter, OC-SVM is used in the classification phase. Results overwhelmingly demonstrate that, for signature datasets, SigNet outperforms other pre-trained models. To verify signatures using the deep learning-based approach with the dataset built for this research (SKU), SigNet is employed as a pre-trained model.

For the purpose of verifying a signature using just audio data, two image files, Spectral Flux Onset Strength Envelope of Signal (SFOSE) and Spectral Centroid (SC) of Signal, are derived from the audio data of the signature. From each of these images, one 1×256-dimensional SigNet deep feature vector (CNN-based deep features utilizing SigNet Model) is produced. For feature fusion based on audio data, 1×256 size SigNet deep feature vectors from both audio-based images (SFOSE, SC) are combined to generate a 1×512 dimensional SigNet deep feature vector. By eliminating the mean and scaling to unit variance, the feature vector is standardized. This 1×512 size vector is employed in the OC-SVM classifier. Using min-max normalization, the classification score is normalized between 0 and 1. Depending on a threshold value, this score indicates whether the signature represented by the combined vector is genuine or a forgery.

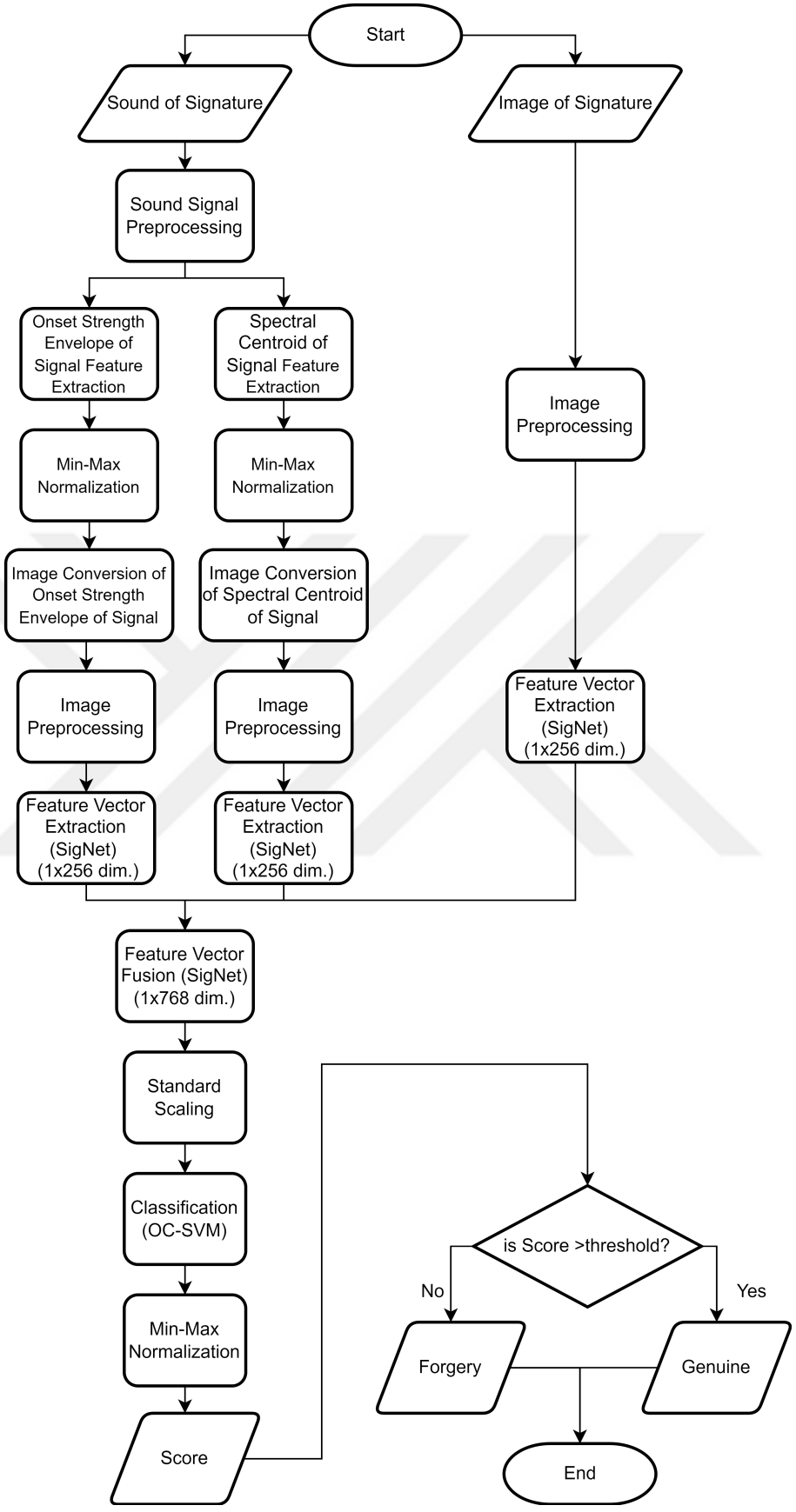


Figure 7.1 Flowchart for deep learning-based approach

For verification using only the static signature image, 1×256 SigNet deep feature vectors are produced for each signature sample. Following the standard scaling phase, the classifier is trained using these vectors. The previously indicated min-max normalization procedure is applied to the test scores of the trained OC-SVM classifier. Whether the score is greater than or less than the threshold value determines if the examined signature is genuine or a forgery.

To improve the success of the verification, audio and image data are fused. Two (SFOSE and SC) 1×256 size SigNet deep feature vectors from the audio data and one 1×256 size SigNet deep feature vector from the image data are combined to form a 1×768-dimensional SigNet deep feature vector. These vectors are used to train the classifier after the standard scaling step. The score corresponding to a 1×768-dimensional SigNet deep feature vector is acquired by testing the OC-SVM classifier. After all, min-max normalization is performed. The examined signature's genuineness is determined by whether the score exceeds the threshold value. As a consequence, the fusion of signature sounds and corresponding signature images has been accomplished, improving classification accuracy.

Using leave-one-out cross-validation, four genuine signatures are used for training, while four forgeries are used for only testing. In the testing stage, each of the four forgeries, and each left-out genuine signature during the leave-one-out cross-validation procedures, are utilized together. The flowchart of the proposed deep learning-based approach is shown in Figure 7.1.

Table 7.1 Pretrained models are compared using offline (static) signature datasets. Run times are based on the verification procedures belonging to each participant in the dataset. In SKU dataset ballpoint pen-plain paper combination is used. (Computer specifications: Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz processor, 32 GB RAM, Windows 10 OS)

Dataset	#Participants	VGG-16		VGG-19		ResNet		SigNet	
		Time(s.)	EER	Time(s.)	EER	Time(s.)	EER	Time(s.)	EER
SKU	93	4.14	24.57%	5.84	9.00%	58.48	10.22%	7.27	**3.23%**
GPDS-100	100	4.40	28.80%	5.72	16.69%	63.19	15.00%	7.53	**3.50%**
MCYT-75	75	4.57	24.00%	5.66	18.00%	67.69	17.14%	8.80	**5.33%**

7.1 Test Results

The values of the False Acceptance Rate (FAR), Genuine Acceptance Rate (GAR), False Reject Rate (FRR), and Equal Error Rate (EER) are measured for 93 participants, 55 male and 38 female, ranging in age from 19 to 64. The EER is defined as the value where the FRR and FAR values are identical. When the FRR and FAR values are not equivalent, the closest FRR and FAR values are picked, and the EER is derived over

these values using linear interpolation. The verification results are based on several pen-paper and phone model combinations. The dataset produced for this research is given the name SKU for convenience of presentation (See Chapter 2).

7.1.1 Results of the Tests for the Deep Learning-Based Approach Using Only Offline (Static) Signature Image Data

The results of using the deep learning-based technique on the static image dataset built from scratch (SKU) are shown in Table 7.2 alongside those of other publicly accessible offline signature datasets.

Table 7.2 Application of the deep learning-based approach to offline signature images in the dataset built from scratch (SKU) and to publicly available offline signature datasets (MCYT, GPDS).

Data Set	#Participants	#Samples	Results
MCYT-75 [29]	75	4	EER: 5.33%
GPDS-100 [18]	100	4	EER: 3.50%
SKU (ballpoint pen and plain paper)	**93**	**4**	**EER: 3.23%**
SKU (rollerball pen and plain paper)	**93**	**4**	**EER: 6.88%**
SKU (ballpoint pen and thin paper)	**93**	**4**	**EER: 3.23%**
SKU (rollerball pen and thin paper)	**93**	**4**	**EER: 6.09%**

Table 7.3 Based on the MCYT-75 dataset, a comparison between the deep-learning approach and some of the state-of-the-art publications.

Study	Feature Extraction	Classification	#Samples	Results
Masoudnia et al. [53]	CNN	SVM	10	EER: 5.85%
Ooi et al. [23]	DRT with PCA	PNN	5	EER: 9.87%
Zois et al. [54]	Poset Grid Features	SVM	5	EER: 6.02%
Maergner et al. [55]	Keypoint Graphs	GED, Bipartite	10	EER: 12.01%
Hafemann et al. [24]	CNN (SigNet)	SVM	10	EER: 2.87%
Proposed method for only signature image	**CNN (SigNet)**	**OC-SVM**	**4**	**EER: 5.33%**

Table 7.4 Based on the GPDS dataset, a comparison between the proposed approach and some of the state-of-the-art publications.

Study	Feature Extraction	Classification	#Samples	Results
Hafemann et al. [56]	CNN	SVM	14	EER: 10.70%
Hafemann et al. [24]	CNN (SigNet)	SVM	5	EER: 2.41%
Xing et al.[57]	Convolutional Siamese	Cosine, Euclidean	54	EER: 10.37%
Narwade et al. [58]	Pixel Matching Features	SVM	12	EER:8.71%
Proposed method for only signature image	**CNN (SigNet)**	**OC-SVM**	**4**	**EER: 3.50%**

Results from the proposed deep learning-based approach and a few state-of-the-art research when they were applied to publicly accessible offline (static) signature datasets (MCYT, GPDS), are shown in Tables 7.3 and 7.4. A variety of study types are covered in depth in paper [52], which provides a thorough summary of current developments on offline signature verification.

The EER values of the verification successes are derived just from the offline signature (static) data from the dataset (SKU), and they are also separately computed according to genders and age ranges (See Table 7.5).

Table 7.5 Equal error rate values for participants' genders (Male (M), Female (F)) and ages (<30, ≥30) when using just static data (signature image).

Combination (SKU dataset)	Gender		Age		Results (Gender)(EER)		Results (Age)(EER)	
	#M	#F	#(<30)	#(≥30)	M	F	<30	≥30
Ballpoint Pen-Plain Paper	55	38	63	30	5.45%	0.00%	0.00%	10.00%
Rollerball Pen-Plain Paper	55	38	63	30	5.45%	7.89%	6.35%	3.34%
Ballpoint Pen-Thin Paper	55	38	63	30	3.64%	2.63%	3.17%	3.34%
Rollerball Pen-Thin Paper	55	38	63	30	3.64%	7.89%	8.73%	0.00%

Based only on offline (static) signature data, Figure 7.2 presents Receiver Operating Characteristic (ROC) curve for averaged results obtained from all pen-paper combinations in Table 7.2.

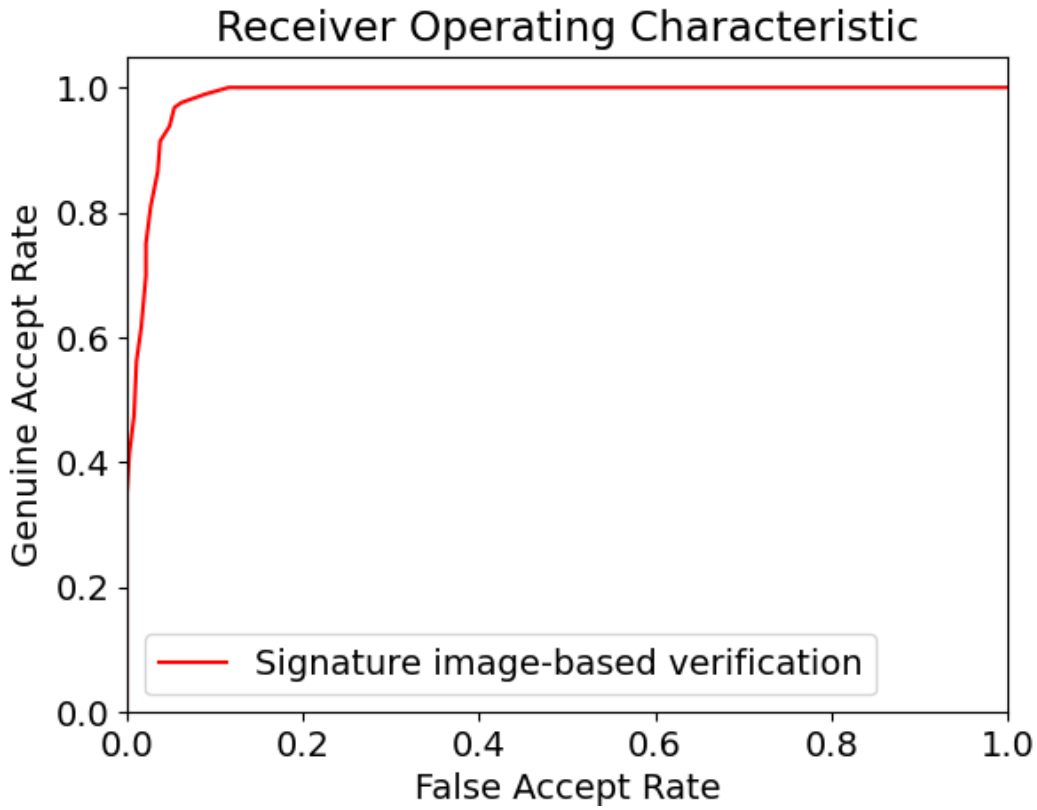


Figure 7.2 ROC curve for static offline signature verification according to averaged error rates obtained from all pen-paper combinations.

7.1.2 Results of the Tests Using Only Dynamic Signature Sound Data

The features extracted from the Spectral Centroid (SC) image of the signature sound alone, the features extracted from the images of the Spectral Flux Onset Strength

Envelope (SFOSE) of the signature sound alone, and the combined feature vectors (combined SigNet features) extracted from these images (SFOSE+SC) concatenated for improving accuracy, are used separately for verification utilizing sound data only. Four genuine and four forged signature sound samples from 93 participants are used to perform signature verification using just audio data. A Signet feature vector is generated from each of the two image files (SFOSE Image, SC Image) obtained from the audio data. (See Figure 7.1). In the classification stage, leave-one-out cross-validation is used together with OC-SVM. The results are shown in Table 7.6.

Table 7.6 Verification results for audio data alone (Number of participants: 93)

Combination (SKU dataset)	Results (SFOSE [59])	Results (SC [60])	Results (SFOSE+SC)
Ballpoint Pen-Plain Paper-Galaxy Note 3	EER: 7.17%	EER: 7.53%	EER: 5.59%
Rollerball Pen-Plain Paper-Galaxy Note 3	EER: 3.23%	EER: 6.09%	EER: 2.15%
Ballpoint Pen-Thin Paper-Galaxy Note 3	EER: 5.16%	EER: 9.68%	EER: 3.23%
Rollerball Pen-Thin Paper-Galaxy Note 3	EER: 3.87%	EER: 6.18%	EER: 1.94%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	EER: 3.23%	EER: 8.60%	EER: 2.15%
Rollerball Pen-Plain Paper-iPhone 7 Plus	EER: 2.22%	EER: 8.31%	EER: 4.57%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	EER: 5.38%	EER: 8.60%	EER: 2.96%
Rollerball Pen-Thin Paper-iPhone 7 Plus	EER: 6.45%	EER: 5.38%	EER: 5.22%

Table 7.7 Verification results for audio data alone according to gender (Number of male (M) participants: 55, number of female (F) participants: 38).

Combination (SKU dataset)	Results (SFOSE)(EER)		Results (SC)(EER)		Results (SFOSE+SC)(EER)	
	M	F	M	F	M	F
Ballpoint Pen-Plain Paper-Galaxy Note 3	8.48%	5.26%	5.45%	7.89%	6.82%	4.39%
Rollerball Pen-Plain Paper-Galaxy Note 3	3.64%	4.39%	3.64%	10.53%	0.00%	5.26%
Ballpoint Pen-Thin Paper-Galaxy Note 3	3.64%	6.58%	10.91%	7.89%	3.64%	3.95%
Rollerball Pen-Thin Paper-Galaxy Note 3	4.73%	2.63%	1.82%	10.53%	0.91%	3.51%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	2.27%	5.45%	9.09%	7.89%	1.82%	2.63%
Rollerball Pen-Plain Paper-iPhone 7Plus	1.14%	4.54%	5.45%	12.28%	3.03%	6.58%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	2.27%	1.82%	7.27%	10.53%	0.00%	7.02%
Rollerball Pen-Thin Paper-iPhone 7 Plus	2.27%	7.27%	3.64%	7.89%	6.36%	3.51%

Table 7.8 Verification results for audio data alone according to age of participants (Number of participants younger than 30 (<30): 63, number of participants older than 30 (≥30): 30).

Combination (SKU dataset)	Results (SFOSE)(EER)		Results (SC)(EER)		Results (SFOSE+SC)(EER)	
	<30	≥30	<30	≥30	<30	≥30
Ballpoint Pen-Plain Paper-Galaxy Note 3	6.35%	8.89%	6.35%	8.67%	4.23%	9.17%
Rollerball Pen-Plain Paper-Galaxy Note 3	3.70%	3.34%	5.56%	6.67%	1.59%	3.34%
Ballpoint Pen-Thin Paper-Galaxy Note 3	5.95%	3.34%	11.11%	6.67%	4.23%	3.34%
Rollerball Pen-Thin Paper-Galaxy Note 3	3.97%	3.34%	7.14%	3.33%	2.12%	1.67%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	1.59%	10.00%	3.97%	20.00%	0.00%	6.67%
Rollerball Pen-Plain Paper-iPhone 7 Plus	1.59%	6.67%	7.94%	9.33%	2.65%	8.34%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	6.35%	3.34%	7.94%	10.00%	4.23%	0.00%
Rollerball Pen-Thin Paper-iPhone 7 Plus	4.76%	10.00%	4.76%	6.67%	2.86%	10.00%

The Equal Error Rate values of the verification outcomes resulting from the audio data are calculated separately for gender and age. EER values for verification using

just audio data by gender are displayed in Table 7.7. EER values for verification using solely audio data by age ranges are listed in Table 7.8.

Figure 7.3 displays the ROC curves for just sound-based signature verification according to the Spectral Flux Onset Envelope-based verification values, the Spectral Centroid-based verification values, and the combination of the Spectral Flux Onset Envelope and Spectral Centroid-based verification values of the sound of signature data.

The verification's success is enhanced by combining the two features (SFOSE and SC) acquired from Sound. This situation is verified using the t-test statistical significance test. P value, as a result, is 0.0004137. This number is less than 0.05, allowing us to accept the premise that the combination of these audio-based features improves the success of verification.

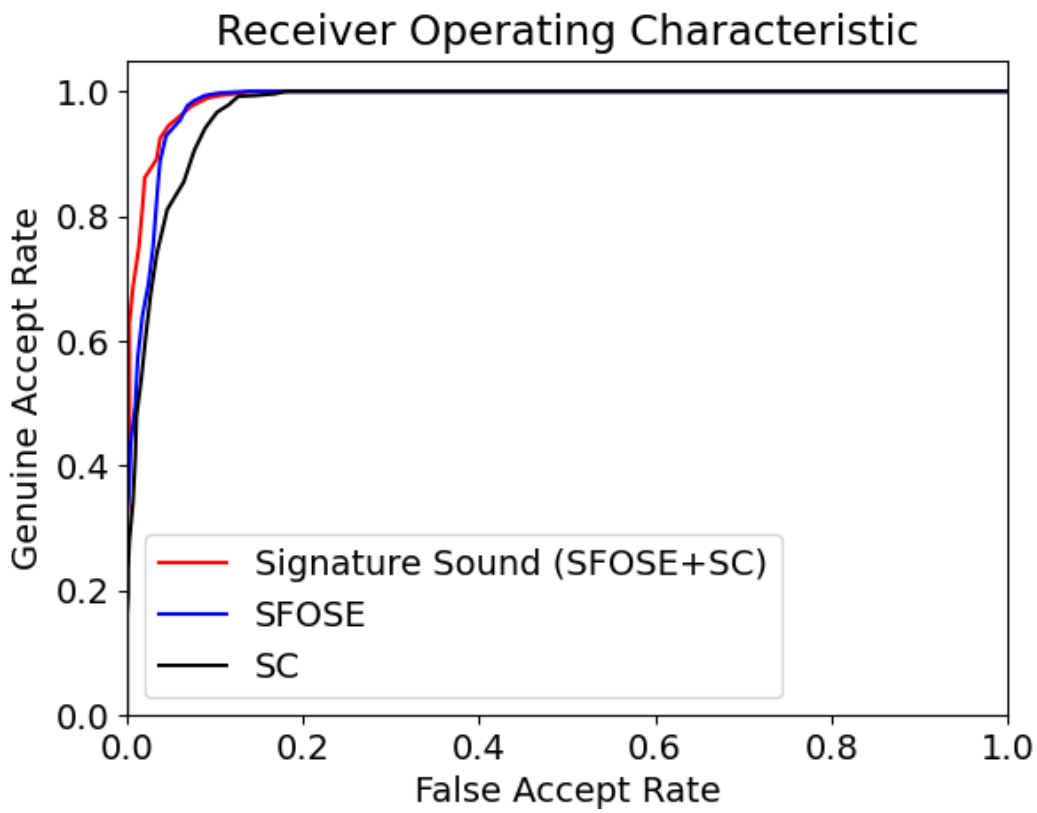


Figure 7.3 ROC curves for spectral flux onset envelope of audio data, spectral centroid of audio data and fusion of spectral flux onset envelope and spectral centroid of audio data

Many studies use audio data for signature verification (See Section 1.1.1). Table 7.9 compares these studies and the deep learning-based approach purely based on audio data.

Table 7.9 Comparison of the proposed approach on sound data only with some of the signature sound verification studies on their own dataset

Study	#Participants	#Samples	Feature Extraction	Classification	Results
Li (2004, 2010) [6] [5]	5	10 genuine, 10 forged	Normalized Hilbert envelope of sounds	multi-layer back propagation neural network	≥ 75% correctness for different scenarios
Khazei et al. (2012) [10]	30	10 genuine	AR coefficients, cepstrum based features	Euclidean, Manhattan, Chessboard	49%≤ EER≤ 50.133%
Armiato et al. (2016) [11]	55	10 genuine, 2 forged	Combination of wavelet-based features	Euclidean classifier, Modified correlation classifier	≥ 80% accuracy
Ding et al. (2019) [13]	14	112 genuine, 60 forged	A chord-based method, to estimate phase-related changes caused by small activities.	Deep CNN	EER: 5.5% AUC: 98.7%
Chen et al. (2020) [14]	35	20 genuine, 20 forged	SSIM, PSNR, MSE, and Hausdorff distance.	LR,NB, RF, SVM	EER: 1.25% AUC: 98.2%
Wei et al. (2021) [15]	12	70 genuine, 60 forged	Zero Crossing Rate, Spectral Centroid, Spectral Spread, Sprectral Flux, Spectral Entropy, Spectral Rolloff	One-Class classifier based on CNN	EER: 5% AUC: 98.4%
Zhao et al. (2021) [16]	40	32 genuine, 28 forged	Spatio-Temporal Features from Channel Impulse Response (CIR)	CNN-based Multi-Modal Siamese Network	EER: 3.27%
Proposed (Ballpoint Pen-Plain Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 5.59%**
Proposed (Rollerball Pen-Plain Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER:2.15%**
Proposed (Ballpoint Pen-Thin Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 3.23%**
Proposed (Rollerball Pen-Thin Paper-Galaxy Note 3)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 1.94%**
Proposed (Ballpoint Pen-Plain Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 2.15%**

Continued on next page

Table 7.9 – continued from previous page

Study	#Participants	#Samples	Feature Extraction	Classification	Results
Proposed (Rollerball Pen-Plain Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 4.57%**
Proposed (Ballpoint Pen-Thin Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 2.96%**
Proposed (Rollerball Pen-Thin Paper-iPhone 7 Plus)	**93**	**4 genuine**	**SC+SFOSE, LBP, SIFT**	**OC-SVM**	**EER: 5.22%**

7.1.3 Results of the Tests Using Fusion of Offline (Static) Signature Data and Signature Sound (Dynamic) Data

The proposed deep learning-based approach intends to improve classification accuracy by incorporating the signature sound and the signature image. The combined feature vectors (SigNet) from the generated Spectral Flux Onset Envelope and Spectral Centroid image graphics from each audio data are also combined with the corresponding offline (static) signature image feature vectors (SigNet) in order to achieve improvement in the verification success rates of results (See Figure 7.1). Table 7.10 lists the EER values for each gender and age group, as determined through testing to verify individuals, using a fusion of offline (Static) signature data and signature sound (Dynamic) data.

Table 7.10 Results of verification for the fusion of offline (static) signature data with dynamic signature sound data, segmented by participant age (<30, ≥30) and gender (Male (M), Female (F)).

Combination (SKU dataset)	Gender		Age		EER (Gender)		EER (Age)	
	#M	#F	#(<30)	#(≥30)	M	F	<30	≥30
Ballpoint Pen-Plain Paper-Galaxy Note 3	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Rollerball Pen-Plain Paper-Galaxy Note 3	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Ballpoint Pen-Thin Paper-Galaxy Note 3	55	38	63	30	1.82%	0.00%	1.59%	0.00%
Rollerball Pen-Thin Paper-Galaxy Note 3	55	38	63	30	0.00%	2.63%	1.59%	0.00%
Ballpoint Pen-Plain Paper-iPhone 7 Plus	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Rollerball Pen-Plain Paper-iPhone 7 Plus	55	38	63	30	0.00%	5.26%	3.17%	0.00%
Ballpoint Pen-Thin Paper-iPhone 7 Plus	55	38	63	30	0.00%	0.00%	0.00%	0.00%
Rollerball Pen-Thin Paper-iPhone 7 Plus	55	38	63	30	0.00%	2.63%	1.59%	0.00%

Table 7.11 summarizes the verification results calculated using offline (static) signature image data only, signature sound data (dynamic) only, and a fusion of both types of data. The p-value for the statistical significance test for the results in Table 7.11 is 0.0006485. This result indicates that the outcomes shown in the table are

reliable.

Table 7.11 Results of verification for the fusion of static (offline) signature image data and signature sound data (dynamic) using various types of used pens, papers, and mobile phone models

Combination (SKU dataset)	Sig. Sound	Sig. Image	Fusion of signature sound and signature image
B.Point Pen-Plain Paper-Galaxy Note 3	EER: 5.59%	EER: 3.23%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
R.Ball Pen-Plain Paper-Galaxy Note 3	EER :2.15%	EER: 6.88%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
B.Point Pen-Thin Paper-Galaxy Note 3	EER: 3.23%	EER: 3.23%	GAR: 97.85% FAR: 1.08% FRR: 2.15% **EER: 1.08%**
R.Ball Pen-Thin Paper-Galaxy Note 3	EER: 1.94%	EER: 6.09%	GAR: 98.92% FAR: 1.08% FRR: 1.08% **EER: 1.08%**
B.Point Pen-Plain Paper-iPhone 7 Plus	EER: 2.15%	EER: 3.23%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
R.Ball Pen-Plain Paper-iPhone 7 Plus	EER: 4.57%	EER: 6.88%	GAR: 97.85% FAR: 2.15% FRR: 2.15% **EER: 2.15%**
B.Point Pen-Thin Paper-iPhone 7 Plus	EER: 2.96%	EER: 3.23%	GAR: 100.00% FAR: 0.00% FRR: 0.00% **EER: 0.00%**
R.Ball Pen-Thin Paper-iPhone 7 Plus	EER: 5.22%	EER: 6.09%	GAR: 98.92% FAR: 1.08% FRR: 1.08% **EER: 1.08%**

The ROC curves for verification findings derived using offline (static) signature data, signature sound (dynamic) data, and fusion of offline (static) signature data with signature sound (dynamic) data are shown in Figure 7.4. (Curves correspond to the averaged error rates obtained from pen-paper-phone combinations given in Tables 7.2, 7.6, and 7.11.)

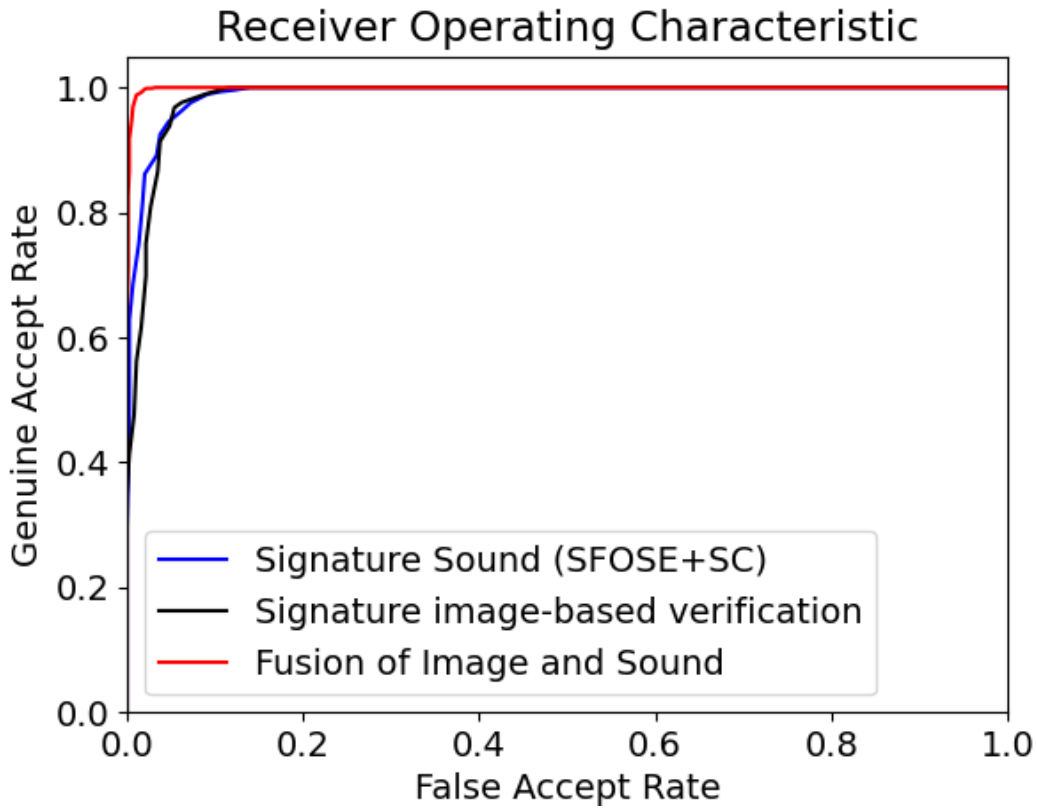


Figure 7.4 ROC Curves for signature sound (dynamic) data only, offline (static) signature data only, and fusion of signature sound (dynamic) data with offline (static) signature data

Tables 7.12-7.13 compare the shallow (non-deep) learning-based approach versus the deep learning-based approach in terms of verification results.

Table 7.12 Verification results according to shallow learning-based approach and deep learning-based approach based on image data only, sound data only and fusion of image and sound data

Approach	Image-Based EER	Sound-Based EER	Image&Sound Fusion-Based EER
Shallow Learning	5.02-11.47% (Average: 9.50%)	1.08-5.38% (Average: 3.02%)	0.00-1.08% (Average: 0.29%)
Deep Learinng	3.23-6.88% (Average: 4.86%)	1.94-5.59% (Average: 3.48%)	0.00-2.15% (Average: 0.67%)

Table 7.13 Comparison of deep learning-based approach and shallow learning-based approach in terms of verification results.

Combination (SKU dataset)	Shallow learning based approach (Chapter 6)			Deep learning based approach (Chapter 7)		
	Sound	Image	Sound+Image Fusion	Sound	Image	Sound+Image Fusion
Ballpoint Pen-Plain Paper-Galaxy Note 3	EER: 5.38%	EER: 11.47%	**EER: 0.00%**	EER: 5.59%	EER: 3.23%	**EER: 0.00%**
Rollerball Pen-Plain Paper-Galaxy Note 3	EER: 1.08%	EER: 11.29%	**EER: 0.00%**	EER: 2.15%	EER: 6.88%	**EER: 0.00%**
Ballpoint Pen-Thin Paper-Galaxy Note 3	EER: 4.30%	EER: 5.02%	**EER: 0.00%**	EER: 3.23%	EER: 3.23%	**EER: 1.08%**
Rollerball Pen-Thin Paper-Galaxy Note 3	EER: 2.15%	EER: 10.22%	**EER: 1.08%**	EER: 1.94%	EER: 6.09%	**EER: 1.08%**
Ballpoint Pen-Plain Paper-iPhone 7 Plus	EER: 4.52%	EER: 11.47%	**EER: 1.08%**	EER: 2.15%	EER: 3.23%	**EER: 0.00%**
Rollerball Pen-Plain Paper-iPhone 7 Plus	EER: 1.79%	EER: 11.29%	**EER: 0.08%**	EER: 4.57%	EER: 6.88%	**EER: 2.15%**
Ballpoint Pen-Thin Paper-iPhone 7 Plus	EER: 3.87%	EER: 5.02%	**EER: 0.09%**	EER: 2.96%	EER: 3.23%	**EER: 0.00%**
Rollerball Pen-Thin Paper-iPhone 7 Plus	EER: 1.08%	EER: 10.22%	**EER: 0.00%**	EER: 5.22%	EER: 6.09%	**EER: 1.08%**

8
RESULTS AND DISCUSSION

In this thesis, a different biometric data—the sound produced when a pen tip rubs against paper while signing—is taken into account, and the improvement in verification performance that results from combining this data with the matching offline (static) signature image data is studied.

The internal microphones of two distinct mobile phone models are used to record the sounds of the signatures throughout the data-collecting phase, which involved samples obtained from 93 people. Participants' signatures are collected using two types of pens and papers, including ballpoint, rollerball, plain paper, and thin paper with an auto-copy feature. For each paper and pen combination, each participant is required to sign their authentic signatures at least four times. The identical participant practiced imitating another participant's signature before being asked to provide four skilled forgery samples for each pen-paper combination. Even though some people's signatures are short and silent, these samples are included in the dataset, meaning that no validation phase took into account the sound loudness, signature duration, et cetera while creating the dataset. By not rejecting any participant or signature sample, the proposed approach seeks to increase the usage area and produce a realistic result.

Two biometric verification approaches are proposed in this study. In the first approach, LBP and SIFT features are extracted from the images generated by converting the Spectral Flux Onset Envelope and Spectral Centroid graphics obtained from the audio data (dynamic). Also, the offline (static) signature images are processed for the extraction of the LBP and SIFT features. Both data types (static and dynamic) underwent feature extraction in a writer-independent fashion. Writer-dependent OC-SVM classifier is trained using the LBP and SIFT features independently with four genuine signatures from each user, with leave-one-out cross-validation. Total verification success is determined separately for signature sound and signature image by averaging the two (LBP and SIFT) classification scores. The classification success is also calculated for the verification based on the fusion of the signature image and signature sound, which is done by combining the feature vectors acquired from

the image and audio data. In the other proposed approach (deep learning-based), deep features extracted from the images using a CNN-based model (SigNet) are utilized rather than LBP and SIFT features. This approach is also subjected to feature extraction in a writer-independent manner.

Table 6.1 and Table 7.2 demonstrate that the outcomes of using the proposed approaches on offline (static) signature images from a dataset produced from scratch (SKU dataset) are comparable to those of using the proposed approaches on the GPDS and MCYT public datasets. These results indicate the coherency of the compiled data set. It can be said that the proposed approach using only offline (static) signature image data is quite comparable to the offline signature verification studies at the state-of-the-art level, given that this study used only four genuine signatures in the training of the classifier. The results obtained are similar to results from studies in Tables 1.2-6.2-6.3-7.3-7.4, and publication [52]. A genuine signature sample and its forged version are depicted in Figure 8.1. Figure 8.2 shows the spectral flux onset strength envelope and spectral centroid images of the genuine signature sounds and the forged signature sounds of the identical signatures. It can be concluded from Figure 8.2 and the whole dataset assembled for this study that graphical images of genuine and forged signature sounds can be identified rather clearly, even by the unaided eye. Due to this, verification attempts that use solely sound signals have a higher success rate as the Tables 6.10-7.11 show. The fact that verification is done using random or simple forgeries in terms of signature sound is one of the factors contributing to the high success rate of sound-based signature verification. Namely, the imitator had never heard the sound of a signature before and had never attempted to imitate it (even if the imitator does hear the sound of a signature, it might be challenging for him or her to recall and replicate the sound he or she hears). Tables 6.1-7.2-7.12 clearly show that when the proposed approach is applied solely to offline (static) signature images from the dataset built, the equal error rates range between 5.02% and 11.47% (Average: 9.50%) for the shallow learning-based approach (Chapter 6) and 3.23% and 6.88% (Average: 4.86%) for the deep learning-based approach (Chapter 7). When verification is done using solely signature sound (dynamic) data, the equal error rate varies between 1.08% and 5.38% (Average: 3.02%) for the shallow learning-based approach and 1.94% and 5.59% (Average: 3.48%) for the deep learning-based approach, according to Table 6.5 and Table 7.6. Table 7.13 shows that the equal error rates for the shallow learning-based approach are reduced from the 5.02-11.47% range to the 0.00-1.08% range (Average: 0.29%) by fusing the offline (static) signature image with the signature sound (dynamic). The equal error rate values in the deep learning-based approach also dropped from a range of 3.23–6.88% to a range of 0.00–2.15% (Average: 0.67%).

The outcomes of the statistical significance tests supported these findings. It is clear from Chapter 7 that the deep learning-based approach produces superior outcomes over the shallow learning-based approach when verification is done using only static signature images (see Tables6.1-7.2). However, the effectiveness of both approaches in verifying the fusion of sound and image is slightly different (Table 7.13). So, the shallow learning-based method proposed in Chapter 6 is a robust substitute for the deep learning-based approach for verification with the fusion of signature image and sound because it does not have the relative drawbacks of the deep learning-based approach, such as run times and the requirement to train the model file beforehand. By analyzing the cases where the pen-paper-phone combinations of the reference and query signatures are different, it can be concluded that the item that affects the signature verification performance the most is the changes in the pen type, then the phone model, and the least in the paper type (See Section 6.1.4). Last of all, Since validation results based on female-male or age groups vary according to pen-paper-phone combinations, it does not seem possible to make an inference about whether verification based on any group is more successful or unsuccessful than the other group.



Figure 8.1 Comparison of genuine signature image with skilled forgery image: a) Genuine signature image b) Forged signature image

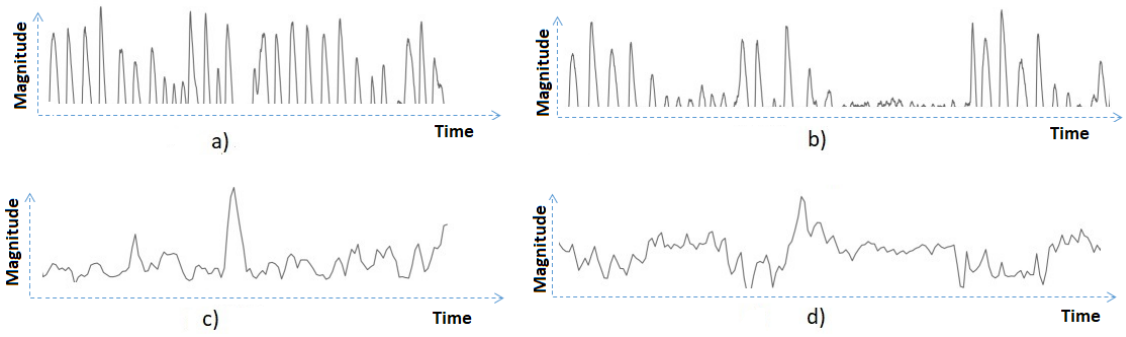


Figure 8.2 Comparison of graphic images obtained from signature sound data a) Image of Onset Strength Envelope of genuine signature sound b) Image of Onset Strength Envelope of forged signature sound c) Image of Spectral Centroid of genuine signature sound d) Image of Spectral Centroid of forged signature sound

8.1 Future Work

According to the findings of this study, it can be concluded that the sound generated from a handwritten signature alone has a biometric value that merits further research. Future research may include more individuals while taking into account various age groups, surroundings, etc. It would make sense to increase audio-based verification accuracy by collecting audio recordings from a sufficient number (for example, 500 or more) of individuals and utilizing that data to build a model in the deep learning context. Higher background noise conditions may be used to gather data, and phones' built-in microphones can be placed farther from the signing position. In order to remove background noise from audio recordings, filtering algorithms or external shotgun microphones might be utilized. It would be helpful to develop mobile applications that begin recording audio as soon as signing begins and stop recording audio as soon as it is complete automatically. Videos of signatures may be recorded with their sounds included, and imitators can watch these videos repeatedly to copy their sound components. Thus, it will be possible to build a dataset that contains a different version of skilled forgeries for the audio data. By taking into account relevant data privacy policies and laws, the acquired datasets—at least audio files—can be made available to the public.

REFERENCES

[1] A. K. Jain, P. Flynn, A. A. Ross, *Handbook of Biometrics*. Springer Science & Business Media, 2007.

[2] D. Impedovo, G. Pirlo, "Automatic signature verification: The state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 5, pp. 609–635, 2008.

[3] D. Bertolini, L. S. Oliveira, E. Justino, R. Sabourin, "Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers," *Pattern Recognition*, vol. 43, no. 1, pp. 387–396, 2010.

[4] A. A. Ross, K. Nandakumar, A. K. Jain, *Handbook of Multibiometrics*. Springer Science & Business Media, 2006, vol. 6.

[5] F. F. Li, "Sound-based multimodal person identification from signature and voice," in *2010 Fifth International Conference on Internet Monitoring and Protection*, 2010, pp. 84–88.

[6] F. F. Li, "Handwriting authentication by envelopes of sound signature," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 1, pp. 401–404.

[7] M. S. Sadak, N. Kahraman, U. Uludağ, "Dynamic and static feature fusion for increased accuracy in signature verification," *Signal Processing: Image Communication*, vol. 108, 2022.

[8] M. S. Sadak, N. Kahraman, U. Uludag, "Handwritten signature verification system using sound as a feature," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2020, pp. 365–368.

[9] A. Seniuk, D. Blostein, "Pen acoustic emissions for text and gesture recognition," in *2009 10th International Conference on Document Analysis and Recognition*, IEEE, 2009, pp. 872–876.

[10] D. Khazaei, K. Maghooli, F. Afdideh, H. Azimi, "A unimodal person authentication system based on signing sound," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2012, pp. 152–154.

[11] D. A. Armiato, Y. Yano, V. Z. de Faveri, R. C. Guido, "Handwritten signatures verification through their acoustic patterns based on the discrete wavelet-packet transform and semantic-matching classifiers," *International Journal of Semantic Computing*, vol. 10, no. 04, pp. 557–567, 2016.

[12] H. Du, P. Li, H. Zhou, W. Gong, G. Luo, P. Yang, "Wordrecorder: Accurate acoustic-based handwriting recognition using deep learning," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, IEEE, 2018, pp. 1448–1456.

[13] F. Ding, D. Wang, Q. Zhang, R. Zhao, "ASSV: handwritten signature verification using acoustic signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–22, 2019.

[14] M. Chen, J. Lin, Y. Zou, R. Ruby, K. Wu, "Silentsign: Device-free handwritten signature verification through acoustic sensing," in *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, IEEE, 2020, pp. 1–10.

[15] Z. Wei, S. Yang, Y. Xie, F. Li, B. Zhao, "SVSV: Online handwritten signature verification based on sound and vibration," *Information Sciences*, vol. 572, pp. 109–125, 2021.

[16] R. Zhao, D. Wang, Q. Zhang, X. Jin, K. Liu, "Smartphone-based handwritten signature verification using acoustic signals," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. ISS, pp. 1–26, 2021.

[17] *CEDAR*. [Online]. Available: `https://cedar.buffalo.edu/NIJ/data/` (visited on 01/26/2022).

[18] *GPDS*. [Online]. Available: `http://www.gpds.ulpgc.es/download` (visited on 01/26/2022).

[19] A. Almehmadi, "A biometric-based verification system for handwritten image-based signatures using audio to image matching," *IET Biometrics*, vol. 11, no. 2, pp. 124–140, 2022.

[20] Y. Guerbai, Y. Chibani, B. Hadjadji, "The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters," *Pattern Recognition*, vol. 48, no. 1, pp. 103–113, 2015.

[21] M. B. Yilmaz, B. Yanikoğlu, "Score level fusion of classifiers in off-line signature verification," *Information Fusion*, vol. 32, pp. 109–119, 2016.

[22] S. Pal, A. Alaei, U. Pal, M. Blumenstein, "Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, IEEE, 2016, pp. 72–77.

[23] S. Y. Ooi, A. B. J. Teoh, Y. H. Pang, B. Y. Hiew, "Image-based handwritten signature verification using hybrid methods of discrete radon transform, principal component analysis and probabilistic neural network," *Applied Soft Computing*, vol. 40, pp. 274–282, 2016.

[24] L. G. Hafemann, R. Sabourin, L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," *Pattern Recognition*, vol. 70, pp. 163–176, 2017.

[25] M. Okawa, "Kaze features via fisher vector encoding for offline signature verification," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, 2017, pp. 10–15.

[26] A. Alaei, S. Pal, U. Pal, M. Blumenstein, "An efficient signature verification method based on an interval symbolic representation and a fuzzy similarity measure," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 10, pp. 2360–2372, 2017.

[27] M. Sharif, M. A. Khan, M. Faisal, M. Yasmin, S. L. Fernandes, "A framework for offline signature verification system: Best features selection approach," *Pattern Recognition Letters*, vol. 139, pp. 50–59, 2020.

[28] R. Ghosh, "A recurrent neural network based deep learning model for offline signature verification and recognition system," *Expert Systems with Applications*, vol. 168, p. 114 249, 2021.

[29] J. Ortega-Garcia *et al.*, "MCYT baseline corpus: a bimodal biometric database," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[30] C. Freitas, F. Bortolozzi, R. Sabourin, J. Facon, "Bases de dados de cheques bancarios brasileiros," 1998.

[31] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[32] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Lladós, U. Pal, "Signet: Convolutional siamese network for writer independent offline signature verification," *arXiv preprint arXiv:1707.02131*, 2017.

[33] J. O. Pinzón-Arenas, R. Jiménez-Moreno, C. G. Pachón-Suescún, "Offline signature verification using dag-cnn," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, p. 3314, 2019.

[34] V. Ruiz, I. Linares, A. Sanchez, J. F. Velez, "Off-line handwritten signature verification using compositional synthetic generation of signatures and siamese neural networks," *Neurocomputing*, vol. 374, pp. 30–41, 2020.

[35] F. E. Batool *et al.*, "Offline signature verification system: A novel technique of fusion of glcm and geometric features using svm," *Multimedia Tools and Applications*, pp. 1–20, 2020.

[36] H. Rantzsch, H. Yang, C. Meinel, "Signature embedding: Writer independent offline signature verification with deep metric learning," in *International symposium on visual computing*, Springer, 2016, pp. 616–625.

[37] M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, B. Found, "ICDAR 2013 competitions on signature verification and writer identification for on- and offline skilled forgeries (SigWiComp 2013)," in *2013 12th International Conference on Document Analysis and Recognition*, 2013, pp. 1477–1483.

[38] M. Liwicki *et al.*, "Signature verification competition for online and offline skilled forgeries (sigcomp2011)," in *2011 International Conference on Document Analysis and Recognition*, IEEE, 2011, pp. 1480–1484.

[39] F. Chollet, J. Allaire, *Deep Learning with R*. Manning Publications, 2018, ISBN: 9781617295546. [Online]. Available: `https://books.google.com.tr/books?id=xnIRtAEACAAJ`.

[40] *Bic-cristal*. [Online]. Available: `https://en.wikipedia.org/wiki/Ballpoint_pen#Guinness_World_Records` (visited on 01/26/2022).

[41] *Bestsellingphones*. [Online]. Available: https://en.wikipedia.org/wiki/List_of_best-selling_mobile_phones (visited on 01/26/2022).

[42] *Audacity(R): Free audio editor and recorder [computer application]. version 2.3.1*. [Online]. Available: https://audacityteam.org/ (visited on 01/26/2022).

[43] A. Buades, B. Coll, J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.

[44] S. Dixon, "Onset detection revisited," in *Proceedings of the 9th International Conference on Digital Audio Effects*, Citeseer, vol. 120, 2006, pp. 133–137.

[45] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[46] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[47] G. E. Hinton, S. Osindero, Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[48] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[49] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[50] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[51] C. Cortes, V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[52] M. M. Hameed, R. Ahmad, M. L. M. Kiah, G. Murtaza, "Machine learning-based offline signature verification systems: A systematic review," *Signal Processing: Image Communication*, p. 116 139, 2021.

[53] S. Masoudnia, O. Mersa, B. N. Araabi, A.-H. Vahabie, M. A. Sadeghi, M. N. Ahmadabadi, "Multi-representational learning for offline signature verification using multi-loss snapshot ensemble of CNNs," *Expert Systems with Applications*, vol. 133, pp. 317–330, 2019.

[54] E. N. Zois, L. Alewijnse, G. Economou, "Offline signature verification and quality characterization using poset-oriented grid features," *Pattern Recognition*, vol. 54, pp. 162–177, 2016.

[55] P. Maergner *et al.*, "Combining graph edit distance and triplet networks for offline signature verification," *Pattern Recognition Letters*, vol. 125, pp. 527–533, 2019.

[56] L. G. Hafemann, R. Sabourin, L. S. Oliveira, "Writer-independent feature learning for offline signature verification using deep convolutional neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 2576–2583.

[57] Z.-J. Xing, F. Yin, Y.-C. Wu, C.-L. Liu, "Offline signature verification using convolution siamese network," in *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*, International Society for Optics and Photonics, vol. 10615, 2018, p. 106151I.

[58] P. N. Narwade, R. R. Sawant, S. V. Bonde, "Offline handwritten signature verification using cylindrical shape context," *3D Research*, vol. 9, no. 4, pp. 1–12, 2018.

[59] S. Böck, G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx)*., vol. 7, 2013.

[60] A. Klapuri, M. Davy, *Signal Processing Methods for Music Transcription*. Springer Science & Business Media, 2007.

[61] J. P. Shaver, "What statistical significance testing is, and what it is not," *The Journal of Experimental Education*, vol. 61, no. 4, pp. 293–316, 1993.

[62] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[63] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

PUBLICATIONS FROM THE THESIS

Papers

1. SADAK, Mustafa Semih; KAHRAMAN, Nihan; ULUDAG, Umut. Dynamic and static feature fusion for increased accuracy in signature verification. Signal Processing: Image Communication, vol. 108, 116823, October 2022.

Conference Papers

1. SADAK, Mustafa Semih; KAHRAMAN, Nihan; ULUDAG, Umut. Handwritten signature verification system using sound as a feature. In: 2020 43rd International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2020. p. 365-368.