T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİYOMEDİKAL GÖRÜNTÜLERDEKİ HAREKETLİ NESNELERİN DERİN ÖĞRENME YÖNTEMLERİ İLE ŞEKİL VE YÖRÜNGE ANALİZİ

Abdulsamet AKTAŞ

DOKTORA TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Doç. Dr. Hamza Osman İLHAN

Eş-Danışman
Dr. Öğretim Üyesi Görkem SERBES

Şubat, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİYOMEDİKAL GÖRÜNTÜLERDEKİ HAREKETLİ NESNELERİN DERİN ÖĞRENME YÖNTEMLERİ İLE ŞEKİL VE YÖRÜNGE ANALİZİ

Abdulsamet AKTAŞ tarafından hazırlanan tez çalışması 27.02.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Hamza Osman İLHAN
Yıldız Teknik Üniversitesi
Danışman

Dr. Öğretim Üyesi Görkem SERBES
Yıldız Teknik Üniversitesi
Eş-Danışman

Jüri Üyeleri

Doç. Dr. Hamza Osman İLHAN, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Ali BULDU, Üye
Marmara Üniversitesi _____

Prof. Dr. Nizamettin AYDIN, Üye
İstanbul Teknik Üniversitesi _____

Doç. Dr. Ali Can KARACA, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Erkan USLU, Üye
Yıldız Teknik Üniversitesi _____

Danışmanım Doç. Dr. Hamza Osman İLHAN sorumluluğunda tarafımca hazırlanan Biyomedikal Görüntülerdeki Hareketli Nesnelerin Derin Öğrenme Yöntemleri İle Şekil Ve Yörünge Analizi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Abdulsamet AKTAŞ

İmza

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) ARDEB 1001 - 122E164 numaralı proje ile Türkiye Sağlık Enstitüleri Başkanlığı (TÜSEB) B Grubu AR-GE Projesi - 27698 numaralı proje kapsamında desteklenmiştir.

Canım Anneme ithafen,

TEŞEKKÜR

Tezimin hazırlanmasında bana bilimsel desteğini ve değerli yönlendirmelerini esirgemeyen saygıdeğer danışmanım Doç. Dr. Hamza Osman İLHAN'a sonsuz teşekkürlerimi sunuyorum. Karşılaştığım her zorlukta gösterdiği derin bilgi birikimiyle bana rehberlik etmiş, vizyon ve akademik yaklaşımıyla ufkumu genişletmiş, bu çalışmayı başarıyla tamamlamama büyük katkı sağlamıştır.

Ayrıca, tezimin her aşamasında değerli katkılarını esirgemeyen eş danışmanım Dr. Öğretim Üyesi Görkem SERBES'e şükranlarımı sunuyorum. Sağladığı yapıcı eleştiriler ve bilimsel bilgi birikimi ile araştırmamın metodolojik derinliğini artırmış, sürecin verimli bir şekilde ilerlemesine önemli katkılar sağlamıştır. Çözüm odaklı yaklaşımı ve sürekli desteği, bu çalışmanın başarılı bir şekilde tamamlanmasını mümkün kılmıştır. Hocalarım ile çalışma fırsatına sahip olduğum için büyük bir ayrıcalık duyuyor ve kendimi şanslı hissediyorum.

Tez çalışmamın tamamlanmasında katkı sağlayan değerli tez komite üyeleri Prof. Dr. Ali BULDU ve Doç. Dr. Ali Can KARACA'ya en içten teşekkürlerimi sunuyorum. Zamanlarını ayırıp değerli görüşlerini paylaşarak bu çalışmanın derinliğini artırdılar. Ayrıca, süreç boyunca sundukları rehberlik ve teşvik dolayısıyla minnettarım.

Tez çalışmalarım süresince sağlamış oldukları desteklerden ötürü TÜBİTAK ve TÜSEB'e içten teşekkürlerimi sunuyorum.

Bu zorlu süreç boyunca her zaman yanımda olan, desteğini her an hissedebildiğim değerli çalışma arkadaşım Arş. Gör. Merve PINAR'a en içten teşekkürlerimi sunuyorum.

Son olarak, tez çalışmam boyunca ve hayatımın her anında bana verdikleri sarsılmaz destek, duydukları inanç ve her zaman yanımda hissettiğim sevgileriyle bana güç veren anneme, babama ve kardeşlerime sonsuz teşekkürlerimi sunarım. İyi ki varsınız.

Abdulsamet AKTAŞ

İÇİNDEKİLER

SİMGE LİSTESİ viii

KISALTMA LİSTESİ ix

ŞEKİL LİSTESİ xiii

TABLO LİSTESİ xvi

ÖZET xviii

ABSTRACT xx

1 GİRİŞ 1

1.1 Literatür Taraması 1

1.1.1 Açık Kaynak Sperm Analiz Veri Setleri 2

1.1.2 Sperm Morfoloji Analiz Yöntemlerine Genel Bakış 10

1.1.3 Sperm Tespit ve Takip Yöntemlerine Genel Bakış 14

1.2 Tezin Amacı 19

1.3 Hipotez 20

2 MATERYALLER 22

2.1 Sperm Morfoloji Analizi Veri Setleri 22

2.1.1 Açık Erişim Veri Setleri 23

2.1.2 Özgün Veri Seti: Hi-LabSpermMorpho 25

2.2 Sperm Yörünge Analizi Veri Setleri 33

2.2.1 Özgün Veri Seti: Hi-LabSpermTracking 34

3 DERİN ÖĞRENME YÖNTEMLERİ İLE MORFOLOJİ ANALİZİ 37

3.1 Açık Erişim Sperm Morfoloji Veri Setlerinin Sınıflandırılması . . . 38

3.1.1 CNN ve ViT Mimarilerinin Karşılaştırmalı Analizi 38

3.1.2 Sperm Morfoloji Analizinde Vision Transformer Mimarilerinin Sınıflandırma Analizi 46

3.2 Hi-LabSpermMorpho Verisetinin Sınıflandırılması 68

3.2.1 Materyal ve Yöntem 69

3.2.2 Deneysel Çalışmalar 72
3.2.3 Tartışma ve Sonuç 82
3.2.4 Gelecek Çalışmalar 90

4 DERİN ÖĞRENME YÖNTEMLERİ İLE SPERM YÖRÜNGE ANALİZİ 92

4.1 Hi-LabSpermTracking Veri Seti ile Sperm Yörünge Analizi 92
4.1.1 Materyal ve Yöntem 94
4.1.2 Deneysel Sonuçlar 109
4.1.3 Tartışma 124
4.1.4 Sonuç ve Gelecek Çalışmalar 130

5 SONUÇ 132

KAYNAKÇA 136

TEZDEN ÜRETİLMİŞ YAYINLAR 147

SİMGE LİSTESİ

$c \in TP$	Doğru Eşleşen Her Bir Nesne
α	Intersection over Union Katsayısı
c	Nesne

KISALTMA LİSTESİ

ABC	Yapay Arı Kolonisi (Artificial Bee Colony)
Acc	Doğruluk (Accuracy)
ANOVA	Varyans Analizi (Analysis of Variance)
AssA	İlişkilendirme Doğruluğu (Association Accuracy)
B	Beraberlik
BCF	Beat Cross Frequency
BDSA	Bilgisayar Destekli Sperm Analizi
BeiT	BERT Pre-Training of Image Transformers
BFA	Bacteria Foraging Algorithm
CASA	Bilgisayar Destekli Sperm Analizi (Computer-Assisted Sperm Analysis)
CLAHE	Kontrast Sınırlı Adaptif Histogram Eşitleme (Contrast Limited Adaptive Histogram Equalization)
CNN	Evrişimsel Sinir Ağı (Convolutional Neural Network)
CPU	Merkezi İşlem Birimi (Central Processing Unit)
CSO	Kedi Sürüsü Optimizasyonu (Cat Swarm Optimization)
DA	Yusufçuk Algoritması (Dragonfly Algorithm)
DeepSORT	Derin Öğrenme Tabanlı Basit Çevrimiçi ve Gerçek Zamanlı Takip (Deep Learning-based Simple Online and Realtime Tracking)
DeiT	Veri Verimli Görüntü Transformatörü (Data-efficient image Transformer)
Det-IoU	Tespitler Arası Kesişim Birliktelik Oranı (Detection Intersection Over Union)
DSÖ	Dünya Sağlık Örgütü (World Health Organization)
DT	Karar Ağaçları (Decision Tree)

DTCWT	Çift Ağaçlı Karmaşık Dalgacık Dönüşümü (Dual-Tree Complex Wavelet Transform)
DWT	Ayrık Dalgacık Dönüşümü (Discrete Wavelet Transform)
F1	F1-Skor (F1-Score)
FN	Yanlış Negatif (False Negative)
FNA	Yanlış Negatif İlişkiler (False Negative Associations)
FP	Yanlış Pozitif (False Positive)
FPA	Yanlış Pozitif İlişkiler (False Positive Associations)
FPN	Özellik Piramidi Ağı (Feature Pyramid Network)
fps	Saniye Başına Kare (frame per second)
GA	Genetik Algoritma
GAN	Çekişmeli Üretici Ağlar (Generative Adversarial Networks)
GD	Görsel Değerlendirme
GPT	Üretici Önceden Eğitilmiş Dönüştürücü (Generative Pre-trained Transformer)
GPU	Grafik İşlem Birimi (Graphics Process Unit)
GT	Gerçek Referans Değer (Ground Truth)
GTDets	Gerçek Referans Tespitler (Ground Truth Detections)
GTIDs	Gerçek Referans Kimlikler (Ground Truth Identities)
HOTA	Yüksek Derece Takip Doğruluğu (High Order Tracking Accuracy)
HSD	Dürüst Önemli Fark (Honest Significant Difference)
HuSHeM	Human Sperm Head Morphology
ID	Kimlik (Identity)
IDP	Kimlik Kesinliği (Identity Precision)
IDR	Kimlik Duyarlılığı (Identity Recall)
IoU	Kesişim Üzerine Birleşim (Intersection over Union)
K	Kazanma
KAZE	Kecskés-Azencott-Zhang-École
KNN	k En Yakın Komşu (k-Nearest Neighbors)

KYP	Kayıp
LIN	Doğrusallık (Linearity)
Loc-IoU	Konumsal Kesişim Üzerine Birleşim (Localization Intersection over Union)
Lr	Öğrenme Oranı (Learning Rate)
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
mAP	Ortalama Hassasiyet (mean Average Precision)
MHSA	Çoklu Başlıklı Kendine Dikkat (Multi-head Self Attention)
MHSMA	Modified Human Sperm Morphology Analysis Dataset
MIM	Maskelenmiş Görüntü Modellemesi (Masked Image Modeling)
MLM	Maskelenmiş Dil Modellemesi (Masked Language Modeling)
MLP	Çok Katmanlı Algılayıcı (Multilayer Perceptron)
MOT	Çoklu Nesne İzleme (Multiple Object Tracking)
MOTA	Çoklu Nesne İzleme Doğruluğu (Multiple Object Tracking Accuracy)
MSER	Maksimum Stabil Ekstremal Bölgeler (Maximally Stable Extremal Regions)
NB	Naif Bayes (Naive Bayes)
NLP	Doğal Dil İşleme (Natural Language Processing)
Opt	Optimizasyon
ORG	Orijinal (Original)
P	Kesinlik (Precision)
PANet	Yol Toplama Ağı (Path Aggregation Network)
PSO	Parçacık Sürü Optimizasyonu (Particle Swarm Optimization)
R	Duyarlılık (Recall)
RGB	Red Green Blue
RLVQ	Öznellik Öğrenme Vektör Kuantizasyonu (Relevance Learning Vector Quantization)
ROI	İlgi Alanı (Region of Interest)

RT-DETR	Gerçek Zamanlı Tespit Transformer (Real Time Detection Transformer)
SA	Uzamsal Dikkat (Spatial Attention)
SeSViD	Semen Specimen Video Dataset
SGD	Stokastik Gradyan İnişi (Stochastics Gradient Descent)
SMA	Sperm Morfoloji Analizi
SMIDS	Sperm Morphology Image Data Set
SOTA	Son Teknoloji (State-of-the-art)
STR	Düzlük (Straightness)
SURF	Hızlandırılmış Gürbüz Özellikler (Speeded-Up Robust Features)
SVIA	Sperm Videos and Images Analysis
SVM	Destek Vektör Makineleri (Support Vector Machine)
TN	Doğru Negatif (True Negative)
TP	Doğru Pozitif (True Positive)
TPA	Doğru Pozitif İlişkiler (True Positive Associations)
VA	Görsel Değerlendirme (Visual Assessment)
VAE	Varyasyonel Otomatik Kodlayıcı (Variational Auto Encoder)
VAP	Ortalama Yol Hızı (Average Path Velocity)
VCL	Eğrisel Hız (Curvilinear Velocity)
ViT	Görü Dönüştürücü (Vision Transformer)
VOAT	Görü Dışarıdan Bakıcı (Vision Out-looker)
VSL	Düz Çizgi Hızı (Straight Line Velocity)
WHO	Dünya Sağlık Örgütü (World Health Organization)
YOLO	Sadece Bir Kez Bak (You Only Look Once)
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 1.1 İnsan sperminde çeşitli anormal ve normal morfolojiler (a)konik, (b) armut, (c) yuvarlak, (d) şekilsiz, (e) vakuollü, (f) küçük akrozomal alan, (g) bükük boyun, (e) asimetrik boyun, (i) kalın bağlantı noktası, (j) ince boyun, (k) kısa kuyruk, (l) bükük kuyruk, (m) kıvrılmış kuyruk, (n) aşırı artık sitoplazma [2] 3

Şekil 2.1 SMIDS, HuSHeM ve SCIAN-Morpho veri setleri için örnek görüntüler 25

Şekil 2.2 (a) Özel tasarım aparat ve (b) mikroskop okülerine entegre edilmiş aparata takılı cep telefonu 27

Şekil 2.3 DB1 veri setinden DB2 veri setini oluşturma adımları 28

Şekil 2.4 BesLab boyama için elde edilen DB3_v1 (18 sınıf) veri kümesinden etiketlenmiş sınıflara ait örnek görüntüler 29

Şekil 2.5 BesLab boyama için elde edilen DB3_v2 (4 sınıf) veri kümesinden etiketlenmiş sınıflara ait örnek görüntüler 30

Şekil 2.6 Boyama kitlerine (BesLab, Histoplus ve GBL) dayalı dağılımlar (a) çalışmaya dahil edilen hasta sayısı ve (b) oküler görüntü sayısı. 31

Şekil 2.7 DB3_v1 veri setinde boyama materyali tabanlı sınıflar içindeki bireysel sperm yamalarının dağılımı: (a) BesLab, (b) Histoplus, (c) GBL 32

Şekil 2.8 DB3_v2 veri setlerindeki sperm yama dağılımları: (a)BesLab, (b) Histoplus, (c) GBL 33

Şekil 2.9 DB3'te (a) BesLab, (b) Histoplus ve (c) GBL boyalarının örnek görüntüleri 33

Şekil 2.10 Hi-LabSpermTracking veri setinin oluşturulma sürecine ilişkin akış diyagramı 36

Şekil 2.11 Hi-LabSpermTracking veri setinde bulunan etiketlenmiş örnek bir video karesi 36

Şekil 3.1 Uygulanan veri artırma tekniklerinin her biri farklı bir veri setinden alınan örnek görüntüler üzerindeki etkileri 41

Şekil 3.2 ViT-L16 modeli HuSHeM-40x veri seti karmaşılık matrisi 44

Şekil 3.3 ViT-L16 modeli SMIDS-10x veri seti karmaşılık matrisi 45

Şekil 3.4 ViT-L16 modeli SCIAN-10x veri seti karmaşılık matrisi 45

Şekil 3.5 Önerilen Transformer tabanlı sperm morfolojisi sınıflandırma yaklaşımının akış şeması 51
Şekil 3.6 ViT akış şeması [74] 52
Şekil 3.7 DEiT akış şeması [83] 54
Şekil 3.8 BEiT akış şeması [103] 55
Şekil 3.9 Örnek karmaşıklık matrisi 57
Şekil 3.10 Transformer mimarilerinin Adamax (a,d), SGD (b,e) ve RMSProp (c,f) optimizasyon yöntemleri ve çeşitli öğrenme oranları kullanılarak orijinal (ORG) HuSHeM(a,b,c) ve orijinal (ORG) SMIDS(d,e,f) veri kümeleri üzerindeki doğrulukları 60
Şekil 3.11 a, c)Her üç optimizasyon yöntemindeki en iyi iki model için optimum hiperparametre yapılandırmaları ve doğruluk değerleri, b) ORG-HuSHeM üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi, d) ORG-SMIDS üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi 61
Şekil 3.12 a, c) Optimum hiperparametre seçim süreciyle elde edilen altı mimarinin veri artırma ölçeklerine göre sınıflandırma doğruluğu değerleri, b) 40×-HuSHeM üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi, d) 10×-SMIDS üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi 63
Şekil 3.13 DB3_v1_BesLab üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu 76
Şekil 3.14 DB3_v1_Histoplus üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu 77
Şekil 3.15 DB3_v1_GBL üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu 77
Şekil 3.16 Üç farklı boyama yöntemiyle elde edilen DB3_v1 veri setleri için sınıflandırma doğruluğunun radar grafiği 78
Şekil 3.17 DB3_v2 veri seti için sınıflandırma doğrulukları 79
Şekil 3.18 18 sınıflı DB3_v1_BesLab (a) ve DB3_v1_Histoplus (b) üzerinde eğitilen EfficientNet v2 medium'un karmaşıklık matrisleri, c) 18 sınıflı DB3_v1_GBL üzerinde eğitilen EfficientNet V2 small'un karmaşıklık matrisi 80
Şekil 3.19 a, c) 4 sınıflı DB3_v2_BesLab ve DB3_v2_GBL üzerinde eğitilen EfficientNet v2 small'ın karmaşıklık matrisleri, b) 4 sınıflı DB3_v2_Histoplus üzerinde eğitilen EfficientNet V2 Medium'un karmaşıklık matrisi 81
Şekil 4.1 Önerilen yönteme ait akış diyagramı 95

Şekil 4.2 DSÖ tarafından tanımlanmış sperm hareketlilik değerlendirme parametreleri [3] 99
Şekil 4.3 Localization Intersection over Union (Loc-IoU) metrik hesaplamasının görselleştirilmesi [137] 102
Şekil 4.4 AssA metriğinin hesaplanmasında kullanılan kavramların görsel gösterimi [137] 106
Şekil 4.5 Senaryo-I akış diyagramı 107
Şekil 4.6 Senaryo-II akış diyagramı 108
Şekil 4.7 Senaryo-III akış diyagramı 109
Şekil 4.8 Önerilen topluluk yöntemlerinin derin öğrenme tabanlı tespit modelleri ile gösterilmesi 117
Şekil 4.9 Vid-34 üzerinde YOLO, RT-DETR ve MEAN topluluk yöntemi kullanılarak elde edilen tespit ve yörünge sonuçlarının karşılaştırılması 117
Şekil 4.10 MEAN ensemble yöntemi kullanılarak (a) Vid-34 ve (b) Vid-36'da tespit ve izleme sonuçları (Çerçeve: 0-Son) 120
Şekil 4.11 Vid-34'e ait örnek sperm yörünge analizi 123
Şekil 4.12 Vid-36'ya ait örnek sperm yörünge analizi 123
Şekil 4.13 Vid-34'de bulunan 44 ID değerine sahip sperme ait yörünge analizi 124
Şekil 4.14 Vid-36 son karesindeki sperm hücrelerinin yörüngeleri ve hareketlilik seviyelerine göre renklendirilmesi 125

TABLO LİSTESİ

Tablo 1.1 Literatürde bulunan sperm morfoloji ve hareket verisetleri 7
Tablo 1.2 Sperm morfoloji sınıflandırmasında geleneksel özellik çıkarma ve derin öğrenme yöntemlerinin performans karşılaştırması . . . 13
Tablo 1.3 Literatürde bulunan sperm tespit ve hareket analiz yöntemleri . . 17
Tablo 2.1 SMIDS, HuSHeM ve SCIAN-Morpho veri setlerinin sınıflara göre dağılımları 25
Tablo 2.2 DB3_v1 veri setlerine dahil edilen toplam örnek sayıları 33
Tablo 2.3 Hi-LabSpermTracking veri setine ait video özellikleri 34
Tablo 3.1 Modellere ait parametre sayıları ve eğitim süreleri 42
Tablo 3.1 Modellere ait parametre sayıları ve eğitim süreleri (devam) . . . 42
Tablo 3.2 T-test sonucunda elde edilen skor tablosu 42
Tablo 3.3 HuSHeM-40x veri seti doğruluk değerleri 43
Tablo 3.4 SMIDS-10x veri seti doğruluk değerleri 43
Tablo 3.5 SCIAN-10x veri seti doğruluk değerleri 43
Tablo 3.6 Deneysel kurulum özellikleri 47
Tablo 3.7 Kullanılan transformer mimarilerin ImageNet top-1 doğruluk değerleri 56
Tablo 3.8 Test edilen hiperparametre listesi 58
Tablo 3.9 BEiT Base-RMSProp-10^{-5}'in Şekil 3.11 ve 3.12'den elde edilen karmaşıklık matrisi tabanlı performans ölçütleri bakımından değerlendirilmesi 62
Tablo 3.10 Önerilen yaklaşım ile literatürdeki yaklaşımlar arasında sperm morfoloji sınıflandırmasının karşılaştırılması 65
Tablo 3.11 DB3_v1_BesLab veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması 73
Tablo 3.12 DB3_v1_Histoplus veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması 74
Tablo 3.13 DB3_v1_GBL veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması 75
Tablo 3.14 Şekil 3.18'deki karmaşıklık matrislerinin sınıf bazında ayrıntılı performans ölçütleri 81

Tablo 3.15 Şekil 3.19'daki karmaşıklık matrislerinin detaylı performans metrikleri 81
Tablo 3.16 Kesinlik (Precision) değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları 83
Tablo 3.17 Duyarlılık değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları . . . 84
Tablo 3.18 F-Skor değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları . . . 85
Tablo 4.1 Witkowski sperm hareketlilik parametreleri referans değerleri [134] 99
Tablo 4.2 Tukey'in Honestly Significant Difference (HSD) sonuçları: telefon ve mikroskop kamera tabanlı video performansının değerlendirilmesi 110
Tablo 4.3 Senaryo II: her videonun son 15 saniyesi (test seti) için tespit ve MOT metrikleri 112
Tablo 4.4 Senaryo-III: görünmeyen test videoları için tespit ve MOT metrikleri 114
Tablo 4.5 Senaryo-III: Ensemble-metotların sonuçları 118
Tablo 4.6 Önerilen sperm yörünge analizi ile gerçek referans değerlerin karşılaştırılması 122

ÖZET

Biyomedikal Görüntülerdeki Hareketli Nesnelerin Derin Öğrenme Yöntemleri İle Şekil Ve Yörünge Analizi

Abdulsamet AKTAŞ

Bilgisayar Mühendisliği Anabilim Dalı
Doktora Tezi

Danışman: Doç. Dr. Hamza Osman İLHAN
Eş-Danışman: Dr. Öğretim Üyesi Görkem SERBES

Bu tez, kısırlık tanısında erkek faktörünün değerlendirilmesinde kullanılan sperm şekil ve hareket analizlerini derin öğrenme tabanlı yaklaşımlar ile iyileştirmeyi amaçlamaktadır. Görsel değerlendirme (GD) olarak adlandırılan manuel sperm değerlendirmesi, zaman alıcı, hata yapmaya açık, subjektif ve büyük ölçüde değerlendiricinin uzmanlığına bağlı bir yöntem olarak görülmektedir. Bu eksiklikleri gidermek amacıyla, bilgisayar destekli sperm analiz (BDSA) sistemleri geliştirilmiştir. Ancak mevcut BDSA sistemleri, yüksek maliyetleri ve erişim kısıtlamaları nedeniyle klinik uygulamalarda yaygın olarak kullanılmamaktadır. Son yıllarda, makine öğrenmesi ve derin öğrenme tabanlı yöntemler, sperm şekli (morfoloji) ve hareketliliğinin otomatik analizi için umut vadeden çözümler sunmaktadır.

Tez kapsamında sperm morfoloji analizi için Vision Transformer (ViT) tabanlı bir sistem önerilmiş ve HuSHeM ile SMIDS veri setlerinde sırasıyla %93,52 ve %92,5 doğruluk elde edilmiştir. Açık erişim veri setlerine ek olarak, tez kapsamında oluşturulup yeni yayımlanan Hi-LabSpermMorpho veri setinde de EfficientNet-V2 mimarisi kullanılarak deneyler gerçekleştirilmiştir. Bu doğrultuda, BesLab, Histoplus ve GBL veri setlerinde sırasıyla %65,05, %67,42 ve %63,58 doğruluk oranlarına ulaşılmıştır.

Tez çalışmasının bir diğer fazını oluşturan sperm hareket analizi için ise

Hi-LabSpermTracking veri seti oluşturulmuş, 12 farklı hastadan alınan, 60 saniye uzunluğunda ve 1920×1080 çözünürlüğünde toplam 66 video, iki uzman biyolog tarafından detaylı şekilde etiketlenmiştir. YOLOv8 ve RT-DETR modelleri eğitilerek sperm tespiti gerçekleştirilmiş, MEAN topluluk yöntemi ile birleştirilerek %87,87 F1-Skor ve %86,55 mAP50 değerlerine ulaşılmıştır. Takip işlemi DeepSORT yöntemi ile gerçekleştirilmiş, HOTA, MOTA ve IDF1 metrikleri kullanılarak değerlendirilmiştir. MEAN yöntemi, sırasıyla %66,60 HOTA, %76,42 MOTA ve %85,79 IDF1 değerleri ile en başarılı sonuçları vermiştir.

Son olarak, sperm hareket özellikleri (VCL, VSL, VAP, LIN, STR) çıkarılarak Dünya Sağlık Örgütü (WHO) kriterlerine göre motilite seviyeleri belirlenmiş, MEAN yöntemi ile elde edilen sonuçların Ground-truth (GT) motilite seviyeleri ile büyük ölçüde uyumlu olduğu doğrulanmıştır.

Tez kapsamında oluşturulan veri setleri ve geliştirilen yöntemler, sperm morfoloji ve motilite analizine önemli katkılar sağlamış, literatürdeki mevcut yaklaşımlardan daha başarılı analizler sunmuştur.

Anahtar Kelimeler: Kısırlık teşhisi, sperm şekil ve hareket analizi, sperm tespiti ve takibi, derin öğrenme, görü dönüştürücü

ABSTRACT

Morphology And Trajectory Analysis Of Motile Objects In Biomedical Images With Deep Learning Methods

Abdulsamet AKTAŞ

Department of Computer Engineering
Doctor of Philosophy Thesis

Supervisor: Assoc. Prof. Dr. Hamza Osman İLHAN
Co-supervisor: Asst. Prof. Dr. Görkem SERBES

This thesis aims to improve sperm morphology and motility analysis used for the assessment of male factors in infertility diagnosis by using deep learning based approaches. Manual sperm evaluation, referred to as visual assessment (VA), is considered a time-consuming, error-prone, subjective method that is highly dependent on the expertise of the expert. Computer-assisted sperm analysis (CASA) systems have been developed to overcome these shortcomings. However, the high costs and accessibility limitations of existing CASA systems have prevented their widespread use in clinical applications. In recent years, machine learning and deep learning-based methods have emerged as promising solutions for automated analysis of sperm morphology and motility.

Within the scope of the thesis, a Vision Transformer (ViT) based system was proposed for sperm morphology analysis and an accuracy of 93.52% and 92.5% was obtained in HuSHeM and SMIDS datasets, respectively. In addition to the open access datasets, experiments were carried out using EfficientNet-V2 architecture on the Hi-LabSpermMorpho dataset, which was created within the scope of the thesis and recently published. Accordingly, accuracy rates of 65.05%, 67.42% and 63.58% were achieved in BesLab, Histoplus and GBL datasets, respectively.

For sperm motility analysis, which constitutes another phase of the thesis study, Hi-LabSpermTracking dataset was created, and a total of 66 videos taken from

12 different patients, 60 seconds long and 1920×1080 resolution, were labelled in detail by two expert biologists. YOLOv8 and RT-DETR models were trained to detect sperm and combined with the MEAN ensemble method to achieve 87.87% F1-Score and 86.55% mAP50 values. Follow-up was performed with the DeepSORT method and evaluated using HOTA, MOTA and IDF1 metrics. MEAN method achieved the most successful results with 66.60% HOTA, 76.42% MOTA and 85.79% IDF1 values, respectively.

Finally, sperm motility characteristics (VCL, VSL, VAP, LIN, STR) were extracted and classified according to the World Health Organization (WHO) motility levels. The results obtained using the MEAN method were remarkably consistent with the Ground-Truth (GT) motility levels.

The datasets and methods developed in this thesis contribute significantly to sperm morphology and motility analysis, leading to systems that outperform existing studies both in practical applications and theoretical advancements.

Keywords: Infertility diagnosis, sperm morphology and trajectory analysis, sperm detection and tracking, deep learning, vision transformer

1
GİRİŞ

1.1 Literatür Taraması

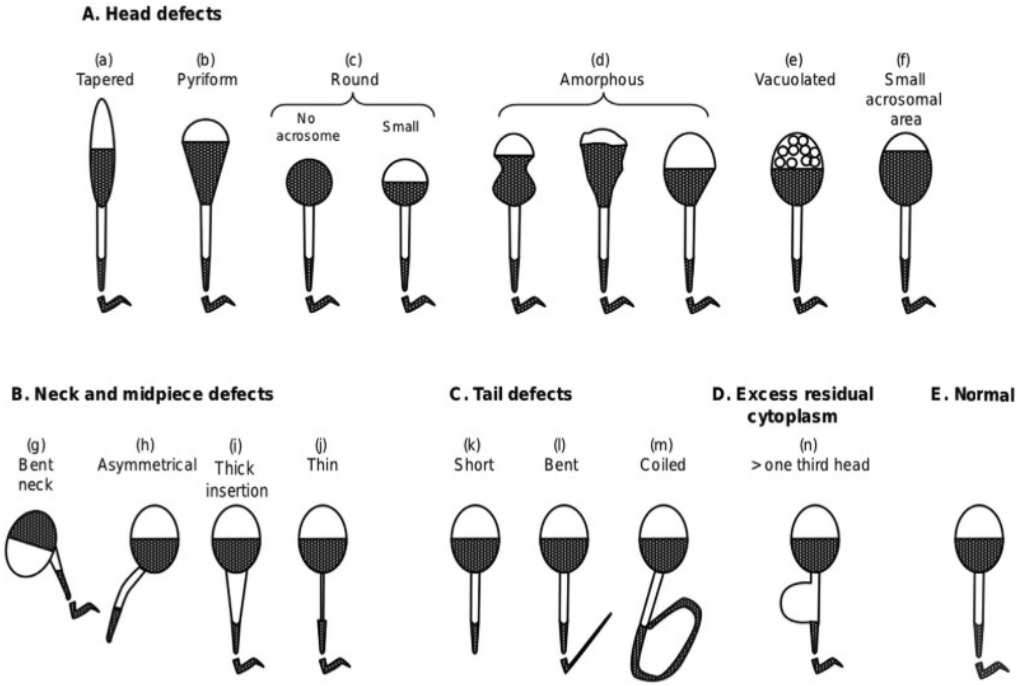
Dünya tarihinde infertilite, ilk kez Asur dönemine ait bir evlilik sözleşmesinin parçası olarak 4000 yıl önce hazırlanmış bir kil tablette bahsedilmektedir [1]. Kısırlık, 12 ay boyunca düzenli ve korunmasız 'cinsel ilişki' sonrasında gebeliğin gerçekleşememesi durumu olarak tanımlanmaktadır [2]. Dünya Sağlık Örgütü'nün (World Health Organization - WHO, DSÖ) son raporuna göre, dünya genelinde her altı yetişkinden biri, yani yaklaşık %17,5'i kısırlık sorunu yaşamaktadır [3]. Çiftlerdeki kısırlık sorununun %50'sinden erkek faktörlerinin sorumlu olduğu düşünülmektedir [4]. Bu nedenle, kısırlık tedavisinde erkek bireylerdeki problemin belirlenmesi oldukça önemlidir. Spermiogram testi, sperm hücrelerini değerlendirmek ve erkek kısırlığına ilişkin sorunları teşhis etmek için kullanılan tıbbi bir inceleme yöntemidir. Bu inceleme sırasında seminal plazmanın hacmi, pH değeri, likidifikasyon süresi ve viskozitesi ölçülerek kaydedilmektedir. İlk incelemenin ardından, spermlerin yoğunluğu, hareketliliği ve morfolojisi mikroskop altında analiz edilmektedir. Bu inceleme temel olarak iki ana aşamadan oluşmaktadır. İlk aşamada, uzmanlar hastalardan toplanan sperm örnekleri üzerinde hem fiziksel hem de kimyasal analizler gerçekleştirmektedir. Değerlendirilen özellikler arasında spermin rengi, akışkanlığı, kokusu, hacmi ve pH değeri bulunmaktadır. İkinci aşamada ise semen örneğinin yoğunluk, morfolojik yapı ve hareketlilik gibi karakteristik özellikleri mikroskop kullanılarak manuel olarak incelenmektedir. Görsel değerlendirme (GD) olarak adlandırılan bu manuel sperm değerlendirmesi, zaman alıcı, hata yapmaya açık, subjektif ve büyük ölçüde değerlendiricinin uzmanlığına bağlı bir yöntem olarak görülmektedir [5]. Görsel değerlendirmede sonucu etkileyen en önemli faktör uzmanların deneyimi olduğundan, farklı biyologlar aynı hasta için farklı inceleme sonuçları sunabilmektedir; bu durum literatürde gözlemci değişkenliği (observer variability) problemi olarak adlandırılmaktadır [6]. İnsan kaynaklı hatayı azaltmak için bilgisayar destekli sperm analiz (BDSA) sistemleri

geliştirilmiştir. BDSA sistemleri, görsel değerlendirmeye göre daha hızlı ve tarafsız sonuçlar sağlasa da, yüksek maliyet, mevcut laboratuvarlara entegrasyon zorlukları ve eğitimli gözlemciler gerektirmesi gibi dezavantajlara sahiptir. Ayrıca, genellikle sperm morfolojisinden ziyade sperm hareketliliğini değerlendirmek için kullanılmaktadırlar. BDSA sistemlerinin morfolojik analiz için sonuçları tutarsız ve/veya yetersiz kalmaktadır [7–9]. BDSA sistemlerinin sonuçları ile GD sonuçları karşılaştırıldığında, BDSA sistemlerine ait sonuçların standart sapmalarının GD sonuçlarına göre daha yüksek olduğu deneysel çalışmalar ile gözlemlenmiştir [10]. Bu nedenle, özellikle sperm morfoloji analizi için sınırlamaları ortadan kaldıran, daha etkili ve tamamen otomatik sistemlere ihtiyaç duyulmaktadır.

Bu bölümde, literatürde sperm morfoloji ve hareketliliği kapsamında oluşturulan veri setlerinin analizi, sperm morfoloji analizi üzerine yapılan çalışmalar ile sperm tespiti ve takibi üzerine gerçekleştirilen araştırmalar detaylı bir şekilde incelenmiştir. Bu kapsamda, ilgili çalışmaların avantajları, dezavantajları ve literatürdeki eksiklikler ayrıntılı bir şekilde ele alınmıştır.

1.1.1 Açık Kaynak Sperm Analiz Veri Setleri

Geleneksel olarak sperm morfoloji analizi, normal morfolojik özelliklere sahip sperm yüzdesinin belirlenmesine odaklanmaktadır. Ancak, son yıllarda yapılan çalışmalar, tedavi stratejilerinin normal dışı (anormal) sınıflara bağlı olarak değişebileceğini vurgulamaktadır [11]. Bu bulgu, spermlerin morfolojik olarak ayrıntılı bir şekilde incelenmesini gerekli kılmaktadır. Anormal sperm morfolojisi, çevresel faktörler, genetik yatkınlıklar ve edinilmiş davranışlarla önemli ölçüde ilişkilendirilmektedir [12–14]. Sperm hücreleri, baş, orta parça (mid-piece) ve kuyruk olmak üzere üç ana bölümden oluşmaktadır. Morfolojik olarak normal kabul edilebilmesi için sperm hücresinin tüm bölümleri, DSÖ tarafından belirlenen spesifik standartlara uygun olmalıdır. Bu yönergelere göre, tipik bir sperm başının uzunluğunun 3,7 ile 4,7 mikrometre, genişliğinin ise 2,5 ile 3,2 mikrometre arasında olması ve düzgün bir akrozom ile kaplanmış oval bir şekle sahip olması gerekmektedir [15]. Orta parçanın uzunluğunun 3,3-5,2 mikrometre, genişliğinin ise 1,3-1,8 mikrometre aralığında olması gerekir. Orta parçadan daha ince olan kuyruk, yaklaşık 45 mikrometre uzunluğunda olmalı ve başın uzunluğunun yaklaşık 10 katı kadar bir boyuta sahip olmalıdır. Bu referans değerlerine uymayan sperm hücreleri anormal olarak kabul edilmektedir. Bu anormallikler, baş, orta parça, kuyruk ve sitoplazmik kalıntı olmak üzere dört ana kategoriye ayrılmaktadır [2]. Her bir kategori, daha ayrıntılı morfolojik analiz gerektiren çeşitli alt sınıfları içermektedir. Bu anormalliklerin detayları Şekil 1.1'de görselleştirilmiştir.



Şekil 1.1 İnsan sperminde çeşitli anormal ve normal morfolojiler (a)konik, (b) armut, (c) yuvarlak, (d) şekilsiz, (e) vakuollü, (f) küçük akrozomal alan, (g) bükük boyun, (e) asimetrik boyun, (i) kalın bağlantı noktası, (j) ince boyun, (k) kısa kuyruk, (l) bükük kuyruk, (m) kıvrılmış kuyruk, (n) aşırı artık sitoplazma [2]

Sperm morfolojisi, kısırlık ile ilgili durumlar hakkında değerli bilgiler sağlasa da, sperm hücrelerinin fonksiyonel kapasitesini tam anlamıyla yansıtmamaktadır. Erkek fertilitesini daha kapsamlı bir şekilde değerlendirebilmek için sperm hareketliliğinin de incelenmesi önemlidir. Sperm hareketliliğinin değerlendirilmesi, düzgün doğrusal hareket kabiliyetine sahip sperm hücrelerinin yüzdesine dayanmaktadır. Sperm hareketliliği, DSÖ tarafından yayınlanan el kitabına göre sırasıyla düzgün doğrusal hareketli (Grade A), hareketli (Grade B), doğrusal olmayan hareketli (Grade C) ve hareketsiz (Grade D) olarak tanımlanan dört kategoriye ayrılmaktadır [2]. Aynı el kitabına göre, sperm konsantrasyonundaki toplam sperm hücrelerinin en az %40'ının hareketli olması ve ölçülen hareketli sperm hücrelerinin en az %4'ünün düzgün doğrusal hareket tipinde (Grade A+B) olması gerekmektedir. Düşük sperm hareketliliği (astenoozospermi), testiküler hasar [16], sperm mitokondriyal disfonksiyonu [17] veya varikosel [18] gibi durumlara işaret edebilmektedir. Derin öğrenme yöntemleri kullanılarak sperm analiz sistemlerinin geliştirilmesinde karşılaşılan en büyük zorluklardan biri, modellerin eğitimi için gerekli olan etiketlenmiş ve açık kaynaklı veri setlerinin yetersizliğidir. Literatür incelemesi, mevcut açık kaynaklı sperm analizi veri setlerinin genellikle sperm morfolojisine odaklandığını ve ağırlıklı olarak sınıflandırma problemlerinde kullanıldığını ortaya koymaktadır. Bu bölümde,

literatürde yer alan açık kaynaklı veri setleri kronolojik sırayla sunulmaktadır.

Yi ve arkadaşları tarafından oluşturulan sperm morfoloji veri seti, 12 farklı bireyden elde edilen 640×480 çözünürlüğünde yaklaşık 300 görüntü içermektedir. Bu çalışmada, Fourier ve dalgacık dönüşümleri kullanılarak sperm başı konturları yeniden yapılandırılmış ve sperm başları dört farklı kategoriye ayrılmıştır: normal, küçük, uzun ve büyük [19]. Dalgacık dönüşümü, konturun yerel detaylarını ve genel formunu ortaya koymak için de kullanılmıştır. Sınıflandırma için her bir görüntü ile anormallik sınıfı arasındaki ortalama kare hatası kullanılmıştır. Bu veri setindeki sınıflar, boyutları bakımından birbirinden kolayca ayırt edilebilir olduğundan sınıflandırma performansı yüksektir. Başka bir küçük veri setinde ise 512×512 çözünürlüğünde 160 normal ve 430 anormal sperm olmak üzere iki sınıfa ait toplam 590 görüntü bulunmaktadır [20]. Ramos ve arkadaşları, bu veri seti üzerinde bilgisayar destekli karyometrik görüntü analiz programı kullanarak sperm analizi gerçekleştirmiştir.

Başka bir çalışmada, Abbiramy ve arkadaşları, küçük bir veri seti üzerinde üç farklı yapay sinir ağı tabanlı sınıflandırıcı kullanarak insan sperm görüntülerini sınıflandırmayı amaçlamışlardır [21]. DSÖ laboratuvar kılavuzuna göre etiketlenmiş toplam 91 görüntü, ikili sınıflandırma için kullanılmıştır. Sınıflandırma işleminden önce, görüntüler üzerinde çeşitli ön işleme adımları uygulanarak görüntü iyileştirme işlemi yapılmıştır. Normal ve anormal örnekler, İleriye Doğru Sinir Ağı (Feed Forward Neural Network), Dairesel Tabanlı Sinir Ağları (Radial Basis Neural Network) ve Elman Geri Yayılım Sinir Ağı (Elman Back Propagation Neural Network) adı verilen üç ağ modeli kullanılarak sınıflandırılmıştır. En yüksek doğruluk oranı, gri seviye eş-oluşum matrisi özellikleri (Grey Level Co-occurrence Matrix) ile eğitilen İleriye Doğru Sinir Ağları yöntemiyle %75 olarak elde edilmiştir.

Spermlerin morfolojik ölçümlerini yapmak için Aksoy ve arkadaşları [22]'da 67 hastadan alınan sperm örneklerinden oluşan küçük bir veri seti oluşturmuştur. Bu çalışmanın amacı, morfolojik ölçümleri almak için bir BDSA sistemi kullanarak sperm morfoloji analizinde en iyi boyama yöntemini belirlemektir. Sonuçlar, sperm uzunluklarının boyama işlemine bağlı olarak değişebileceğini göstermiştir. BDSA'ne dayalı bir başka çalışmada, 250 hastadan alınan örneklerden oluşan bir veri seti, otomatik BDSA sistemlerini (SQA-V GOLD ve CASA CEROS) doğruluk ve kesinlik açısından geleneksel manuel değerlendirme yöntemleriyle karşılaştırmak için kullanılmıştır [23]. Bu çalışmada sperm morfolojisine ek olarak motilite ve yoğunluk da değerlendirilmiştir. Sonuç olarak, BDSA sistemleri ile görsel değerlendirme arasında istatistiksel olarak anlamlı bir fark gözlenmemiştir.

Ghasemian ve arkadaşları, 235 hastadan alınan 1457 sperm görüntüsünden oluşan ve 576×764 çözünürlüğe sahip olan Human Sperm Morphology Analysis Dataset (HSMA-DS) adı verilen daha büyük bir veri seti üzerinde deneyler yapmışlardır [24]. Çalışmada, sperm morfoloji analizi (SMA) adı verilen bir sistem geliştirilmiştir. SMA sisteminde, ilk aşamada görüntülerin kontrastı iyileştirilmiş ve ardından gürültü bileşenleri kaldırılmıştır. Ön işleme sonrasında, sistem sperm örneklerinin baş ve kuyruk gibi farklı parçalarını tespit etmiş ve her bir tespit edilen parçayı şekil ve boyut açısından analiz etmiştir. Son olarak, sistem her bir spermi normal ya da anormal olarak sınıflandırarak, sperm anormallik sınıflandırmasında %90'ın üzerinde bir doğruluk oranına ulaşmıştır.

İnsan sperm başı anormalliklerine odaklanan Human Sperm Head Morphology (HuSHeM) veri setinde, Shaker ve arkadaşları sözlük öğrenme tekniği kullanarak sperm anormalliklerini sınıflandırmışlardır [7]. Bu veri seti, 131×131 çözünürlüğünde 216 görüntüden oluşmakta ve dört farklı sınıfa ayrılmaktadır: Normal, Konik (Tapered), Priform (Pyriform) ve Amorf (Amorphous). HuSHeM veri seti, birçok derin öğrenme çalışmasında yöntemlerin birbirleri ile karşılaştırılması için referans olarak kullanılmıştır [25–28].

HuSHeM veriseti ile benzer sınıf yapılarına sahip olan SCIAN-MorphoSpermGS veri seti, insan sperm başlarını sınıflandırmayı standartlaştırmak amacıyla Chang ve arkadaşları tarafından oluşturulmuştur [29]. SCIAN-MorphoSpermGS veri setinde beş sınıf bulunmaktadır: Normal, Konik (Tapered), Priform (Pyriform), Küçük (Small Head) ve Amorf (Amorphous). Spermlerin etiketlenmesi, uzman görüşlerine dayanarak çoğunluk oyu ile gerçekleştirilmiştir. Veri seti, 35×35 piksel çözünürlüğünde toplam 1132 gri tonlamalı görüntü içermektedir. Bu veri setinde, sınıflandırma temel performanslarını göstermek amacıyla dört farklı denetimli öğrenme yöntemi (Destek Vektör Makinesi (Support Vector Machines - SVM) [30], Karar Ağaçları (Decision Tree -DT) [31], Saf Bayes (Naive Bayes - NB) [32] ve K-En Yakın Komşu (k- Nearest Neighbors - KNN) [33]) ve üç farklı şekil temelli tanımlayıcı (Hu momentleri, Zernike momentleri ve Fourier) karşılaştırılmıştır. SCIAN-MorphoSpermGS, HuSHeM veri setine benzer şekilde, sperm morfoloji sınıflandırması üzerine yapılan birçok çalışmada referans olarak kullanılmaktadır [7, 25–27].

Javadi ve arkadaşları tarafından, sperm morfolojisini analiz etmek amacıyla HSMA-DS [24] veri setinin değiştirilmiş bir versiyonu oluşturulmuştur [34]. Sperm başlarının manuel olarak işaretlendiği ve görüntülerin sperm başları merkeze alınacak şekilde kırpıldığı bu yeni veri seti, "Modified Human Sperm Morphology Analysis Dataset (MHSMA)" olarak adlandırılmıştır. MHSMA veri seti, 128×128

piksel çözünürlüğünde toplam 1540 gri tonlamalı görüntü içermekte olup sperm başı, kuyruk-boyun, akrozom ve vakuol anormalliklerini içermekte, her bir örnek normal veya anormal olarak sınıflandırılmaktadır. Ancak, detaylı anormal sınıfları etiketlenmemiştir. Tasarladıkları evrişimli sinir ağı (Convolutional Neural Network - CNN) mimarisi kullanılarak, akrozom, baş ve vakuol anormallikleri için sırasıyla %84,74, %83,86 ve %94,65 F-skor değerleri elde edilmiştir.

Mohammed ve arkadaşları, kendi oluşturdukları veri setini kullanarak insan sperm başı görüntülerini Twin SVM ve sinir ağları yardımıyla sınıflandırmıştır [35]. Bu veri seti, her biri 20×20 piksel çözünürlüğünde 60 sperm görüntüsünden seçilen 192 normal ve 192 anormal sperm başı görüntüsü içermektedir. Önerilen Yapay Sinir Ağı (YSA) modeli bu veri setine uygulanarak %100 doğruluk oranı elde edilmiştir. Başka bir çalışmada, sperm DNA kalitesini tahmin etmek amacıyla CNN mimarisi kullanılarak farklı bir sperm görüntü veri seti incelenmiştir [36]. Çalışmada, DNA kalitesi uzmanlar tarafından etiketlenmiş 1000 sperm hücresi görüntüsü üzerinde, derin öğrenme tabanlı modellerle DNA kalitesinin tahmin edilmesine yönelik eğitim ve test süreçleri gerçekleştirilmiş, ayrıca sperm hücresinin morfolojik ve hareket özelliklerinin DNA kalitesiyle olan ilişkisi incelenmiştir. Sonuçlar, sperm görüntüleri ile DNA kalitesi arasında orta derecede bir korelasyon olduğunu göstermiştir.

İlhan ve arkadaşları [37], 190×170 çözünürlüğe sahip ve normal, anormal ve sperm olmayan olmak üzere üç sınıfta toplam 3.000 örnek içeren Sperm Morphology Image Data Set (SMIDS) verisetini tanıtmışlardır. Çalışma kapsamında, sperm morfoloji sınıflandırma sistemi geliştirmek amacıyla geleneksel makine öğrenimi yöntemleri ve derin öğrenme yaklaşımları uygulanmıştır. Dalgacık tabanlı (wavelet-based) istatistiksel özellik çıkarımı ve tanımlayıcı tabanlı (descriptor-based) mekansal özellik çıkarımı yöntemleri kullanılarak, çeşitli makine öğrenimi ve topluluk öğrenme yöntemlerine beslenecek gürbüz özellikler elde edilmiştir. Ayrıca, VGG19, InceptionV3 ve MobileNet gibi evrişimli sinir ağı mimarileri ile SMIDS veriseti üzerinde sınıflandırma gerçekleştirilmiştir. Sperm morfoloji alanında önemli bir yere sahip olan SMIDS veri seti, birçok çalışmada yöntemlerin birbirleri ile kıyaslanması açısından standart bir referans olarak kullanılmıştır [26–28, 43].

Sperm morfolojisine dayalı anormalliklerin sınıflandırmasının yanı sıra, sperm ve kirlilik görüntülerin sınıflandırması için yeni bir veri seti olan Sperm Videos and Images Analysis (SVIA) veri seti oluşturulmuştur [38]. SVIA, üç farklı alt küme olarak yapılandırılmıştır: Alt küme-A, alt küme-B ve alt küme-C. Alt küme-C'de, 698×528 çözünürlüğe sahip 121.401 sperm sınıfına ait görüntü ve

Tablo 1.1 Literatürde bulunan sperm morfoloji ve hareket verisetleri

Yayın	Veriseti Adı	Problem	Görüntüler	Videolar	Özet
Yi ve ark. (1998) [19]	-	Sınıflandırma	300	-	Veri seti, 12 farklı bireyden toplanan 640×480 çözünürlükte 300 görüntü içermektedir.
Ramos ve ark. (2002) [20]	-	Sınıflandırma	590	-	Veri seti 512×512 çözünürlükte 590 görüntüden oluşmaktadır. 160 normal ve 430 anormal sperm görüntüsü içermektedir ve ikili (iki sınıflı) Sınıflandırma problemleri için kullanılmaktadır.
Abbiramy ve ark. (2011) [21]	-	Sınıflandırma	91	-	Veri kümesi 86×100 çözünürlükte 91 görüntü içermektedir.
Ghasemian ve ark. (2015) [24]	HSMA-DS: Human Sperm Morphology Analysis DataSet	Sınıflandırma	1457	-	Veri seti, 235 hastadan elde edilen örnekleri içeren 576×764 çözünürlüğe sahip toplam 1457 görüntüden oluşmaktadır. Baş ve kuyruk gibi sperm yapılarının normal veya anormal olarak etiketlenmesiyle oluşturulmuştur.
Shaker ve ark. (2017) [7]	HuSHeM: Human Sperm Head Morphology	Sınıflandırma	216	-	Bu veri seti 131×131 çözünürlükte 216 sperm başı görüntüsü içerir ve dört sınıfa ayrılmaktadır.
Chang ve ark. (2017) [29]	SCIAN-MorphoSpermGS	Sınıflandırma	1132	-	Bu veri seti, beş farklı sperm anormallik sınıfına ait 35×35, çözünürlükte toplam 1132 görüntüden oluşmaktadır.
Javadi ve ark. (2019) [34]	MHSMA: Modifed Human Sperm Morphology Analysis Dataset	Sınıflandırma	1540	-	Bu değiştirilmiş veri seti 128×128 çözünürlükte toplam 1540 görüntü içermektedir.
Mohammed ve ark. (2020) [35]	-	Sınıflandırma	384	-	Veri seti, 20×20 çözünürlüğe sahip 384 sperm head görüntüsünden oluşmaktadır.
Ilhan ve ark. (2020) [37]	SMIDS: Sperm Morphology Image Data Set	Sınıflandırma	3000	-	Bu veri kümesi 190×170 çözünürlükte toplam 3000 görüntüden oluşmaktadır ve üç sınıfa ayrılmıştır: normal, anormal ve non_sperm.
Chen ve ark. (2022) [38]	SVIA: Sperm Videos and Images Analysis dataset	Tespit, Segmentasyon, Sınıflandırma	4041	101 (1-2 sec)	Subset-A 101 video klipten oluşur ve toplam 125.000 nesne konumu içerir. Subset-B, her biri 451 temel gerçek segmentasyon maskesi içeren 10 video içerir. Subset-C kırpılmış görüntülerden oluşmaktadır ve iki kategori mevcuttur: impurity ve sperm.
Yuzkat ve ark. (2023) [39]	SeSViD: Semen Specimen Video Dataset	Tespit	1250	12 (20-30 sec)	Her biri 20 ila 30 saniye arasında değişen 12 video içerir. Toplam kare sayısı yaklaşık 1250'dir.
Haugen ve ark. (2019) [40]	VISEM	Regresyon	-	85	VISEM veri seti, 18 yaşın üzerindeki toplam 85 hastadan elde edilen verilere dayanmaktadır. Süreleri 2 ila 7 dakika arasında değişen videolar 640×480 çözünürlükte kaydedilmiştir.
Thambawita ve ark. (2023) [41]	VISEM-Tracking	Tespit, İzleme, Regresyon	29196	20 (30 sec)	Veri kümesindeki videolar 30 saniye uzunluğunda, 640×480 çözünürlüğünde ve toplam 29.196 kare içermektedir.
Tez Kapsamı (2024) [42]	Hi-LabSpermMorpho	Sınıflandırma	49345	-	Veri kümesi, 3 farklı boyama tekniği kullanılarak elde edilen 18 farklı sınıfa ait 49345 görüntü içermektedir.
Tez Kapsamı (2024)	Hi-LabSpermTracking	Tespit, Tracking	41808	66 (60 sec)	Veri kümesindeki videolar 60 saniye uzunluğunda, 1920×1080 çözünürlüğünde, ve toplam 41.808 kare içeriyor.

4.479 kirlilik (impurity) sınıfına ait görüntü bulunmaktadır. Ancak, sınıf dengesini sağlamak amacıyla, öğrenme modellerinin oluşturulmasında yalnızca rastgele seçilen 5.058 sperm görüntüsü ve 4.479 safsızlık görüntüsü kullanılmıştır. Kirlilik sınıfına ait görüntüler, sperm hücreleri dışında yer alan ve modelin doğruluğunu etkileyebilecek sperm olmayan hücreler, yabancı partiküller, sıvı kalıntıları veya analizdeki diğer kirleticiler gibi unsurları temsil etmektedir. Çalışmanın temel

amacı, sperm sınıflandırmasında kullanılan farklı CNN ve Görü Dönüştürücü (Vision Transformer, ViT) mimarilerinin, gürültülü senaryolarda performanslarını karşılaştırmaktır. Bu bağlamda, 12 farklı CNN mimarisi ve 6 farklı ViT mimarisi eğitilmiş ve elde edilen sonuçlar ayrıntılı bir şekilde analiz edilmiştir.

Sperm morfoloji analizi için normal veya anormal olarak sınıflandırmayı hedefleyen farklı bir verisetinde, Yang ve arkadaşları çeşitli CNN mimarilerini kullanmışlardır [44]. Çalışmada, sperm başları, kuyrukları ve orta kısımları segmentlere ayrılmış ve normal ya da anormal olarak sınıflandırılmıştır. Uzmanlar tarafından toplamda 30.000 görüntü normal veya anormal olarak etiketlenmiş ve sperm morfoloji analizi için kullanılmıştır. Araştırmacılar, sınıflandırma işlemini gerçekleştirmek için EfficientNetB0 [45] mimarisini kullanmış ve sistemin sonuçlarının manuel değerlendirme bulguları ile yüksek derecede tutarlı olduğunu bildirmiştir.

Yüzkat ve arkadaşları, spermlerin tespitini otonom olarak gerçekleştirebilmek amacıyla Semen Specimen Video Dataset (SeSViD) adlı veri seti oluşturmuşlardır [39]. SeSViD, yaşları 19 ile 37 arasında değişen 12 farklı hastadan toplanan 12 videodan oluşmaktadır. Başlangıçta 30 fps (frame per second) hızında kaydedilen videolar, etiketleme amacıyla 5 fps'ye düşürülmüştür. Orijinalde 1920×1200 çözünürlüğünde olan videolar, derin öğrenme tabanlı tespit algoritmalarında model gerekliliklerini sağlamak amacıyla 640×640 piksel çözünürlüğüne yeniden ölçeklendirilmiştir. Yaklaşık 1250 kareden oluşan bu veri seti, hem literatürde yer alan teorik veri bölütleme yöntemlerinde hem de gerçek dünya test senaryolarını simüle eden, tamamen birbirinden ayrılmış eğitim ve test videoları kullanılarak oluşturulan pratik sistemlerde kullanılabilmektedir.

Haugen ve arkadaşları tarafından sunulan VISEM veri seti, 18 yaşın üzerindeki 85 hastadan toplanan veriler kullanılarak oluşturulmuştur [40]. Orijinal videolar, her biri 2 ila 7 dakika arasında değişen sürelerde ve 640×480 çözünürlüğünde kaydedilmiştir. Bu videoların her birinden 250 kare çıkarılmış ve uzmanlar tarafından sağlanan semen_analysis_data ile eşleştirilerek veri setinin bir regresyon görevi olarak kullanılması sağlanmıştır. Toplamda 21.250 görüntü içeren veri seti, üç katlı çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. Haugen ve arkadaşları tarafından oluşturulan VISEM veri setindeki videoları kullanarak, Vajira ve arkadaşları VISEM-Tracking veri setini oluşturmuşlardır [41]. Orijinal VISEM videolarından 20'si detaylı etiketleme için seçilmiş ve bu videolardan 30 saniyelik klipler çıkarılmıştır. Bu işlem sonucunda toplam 29.196 etiketli kare elde edilmiştir. Çözünürlük (640×480) orijinal VISEM veri setiyle aynı kalmıştır. Önceki veri setinden farklı olarak, VISEM-Tracking veri seti bireysel sperm hücreleri için sınırlayıcı kutu koordinatları ve kimlik bilgilerini içermektedir. Hem kimliklerin

hem de sınırlayıcı kutuların etiketlenmesiyle VISEM-Tracking veri seti, hem tespit hem de takip görevlerini destekleyen faydalı bir kaynak haline gelmiştir.

Tablo 1.1, literatürde yaygın olarak kullanılan açık erişimli sperm verisetlerinin özetini sunmaktadır. Mevcut verisetlerinin çoğu, genellikle sperm morfolojisini vurgulamakta olup, çoğunlukla sınıflandırma görevleri için kullanılmaktadır [7, 19–21, 24, 29, 35, 37, 38, 42]. İncelenen çalışmalar, sperm morfoloji alanında belirli verisetlerinin sıklıkla kullanıldığını göstermektedir [7, 29, 37]. Bu verisetleri, yöntem karşılaştırmaları açısından birçok çalışmada referans noktası olarak kullanılmıştır [7, 25–28, 43]. Ancak, sınırlı sınıf sayısı [24, 37, 38], küçük örneklem boyutları [7, 29], sınıf dengesizliği [38] ve sperm başı anormalliklerine odaklanılması [29] gibi sınırlamalar, kapsamlı analiz sistemlerinin geliştirilmesini engellemektedir. Mevcut veri setlerinin çoğu en fazla beş sınıf içermekte ve genellikle orta parça ve kuyruk bozuklukları için ayrıntılı etiketleme sağlamamaktadır. Ayrıca, farklı boyama tekniklerinin sperm morfolojisi üzerindeki etkileri manuel olarak değerlendirilmiş olmasına rağmen, bu teknikleri tam otomatik sistemler kullanarak karşılaştırmaya olanak tanıyacak bir veri kümesi literatürde mevcut değildir [22].

Mevcut sperm tespiti ve takibi verisetleri incelendiğinde, SeSViD verisetinin sperm tespiti üzerine odaklandığı ve etiketli tespit verileri sunduğu, ancak bireysel sperm hücrelerinin takibini gerçekleştirebilmek için gerekli kimlik etiketlerini içermediği gözlemlenmiştir [39]. Bir diğer veriseti VISEM de ise, sperm tespiti ve takibine odaklanmak yerine problem bir regresyon görevi olarak ele alınmıştır [40]. VISEM-Tracking veriseti, detaylı sınırlayıcı kutu ve kimlik etiketlerini içerse de [41], yalnızca 20 video ile sınırlıdır ve her biri yalnızca 30 saniyedir. Kısa kayıtlar, sperm hareketliliğini ve davranışını uzun süreli analiz etmeyi kısıtlamaktadır. Uzun süreli videolar, takip algoritmalarının etkinliğini daha doğru bir şekilde değerlendirmek için kritik öneme sahiptir. Ayrıca, verisetindeki bazı videoların (örneğin, 14, 15, 21, 23, 35, 47, 54, 82) iki farklı videonun birleştirilmesiyle oluşturulduğu, diğerlerinin (örneğin, 29, 52, 60) ise üç farklı videonun birleştirilmesiyle oluşturulduğu belirlenmiştir. Bu durum, bir sperm hücresinin bütün video boyunca izlediği rotayı tutarlı bir şekilde analiz etmeyi imkansız hale getirmektedir. Ayrıca, VISEM-Tracking verisetindeki videolar, 400$\times$ büyütme ile kaydedilmiş ve 640$\times$480 çözünürlüğe sahiptir. Bu teknik özellikler, sperm hücrelerinin dar bir görüş açısıyla kaydedilmesine yol açmakta ve bunun sonucunda sperm yörüngelerinin detaylı ve geniş açılı analizini sınırlamaktadır. Ayrıca, hem VISEM hem de VISEM-Tracking, erkek üreme fonksiyonları üzerinde aşırı kilo ve obezitenin etkilerini incelemek amacıyla tasarlanmıştır. Bu durum genetik ve biyolojik çeşitliliği sınırlamaktadır. Daha kapsamlı ve genellenebilir

bilgiler elde edebilmek için, yalnızca belirli bir rahatsızlığa sahip bireylerle sınırlı kalmayıp, genel popülasyondan alınan verileri temsil eden daha geniş veri setlerinin kullanılması oldukça faydalı olacaktır.

1.1.2 Sperm Morfoloji Analiz Yöntemlerine Genel Bakış

Manuel sperm morfoloji analizi, hem zaman alıcıdır hem de uzman kişinin deneyimine büyük ölçüde bağlıdır [5]. Bu nedenle, insan faktörünün en aza indirildiği ve tanının otomatik olarak gerçekleştirildiği sistemlerin geliştirilmesine yönelik artan bir ihtiyaç bulunmaktadır. Sperm morfoloji analiz sistemleri, geleneksel (klasik özellik çıkarımı ve makine öğrenimi) ve derin öğrenme yaklaşımları olmak üzere iki kategori altında incelenmiştir.

1.1.2.1 Geleneksel Sperm Morfoloji Analiz Yöntemleri

İlk sperm morfoloji verisetleri, derin ağ mimarilerini eğitmek için yeterli etiketlenmiş görüntü içermemekteydi. Bu nedenle, araştırmacılar sperm hücrelerinin morfolojik özelliklerini çıkarmak için geleneksel özellik çıkarım yöntemlerini kullanmışlardır.

Alegre ve arkadaşları, geleneksel özellik çıkarım yöntemlerini kullanarak sperm morfoloji sınıflandırması gerçekleştirmiştir [46]. Otsu eşikleme yöntemi ile sperm konturlarını tespit etmiş ve tespit edilen spermleri Haralick ve kontur özellikleri kullanarak sınıflandırmışlardır. Sınıflandırıcı performansını değerlendirmek için İlgi Öğrenmeli Vektör Niceleme (Relevance Learning Vector Quantization - RLVQ), sınıf koşullu ortalamalar ve KNN yöntemlerini kullanarak çapraz doğrulama uygulamışlardır. En iyi sonuçlar, gradyan büyüklüğü verileri kullanılarak elde edilmiş ve toplam test hatası yalnızca %1 olarak rapor edilmiştir. Bir diğer çalışmada, Alegre ve arkadaşları aynı veri seti üzerinde sperm doku özelliklerini çıkararak bir sınıflandırma gerçekleştirmiştir [47]. Sperm dokuya dayalı özellikler, birinci dereceden istatistikler ve eş-oluşum matrisinden türetilmiştir. Çıkarılan doku özellikleriyle beslenen Çok Katmanlı Algılayıcı (Multi Layer Perceptron - MLP) ve KNN sınıflandırıcılar ile %94 doğruluk oranına ulaşılmıştır. Sperm baş, orta parça ve kuyruk olarak segmente edilerek başlatılan bir yaklaşımda; büyük ve küçük eksenler, bölgesel alan ve çevre uzunluğu gibi uzamsal özellikler kullanılarak, bulanık mantık ifadelerinin tanımlandığı ayrıntılı bir anormallik sınıflandırması gerçekleştirilmiştir [48].

Olalla ve arkadaşları, faz kontrast mikroskopisi ile elde edilen gri tonlamalı sperm görüntülerini akrozom-intakt ve akrozom-hasar görmüş olarak sınıflandırmıştır

[49]. Otsu eşikleme, bölgesel minimumlar ve watershed dönüşümü kullanarak sperm başlarını bölütlemiş ve hizalamışlardır. Sınıflandırma işlemini ise en küçük kareler algoritması ve lineer çekirdek destekli SVM yöntemiyle gerçekleştirmişlerdir. Bu yöntemle %99,13 F1-skoru elde edilmiştir.

İlhan ve arkadaşları, SMIDS veri seti ile yaptıkları çalışmalarında, sperm görüntülerine klasik ayrık dalgacık dönüşümü (Discrete Wavelet Transform - DWT) ve çift ağaç karmaşık dalgacık dönüşümü (Dual-Tree Complex Wavelet Transform - DTCWT) yöntemlerini uygulayarak öznitelik çıkarımı gerçekleştirmişlerdir [50]. Dalgacık katsayılarından istatistiksel öznitelikler elde etmiş ve bu öznitelikleri sınıflandırmak için farklı çekirdek fonksiyonlarıyla SVM algoritmasını kullanmışlardır. Sonuçlar, DTCWT'nin kayma değişmezliği ve daha iyi yön seçimi özellikleri sayesinde geleneksel DWT'ye kıyasla daha yüksek doğruluk ve F-skor metrikleri sağladığını göstermiştir. Ayrıca, İlhan ve arkadaşları, sperm morfoloji görüntülerini, betimleyici tabanlı öznitelikler (Kecskés-Azencott-Zhang-École (KAZE) [51], Hızlandırılmış Sağlam Öznitelikler (Speeded Up Robust Features - SURF) [52], Maksimum Stabil Ekstremal Bölgeler (Maximally Stable Extremal Regions - MSER) [53]) ve doku öznitelikleri (DWT, DTCWT) kullanarak sınıflandırmayı hedeflemişlerdir [37]. Bu öznitelikler, geleneksel makine öğrenmesi yöntemleri (SVM, DT, KNN) ve topluluk öğrenme yöntemleri (Adaboost, Bagging, RusBoost) ile sınıflandırılmıştır. Geleneksel öznitelik çıkarım yöntemlerinde, dalgacık ve betimleyici tabanlı özniteliklerle beslenen SVM ile sırasıyla %80,5 ve %83,8 en yüksek doğruluk değerlerine ulaşılmıştır. Bunun yanı sıra, öznitelik çıkarımı ve manuel müdahale gerektirmeyen tamamen otomatik çalışan derin öğrenme mimarileri (VGG19, Inception, MobileNet) ile yapılan deneylerde sırasıyla %86, %87,27 ve %87 doğruluk değerleri elde edilmiştir.

1.1.2.2 Derin Öğrenme Tabanlı Sperm Morfoloji Analiz Yöntemleri

Geleneksel öznitelik çıkarımı yöntemleri ve sonrasında öğrenme modelleriyle gerçekleştirilen sperm morfoloji analizleri, tam otomatik sınıflandırma yapılmasını mümkün kılmamaktadır. Yukarıda yer alan çeşitli çalışmalarda görüldüğü üzere, görüntüler sınıflandırılmadan önce ön işleme adımlarından geçirilerek özellik çıkarımı işleminin yapılması gerekmektedir. Ancak, derin öğrenme yöntemleri, ek öznitelik çıkarımı adımı olmaksızın tamamen otomatik çalışan sistemlerin tasarlanmasına olanak tanımaktadır. Derin öğrenme mimarileri, çalışma prensipleri gereği büyük miktarda veriye ihtiyaç duymaktadır. Sperm morfoloji veri setlerinde etiketli örnek sayısının artması, bu alanda derin öğrenme mimarilerinin kullanımını mümkün kılmıştır. Derin öğrenme yöntemleriyle yapılan sperm morfoloji sınıflandırmasıyla ilgili dikkat çeken bazı çalışmalar bu başlık altında özetlenmiştir.

Riordon ve arkadaşları, HuSHeM [7] ve SCIAN-MorphoSpermGS [29] sperm morfoloji verisetlerinde sınıflandırma yapmak için VGG16 ağını kullanmışlardır [25]. Model, başlangıçta 100 epoch boyunca eğitilmiş ve ardından 200 ek epoch ile ince ayar yapılmıştır. Eğitim öncesinde, görüntüler çeşitli yöntemlerle artırılmış ve manuel olarak sperm başları aynı yöne bakacak şekilde döndürülmüştür. Bu manuel müdahale, benzer desenlerin model tarafından daha iyi algılanmasını ve daha anlamlı özniteliklerin çıkarılmasını sağlayarak öğrenme sürecine katkıda bulunmaktadır. Böylece, modellerin sınıflandırma doğruluğu artırılmıştır. Önerilen yöntemle sırasıyla HuSHeM ve SCIAN-MorphoSpermGS veri setlerinde %94,1 ve %62 doğruluk değerleri elde edilmiştir.

Tortumlu ve arkadaşları, HuSHeM ve SMIDS veri setlerini kullanarak sperm morfoloji sınıflandırma performanslarını karşılaştırmışlardır [54]. Çalışmada, hesaplama gücü sınırlı platformlarda yaygın olarak kullanılan MobileNet-V1 ve MobileNet-V2 modelleri sınıflandırma yöntemleri olarak seçilmiştir. Ayrıca, veri arttırma, eğitim süresi ve modellerin yakınsama sağlamak için gereken epoch sayısı gibi hiperparametreler detaylı bir şekilde karşılaştırılmıştır. MobileNet-V1 modeli ile SMIDS veri seti üzerinde %88, MobileNet-V2 modeli ile ise HuSHeM veri seti üzerinde %77 doğruluk oranlarına ulaşılmıştır.

Spencer ve arkadaşları, sperm başı anormalliklerini sınıflandırmak için meta sınıflayıcılarla birleştirilmiş topluluk derin öğrenme yöntemlerini kullanmışlardır [55]. HuSHeM ve SCIAN-SpermMorphoGS veri kümeleri üzerinde VGG16, VGG19, modifiye ResNet-34 ve DenseNet-161 mimarilerini kullanarak önerdikleri yaklaşımın sperm morfoloji performansını ölçmüşlerdir. Bu derin öğrenme modelleri ile elde edilen özellik vektörleri birleştirilerek nihai tahminler için meta sınıflayıcılar beslenmiştir. [25]'e benzer şekilde, tüm sperm başlarının aynı yöne bakmasını sağlamak için sperm görüntüleri manuel olarak döndürülmüş ve kırpılmıştır. Bu yarı manuel veri artırma yaklaşımı ile HuSHeM veri kümesi için %98,2, SCIAN-SpermMorphoGS veri kümesi için ise %63,3 doğruluk değerleri elde edilmiştir.

Yüzkat ve arkadaşları, altı farklı CNN mimarisini topluluk öğrenme yöntemleriyle birleştirerek sperm morfoloji sınıflandırması gerçekleştirmiştir [28]. Sonuçlar, olasılığa dayalı (soft voting) ve çoğunluğa dayalı (hard voting) oylama yöntemleri kullanılarak birleştirilmiştir. HuSHeM, SMIDS ve SCIAN-MorphoSpermGS veri kümelerinde sırasıyla %90,73, %85,18 ve %71,91 doğruluk değerleri elde edilmiştir. Füzyon fikrine dayanan bir başka çalışmada, İlhan ve arkadaşları VGG16 ve GoogleNet modellerini karar seviyesinde olasılığa dayalı (soft voting) ve çoğunluğa dayalı (hard voting) oylama ile birleştirerek yeni bir yaklaşım

Tablo 1.2 Sperm morfoloji sınıflandırmasında geleneksel özellik çıkarma ve derin öğrenme yöntemlerinin performans karşılaştırması

Yöntemler	Döndürme ve Kırpma	Yayın	Yöntem	Doğruluk (%)
Özellik Çıkarım Tabanlı	Uygulanmadı	Alegre ve ark. 2012 [47]	DWT,MLP ve KNN ile özellik çıkarma	94
	Uygulanmadı	Alegre ve ark. 2013 [46]	Otsu eşikleme + Haralick ve Kontur Özellikleri, RLVQ, sınıf koşullu ortalamalar ve KNN	99
	Uygulanmadı	Ollalla ve ark. 2015 [49]	Otsu eşikleme, bölgesel minimumlar ve watershed dönüşümü, en küçük kareler algoritması ve SVM sınıflandırması	99
	Manuel	Chang ve ark. 2017 [29]	Fourier tanımlayıcısı ile özellik çıkarımı ve SVM sınıflandırması	SCIAN: 49
	Manuel	Chang ve ark. 2017 [56]	Fourier, Zernike ve şekil tanımlayıcıları topluluğu ile özellik çıkarımı, SVM sınıflandırması topluluğu	SCIAN: 58
	Manuel	Shaker ve ark. 2017 [7]	Renk alanı dönüştürme ve sözlük öğrenimi	SCIAN: 62
	Manuel	Shaker ve ark. 2017 [7]	Elle sperm başı döndürme ve kırpma 50×76 piksel renk uzayına dönüştürme sözlük öğrenme	HuSHeM: 92,2
	Otomatik	Ilhan ve ark. 2018 [50]	DWT, DTCWT ve DVM ile özellik çıkarımı ve SVM sınıflandırması	SMIDS: 82,33
	Otomatik	Ilhan ve ark. 2020 [37]	KAZE, SURF, MSER, DWT ve DTCWT ile özellik çıkarma SVM, DT, KNN, Adaboost, Bagging, RusBoost	SMIDS: 83,8
Derin Öğrenme Tabanlı	Manuel	Riordon ve ark. 2019 [25]	VGG16	HuSHeM: 94,1 SCIAN : 62
	Uygulanmadı	Tortumlu ve ark. 2020 [54]	MobileNet-V1, MobileNet-V2	SMIDS: 88 HuSHeM 77
	Uygulanmadı	Yuzkat ve ark. 2021 [28]	6 CNN mimarisi ile Topluluk öğrenme yöntemleri	SMIDS: 85,18 HuSHeM: 90,73 SCIAN: 71,91
	Yarı-Manuel	Spencer ve ark. 2022 [55]	VGG16, VGG19, ResNet-34, DenseNet-161 + Meta-Classifier	HuSHeM: 98,2 SCIAN : 63,3
	Uygulanmadı	Ilhan ve ark. 2022 [26]	Olasılığa ve çoğunluğa dayalı oylama ile VGG16 ve GoogleNet modelleri	SMIDS: 90,87 HuSHeM: 92,1 SCIAN: 73,2
	Uygulanmadı	Mahali ve ark. 2023 [43]	Uygunsuz Veri Artırımı SwinTransformer, MobileNetv3 ve Otomatik kodlayıcı	SMIDS: 91,7 HuSHeM: 97,6
	Uygulanmadı	Jabbari ve ark. 2023 [57]	Kapsül sinir ağları + GAN	HuSHeM: 97,8
	Uygulanmadı	Aktas ve ark. 2023 [27]	Vision Transformer, CNN	SMIDS: 89,43 HuSHeM: 90,85 SCIAN: 65,90

önermişlerdir [26]. Modelleri sperm morfoloji görüntülerine daha iyi uyarlamak için iki aşamalı bir ince ayar süreci uygulanmıştır. SMIDS, HuSHeM ve SCIAN-MorphoSpermGS veri kümeleri üzerinde yapılan deneysel çalışmalarda sırasıyla %90,87, %92,1 ve %73,2 doğruluk değerleri elde edilmiştir.

Literatür incelendiğinde, sperm morfoloji analizi alanında daha önceki veri setlerinin kısıtlamalarından arındırılmış kapsamlı bir veri setine ihtiyaç duyulduğu ortaya çıkmaktadır. Özellikle, geleneksel özellik çıkarım yöntemlerinde kullanılan manuel ön işleme, tam otomatik sperm analiz sistemlerinin geliştirilmesini engellemektedir. Derin öğrenme tabanlı sistemler, hedef kalıpları etkili bir

şekilde genelleştirebilmek için büyük veri kümelerine ihtiyaç duysa da, çoğu çalışma, sperm morfoloji analizi literatüründe halka açık veri kümelerinin azlığı nedeniyle boyutları sınırlı olan HuSHeM, SMIDS ve SCIAN-MorphoSpermGS gibi veri setlerinde gerçekleştirilmiştir [26, 28, 34, 55]. Bazı çalışmalarda [34, 55], derin öğrenme yöntemlerinde daha anlamlı özellikler çıkarmak ve böylece sonuçları iyileştirmek için döndürme ve kırpma gibi manuel ön işleme adımlarıyla birleştirilmiştir. Ancak bu manuel yaklaşım, sistemlerin tam otomatik olmasını kısıtlamaktadır. Literatürdeki mevcut sperm morfoloji analiz çalışmalarının detaylı özeti Tablo 1.2'de sunulmuştur.

1.1.3 Sperm Tespit ve Takip Yöntemlerine Genel Bakış

Bu başlık altında, literatürdeki sperm tespiti ve takip sistemleri sistematik bir şekilde incelenmiş, alandaki son gelişmeler vurgulanarak mevcut araştırma boşlukları belirlenmiştir.

Dewan ve arkadaşları, çalışmalarında sperm konsantrasyonu ve toplam hareketliliği tahmin etmeyi amaçlamışlardır [58]. Çalışmada her bir karede sperm hücreleri, blob algılama yöntemi ile tespit edilmiş ve bu spermlerin hareketini temsil eden yörüngeler oluşturulmuştur. Sperm konsantrasyonu, her bir karede tespit edilen sperm sayısına göre hesaplanmıştır. Yörüngelerin sperm hücrelerine ait olup olmadığını belirlemek için, araştırmacılar tarafından geliştirilen CNN modeli kullanılmıştır. Ayrıca, çıkarılan yörüngeler video stabilizasyonu için de kullanılmıştır. Ancak çalışmada bazı kısıtlar bulunmaktadır. İlk olarak, nesne önerilerini belirlemek için kullanılan blob algılama yöntemi, sperm örneklerindeki kalıntılar ve aydınlatma koşullarından etkilenebilmektedir. Hareket tahmininde ise belirli bir eşik değerine dayalı sert eşikleme uygulanmıştır. Bu yaklaşım, sperm hareketlerinde meydana gelen ince farkları göz ardı ederek sınıflandırma doğruluğunu sınırlamaktadır. Çalışmada kullanılan veri seti, 100'den fazla hastadan elde edilen toplam 3.800 adet, 50×50 boyutlarında görüntüler içermektedir. Bununla birlikte, yapılan deneysel çalışmaların sonuçları incelendiğinde, önerilen yöntemin performansını değerlendirirken oldukça önemli olan Çoklu Nesne Takip Doğruluğu (Multiple Object Tracking Accuracy - MOTA) veya Daha Yüksek Mertebe Takip Doğruluğu (Higher Order Tracking Accuracy - HOTA) gibi metriklerin sunulmadığı gözlemlenmiştir.

Thambawita ve arkadaşları, sperm hareket türü ve şekil bozukluklarını tahmin etmek için derin öğrenme tabanlı iki aşamalı bir yöntem önermişlerdir [59]. İlk aşamada, sperm videolarından zamansal özellikleri çıkarmak için bir otokodlayıcı (autoencoder) kullanılmıştır. Bu özellikler, zamansal bilgileri uzamsal bir alana

yansıtarak "özellik-görüntüleri" adı verilen formatlara dönüştürülmüştür. İkinci aşamada ise, oluşturulan özellik-görüntüleri üzerinde transfer öğrenme kullanılarak önceden eğitilmiş bir ResNet-34 modeli ince ayar ile yeniden eğitilmiştir. Hareket ve şekil analizini bir regresyon problemi olarak ele alan araştırmacılar, hareket tahmini için 9,427, şekil anormalliği tahmini için ise 5,606 ile en düşük Ortalama Mutlak Hata (Mean Absolute Error - MAE) değerlerini elde etmişlerdir.

Hicks ve arkadaşları, çalışmalarında hem geleneksel hem de modern makine öğrenimi tekniklerini kullanarak sperm hareketliliğini tahmin etmeyi amaçlamışlardır [60]. Çalışmada üç farklı özellik çıkarım yöntemi kullanılmış ve sperm hareket tahmini bir regresyon problemi olarak ele alınmıştır. Kullanılan yöntemler arasında ham kareler (raw frames), seyrek ve yoğun optik akış (sparse and dense optical flow) özellikleri ve her iki özelliğin birleşimi yer almaktadır. Basit doğrusal regresyondan CNN tabanlı derin öğrenme mimarilerine kadar çeşitli modeller değerlendirilmiştir. En etkili sonuç, ResNet-50 ağında optik akış özelliklerinin kullanılmasıyla elde edilmiş ve Ortalama Mutlak Hata (Mean Absolute Error - MAE) değeri 8,740 olarak rapor edilmiştir.

Başka bir çalışmada, Thambawita ve arkadaşları, önceki çalışmalarına benzer şekilde, sperm hareket ve şekil bozukluğu tahminini ham video kare yığınları ve optik akış tabanlı temsiller kullanarak gerçekleştirmişlerdir [61]. Hicks ve arkadaşlarının [60] ham kareler ile optik akışı birleştirdiği yaklaşımdan farklı olarak, Thambawita ve arkadaşları bu yöntemleri ayrı ayrı değerlendirmişlerdir. İlk ham veri işleme yaklaşımında, dokuz ardışık gri tonlamalı kare bir araya getirilerek bir yığın oluşturulmuş, ikinci yaklaşımda ise farklı adımlarda hesaplanan RGB kareler ve optik akış haritalarını entegre eden karmaşık bir temsil kullanılmıştır. En iyi performans, ikinci yaklaşımı takip ederek, MLP'yi dropout ile bir ResNet-34 modeline entegre eden özel bir derin sinir ağıyla elde edilmiştir. Bu özel ağ ile hareket analizi için 8,825 ve şekil analizi için 5,293 MAE değerleri elde edilmiştir.

Valiuskaite ve arkadaşları, sperm motilitesini değerlendirmek için Bölge Tabanlı CNN (Region based -CNN) yaklaşımına dayalı bir derin öğrenme yöntemi geliştirmişlerdir [62]. VISEM veri setinin (elle etiketlenmiş 650 kareden oluşan bir alt küme) kullanıldığı bu yöntemde, ilk olarak sperm başları segmente edilerek elde edilen spermler ardışık kareler (frames) boyunca izlenmektedir. Kimlik atamaları, Öklid mesafesine dayalı bir eşikleme yöntemi ile gerçekleştirilmektedir. Önerilen yöntem, sperm başları için %91,77'lik bir tespit doğruluğu elde etmiş olsa da, VISEM veri setinin çok küçük bir alt kümesi üzerinde test edilmiş ve spermlerin çarpışma veya örtüşmesi senaryolarında ortaya çıkabilecek kimlik değişim (ID switching) problemini kapsamlı bir şekilde ele almamıştır. Bu nedenle, yöntemin

gerçek hayat problemlerinde pratik uygulanabilirliği sınırlı kalmıştır.

VISEM veri setini kullanan bir diğer çalışmada, Ottl ve arkadaşları, videolardan yer değiştirme özelliklerini ve hareket istatistiklerini çıkarmak için denetimsiz izleme algoritmalarını (Lucas-Kanade ve Crocker-Grier yöntemleri) kullanmıştır [63]. Bu özellikler, çeşitli sinir ağlarını ve destek vektör regresyonu (Support Vector Regression - SVR) modellerini eğitmek için kullanılmıştır. Çalışmada en iyi sonuç, her bir sperm hücresi için bireysel yer değiştirme özelliklerinin temsili ile eğitilen doğrusal bir SVR modeli ile elde edilmiş ve MAE değeri 7,31 olarak raporlanmıştır.

Benzer şekilde, Adinugroho ve Nakazawa, sperm hareketliliği ve anormal şekilleri tahmin etme problemlerini regresyon problemleri olarak ele alarak MotionFlow adında yeni bir hareket temsili tekniği önermişlerdir [64]. Tüm çerçeve için akışı hesaplayan geleneksel optik akış yöntemlerinin aksine MotionFlow, bölütlenmiş her bir sperm hücresi için optik akış vektörlerini hesaplamaktadır. Spermleri arka plandan ayırmak için VISEM-Tracking veri setinden 10.000 görüntü (29.196 görüntü arasından seçilen) kullanılarak semantik bir bölütleme yöntemi olan U-NET, 49 epoch eğitilmiştir. Ardından, ResNet3D-18 modeli, hareketliliği ve morfolojiyi tahmin etmek amacıyla U-Net ve MotionFlow yöntemleri ile elde edilen hareket vektörleri kullanılarak eğitilmiştir. Eğitim sonunda ResNet3D-18 modeli ile hareketlilik için 6,842 ve morfoloji için 4,148 MAE değerleri elde edilmiştir.

İlhan ve arkadaşları, sperm analizi için maliyet ve modülerlik açısından avantajlar sunan hibrit bir uzman sistem geliştirmişlerdir [65]. Mikroskoba entegre edilmiş akıllı telefon tabanlı veri toplama yaklaşımı kullanarak sperm analizini daha taşınabilir, modüler ve pratik hale getirmeyi amaçlamışlardır. Önerdikleri veri toplama yöntemini kullanarak 6 bireyden 18 farklı video elde etmiş ve toplam 86.438 sperm yörüngesi dört farklı kümeye ayrılmıştır. Videolar Speeded-Up Robust Features (SURF) tabanlı eşleştirme tekniği kullanılarak stabilize edilmiş ve izleme için Kalman filtreleme uygulanmıştır. Alternatif bir uzman sistemle (SQA-Vision) karşılaştırıldığında, önerilen yöntem ile Grade A, Grade B ve Grade C spermler için sırasıyla %2,1, %7,4 ve %5,3 hata oranları elde edilmiştir.

Zhu ve arkadaşları, sperm tespit performansını iyileştirmek amacıyla modifiye edilmiş bir YOLOv5s tabanlı model olan YOLOv5s-SA'yı geliştirmişlerdir [66]. Modelin başlangıcında bulunan omurga ağındaki bazı geleneksel evrişimsel katmanlar, derinlemesine ayrılabilir evrişimsel yapılarla değiştirilerek doğruluk korunurken parametre sayısı azaltılmıştır. Modelin tespit kabiliyetini geliştirmek için, özellik haritalarındaki konumsal bilgileri ve ayrıntılı özellikleri kullanmak üzere çok ölçekli bir füzyon modülü tasarlanmıştır. Son olarak, çıkış özellik

haritalarından önce boyun ağına, özellik haritası kanalları arasındaki korelasyonu güçlendiren ve ayrıntılı özelliklerin daha iyi füzyonunu sağlayan bir Uzamsal Dikkat (Spatial Attention - SA) mekanizması eklenmiştir. Bu çalışmada, 10 farklı hastadan toplanmış 10 farklı sperm videosu kullanılmıştır. Videoların uzunluğu 10 ila 130 saniye arasında değişmektedir. Videolar 1280×960 piksel çözünürlüğe sahiptir. İlk 6 hastadan alınan videoların ilk 100 karesi eğitim ve doğrulama setlerini oluşturmak için kullanılmıştır. Karelerin %80'i (480 kare) eğitim için, kalan %20'si (120 kare) ise doğrulama için kullanılmıştır. Diğer 4 hastanın her bir videosundan rastgele seçilen 30 kare test seti olarak kullanılmıştır. Önerilen YOLOv5s-SA modeli %87,5 ortalama doğruluk değeriyle YOLOv3, YOLOv3-spp, YOLOv5s ve YOLOv5m modellerine kıyasla sırasıyla %18,1, %15,2, %6,9 ve %1,9 daha iyi performans göstermiştir.

Tablo 1.3 Literatürde bulunan sperm tespit ve hareket analiz yöntemleri

Yayın	Veri seti	Görev(ler)	Problem	Yöntem	Performans
Dewan ve ark. (2018)[58]	Kendi Veri seti	Motilite Tahmini	Regresyon	Sürüklenme tespiti ve tahmini	Pearson Coefficient 0,818 (toplam motilite)
Thambawita ve ark. (2019)[59]	VISEM	Motilite Tahmini	Regresyon	Görüntü özellik çıkarımı için Autoencoder +ResNet-34 ve Motilite Tahmini	MAE (9,427)
Hicks ve ark. (2019) [60]	VISEM	Motilite Tahmini	Regresyon	OpticalFlow + ResNet50 görüntü özelliği çıkarma ve Motilite Tahmini	MAE (8,740)
Thambawita ve ark. (2019)[61]	VISEM	Motilite Tahmini	Regresyon	Stacking OpticalFlow + ResNet-34 + MLP ile özellik çıkarımı ve hareket tahmini	MAE (8,825)
Valiuvskaite ve ark. (2020) [62]	VISEM	Sperm tespiti	Tespit	Stacking OpticalFlow + Sperm kafa tahmini için Faster R-CNN	91,77 Acc.
Ottl ve ark. (2022) [63]	VISEM	Motilite Tahmini	Regresyon	Makine Öğrenmesi Algoritmaları için Lucas-Kanade ve Crocker-Grier algoritmaları	MAE (7,31)
Adinugroho ve ark. (2024) [64]	VISEM	Motilite Tahmini	Regresyon	Görüntü özellik çıkarımı için MotionFlow ve motilite tahmini için ResNet3D-18	MAE (6,842)
Ilhan ve ark (2021) [65]	Kendi Veri seti	Motilite Tahmini	İzleme	Rekürsif Kalman Filtresi ve Macar Algoritması ile motilite tahmini	Ort. Hata 4,93
Zhu ve ark. (2023) [66]	Kendi Veri seti	Sperm Tespiti	Tespit	Sperm tespiti için YOLOv5s-SA	87,5 mAP
Yuzkat ve ark. (2023) [39]	SeSVid	Sperm Tespiti	Tespit	Sperm tespiti için Model birleştirme (YOLOv5 + Mask RCNN + Detectron2)	85,2 mAP
Thambawita ve ark. (2023) [41]	VISEM-Tracking	Sperm Tespiti	Tespit Regresyon Takip	Sperm tespiti için YOLOv5	22,31 mAP
Zhang ve ark. (2024) [67]	VISEM-Tracking	Sperm Tespiti Sperm Takibi	Tespit Regresyon Takip	Sperm tespit ve takibi için YOLOv8E-TrackEVD	74.303 HOTA 71.167 MOTA
Tez kapsamı	Hi-LabSpermTracking	Sperm Tespit Sperm Takip	Tespit Takip	Sperm tespit, takip ve yörünge analizi için topluluk yöntemi (YOLOv8+ RT-DETR+ DeepSORT)	66.60 HOTA 76.42 MOTA

Yüzkat ve arkadaşları, derin öğrenme tabanlı nesne algılama algoritmalarını birleştirerek sperm algılama doğruluğunu ve tutarlılığını artırmayı amaçlamışlardır [39]. SeSVid veriseti kullanılarak derin öğrenme modelleri iki farklı senaryo altında test edilmiştir. İlk senaryoda, veri setindeki her bir video için farklı eğitim/test oranlarıyla nesne tespiti algoritmaları eğitilmiş ve modellerin bireysel performansları analiz edilmiştir. Bu senaryoda, model performanslarını değerlendirmek için hasta odaklı bir yaklaşım benimsenmiştir. İlk senaryoda en

düşük mean average Precision (mAP) değerlerine sahip videolar, ikinci senaryo için hedef videolar olarak seçilmiştir. İkinci senaryoda, düşük mAP değerlerine sahip zorlu videolardaki tespit performansını artırmak amacıyla, hasta odaklı bir yaklaşım olmaksızın, düşük mAP değerine sahip videolar dışındaki tüm diğer videolar birleştirilerek modeller eğitilmiştir. Bu yöntemle, farklı videolardan elde edilen görüntüler sayesinde daha gürbüz ve genelleştirme kapasitesi yüksek modeller eğitilmiştir. Ayrıca, yazarlar modellerin tespit performansını geliştirmek amacıyla topluluk öğrenme yöntemlerini kullanarak birleştirme işlemi gerçekleştirmiştir. Bu yaklaşım sonucunda, test setindeki iki videoda sırasıyla %86,1 ve %84,3 mAP skorları elde edilmiştir. Sadece YOLOv5 kullanılarak elde edilen sonuçlarla karşılaştırıldığında, önerilen yöntemin sırasıyla %1,5 ve %0,8 oranında daha yüksek mAP değerleri sağladığı görülmüştür.

Zhang ve arkadaşları, sperm tespiti ve takibi için YOLOv8E-TrackEVD modelini geliştirmişlerdir. Model eğitiminde VISEM-Tracking veri kümesinin artırılmış versiyonu, MHSMA veri kümesiyle birleştirerek kullanılmıştır [67]. Orijinal VISEM-Tracking veri seti, 29.196 görüntüden oluşmakta olup, 19.530 eğitim, 5.580 test ve 2.790 doğrulama görüntüsüne ayrılmıştır. Bu kümeler, MHSMA veri kümesi kullanılarak önemli ölçüde veri artırımı yapılmış ve 459.291 eğitim, 128.707 test ve 65.110 doğrulama görüntüsü elde edilmiştir. Veri artırımı yalnızca eğitim setine değil, aynı zamanda doğrulama ve test setlerine de uygulanmıştır; bu durum, yanlılığa yol açabilecek ve yapay olarak yüksek performans tahminlerine neden olabilecek bir yaklaşım olarak değerlendirilmektedir. Önerilen yöntemle VISEM-Tracking veri setinde sırasıyla %74,303 HOTA ve %71,167 MOTA değerleri elde edilmesine rağmen, bu metrikleri elde etmek için hangi videoların kullanıldığı çalışmada belirtilmemiştir. Eğitim aşamasında tüm VISEM-Tracking videoları kullanıldığı için bağımsız bir test seti bulunmamaktadır. Bu durum, bildirilen HOTA ve MOTA değerlerinin gerçek dünya performansını doğru şekilde yansıtmadığına işaret etmekte olup, gerçek bir test setinin olmaması, aşırı uyum (overfitting) ve yanıltıcı sonuçlar konusunda ciddi endişelere neden olmaktadır.

Bu başlık altında, mevcut sperm hareketliliği çalışmalarına ve bunların analiz sistemlerine dair kapsamlı bir genel bakış sunulmuş ve alandaki eksiklikler vurgulanmıştır. Sperm tespit ve takip yöntemlerinde yapılan incelemeler, net bir karşılaştırma ve ayrıntılı bir perspektif sunmak amacıyla Tablo 1.3'de özetlenmiştir.

1.2 Tezin Amacı

Erkek faktörlü kısırlık tedavilerinde problemin tespiti için uygulanan temel prosedürlerden biri spermiogram testidir. Spermiogram testi, erkek bireylerden alınan semen örneklerinin, belirli kriterler doğrultusunda laboratuvar ortamında uzman kişiler tarafından değerlendirilmesi işlemidir. Bu test kapsamında alınan semen örneklerinin pH, viskozite ve renk gibi fiziksel özelliklerinin yanı sıra, mikroskop altında sperm hücrelerinin şekilsel ve hareket analizleri de gerçekleştirilmektedir. Mikroskop altında gerçekleştirilen analizlerde, GD ve BDSA olmak üzere iki temel değerlendirme yöntemi kullanılmaktadır. GD yönteminde, sperm hücreleri mikroskop altında manuel olarak bir laboratuvar asistanı, biyolog veya medikal doktor tarafından incelenmektedir. Bu yöntem, pratik ve sezgisel bir yaklaşım sunmasına rağmen, değerlendirme sonuçları büyük ölçüde incelemeyi yapan kişinin deneyimine ve uzmanlık düzeyine bağlı olarak değişkenlik gösterebilmektedir. GD yöntemine alternatif olarak geliştirilen BDSA sistemleri, daha objektif ve tutarlı sonuçlar sunmaktadır. Ancak, bu sistemler yüksek maliyetli olup, genellikle sperm hareket analizlerinde kullanılmaktadır. Ayrıca, doğru kullanım için eğitimli personel gerekliliği bulunmaktadır. Bu nedenle, tamamen insandan bağımsız bir sistem olarak nitelendirmek uygun değildir.

Son yıllarda, BDSA ve GD sperm analiz sistemlerine alternatif olarak, hem sperm morfoloji analizinde hem de sperm tespiti ve takibi alanlarında derin öğrenme yöntemleri kullanılarak pek çok çalışma geliştirilmiştir. Literatür taraması kısmında, bu alanlarda sistem geliştirilmesine yönelik oluşturulmuş açık kaynaklı verisetleri, sperm morfoloji ve motilite analiz sistemleri ayrıntılı bir şekilde sunulmuştur. Mevcut veri setlerinin genellikle sperm morfoloji alanında sınıflandırma problemleri için tasarlandığı gözlemlenmektedir. Ancak bu veri setlerine bakıldığında da, sınırlı sayıda sınıf, küçük örneklem boyutları, sınıflar arasındaki dengesizlik ve özellikle sperm başı anormalliklerine odaklanılması gibi sınırlamaların, kapsamlı analiz sistemlerinin geliştirilmesini engellediği görülmektedir. Öte yandan, sperm tespiti ve takibi ile ilgili oluşturulmuş veri setleri incelendiğinde, açık kaynak olarak sunulmuş veri setlerinin sayısının oldukça sınırlı olduğu gözlemlenmektedir. Bu veri setleri incelendiğinde, VISEM-Tracking veriseti dışında, diğer tüm veri setlerinde yalnızca spermlere ait sınırlayıcı kutuların etiketlendiği görülmektedir. VISEM-Tracking veri setini oluşturan videolar incelendiğinde ise, bu videoların birden fazla video kaydının kesilip birleştirilmesiyle oluşturulduğu ve sperm takibi ile sperm yörüngelerinin çıkarılması konusunda yetersiz kaldığı görülmektedir.

Bu tez çalışmasında, literatürdeki eksiklikler dikkate alınarak, sperm morfoloji sınıflandırma alanında kullanılmak üzere Hi-LabSpermMorpho ve sperm tespiti ile takibi alanında kullanılmak üzere Hi-LabSpermTracking veri setleri oluşturulmuştur. Özellikle sperm tespiti ve takibi problemlerinde, CNN tabanlı sperm tespit yöntemleri ile Transformer tabanlı sperm tespit yöntemleri, tez kapsamında önerilen çeşitli topluluk öğrenme yaklaşımları ile birleştirilmiş ve bu birleşimin spermlere ait yörüngelerin çıkarılmasındaki etkileri ayrıntılı bir şekilde analiz edilmiştir. Ayrıca, literatürde yaygın olarak kullanılan veri setlerinin yanı sıra, tez kapsamında oluşturulan verisetleri de kullanılarak CNN tabanlı ve Transformer tabanlı mimariler hem sperm morfoloji hem de sperm tespit ve takip alanlarında uygulanmış; bu doğrultuda, GD ve BDSA sistemlerine alternatif olabilecek çözümler geliştirilmesi amaçlanmıştır.

1.3 Hipotez

Bu tez çalışmasında, GD ve BDSA sistemlerine alternatif olarak, daha ucuz, güvenilir, tarafsız ve daha doğru sonuçlar sağlayabilen derin öğrenme tabanlı sperm morfoloji ve hareket analiz sistemlerinin geliştirilebileceği hipotez olarak ortaya konulmuştur. Tez kapsamında sunulan hipotezler, aşağıda detaylı bir şekilde listelenmiştir.

- GD ve BDSA sistemlerine alternatif olarak, cep telefonu tabanlı görüntüleme yöntemleri kullanılarak sperm tespiti ve takibi için derin öğrenme tabanlı sistemler geliştirilebilir.
- Literatürde yaygın olarak kullanılan sperm morfoloji veri setleri, CNN ve Transformer tabanlı sınıflandırıcı modeller ile eğitilerek, sperm morfoloji alanında hangi modellerin daha başarılı performans sergilediği detaylı bir hiperparametre konfigürasyonu ve karşılaştırmalı analiz ile değerlendirilebilir.
- Veri setlerindeki sınıf dengesizliği ve veri yetersizliği problemlerinin model performansına etkisini değerlendirmek amacıyla, farklı veri artırma tekniklerinin ve ölçeklerinin model sonuçlarına olan etkileri karşılaştırmalı bir şekilde analiz edilebilir.
- Geleneksel makine öğrenmesi ve CNN tabanlı mimarilerin yerine, Transformer tabanlı mimariler kullanılarak daha başarılı ve etkili sperm morfoloji analiz sistemleri geliştirilebilir.

- Literatürde mevcut sperm morfoloji analiz veri setleri, hem nitelik hem de nicelik açısından önemli ölçüde yetersizdir. Bu eksikliklere alternatif olarak, sperm başı, boynu ve kuyruğundaki anormallikleri kapsayan daha kapsamlı ve detaylı bir veri seti oluşturulabilir.

- Sperm morfolojik yapılarının mikroskop altında detaylı bir şekilde incelenmesi sırasında farklı boyama maddeleri kullanılmaktadır. Bu boyama yöntemleriyle elde edilen görüntüler, derin öğrenme yöntemleri kullanılarak morfoloji analizi açısından değerlendirilerek, hangi boyama yönteminin daha etkili olduğu sınıf bazında istatistiksel yöntemlerle analiz edilebilir.

- Derin öğrenme yöntemleriyle hem sperm tespiti hem de takibini gerçekleştirme imkanı sağlayan yeni ve kapsamlı bir veri seti oluşturulabilir. Bu çalışma, sperm hareket analizi literatürüne özgün ve benzersiz bir katkı sunma potansiyeline sahiptir.

- CNN tabanlı sperm tespit yöntemleri ile Transformer tabanlı sperm tespit yöntemleri, çeşitli topluluk öğrenmesi yaklaşımlarıyla birleştirilerek, daha gürbüz ve daha yüksek doğruluk oranına sahip sonuçlar üretebilen yeni bir çerçeve tasarlanabilir.

- Sperm tespit yöntemleriyle tespit edilen spermler, tüm video boyunca kimliklendirilerek spermlere ait yörüngeler elde edilebilir. Bu yörüngeler üzerinden, spermlerin hareket türleri tamamen otomatik bir şekilde sınıflandırılabilir.

- Tez kapsamında oluşturulan veri setleri, açık kaynak olarak sunularak dünya çapında araştırmacılara sperm morfoloji ve motilite alanında daha kapsamlı ve etkili sperm analiz sistemleri geliştirme imkanı tanınabilir.

2
MATERYALLER

Tez çalışması kapsamında, spermlerin hem şekilsel yapıları hem de hareket yörüngeleri detaylı bir şekilde incelenmiştir. Literatür taramaları sonucunda, sperm şekil ve hareket analizi ile ilgili çeşitli açık kaynak veri setlerinin bulunduğu görülmüştür. Ancak, sperm hareketiyle ilgili veri setlerinde genellikle yalnızca sperm tespitine yönelik sınırlayıcı kutularla etiketleme yapıldığı gözlemlenmiştir. Sperm morfoloji alanında, sperm hareketliliğine kıyasla daha fazla veri seti bulunsa da, belirli kısıtlamalar mevcuttur. Bu bağlamda, tez kapsamında hem sperm morfoloji hem de hareketi alanlarında daha kapsamlı veri setleri oluşturulmuş ve literatüre önemli katkılar sağlanmıştır. Bu doğrultuda, Hi-LabSpermMorpho ve Hi-LabSpermTracking adında iki yeni veri seti oluşturulmuştur. Veri setlerinin oluşturulmasından önce, literatürde yaygın olarak kullanılan veri setleri çeşitli derin öğrenme yöntemleriyle değerlendirilmiş ve bu analizler sonucunda alanın ihtiyaç duyduğu eksiklikler tespit edilmiştir. Yeni veri setleri, bu eksiklikler göz önünde bulundurularak oluşturulmuş ve literatürdeki mevcut sınırlamaların aşılması amaçlanmıştır.

Bu başlık altında, tez kapsamında kullanılan açık kaynaklı veri setlerinin yanı sıra oluşturulan veri setleri detaylı bir şekilde açıklanmaktadır. Veri setlerinin oluşturulmasında kullanılan numune alma, inceleme ve veri toplama teknikleri ayrıntılı bir biçimde sunularak, elde edilen veri setlerinin bilimsel geçerliliği ve kapsamı vurgulanmaktadır.

2.1 Sperm Morfoloji Analizi Veri Setleri

Tez kapsamında, sperm morfoloji analizi için literatürde yaygın olarak kullanılan SMIDS [37], HuSHeM [7] ve SCIAN-Morpho [29] veri setleri tercih edilmiştir. Bu veri setleri, literatürdeki birçok çalışmada sunulan yöntemlerin karşılaştırılması için referans olarak kullanılan güvenilir ve yaygın şekilde kabul görmüş kaynaklar olmaları nedeniyle tercih edilmiştir.

Bunlara ek olarak, tez çalışması kapsamında özel olarak oluşturulan Hi-LabSpermMorpho veri seti, sperm morfoloji analizi çalışmalarında kullanılmıştır. Hi-LabSpermMorpho veri seti ile literatürdeki mevcut veri setlerinin eksiklikleri giderilmiş ve sperm morfoloji analizi alanında daha kapsamlı çalışmalar yapılmasına olanak tanınmıştır. Bu doğrultuda, hem literatürdeki mevcut veri setlerinin değerlendirilmesi hem de yeni oluşturulan veri setinin performans ve kapsam açısından katkıları detaylı bir şekilde incelenmiştir.

2.1.1 Açık Erişim Veri Setleri

2.1.1.1 HuSHeM

Shaker ve arkadaşları ilk olarak HuSHeM'i insan sperminin baş kısımlarını sınıflandırmak için yaptıkları çalışmada tanıtmışlardır [7]. Veri kümesi toplam 216 görüntüden oluşmaktadır ve 4 sınıf içermektedir: normal (54), pyriform (57), tapered (53) ve amorphous (52). Görüntüler Olympus BX50 mikroskobuna entegre edilmiş bir kamera sistemi kullanılarak elde edilmiştir.Uzmanların sperm görüntülerini daha kolay analiz edebilmesi için, görüntüler elde edilmeden önce Diff-Quik boyama yöntemi kullanılarak sperm hücrelerinin morfolojik özellikleri belirginleştirilmiştir. Mikroskop yakınlaştırma ölçeği kayıt sırasında 20×'e ayarlanmıştır. Veri setindeki ham görüntüler, karışık kuyruklar, birden fazla sperm başı, gürültü ve bozulmuş sperm şekilleri gibi unsurlar içermektedir. Bu tür görüntüler, sınıflandırma başarısını olumsuz yönde etkileyebileceğinden, araştırmacılar tarafından manuel ön işleme süreci ile düzenlenmiştir. Ön işleme aşamasında, istenmeyen öğelerin ortadan kaldırılması için sperm gövdeleri kırpılarak yalnızca sperm başlarına odaklanılmış ve ardından tüm sperm başları, analiz için uygun hale gelmesi amacıyla sola döndürülmüştür. Bu işlemlerden sonra, görüntülerden elde edilen özellikler daha belirgin ve anlamlı hale getirilmiş, böylece sınıflandırma ve analiz süreçleri için daha uygun bir veri seti oluşturulmuştur [7, 25]. Literatürde bulunan birçok çalışma, bu tür manuel ön işleme adımları uygulayarak daha yüksek performanslı modeller geliştirmeyi hedeflemiştir [7, 25, 55]. Ancak, veri setinin bu şekilde manipüle edilmesi, tamamen otomatik bir sperm morfoloji analiz sistemi tasarımını olumsuz yönde etkilemektedir. Bu nedenle, tez kapsamında HuSHeM veri seti ham haliyle kullanılmış ve herhangi bir manuel ön işlem uygulanmamıştır.

2.1.1.2 SMIDS

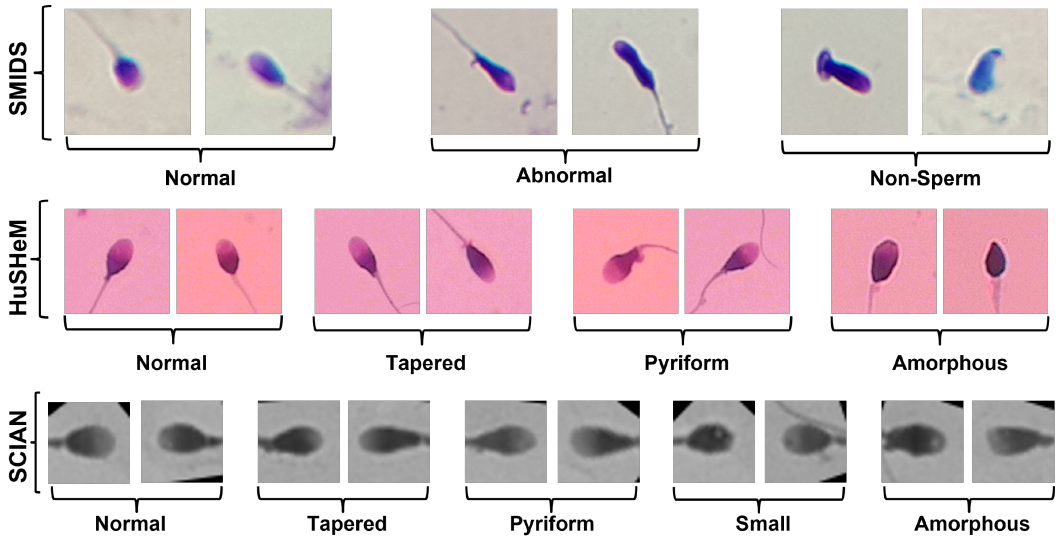
Halka açık bir başka veri seti olan SMIDS, ilk olarak İlhan ve arkadaşları tarafından [68] çalışmasında akıllı telefon tabanlı bir veri toplama yöntemi

kullanılarak oluşturulmuştur. SMIDS'in hazırlanması sırasında ilk olarak, akıllı telefon tabanlı veri toplama yaklaşımı kullanılarak daha az sayıda sperm örneği içeren küçük bir versiyon elde edilmiştir. Daha sonra, SMIDS'in tam sürümü, [37]'de tanıtılan sperm morfoloji analiz sisteminin tasarımında kullanılmak üzere oluşturulmuştur. Bu sistemde, uzmanların görüntüleri etiketlemesi için ham sperm görüntülerinden İlgi Bölgesi (Region of Interest - ROI) çıkarılmıştır. Etiketleme sürecinin ardından, sperm olmayan (974), anormal (1005) ve normal (1021) olmak üzere üç sınıfa ayrılmış 3000 görüntüden oluşan SMIDS'in tam sürümü oluşturulmuştur. HuSHeM'e benzer şekilde, SMIDS de 190×170 piksel çözünürlüğünde, çeşitli boyutlardaki RGB formatında görüntüler içermektedir. SMIDS veri setinde gürültü, karışık kuyruk ve çoklu sperm başı görüntüleri de bulunmaktadır. Veri setinin oluşturulma aşamasında etik kurallar titizlikle gözetilmiştir. İstanbul Üniversitesi Tıp Fakültesinde gerçekleştirilen çalışma, İstanbul Üniversitesi Etik Kurulu onayıyla yürütülmüştür. Ayrıca, veri toplama sürecinde örnek veren bireylerden bilgilendirilmiş onam alınmıştır. Standart bir prosedürün uygulanabilmesi amacıyla, bireylere örnek alınmadan önce 2 ila 7 gün arasında cinsel perhiz yapmaları ve örneklerini sabah saat 9 ile 11 arasında vermeleri gerektiği konusunda bilgilendirme yapılmıştır. Bu yaklaşımla, veri setinin kalitesi ve tutarlılığı artırılarak daha güvenilir sonuçlar elde edilmesi amaçlanmıştır.

2.1.1.3 SCIAN-Morpho

V. Chang ve arkadaşları [29] tarafından oluşturulan veri seti, toplamda 1132 adet sperm görüntüsünden oluşmaktadır. Veri setinde, normal (100), tapered (228), pyriform (76), amorphous (656) ve small (72) olmak üzere sağlıklı ve 4 farklı morfolojik bozukluk içeren toplam 5 sınıf bulunmaktadır. Görüntüler gri ölçekli formatta ve 35×35 piksel çözünürlüğünde kaydedilmiştir. Ancak veri setindeki örnek-sınıf dağılımının dengesiz olması ve görüntü çözünürlüğünün oldukça düşük seviyede olması, literatürde bu veri seti üzerinde yapılan çalışmalarda düşük başarımların elde edilmesine yol açmıştır [27, 37, 55].

Tez kapsamında kullanılan üç farklı verisetine ait örnek görüntüler Şekil 2.1'de sunulmuştur. Bu görüntüler, veri setlerinin yapısını ve içerdikleri örneklerin çeşitliliğini görsel olarak anlamak için önemli bir referans sağlamaktadır. Ayrıca, her bir veri setindeki örnek dağılımları ve toplam örnek sayıları ise Tablo 2.1'de detaylı bir şekilde sunulmuştur.



Şekil 2.1 SMIDS, HuSHeM ve SCIAN-Morpho veri setleri için örnek görüntüler

Tablo 2.1 SMIDS, HuSHeM ve SCIAN-Morpho veri setlerinin sınıflara göre dağılımları

SMIDS		**HuSHeM**		**SCIAN-Moprho**	
Etiket	Görüntü Sayısı	Etiket	Görüntü Sayısı	Etiket	Görüntü Sayısı
Normal	1021	Normal	54	Normal	100
Abnormal	1005	Tapered	53	Tapered	228
NonSperm	974	Pyriform	57	Pyriform	76
		Amorphous	52	Small	72
				Amorphous	656
Toplam	3000	Toplam	216	Toplam	1132

2.1.2 Özgün Veri Seti: Hi-LabSpermMorpho

Sperm morfoloji alanındaki en büyük zorluklardan biri, kamuya açık veri setlerinin çok fazla olmamasıdır. Sperm morfoloji görüntülerini elde etmek mikroskobik ölçeklerde çalışmayı gerektirmektedir. Sperm hücreleri ışık mikroskobu altında şeffaf olduklarından, şekillerini ayrıntılı olarak görselleştirmek için çeşitli ön hazırlık işlemleri gerekmektedir. Bu zahmetli hazırlık adımları nedeniyle, baş, boyun, orta parça, kuyruk ve sitoplazma anormallikleri gibi detaylarla etiketlenmiş ve farklı boyama yöntemleri ile boyanmış bir veri kümesi oluşturmak oldukça zordur. Bu tez kapsamında, otomatik sınıflandırma sistemlerinin geliştirilmesinde kullanılmak üzere oluşturulmuş olan Hi-LabSpermMorpho [42] sperm morfoloji veri seti tanıtılmaktadır. Aşağıdaki bölümlerde, numune hazırlama yöntemleri, görüntü alma ekipmanları, ön işleme ve etiketleme süreçleri, sperm numune kaynakları ve veri setinin ayrıntılı analizi ele alınmaktadır.

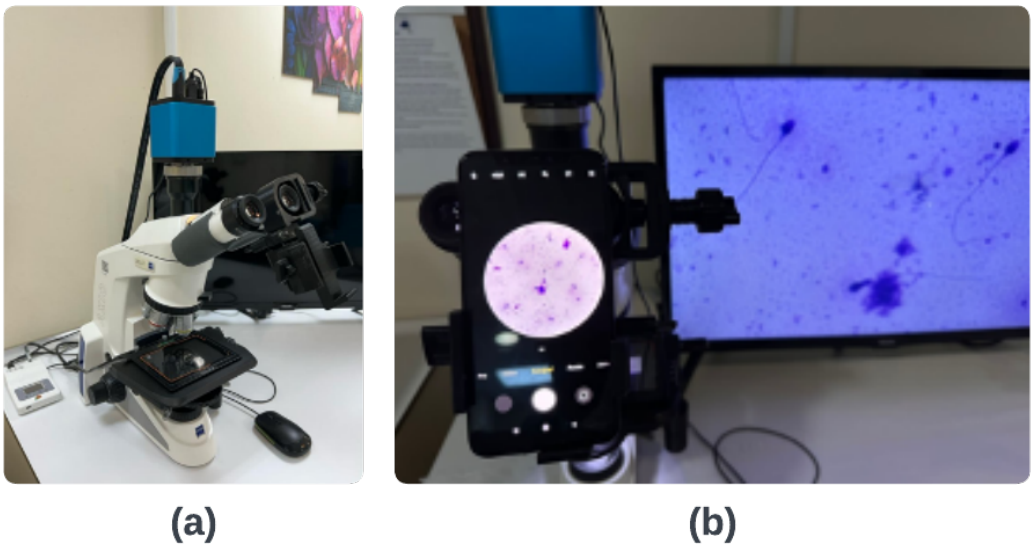
2.1.2.1 Örnek Hazırlama Prosedürleri

Hi-LabSpermMorpho veri setini oluşturmak için 2021 yılında DSÖ tarafından yayımlanan rehberler ve öneriler takip edilmiştir [2]. Sperm morfolojisinin detaylı bir şekilde incelenebilmesi için fiksasyon ve boyama işlemleri gereklidir; bu işlemler, anormalliklerin vurgulanması açısından büyük önem taşımaktadır. Aksoy ve arkadaşlarının çalışmasında da değinildiği gibi, farklı boyama yöntemleri spermatozoa morfolojisini önemli ölçüde etkileyebilmektedir [22]. Bununla birlikte, mevcut veri setlerinin hiçbirinde farklı boyama teknikleri kullanılarak elde edilmiş ve hassas bir şekilde etiketlenmiş sperm morfoloji görüntüleri bulunmamaktadır. Bu nedenle, tez kapsamında farklı boyama yöntemlerinin sperm şekilleri üzerindeki etkilerini incelemek amacıyla üç farklı Diff-Quick boyama seti—BesLab, Histoplus ve GBL— kullanılarak görüntüler elde edilmiştir. Bu boyama setleri, metanol içinde çözünmüş triarylmethane içeren bir fiksatif çözeltisi, eozinofilik xanten içeren boyama çözeltisi 1 (reaktif 1) ve bazofilik tiazini içeren boyama çözeltisi 2 (reaktif 2) içermektedir. Boyama çözeltileri benzer içeriklere sahip olmakla birlikte, dozajlarındaki farklılıklar nedeniyle renk yoğunluklarında değişiklikler meydana gelmektedir.

Örnek hazırlama sırasında, hava ile kurutulmuş sperm yayması, Diff-Quick fiksatif çözelti kitine daldırılmıştır. Bu işlem, her saniye daldırma ve geri çekme şeklinde toplam 10 saniye boyunca tekrarlanmıştır. Ardından lamlar, fazla çözeltinin süzülmesi için emici filtre kağıdı üzerine dikey olarak yerleştirilmiştir. Fiksatif işleminden sonra, sperm yayması hızlı boyama çözeltisi 1'e (reaktif 1) 8 saniye boyunca daldırılmış, aynı daldırma ve çıkarma işlemi her saniyede uygulanmıştır. Bu işlem, hızlı boyama çözeltisi 2 (reaktif 2) için de tekrarlanmıştır. Her boyama adımının ardından, lamlar fazla boyanın süzülmesi için emici filtre kağıdı üzerine dikey olarak yerleştirilmiştir. Son olarak, fazla boyayı temizlemek amacıyla lamlar 1 dakika boyunca musluk suyu altında yıkanmış ve kurumaya bırakılmıştır. Kuruduktan sonra, gözlem için daha net görüntüler elde edebilmek amacıyla lamların üzerine immersion yağı uygulanmıştır.

2.1.2.2 Görüntülerin Elde Edilmesi

Sperm morfolojisinin ayrıntılı görüntülerini elde etmek için yaygın olarak kullanılan parlak alan mikroskop ekipmanının bir uzantısı olan ZEISS AXIO LAB A1 ve Olympus BX43 mikroskopları görüntü elde etme sürecinde kullanılmıştır. Spermler parlak alan optikleri kullanılarak 100× büyütmede görüntülenmiştir. Görüntü alma işlemi, numune hazırlandıktan sonra mikroskobun okülerine entegre edilmiş özel tasarım bir aparat ile cep telefonu monte edilerek gerçekleştirilmiştir.

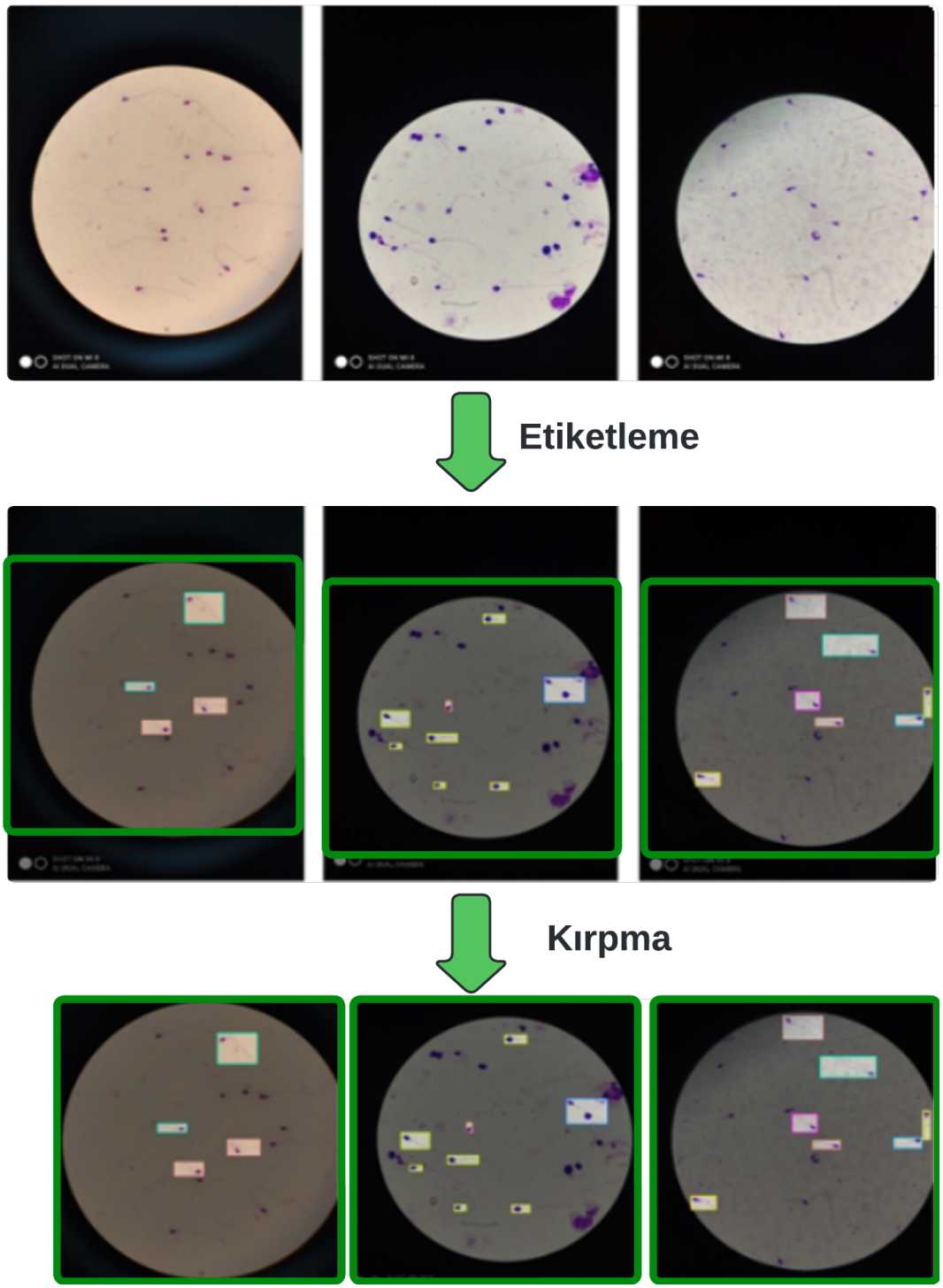


Şekil 2.2 (a) Özel tasarım aparat ve (b) mikroskop okülerine entegre edilmiş aparata takılı cep telefonu

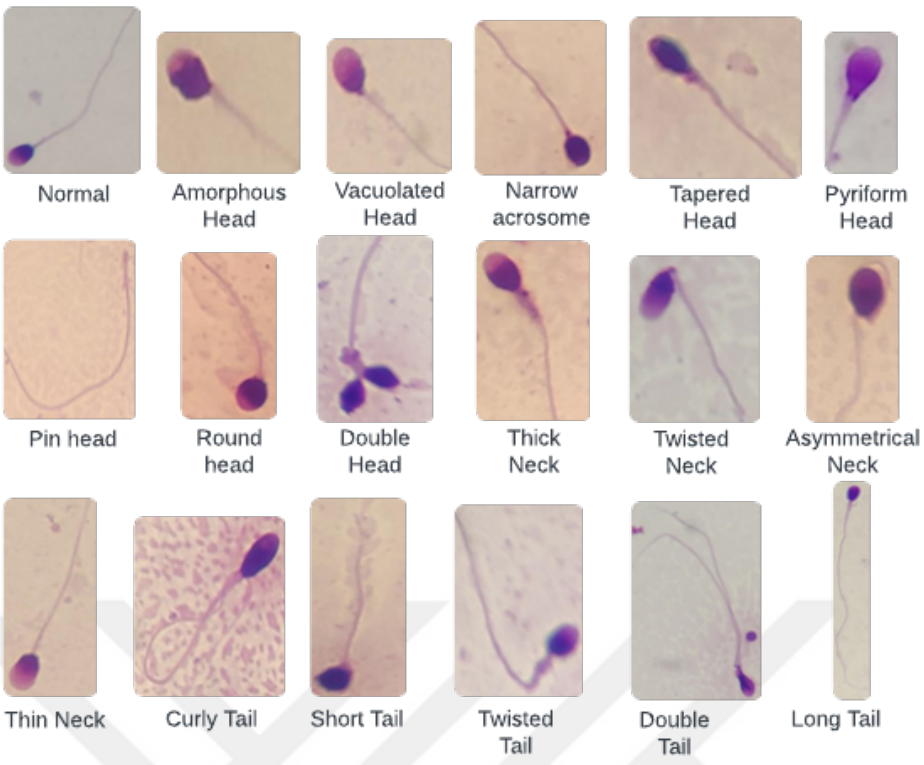
Bu veri toplama yaklaşımı Şekil 2.2'de gösterilmektedir. Veri toplama süreci, BesLab, Histoplus ve GBL boyama yöntemleri için ayrı ayrı tekrarlanmış ve elde edilen ham veriler, uzak bir sunucuda DB1 adıyla depolanmıştır. Görüntüler, hasta tanımlayıcısı olmadan, verilerin anonimleştirilmesi sağlanarak sıralı bir kimlikle uzak sunucuya kaydedilmiştir. DB1 veri kümesindeki ham görüntüler RGB renk uzayında 3024×4032 çözünürlüğe sahiptir. DB1 veri kümesi, 1001 hastanın (379 BesLab, 321 Histoplus, 301 GBL) boyanmış sperm örneklerinin farklı bölgelerinden elde edilen toplam 14.916 görüntüden (5995 BesLab, 4852 Histoplus, 4096 GBL) oluşmaktadır. DB1 veri kümesine ait örnek görüntüler, Şekil 2.3'nin üst kısmında gösterilmektedir.

2.1.2.3 Ön İşleme Adımları ve Etiketleme Prosedürü

DB1 veri seti, 14916 bölgesel sperm gözlem görüntüsü içermektedir ve uzmanlar tarafından DSÖ standartlarına [2] uygun olarak etiketlenmiştir. Ancak, veri toplama sırasında cep telefonu kullanıldığından, kayıtlar sırasında Şekil 2.3'ün üst kısmında gösterildiği gibi yuvarlak bir bölge görünmektedir ve bu bölgenin konumu telefonun açısına bağlı olarak değişebilmektedir. Bu sorunu gidermek için, etiketli görüntülerin siyah bölgeleri kırpılarak yalnızca oküler bölgenin görüntüleri korunmuş ve bu işlenmiş görüntüler DB2 olarak bulut sunucusunda depolanmıştır. Kırpma ve hizalama işlemi sırasında, oküler görüntü alanını tespit etmek için basit bir histogram tabanlı eşikleme kullanılmıştır. Ardından, 1440×1440 piksel pencere, bu bölgeyi kapsayacak şekilde merkezlenmiş ve tüm görüntülerde bu alanın dışındaki kısımlar kaldırılmıştır. Bu ön işleme, DB2'deki etiketli görüntülerin



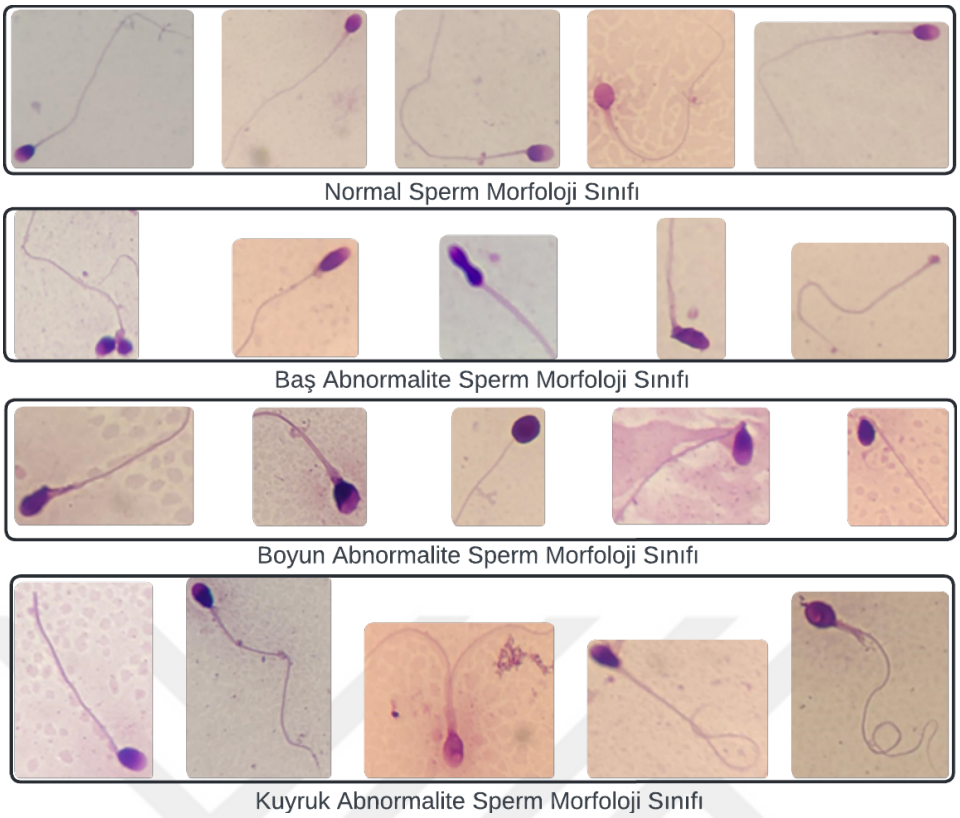
Şekil 2.3 DB1 veri setinden DB2 veri setini oluşturma adımları



Şekil 2.4 BesLab boyama için elde edilen DB3_v1 (18 sınıf) veri kümesinden etiketlenmiş sınıflara ait örnek görüntüler

boyutlarının, Şekil 2.3'ün alt kısmında gösterildiği gibi 1440×1440 piksel olarak sabitlenmesini sağlamaktadır. Şekil 2.3, üç boyama yönteminden elde edilen örnek görüntüler üzerinde yapılan ön işleme ve etiketleme adımlarını göstermektedir.

Veri seti oluşturma sürecinin son aşamasında, literatürdeki mevcut sperm morfoloji veri setleri, HuSHeM [7], SMIDS [37], SCIAN-MorphoSpermGS [69] ve MHSMA [34] gibi veri setleri ayrıntılı bir şekilde incelenmiştir. DB3 veri seti, bu veri setlerinin yapısına benzerlik gösterirken, aynı zamanda mevcut veri setlerinin karşılaştığı sınırlamaları ortadan kaldıracak şekilde oluşturulmuştur. DB3 veri seti, iki uzmanın belirlediği etiket bilgilerine göre kategorize edilmiş bireysel sperm hücrelerini (sperm yamaları) içermektedir. Sperm yamaları, DB2 veri setindeki gözlem görüntülerinden, iki uzmanın sağladığı etiket koordinatlarına dayanarak çıkarılmıştır. DB3 veri seti, GBL, Histoplus ve BesLab boyama kitlerinin her biri için oluşturulmuştur ve iki sürüm halinde düzenlenmiştir. DB3_v1 veri kümesi, 18 detaylı morfolojik anormallik sınıfı içerirken, DB3_v2 veri kümesi, tanımlanan tüm morfolojik anormallikleri baş, boyun ve kuyruk ana kategorileri altında birleştirerek, normal örnekleri de dahil eden 4 sınıflı bir yapı olarak sunulmuştur. DB3_v1 ve DB3_v2 veri setlerinden örnek görüntüler sırasıyla Şekil 2.4 ve Şekil 2.5'de sunulmaktadır.

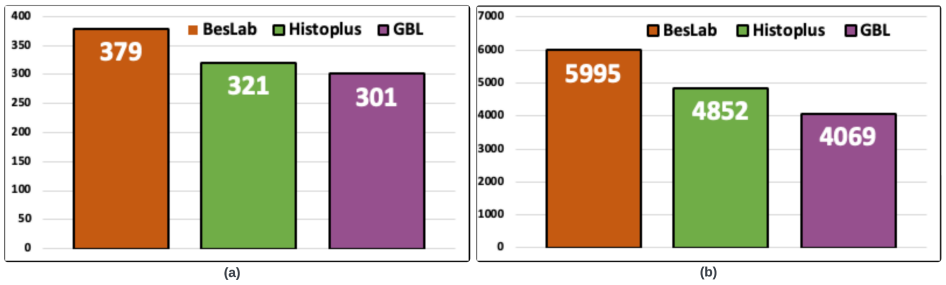


Şekil 2.5 BesLab boyama için elde edilen DB3_v2 (4 sınıf) veri kümesinden etiketlenmiş sınıflara ait örnek görüntüler

2.1.2.4 Hastalardan Semen Örneği Toplama Prosedürü

Sperm örnekleri, Rize'de bulunan Recep Tayyip Erdoğan Üniversitesi Eğitim ve Araştırma Hastanesi Androloji Laboratuvarı'na kısırlık şikayetleri ile başvuran ve belirli kriterleri karşılayan hastalardan toplanmıştır. Örnekler alınmadan önce her katılımcıya ayrıntılı bilgi verilerek, bilgilendirilmiş onam veren gönüllülerden örnekler alınmıştır. Katılım kriterleri aşağıdaki gibidir:

- Katılımcıların 20-40 yaş aralığında olması gerekmektedir.
- En az bir yıl düzenli cinsel ilişkiye rağmen çocuk sahibi olamama.
- En az bir yıldır kontraseptif kullanılmamış olması.
- Kısırlık şikayetleri dışında ürogenital sistem hastalığı olmaması.
- Kısırlık dışında sistemik hastalık olmaması.
- Kısırlık nedeniyle daha önce herhangi bir cerrahi operasyon geçirmemiş olmak.
- Daha önce kısırlık için herhangi bir tıbbi tedavi almamış olmak.



Şekil 2.6 Boyama kitlerine (BesLab, Histoplus ve GBL) dayalı dağılımlar (a) çalışmaya dahil edilen hasta sayısı ve (b) oküler görüntü sayısı.

- Sistemik hastalıklar nedeniyle uzun süreli ilaç kullanmamış olmak.

Tüm katılımcılar için standart bir prosedür olarak, semen örneği üç günlük cinsel perhiz dönemini takip eden dördüncü günün sabahında, saat 9-11 arasında toplanmıştır. Toplanan örnekler, sıcaklık dalgalanmalarından kaynaklanan olumsuz etkileri önlemek amacıyla 37 santigrat derecede inkübatörde tutulmuştur.

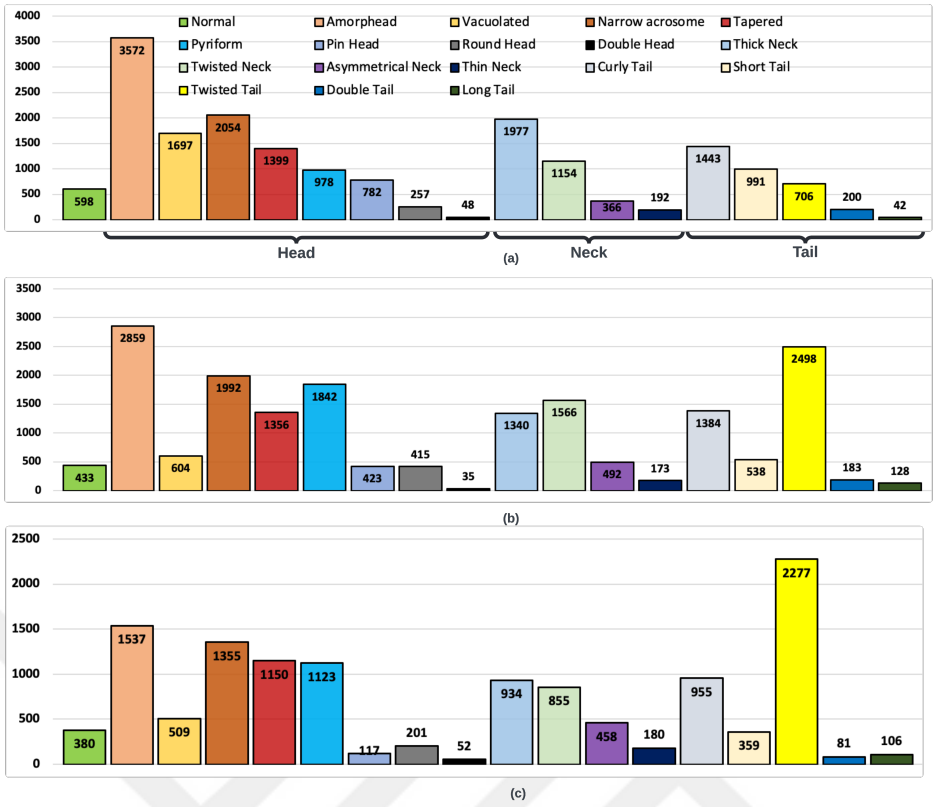
2.1.2.5 Hi-LabSpermMorpho Veri Setinin Yapısı

Yukarıdaki teknikler kullanılarak, üç farklı boyama kiti ile toplam 1001 hastadan sperm görüntüsü elde edilmiştir. Şekil 2.6.a, dahil edilen hastaların dağılımını, Şekil 2.6.b ise bu hastalardan toplanan çeşitli semen bölgelerinin oküler görüntülerinin sayısını boyama materyallerine göre göstermektedir.

DSÖ tarafından tanımlanan neredeyse tüm sperm morfoloji anormallik sınıflarını içeren DB3_v1 veri kümesinin 18 sınıf için kırpılmış sperm görüntülerinin dağılımı, Şekil 2.7'de gösterilmektedir. Üç farklı boyama kiti kullanılarak elde edilen görüntüler 18 sınıfta etiketlenerek üç ayrı DB3_v1 veri kümesi elde edilmiştir: DB3_v1_GBL, DB3_v1_Histoplus ve DB3_v1_BesLab.

Örnek toplama sırasında, hastalar arasındaki farklılıklar nedeniyle belirli bir morfolojik anormallik tipine ait sperm yamalarının sayısında değişiklikler gözlenmiştir. Bu durum, sperm yamalarının anormallik sınıfları arasında dengesiz dağılımına yol açmaktadır. Baş anormallikleri arasında en sık gözlenen tip amorphous olmuştur. Boyun anormalliklerinde ise Thick ve Twisted Neck'lere sıkça rastlanmıştır. Kuyruk anormalliklerinde ise Twisted Tail önemli bir çoğunluğu oluşturmuştur. Uygulanan boyama kitlerine göre DB3_v1 veri setindeki toplam örnek sayısı Tablo 2.2'de sunulmuştur.

Her ne kadar 18 sınıflı DB3_v1 veri kümesi detaylı morfolojik analizlerin önünü açsa da, anormallik türlerindeki yüksek çeşitlilik, derin öğrenme modellerinin



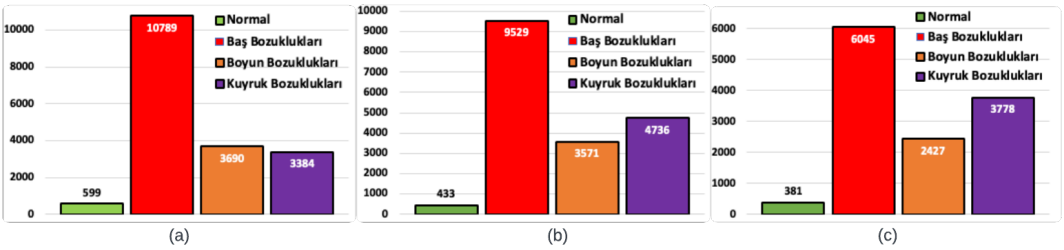
Şekil 2.7 DB3_v1 veri setinde boyama materyali tabanlı sınıflar içindeki bireysel sperm yamalarının dağılımı: (a) BesLab, (b) Histoplus, (c) GBL

eğitimi için dengesiz bir veri dağılımına yol açmıştır. Bu sorunu gidermek amacıyla, benzer morfolojik anormallikler spermin temel bileşenleri olan baş, boyun ve kuyruk kategorileri altında gruplandırılarak daha dengeli bir veri kümesi olan DB3_v2 oluşturulmuştur. Literatürde de benzer yaklaşımlar benimsenmiş olup, bazı veri kümeleri bölgesel anormallikleri bir arada ele almıştır (örneğin, HuSHeM [7] ve SCIAN-MorphoSpermGS [29]). Şekil 2.8, DB3_v2 veri kümesinin kullanılan boyama malzemelerine göre örnek dağılımını göstermektedir. DB3_v1'de olduğu gibi, farklı boyama kitlerine dayalı olarak üç ayrı DB3_v2 veri kümesi oluşturulmuştur.

Kullanılan üç boyama kiti görsel olarak karşılaştırıldığında, standart uygulamada yaygın olarak kullanılan Şekil 2.9.a.'daki BesLab boyasının daha yumuşak geçişler sağladığı ve dolayısıyla daha az gürültüye neden olduğu gözlemlenmiştir. Ancak, bazı durumlarda belirli sperm yapılarının BesLab ile boyanmadan kalabileceği tespit edilmiştir. Şekil 2.9.c.'de gösterildiği gibi, GBL boyası daha sert geçişler yaratarak sperm başlarında aşırı emilime yol açmakta ve daha koyu yapılara neden olmaktadır. Bu, belirli baş anormalliklerini öne çıkarırken, diğer anormalliklerde yanlış bilgilere yol açabilmektedir. Histoplus boyasının kullanımında, Şekil 2.9.b.'de arka planda semen sıvısında yoğun sedimantasyon (çökelme) olduğu ve

Tablo 2.2 DB3_v1 veri setlerine dahil edilen toplam örnek sayıları

Boyama Yöntemi	Benzersiz Sperm Görüntü Sayısı
BesLab (DB3_v1_BesLab)	18456
Histoplus (DB3_v1_Histoplus)	18260
GBL (DB3_v1_GBL)	12629
Toplam	49345

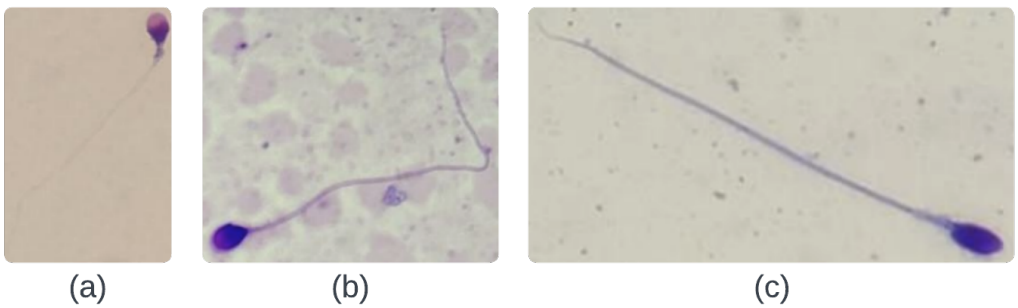


Şekil 2.8 DB3_v2 veri setlerindeki sperm yama dağılımları: (a)BesLab, (b) Histoplus, (c) GBL

bunun da görüntüyü daha karmaşık hale getirdiği görülmektedir. Sperm sayısı yüksek olan örneklerde, Histoplus boyasının kullanımından kaynaklanan artefaktlar (kalıntılar) daha belirgin hale gelmektedir. Bununla birlikte, sperm tarafından boyanın emilme derecesi, kuyruk anormalliklerinin belirlenmesinde önemli bir rol oynayabilmektedir.

2.2 Sperm Yörünge Analizi Veri Setleri

Bu tez kapsamında sperm morfoloji analizinin yanı sıra sperm yörünge analizi de yapılmıştır. Literatürde yer alan sperm tespit ve takip verisetleri incelendiğinde, çoğunluğunun yalnızca sperm tespiti problemi için oluşturulduğu gözlemlenmektedir [39, 62, 66]. Literatürde yaygın olarak kullanılan VISEM



Şekil 2.9 DB3'te (a) BesLab, (b) Histoplus ve (c) GBL boyalarının örnek görüntüleri

[40] veri seti incelendiğinde ise sperm tespiti ve takibi problemlerinin regresyon problemi olarak ele alındığı görülmektedir. VISEM veri setinde bulunan 20 adet video kullanılarak oluşturulan VISEM-Tracking veri seti, hem tespit hem de takip problemlerinde kullanılabilen, açık kaynak olarak sunulmuş önemli bir veri setidir [41]. Ancak, veri setinde bulunan videoların kısa süreli olması ve birden fazla videonun birleştirilerek elde edilmesi gibi sperm yörünge analizini olumsuz yönde etkileyen özellikler veri setinin kullanım alanlarını oldukça kısıtlamaktadır. Literatürde bulunan bu boşluğu doldurmak amacıyla, sperm tespiti ve takibi alanlarında kullanım imkanı sağlayan Hi-LabSpermTracking veri seti bu tez kapsamında oluşturulmuştur.

2.2.1 Özgün Veri Seti: Hi-LabSpermTracking

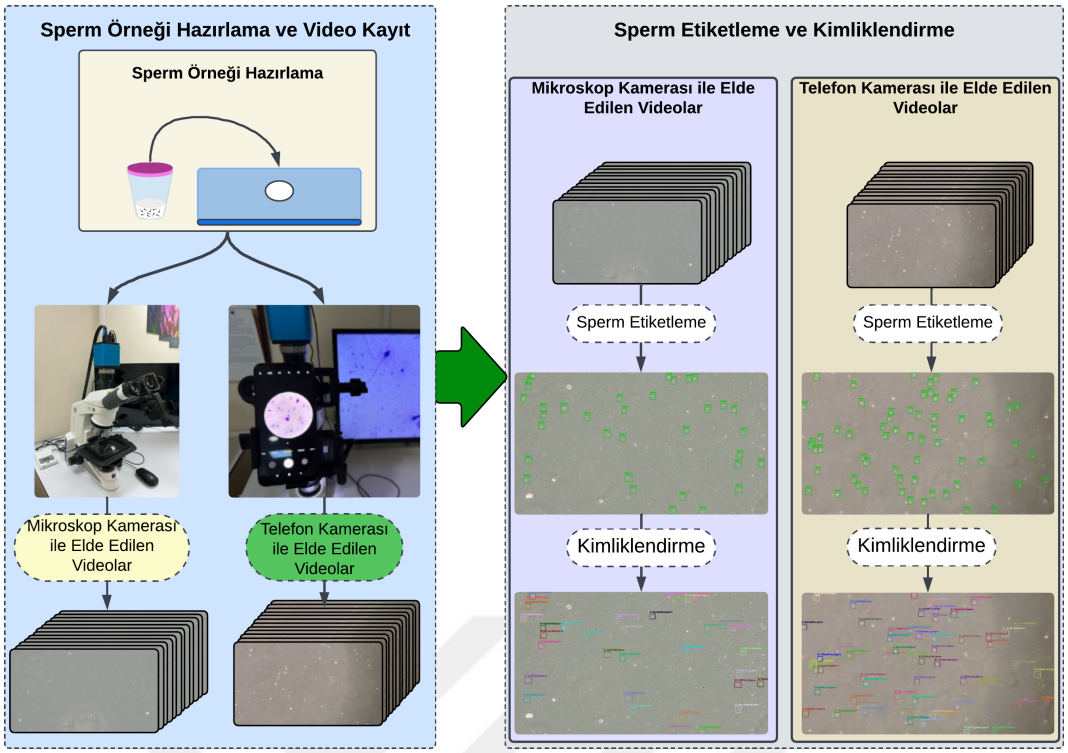
Hi-LabSpermTracking, sperm tespiti ve takibi üzerine çalışan derin öğrenme modelleri için yüksek kaliteli kaynak sağlamayı amaçlayan kapsamlı bir video veri setidir. Bu veri seti, Recep Tayyip Erdoğan Üniversitesi Eğitim ve Araştırma Hastanesi (Rize, Türkiye) Androloji Laboratuvarı'na başvuran 523 hastadan alınan teşhis numuneleri kullanılarak oluşturulmuştur. Her hasta için, araştırma ekibimiz tarafından geliştirilen akıllı telefon tabanlı görüntüleme sistemi ile dört video kaydedilmiştir. Ayrıca aynı hastadan, geleneksel laboratuvar mikroskobuna entegre edilmiş kamera kullanılarak ek dört video daha elde edilmiştir [65]. Akıllı telefon tabanlı sistem ile çekilen videolar, gelişmiş laboratuvar ekipmanlarına duyulan ihtiyacı en aza indirerek, sperm hareketliliği analizi için erişilebilir, uygun maliyetli ve taşınabilir bir çözüm sunma potansiyeline sahiptir. Bu yönüyle, mikroskop türleri arasında küresel uyumluluk sağlamaya yönelik çözümler geliştirilmesine katkı sunarak, farklı klinik ve araştırma ortamlarında sistemin uygulanabilirliği artmaktadır. Önerilen veri toplama yöntemleri ile çalışma kapsamında toplam 4.184 video elde edilmiştir. Hi-LabSpermTracking, hem akıllı telefon hem de mikroskoba entegre edilmiş görüntüleme sistemlerinden elde edilen verileri içermesiyle, hücre algılama ve izleme yöntemlerinin karşılaştırmalı değerlendirmelerine olanak tanımaktadır. Bu yönleriyle veri seti, Metodoloji bölümünde detaylandırılacağı üzere, sperm hareketlilik analizi üzerine yapılan çalışmalar için önemli bir referans olacaktır.

Tablo 2.3 Hi-LabSpermTracking veri setine ait video özellikleri

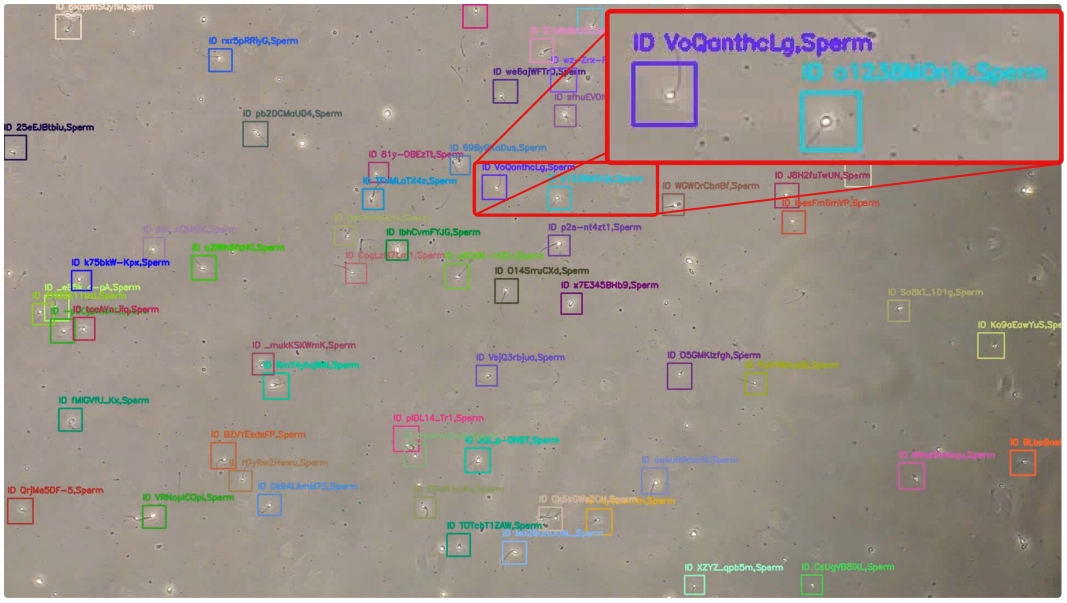
Tanım	Tlf. Kamerası Videoları	Tlf. Kamerası Test Videoları	Mik. Kamerası Videoları
Etiketlenmiş Video	30	6	30
FPS	10	10	10
Etiketlenmiş çerçeveler	18.876	3.989	18.943
En az bir sperm olan çerçeveler	18.876	3.989	18.943
Sınırlayıcı kutular	275.778	116.666	334.849
Kimliklendirilmiş Sperm	588	248	736

Videolar elde edilmeden önce tüm katılımcılara kapsamlı bilgi verilmiş ve bilgilendirilmiş onamları alınmıştır. Semen örnekleri, dördüncü gün sabah 09:00 ile 11:00 arasında örnek toplanmadan önce üç günlük cinsel perhiz dönemini içeren standartlaştırılmış bir protokol izlenerek elde edilmiştir. Sıcaklık dalgalanmalarının etkisini azaltmak için numuneler ya oda sıcaklığında (20-37°C) ya da 37°C'de inkübatörde muhafaza edilmiştir. Her bir semen örneğinden 5 µL'lik bir alikot, lam ve lamel kullanılarak, 100× büyütmeli Olympus BX43 mikroskobuna entegre edilmiş akıllı telefon tabanlı görüntüleme sistemi ile kaydedilmiştir. Videolar başlangıçta 30 fps hızında kayıt edilmiştir. Ancak, ardışık kareler arasındaki minimum değişiklikler gereksiz verilere ve etiketleme süresinin artmasına neden olmaktadır.Bu nedenle, kare hızı 10 fps'ye düşürülmüş olup, çözünürlük 1920×1080 piksel olarak korunmuştur.

Elde edilen videolardaki sperm hücreleri iki uzman tarafından yoğun ve zahmetli bir süreçle manuel olarak etiketlenmiştir. Bu etiketleme sürecinin son derece zorlu ve zaman alıcı olması nedeniyle, mevcut tüm videoların etiketlenmesi mümkün olmamıştır. Hi-LabSpermTracking veri setini oluşturmak için 14 hastadan toplam 66 video (36'sı akıllı telefonla çekilmiş ve 30'u mikroskoba entegre kamera kullanılarak elde edilmiş) titizlikle etiketlenmiştir. Her karedeki bütün sperm sınırlayıcı kutular kullanılarak etiketlenmiştir. Ayrıca, her bir sperm hücresine benzersiz bir kimlik atanmış ve bu kimliklerin tüm karelerde aynı hücreyle tutarlı kalması sağlanmıştır. Bu yapılandırma, veri kümesini hem nesne algılama hem de nesne izleme görevleri için değerli bir kaynak haline getirmektedir. Tablo 2.3'te veri kümesi hakkında ayrıntılı bilgiler sağlanmaktadır. Şekil 2.10'da Hi-LabSpermTracking veri setinin oluşturulma aşamaları akış diyagramı olarak verilmiştir. Spermlerin etiketlenerek kimliklendirildiği örnek bir çerçeve Şekil 2.11'de sunulmuştur.



Şekil 2.10 Hi-LabSpermTracking veri setinin oluşturulma sürecine ilişkin akış diyagramı



Şekil 2.11 Hi-LabSpermTracking veri setinde bulunan etiketlenmiş örnek bir video karesi

3
DERİN ÖĞRENME YÖNTEMLERİ İLE MORFOLOJİ ANALİZİ

Sperm hareketliliği, konsantrasyonu ve morfolojisi, yumurtanın döllenmesinde önemli bir rol oynamaktadır. Özellikle anormal morfolojiye sahip spermler, düşük döllenme ve gebelik oranlarıyla doğrudan ilişkilidir [70]. Bu nedenle, spermiogram testinde sperm morfolojisinin doğru ve tutarlı bir şekilde değerlendirilmesi, kısırlık tanısında oldukça önemlidir. DSÖ'ne göre, yüksek doğruluk ve güvenilirlikte bir teşhis koymak için en az 200 spermatozoanın incelenmesi gerekmektedir [71]. Ancak, 200 spermin görsel olarak değerlendirilmesi, zaman açısından çok maliyetlidir ve görsel yorumlamanın başarısı, uzman kişinin deneyimine bağlı olarak değişkenlik gösterebilmektedir. Bu nedenle, uzmanlar genellikle her bir hastanın sperm morfolojisini genel olarak değerlendirmek amacıyla 50'den az spermatozoayı incelemektedir. Ancak, bu tür değerlendirmelerin sonuçları incelendiğinde, dramatik bir değişkenlik ve hata olduğu gözlemlenmiştir [72].

Bu sorunların üstesinden gelmek ve sperm morfoloji analizini otomatikleştirmek için makine öğrenimi ve derin öğrenme mimarileri yaygın bir şekilde kullanılmaktadır [7, 27, 29, 34, 42, 50, 56, 73]. Görsel değerlendirme ile karşılaştırıldığında, derin öğrenme yaklaşımları daha tutarlı, daha ucuz, daha hızlı ve daha objektif sonuçlar vermektedir [26]. Geleneksel makine öğrenimi ve CNN tabanlı derin öğrenme yaklaşımlarına ek olarak, son yıllarda popüler hale gelen Görü Dönüştürücü (Vision Transformer - ViT) [74] mimarileri de morfoloji analizi alanında kullanılmaya başlanmıştır [27].

Sperm morfoloji alanında açık erişim olarak sunulan veri setlerinden SMIDS, HuSHeM ve SCIAN-Morpho veri setleri kullanılarak tez kapsamında CNN mimarileri ve ViT mimarileri detaylı bir şekilde istatistiksel yöntemler ile karşılaştırılmıştır. Elde edilen bulgular doğrultusunda, veri artırma ölçeklerinin ViT mimarileri üzerindeki etkisi, hiperparametre konfigürasyonlarının detaylı bir analizi ile incelenmiştir. Sonuçlar sperm morfoloji analizi alanında literatürde yer

alan en güncel yöntemlerle (state-of-the-art, SOTA) karşılaştırılarak kapsamlı bir şekilde değerlendirilmiştir. Tez kapsamında oluşturulan Hi-LabSpermMorpho veri seti üzerinde 33 farklı CNN modeli ve iki farklı ViT modeli eğitilerek verisetine ait sınıflandırma temel çizgisi (baseline) oluşturulmuştur. Kullanılan derin öğrenme yöntemleri, eğitim süreçleri, performans metrikleri ve elde edilen sonuçlar detaylı bir şekilde analiz edilerek ilerleyen başlıklarda sunulmuştur.

3.1 Açık Erişim Sperm Morfoloji Veri Setlerinin Sınıflandırılması

3.1.1 CNN ve ViT Mimarilerinin Karşılaştırmalı Analizi

Derin öğrenme yöntemlerindeki gelişmeler, birçok farklı alanda görüntü işleme uygulamalarının geliştirilmesine olanak sağlamıştır. Bu yöntemler, sperm hücrelerinin sınıflandırılması, tespiti ve takibi gibi alanlarda yaygın bir şekilde kullanılmaktadır. Literatürdeki mevcut çalışmalar incelendiğinde, sperm morfoloji sınıflandırma alanında genellikle geleneksel derin öğrenme yöntemlerinin tercih edildiği görülmektedir. VGG-16, VGG-19[75], DenseNet[76], ResNet[77] gibi CNN mimarileri bu yöntemler arasında ön plana çıkmaktadır. Bununla birlikte, son yıllarda popülerlik kazanan ViT mimarilerinin sperm sınıflandırma çalışmalarında kullanımının ise nispeten daha yeni olduğu görülmektedir [74].

Bu başlık altında, CNN ve ViT mimarilerinin sperm sınıflandırma alanındaki performansları karşılaştırılmıştır. HuSHeM, SCIAN-Morpho ve SMIDS veri setleri üzerinde, transfer öğrenme yöntemi kullanılarak geleneksel CNN ve ViT mimarilerinin performans analizi yapılmıştır. Bu sayede, her iki mimarinin sperm sınıflandırma görevindeki etkinliği karşılaştırılarak alandaki mevcut boşluk doldurulmuştur.

3.1.1.1 Kullanılan Modeller

CNN mimarileri, ilk kullanımlarından itibaren özellikle görüntü sınıflandırma alanında büyük bir popülarite kazanmıştır. Ekran kartı teknolojilerindeki ilerlemeler sayesinde daha karmaşık ve hesap gücü gerektiren modeller tasarlanabilmiş ve sinir ağı mimarilerinde daha detaylı ve verimli tasarımlar yapılabilmiştir. ImageNet [78] veri seti üzerinde gerçekleştirilen yarışmalar, CNN mimarilerine olan ilgiyi artırmıştır. ImageNet üzerinde eğitilen modellere ait ağırlıklar, farklı veri setlerinde başlangıç ağırlığı olarak kullanılabilmektedir. Bu yöntem, transfer öğrenme olarak adlandırılmakta olup, eğitim sürelerinin dramatik şekilde kısalmasını sağlamaktadır [79].

CNN mimarilerinin yanı sıra, ViT mimarileri, doğal dil işleme alanındaki dikkat tabanlı modellerden [80] esinlenerek geliştirilmiş ve görüntü sınıflandırma alanında kullanılan yeni bir derin öğrenme yaklaşımı olarak dikkat çekmiştir. Özellikle çok boyutlu veri setlerinde, ViT mimarilerinin CNN modellerine kıyasla daha iyi sonuçlar verdiği gözlemlenmiştir [74]. Ancak, sperm veri setleri üzerinde ViT mimarilerinin genel performans değerlendirmesine dair bir çalışma literatürde bulunmamaktadır.

Tez kapsamında, hem CNN mimarilerinde hem de ViT mimarilerinde, ImageNet üzerinde eğitilen modellere ait ağırlıklar kullanılarak sperm veri setleri üzerinde ince-ayar (fine-tuning) işlemi yapılmış ve sınıflandırma gerçekleştirilmiştir.

- **VGG16:** VGG16 ağı [75], 13 evrişimsel katman ve 3 tam bağlı katman olmak üzere toplamda 16 katmandan oluşan bir CNN mimarisidir. Bu mimari, 3×3 boyutunda küçük evrişimsel filtreler kullanarak, daha derin ve daha verimli bir ağ yapısı oluşturmayı başarmıştır. VGG16, kendi zamanında, en derin yapay sinir ağlarından biri olmayı başarmış ve görüntü işleme alanında önemli bir kilometre taşı olmuştur.

- **VGG19:** VGG19 ağı [75], VGG16 ağıyla benzer bir yapıya sahip olup, aralarındaki temel fark, katman sayısının farklı olmasıdır. VGG19, 16 evrişimsel katman ve 3 tam bağlı katman olmak üzere toplamda 19 katmandan oluşmaktadır. Bu ek katmanlar, daha derin özellikleri öğrenme kapasitesini artırarak, ağın daha karmaşık görsel bilgileri işlemeye olanak tanımaktadır.

- **DenseNet121:** DenseNet mimarisinde [76], her bir katman diğer tüm katmanlarla ileri beslemeli bir şekilde birbirine bağlıdır. Geleneksel CNN mimarilerinde, L adet katman varsa L adet bağlantı bulunurken, DenseNet modelinde bu sayı $L(L+1)/2$'dir, yani her bir katman tüm önceki katmanlarla doğrudan bağlantı kurmaktadır. Bu yoğun bağlantılar, özniteliklerin ağ boyunca daha verimli bir şekilde kullanılmasını sağlayarak kaybolan gradyan problemi hafifletilmektedir. Bu yapı, daha derin ağlarda bile etkili öğrenme sağlayarak modelin performansını artırmaktadır.

- **ResNet50:** CNN mimarilerinde, katman sayısının artmasına rağmen test verilerindeki başarı oranı düşüş göstermekteydi. Bu sorunu çözmek amacıyla, her iki önceki katmanda elde edilen vektörler, ileriye doğru birikimli bir şekilde aktarılarak öznitelik kaybı azaltılmış ve ResNet mimarileri geliştirilmiştir. Bu yaklaşım, ağın derinleşmesiyle birlikte özniteliklerin daha etkin bir şekilde korunmasına olanak tanımış ve kaybolan gradyan

problemine karşı bir çözüm sunmuştur. Elde edilen çok katmanlı kalıntılı ağlar, aynı sayıda katman içeren düz ağlara göre çok daha iyi sonuçlar vermektedir [77].

- **EfficientNetB7:** EfficientNet, CNN mimarilerindeki katmanların genişlik, derinlik ve görüntü çözünürlüğü özelliklerini daha verimli bir şekilde kullanmak amacıyla geliştirilmiş mimarilerdir. Geleneksel yöntemlerin aksine, bu mimari her bir boyutu (genişlik, derinlik ve çözünürlük) keyfi bir şekilde değiştirmek yerine, her boyutu sabit bir kat sayı ile eşit oranda ölçeklendirmektedir [45]. Bu sayede, ağın performansı ve verimliliği artırılırken, hesaplama maliyetleri de optimize edilmektedir.

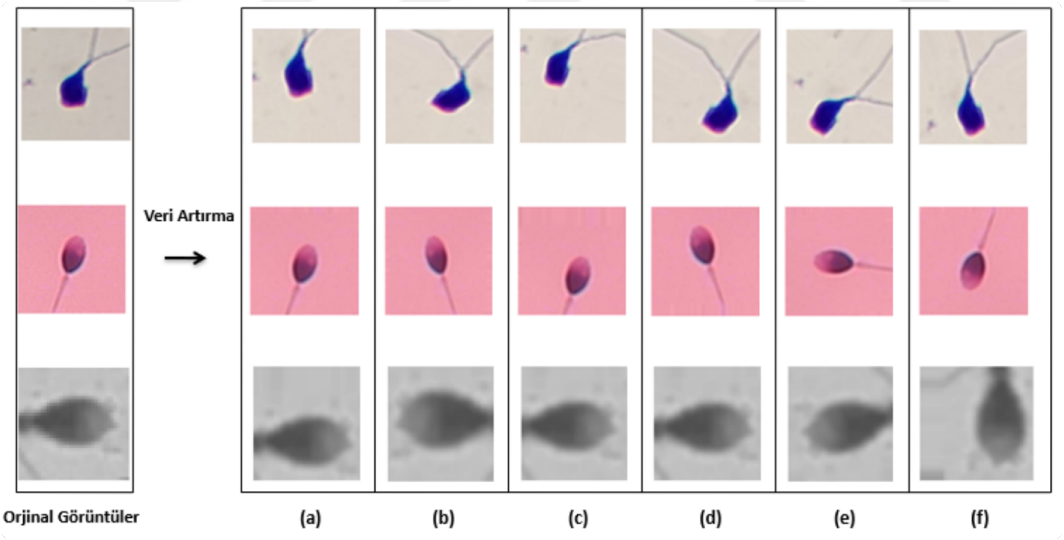
- **ViT-L16:** ViT-L16 modeli, doğrudan BERT [81] modelinden türetilmiş bir Vision Transformer (ViT) varyantıdır. İsmindeki "L16", giriş görüntüsünün 16×16 boyutlarında yamalar (patches) olarak bölündüğünü "Large" ise varyantını ifade etmektedir. Bu mimaride giriş görüntüsü, 16×16 boyutlarında küçük yamalara (patches) ayrıldıktan sonra, her bir parça düzleştirilerek eğitilebilir doğrusal projeksiyona verilerek, yamalarla konum bilgileri birleştirilmektedir. ViT-L16 modeli, toplamda 24 katmandan oluşmaktadır. Her katmanda 1024 gizli boyut ve 16 dikkat kafası bulunmaktadır. Modelin, 4096 boyutunda çok katmanlı algılayıcıya sahip olması ve toplamda 304 milyon parametreye sahip olması, modelin kapasitesini ve hesaplama kabiliyetini yansıtmaktadır [82].

- **ViT-B16:** ViT-B16 modeli toplamda 12 katmandan oluşmaktadır. 768 gizli boyuta sahip olup, 12 adet dikkat kafası bulunmaktadır. Modelin çok katmanlı algılayıcı boyutu 3072'dir ve toplam parametre sayısı 86 milyondur [82]. Bu yapı, modelin orta ölçekli bir kapasiteye sahip olduğunu ve daha büyük veri setlerinde etkili bir şekilde çalışabileceğini göstermektedir.

- **ViT-S16:** ViT-S16 modeli, diğer Vision Transformer (ViT) modellerine kıyasla daha küçük bir mimariye sahiptir. Toplamda 8 katmandan oluşan bu model, 1000 gizli boyuta sahip olup, 10 adet dikkat kafasından oluşmaktadır. Modelin çok katmanlı algılayıcı boyutu 1024'dür ve toplam parametre sayısı 21 milyondur [82]. Bu özellikler, ViT-S16 modelinin daha az hesaplama gücü gerektiren, daha hızlı eğitim ve tahmin süreçlerine olanak tanıyan kompakt bir yapıya sahip olduğunu göstermektedir.

3.1.1.2 Veri Artırımı

Veri setlerinin incelenmesi sonucunda, özellikle HuSHeM veri setinde örnek sayısının oldukça düşük olduğu tespit edilmiştir. Eğitim aşamasında modellerin aşırı öğrenmesini önlemek amacıyla çeşitli veri artırımı teknikleri uygulanmıştır. Bu teknikler arasında döndürme, yatay ve dikey çevirme gibi yöntemler bulunmaktadır. Veri artırım işleminden önce, veri setleri 5 katlı çapraz doğrulama yöntemiyle eğitim ve test setlerine ayrılmıştır. Veri artırımı sadece eğitim verileri üzerinde uygulanmış ve test verilerinde herhangi bir artırım yöntemi uygulanmamıştır. Veri artırım sonrasında, HuSHeM veri setindeki eğitim örneklerinin sayısı 40 katına, SMIDS ve SCIAN veri setlerindeki eğitim örneklerinin sayısı ise 10 katına çıkarılmıştır. Bu yöntem, modellerin daha genellenebilir özellikler öğrenmesine olanak sağlamıştır. Üç farklı veri seti üzerinde gerçekleştirilen veri artırma yöntemleri, Şekil 3.1'de detaylı bir şekilde gösterilmektedir. Bu işlemler arasında (a) dikey kaydırma ve kesme, (b) yatay çevirme ve yakınlaştırma, (c) yatay ve dikey kaydırma, (d) yatay ve dikey çevirme, (e) döndürme ve (f) dikey kaydırmayı temsil etmektedir. Görüntülerin kalitesini korumak amacıyla, yakınlaştırma, kesme ve kaydırma işlemleri, her üç veri seti için orijinal görüntü boyutları sırasıyla SMIDS, HuSHeM ve SCIAN-Morpho için %10, %5 ve %15 oranları ile sınırlandırılmıştır.



Şekil 3.1 Uygulanan veri artırma tekniklerinin her biri farklı bir veri setinden alınan örnek görüntüler üzerindeki etkileri

3.1.1.3 Deneysel Sonuçlar

Modeller, beş katlı çapraz doğrulama uygulanan veri setleri üzerinde ayrı ayrı eğitilmiştir. Eğitim sürecine dair işlem süreleri ve parametre sayıları, Tablo 3.1'de verilmiştir. HuSHeM-40x veri seti için 25 epoch, SMIDS-10x ve SCIAN-10x veri

setleri için ise toplam 50 epoch boyunca eğitim yapılmıştır. Tabloda yer alan toplam süreler, belirtilen epoch sayıları doğrultusunda hesaplanmıştır. Eğitim işlemleri, NVIDIA 4070-Ti ekran kartı ile gerçekleştirilmiştir.

Tablo 3.1 Modellere ait parametre sayıları ve eğitim süreleri

		HuSHeM-40x	
Model	**Parametre Sayısı**	**Süre (sn)***	**T. Süre (dk)****
Vgg16	14,9M	8	3
Vgg19	20,2M	9	3
DenseNet201	18.9M	10	3
ResNet-50	24,2M	7	2
EfficientNetb7	64,8M	27	9
ViT-L16	304M	37	13
ViT-s16	21M	5	2
ViT-b16	86,4M	13	5

Tablo 3.1 Modellere ait parametre sayıları ve eğitim süreleri (devam)

	SCIAN-10x		**SMIDS-10x**	
Model	**Süre (sn)***	**T. Süre (dk)****	**Süre (sn)***	**T. Süre (dk)****
Vgg16	22	15	26	18
Vgg19	29	20	29	20
DenseNet201	37	26	36	25
ResNet-50	23	16	22	15
EfficientNetb7	86	60	98	68
ViT-L16	124	86	137	95
ViT-s16	16	11	18	13
ViT-b16	39	27	46	32

*İlgili veri setinde 1 epoch eğitim için harcanan süre
**İlgili veri setinde 5 kat eğitim için harcanan toplam süre

Tablo 3.2 T-test sonucunda elde edilen skor tablosu

Model	**V-16**	**V-19**	**D-201**	**R-50**	**Eff-7**	**V-L16**	**V-S16**	**V-B16**
V-16	-	0/3/0	0/2/1	0/2/1	1/2/0	2/1/0	0/2/1	1/1/1
V-19	0/3/0	-	0/2/1	0/2/1	0/3/0	2/1/0	0/3/0	0/3/0
D-201	1/2/0	1/2/0	-	1/2/0	1/2/0	2/1/0	1/2/0	1/2/0
R-50	1/2/0	1/2/0	0/2/1	-	1/2/0	1/2/0	1/2/0	1/2/0
Eff-7	0/2/1	0/3/0	0/2/1	0/2/1	-	1/2/0	0/2/1	0/3/0
V-L16	0/1/2	0/1/2	0/1/2	0/2/1	0/2/1	-	0/1/2	0/1/2
V-S16	1/2/0	0/3/0	0/2/1	0/2/1	1/2/0	2/1/0	-	0/3/0
V-B16	1/1/1	0/3/0	0/2/1	0/2/1	0/3/0	2/1/0	0/3/0	-
K	4	2	0	1	4	12	2	3
B	13	17	13	14	16	9	17	17
Kyp	4	2	8	6	1	0	2	1

Model eğitim sürecinin tamamlanmasının ardından, elde edilen modeller test veri setleri üzerinde sınıflandırma işlemleri gerçekleştirilmiştir. Her bir veri seti için

Tablo 3.3 HuSHeM-40x veri seti doğruluk değerleri

	HuSHeM-40x					
	kat1	kat2	kat3	kat4	kat5	Ort
Vgg16	0,8440	0,8000	0,7670	0,9048	0,8780	0,8388
Vgg19	0,7556	0,7556	0,8372	0,9048	0,8537	0,8214
DenseNet201	0,6444	0,5778	0,7209	0,9286	0,8049	0,7353
ResNet-50	0,7333	0,7556	0,8837	0,9524	0,9286	0,8507
EfficientNetb7	0,7111	0,7556	0,7907	0,8810	0,8780	0,8033
ViT-L16	**0,8222**	**0,9111**	**0,9302**	**0,9524**	**0,9268**	**0,9085**
ViT-s16	0,7333	0,6889	0,8140	0,9048	0,8780	0,8038
ViT-b16	0,8222	0,8222	0,8605	0,8810	0,8537	0,8479

Tablo 3.4 SMIDS-10x veri seti doğruluk değerleri

	SMIDS-10x					
	kat1	kat2	kat3	kat4	kat5	Ort
Vgg16	0,8731	0,8719	0,8633	0,8583	0,8683	0,8670
Vgg19	0,8798	0,8819	0,8717	0,8600	0,8633	0,8713
DenseNet201	0,8397	0,8419	0,8200	0,8300	0,8600	0,8383
ResNet-50	0,8497	0,8619	0,8433	0,8567	0,8633	0,8550
EfficientNetb7	0,8932	0,8669	0,8833	0,8800	0,8850	0,8817
ViT-L16	**0,8831**	**0,8935**	**0,8967**	**0,8833**	**0,9150**	**0,8943**
ViT-s16	0,8731	0,8636	0,8667	0,8567	0,8767	0,8674
ViT-b16	0,8932	0,8785	0,8800	0,8583	0,8967	0,8813

Tablo 3.5 SCIAN-10x veri seti doğruluk değerleri

	SCIAN-10x					
	kat1	kat2	kat3	kat4	kat5	Ort
Vgg16	**0,6578**	**0,6594**	**0,6608**	**0,6637**	**0,6533**	**0,6590**
Vgg19	0,6311	0,6507	0,6388	0,6814	0,6400	0,6484
DenseNet201	0,6267	0,6026	0,6872	0,708	0,6667	0,6582
ResNet-50	0,6667	0,6245	0,6388	0,5973	0,6756	0,6406
EfficientNetb7	0,6311	0,6201	0,6432	0,6150	0,6978	0,6414
ViT-L16	0,6933	0,6332	0,6300	0,6549	0,6756	0,6574
ViT-s16	0,6400	0,6026	0,6211	0,6195	0,6667	0,6300
ViT-b16	0,6356	0,6114	0,6344	0,6239	0,6440	0,6299

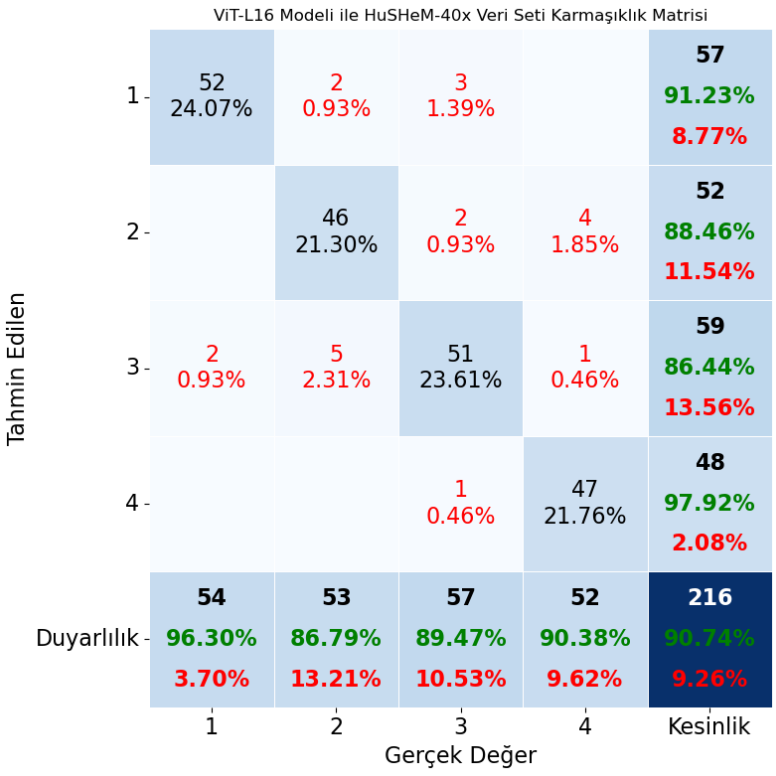
modellerin doğruluk (accuracy) değerleri, %90 güven düzeyinde iki kuyruklu t-testi analizi ile karşılaştırılmıştır. Bu analiz, modeller arasında istatistiksel olarak anlamlı farklar olup olmadığını belirlemek amacıyla yapılmıştır. İstatistiksel olarak anlamlı bir fark tespit edilen modeller için, daha başarılı olma durumlarına göre "Kazanma", "Kaybetme" ve "Berabere" skorları hesaplanmıştır. Tüm modellerin birbiriyle kıyaslanmasından elde edilen bu skorlar, Tablo 3.2'de özetlenmiştir. T-test analizinde kullanılan HuSHeM, SMIDS ve SCIAN veri setlerine ait doğruluk

değerleri ise sırasıyla Tablo 3.3, 3.4 ve 3.5'te sunulmuştur.

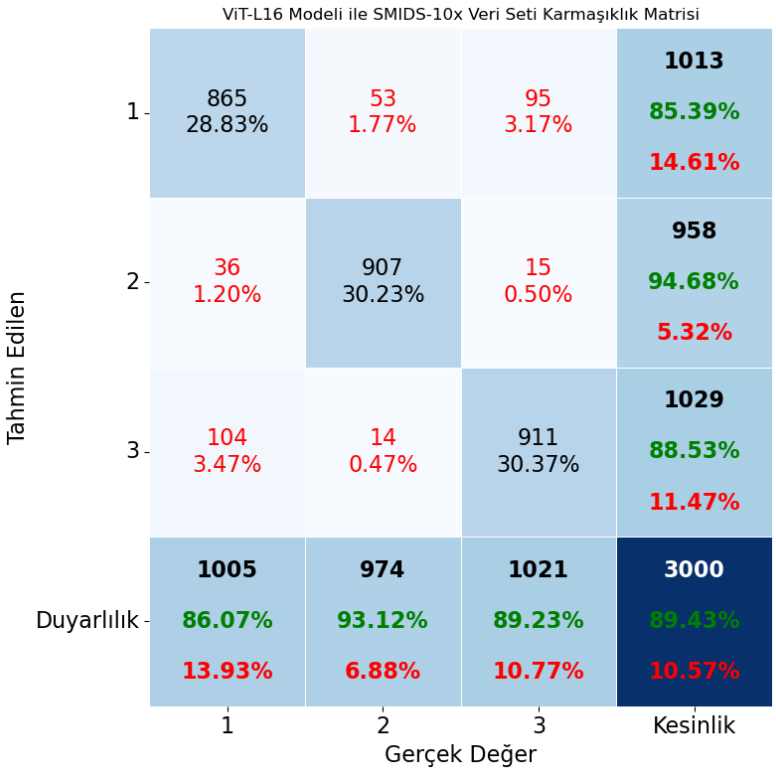
3.1.1.4 Sonuçlar

Tablo 3.2 incelendiğinde, ViT-L16 modeli ile VGG-16 modelinin karşılaştırmasına ilişkin olarak "2/1/0" veya "0/1/2" ifadeleri görülebilmektedir. Örneğin, "2/1/0" hücresi değerlendirildiğinde, analiz edilen üç veri setinden ikisinde ViT-L16 modelinin VGG-16 modeline üstünlük sağladığı, bir veri setinde iki model arasında istatistiksel olarak anlamlı bir fark bulunmadığı için "beraberlik" olduğu ve ViT-L16 modelinin VGG-16 karşısında hiçbir veri setinde başarısız olmadığı (0 kayıp) görülmektedir.

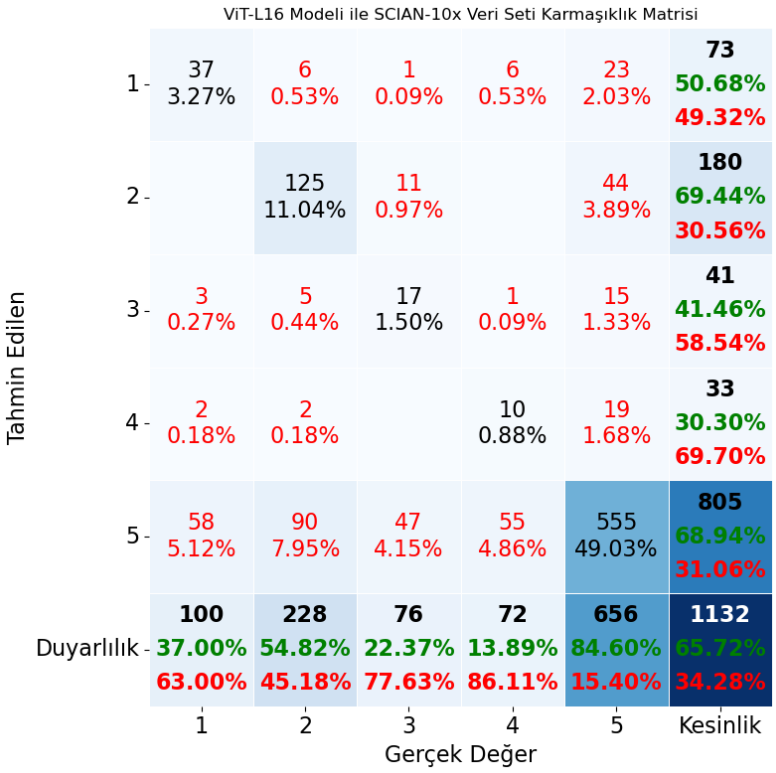
Tablo 3.2'deki genel sonuçlar incelendiğinde, ViT-L16 modeli, HuSHeM veri setinde 12 kazanma (K), 9 beraberlik (B) ve 0 kayıp (KYP) ile, en iyi CNN mimarisinden yaklaşık %6, SMIDS veri setinde ise yaklaşık %2 daha başarılı olduğu gözlemlenmiştir. Bununla birlikte, SCIAN veri setinde Tablo 3.5 değerlendirildiğinde, VGG-16'nın en yüksek ortalama doğruluğa ulaştığı tespit edilmiştir. Ancak, yapılan t-test analizine göre, ViT-L16 modeli ile VGG-16 modeli arasında istatistiksel olarak anlamlı bir fark bulunmadığı için bu iki model "berabere" olarak değerlendirilmiştir.



Şekil 3.2 ViT-L16 modeli HuSHeM-40x veri seti karmaşılık matrisi



Şekil 3.3 ViT-L16 modeli SMIDS-10x veri seti karmaşılık matrisi



Şekil 3.4 ViT-L16 modeli SCIAN-10x veri seti karmaşılık matrisi

En başarılı mimari olan ViT-L16'nın 5 katın her birinde sıfırdan eğitilen modellerinin, test verileri tahminlerine göre karmaşıklık matrisleri oluşturulmuş ve birleştirilmiştir. HuSHeM, SMIDS ve SCIAN-Morpho veri setlerine ait karmaşıklık matrisleri sırasıyla Şekil 3.2, 3.3 ve 3.4'te sunulmuştur. Bu karmaşıklık matrislerinin en sağ sütununda sınıf bazında Kesinlik (Precision) değerleri, en alt satırında sınıf bazında Duyarlılık (Recall) değerleri ve en alt sağ hücresinde genel doğruluk oranları verilmiştir.

Yapılan deneysel çalışmalar sonucunda, ViT mimarilerinin özellikle yüksek çözünürlüklü görüntü veri setlerinde CNN mimarilerine kıyasla daha üstün performans sergilediği tespit edilmiştir. Ancak, çözünürlüğün 35×35 piksel olduğu SCIAN-Morpho veri setinde, ViT-L16 modelinin diğer CNN mimarileri ile benzer sonuçlar verdiği gözlemlenmiştir. Çalışma zamanı açısından değerlendirildiğinde, ViT modellerinin eğitimi için CNN mimarilerine oranla daha fazla hesaplama kaynağı ve işlem süresi gereksinimi olduğu belirlenmiştir. Bu durum, ViT modellerinin uygulanabilirliği açısından dikkate alınması gereken önemli bir faktördür.

3.1.2 Sperm Morfoloji Analizinde Vision Transformer Mimarilerinin Sınıflandırma Analizi

Literatür taraması bölümünde, HuSHeM, SMIDS ve SCIAN-Morpho veri setlerini sınıflandırmada ağırlıklı olarak CNN mimarilerinin kullanıldığı, buna karşın ViT mimarileri kullanan çalışmaların nispeten daha az olduğu ortaya konmuştur.Bu durum, ViT modellerinin performansının, özellikle sperm morfoloji sınıflandırması bağlamında, daha derinlemesine ve titizlikle değerlendirilmesi gerekliliğini ortaya koymaktadır. Bir önceki başlıkta yaygın olarak kullanılan CNN mimarileri ile ViT mimarileri her üç veriseti kullanılarak karşılaştırılmıştır. Deneysel çalışmalar, ViT mimarilerinin düşük çözünürlüklü görüntülerde etkili bir performans sergilemediğini ortaya koymaktadır. Bu nedenle, ViT mimarilerinin daha kapsamlı bir analizinin yapıldığı bu başlık altında, SCIAN-Morpho veri setine kıyasla daha yüksek çözünürlükte görüntüler içeren HuSHeM ve SMIDS veri setleri kullanılmıştır.

ViT [74], BeiT [81] ve DeiT [83] gibi son teknoloji ViT mimarileri ve bunların varyasyonları, farklı optimizasyon yöntemleri ve öğrenme oranları kullanılarak seçilen iki veri seti üzerinde eğitilmiştir. Bu modellerin performansları, hem birbirlerine karşı hem de literatürdeki mevcut çalışmalarla karşılaştırmalı olarak kapsamlı bir şekilde değerlendirilmiştir. Bu çerçevede, tez kapsamında, ViT mimarilerinin sperm morfoloji sınıflandırmasındaki etkinliği detaylı bir biçimde

analiz edilmiştir.

Literatürde bulunan diğer yaklaşımların sınırlamaları ve eksiklikleri, aşağıda belirtilen iyileştirmeler önerilerek ele alınmıştır:

- Literatürde ilk kez, en başarılı öğrenme mimarisini belirlemek amacıyla, iki farklı halka açık çok sınıflı RGB sperm morfoloji veri setine, bu kadar çeşitli ViT modeli seti sistematik bir şekilde uygulanmıştır.
- ViT mimarileri için farklı optimizasyon yöntemleri, öğrenme oranları ve veri artırma oranları, kapsamlı bir hiperparametre analiz süreci ile gerçekleştirilmiştir.
- ViT modellerinin, çok sınıflı sperm morfoloji analizinde CNN modellerine karşı üstünlüğü ayrıntılı bir şekilde ortaya konulmuştur.

3.1.2.1 Materyal ve Yöntem

- **Deneysel Kurulum:**
  Kurulum detayları, bu çalışmayı tekrar etmek veya üzerine inşa etmek isteyenler için oldukça önemlidir. İşletim sistemi, donanım bileşenleri ve ilgili yazılım kütüphaneleri gibi teknik özellikler, Tablo 3.6'da özetlenmiştir.

Tablo 3.6 Deneysel kurulum özellikleri

Kategori	Detaylar
İşletim Sistemi	Ubuntu 22.04 LTS
CPU	AMD Epyc 7313, 3.0GHz 16 Core /32 Thread
GPU	NVIDIA GeForce RTX 4090 24 GB GDDR6X x 2
RAM	Samsung 128 GB DDR4 3200 MHz
Python Versiyon	3.10
Kütüphaneler ve Versiyonları	pytorch 2.4.0, transformers 4.10.0, datasets 1.11.0, scikit-learn 0.24.2, matplotlib 3.4.3, seaborn 0.11.2
Eğitim Çerçevesi	Huggingface Transformer Modeller ve Verisetleri

- **Veri Artırımı:**
  Androloji alanında, sperm hücrelerini içeren tıbbi görüntü veriseti hazırlamak, mikroskobik boyutlar nedeniyle oldukça zor bir iştir [84]. Spermlere ait elde edilen görüntü örneklerinin, uzmanlar tarafından dikkatlice kontrol edilip etiketlenmesi gerekmektedir. Bu nedenle çok sayıda etiketlenmiş insan sperm görüntüsü içeren açık erişimli sperm morfoloji veriseti bulmak oldukça zordur. Sperm morfoloji analizinde derin öğrenme yöntemleri uygulandığında, açık kaynak sperm morfoloji veri setlerinin

sınırlı örnek sayısı, aşırı uyum (overfitting) sorununa yol açabilmektedir [27, 50]. Derin öğrenme modellerinde, derin katman mimarilerinin kullanımı ile gerçekleştirilen hiyerarşik öğrenme süreci, iyi bir şekilde optimize edilmesi gereken çok sayıda model ağırlığına sahiptir. Bu nedenle, sınırlı veri ile eğitilen derin ağlar, aşırı uyum (overfitting) yapma eğiliminde olup, test veri setinde düşük performans göstererek genelleme yeteneklerini olumsuz yönde etkileyebilmektedir. CNN'ler gibi, ViT mimarileri de veriye aç olan mimarilerdir. Steiner ve ark. ViT modellerinde veriseti boyutunun model performansını nasıl etkileyebileceğini araştırmışlardır [82]. Bu araştırmanın bir bulgusu olarak, veri artırma yöntemlerinin modelin genelleme yeteneği üzerinde önemli olumlu etkiler yarattığı gözlemlenmiştir. Veri artırma [85], öğrenilmeye çalışılan desenlerin özelliklerini koruyarak bir veri setinde sentetik örnekler üretme sürecidir. Bu süreç, ham örnekler üzerinde uygulanan yatay çevirme, yansıtma, döndürme, kırpma ve yeniden ölçeklendirme gibi çeşitli işlemlerle sağlanabilmektedir. Veri artırma sonucunda, modelin genelleme kapasitesini arttıracak daha iyi model uyumu sağlayacak yeni bir eğitim seti elde edilmektedir. Bu sayede olası aşırı uyum senaryoları önlenebilmektedir [55]. Her iki veri setinde de sınıflara ait örnek sayısının dağılımı dengeli olduğundan, sınıflar arasında herhangi bir ek dengeleme işlemi yapılmadan doğrudan veri artırma işlemi uygulanmıştır. HuSHeM veri seti için artırma işlemi $1\times$ ile $40\times$ arasında gerçekleştirilmiştir. $1\times$ orijinal veri setini, $40\times$ ise toplam örnek sayısının, yeni sentetik örnekler dahil olmak üzere 40 kat arttığı yeni veri setini ifade etmektedir. Aynı artırma işlemi SMIDS veri seti için de uygulanmış ve artırma oranları $1\times$ ile $10\times$ arasında değişmektedir. ViT modelleri, orijinal ve artırılmış veri setleriyle ayrı ayrı eğitilerek, görüntü artırma ölçeğinin model başarısı üzerindeki etkisi analiz edilmiştir.

- **Önerilen Yaklaşım:**
  CNN mimarileri, LeCunn ve arkadaşlarının [86] el yazısı karakterlerini sınıflandırmak için ilk kez kullandığı günden bu yana bilgisayarla görme alanında yaygın olarak kullanılmıştır. AlexNet modeli, Krizhevsky ve arkadaşları tarafından [87] büyük ölçekli görüntü sınıflandırma veri setlerinde (örneğin, ImageNet) geleneksel makine öğrenmesi ve sığ katmanlı yapay sinir ağlarına kıyasla daha iyi performans gösterdikten sonra CNN mimarilerine olan ilgi artmıştır. CNN mimarilerinde, her katmanda görüntüler evrişim işleminden geçirilmekte ve bir sonraki katmana aktarılmaktadır. Görüntülerdeki her bir yama (patch), aynı ağırlıklarla eşlenmektedir. Bu işlem, insan beyninin belirli bir uyarana tepki veren görsel korteksinin

çalışma prensibiyle benzer şekilde çalışmaktadır [86]. CNN mimarileri, her sınıf için görüntü temsillerini öğrenebilmekte ve görsel korteksin çalışma prensibiyle paralel olarak çalışmaktadır [88]. Bu benzerlik sayesinde, özellikle görsel görevlerde, geleneksel makine öğrenmesi yöntemlerine kıyasla önemli ölçüde daha iyi performans sergilemişlerdir [87]. Ancak, ağırlıklar ve yamalar arasındaki bu yerel bağlantı, modellerin genelleme kapasitesini düşürebilmektedir. Bu tür problemlere örnek olarak, CNN mimarilerinin görüntü sınıflandırmada bir nesnenin şekli yerine dokusuna odaklanması verilebilir [89]. Cao ve arkadaşları, her RELU katmanına geri besleme katmanı ekleyerek bu tür problemleri ortadan kaldırmaya çalışmışlardır [90]. Aşağıdan yukarıya ve yukarıdan aşağıya olmak üzere her iki yönde de nöron tepkisini aktive edebilen bir kontrol ünitesi oluşturmuşlardır. Aşağıdan yukarıya doğru işleyiş, RELU katmanlarının çalışma prensibini oluşturmaktadır ve baskın özellikler seçilerek bir sonraki katmana aktarılmaktadır. Geri besleme katmanları ise, yüksek düzeydeki semantik bilgilerin yukarıdan aşağıya doğru yayılmasını kontrol ederek, görüntü temsilleri ile geri beslenmektedir. Ancak yalnızca belirli hedef nöronlarla ilişkili kapılar aktif edilmektedir. Bu şekilde, ağırlıklar ve yamalar arasındaki yerel bağlantı, belirli hedef etiketler bazında kontrol edilebilmektedir.

Transformerlar ilk duyurulduğundan bu yana doğal dil işleme (Natural Language Processing - NLP) alanında kayda değer bir başarı göstermişlerdir. Konuşma tanıma [91], metinden konuşmaya çeviri [92] ve doğal dil üretimi [93] gibi çeşitli NLP alanlarında araştırmalar yapılmıştır. Transformer mimarilerin NLP'deki yüksek performansları ile birlikte belirli problemlere özgü özel mimariler tasarlanmaya başlanmıştır. The Bidirectional Encoder Representations from Transformer (BERT) [81] ve The Generative Pre-trained Transformer (GPT-3) [94] çeşitli NLP problemlerini çözmek için kullanılan sadece iki popüler örnektir. Bilgisayarla görme alanında, CNN mimarileri ve türevleri oldukça başarılı performanslar elde etmiş olsa da, görüntülerdeki nesnelerin yerel olmayan korelasyonları gibi uzun vadeli bağımlılıklar genellikle ihmal edilmektedir [95]. Buna karşın, transformer mimarileri, özellikle uzun vadeli bağımlılıkları öğrenme konusunda CNN mimarilerine kıyasla daha yüksek bir yatkınlığa sahip olup, bu özellikleriyle dikkat çekmektedir. Dosovitskiy ve arkadaşları [74], NLP'deki dönüştürücülerin üstün performansından esinlenerek ViT mimarilerini geliştirmiştir. Önerdikleri yöntemde, görüntüler yamalara (bölgelere) bölünmüş ve görüntüdeki uzun vadeli bağımlılıkları yakalayabilen bir dizi görüntü dizisi elde edilmiştir. Bu yaklaşım, modelin odağını, görüntülerdeki

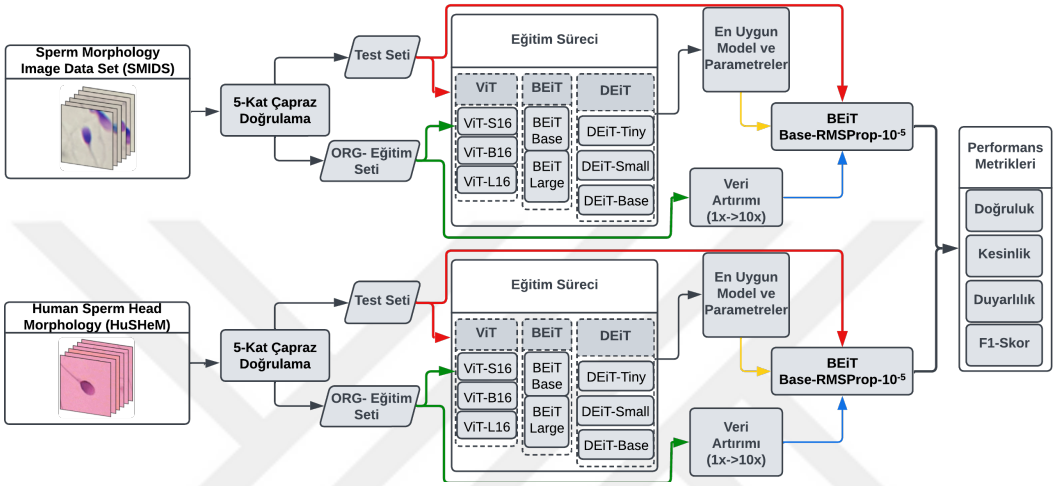
nesnelerin dokusunu öğrenmekten, nesnelerin şekillerine odaklanarak bu şekiller arasındaki bağımlılıkları daha etkili bir şekilde temsil etmeye kaydırmaktadır [96].

Tıbbi görüntü sınıflandırmanın çeşitli alanlarında, ViT modelleri güçlü bir öğrenme aracı olarak kullanılmaktadır [95]. COVID-19 teşhisinde büyük öneme sahip olan X-Ray görüntülemede ViT tabanlı sınıflandırma yaklaşımlarından faydalanılmıştır. Örneğin Liu ve arkadaşları, X-Ray görüntülerinde sınıflandırma yapmak için ViT tabanlı bir sınıflandırma yaklaşımı olan vision out-looker (VOAT) yöntemini geliştirmiştir [97]. Ayrıca, Than ve arkadaşları [98] bilgisayarlı tomografi ile elde edilen görüntüler üzerinde COVID-19 ve hastalıklı akciğer sınıflandırması için kullanılan ViT modellerinde yama boyutunun sınıflandırma üzerindeki etkisini araştırmıştır. Sperm morfoloji analizi bağlamında, Mahali ve arkadaşları insan sperm başlarını sınıflandırmak için Swin Transformer ve MobileNetV3 modellerini birleştirerek Swin-Mobile modelini tasarlamıştır [43]. Ayrıca, Aktaş ve arkadaşları ViT modellerinin farklı varyasyonlarının ve geleneksel CNN mimarilerinin insan spermini sınıflandırmadaki performansını karşılaştırmıştır [27]. Chen ve arkadaşları, gürültülü sperm görüntü sınıflandırması için CNN ve Transformer mimarilerinin kapsamlı bir karşılaştırmasını yapmışlardır [99].

Yapılan literatür taraması sonucunda ViT mimarilerin çeşitli tıbbi görüntü veri kümelerine uygulandığı görülmüştür [95]. İnsan sperminin morfolojik analizi alanında ise literatürde genellikle CNN mimarilerinin kullanıldığı görülmektedir. Ancak, CNN ve ViT mimarileri nesne sınıflandırma problemlerinde karşılaştırıldığında, ViT modellerinin görüntülerdeki nesnelerin dokusundan ziyade şekline daha bağımlı olduğu yapılan çalışmalar ile gösterilmiştir [96]. Sperm morfoloji analizinde, spermin dokusundan ziyade şekline odaklanmak çok daha önemlidir. Anormallikler doğrudan sperm yapılarının şekli ile ilişkilendirilirken, benzer dokuya sahip sperm örnekleri farklı anormallik türüne ait olabilmektedir. Sperm başlarının büyüklüğü, kuyruğun gövdeye bağlanma şekli, akrozom yapısı ve sperm gövdesinin nasıl birleştiği gibi morfolojik bağımlılıklar sperm morfoloji sınıflandırmasında oldukça önemlidir.

CNN ve Transformer mimarileri karşılaştırılırken iki husus dikkate alınmalıdır; i) CNN mimarileri evrişim katmanları aracılığıyla bir görüntüdeki uzamsal ilişkileri ortaya çıkarabilir, ancak genellikle uzun vadeli bağımlılıkları göz ardı etmektedirler. Öte yandan Transformer mimarileri, dikkat mekanizması aracılığıyla uzun vadeli bağımlılıkları daha etkili bir

şekilde ele almaktadır. Bu nedenle, dokudan ziyade nesnelerin şekil bilgilerinin daha önemli olduğu veri kümelerinde Transformer mimarilerinin kullanılması daha iyi sonuçlar vermektedir. ii) CNN mimarileri evrişim katmanlarına dayandığı için Transformer mimarilerine kıyasla daha katı bir yapıya sahiptir. Fakat, Transformer mimarileri evrişim katmanı kullanmadıkları için farklı boyutlardaki girdi verileriyle çalışabilecek esnek bir yapıya sahiptir.

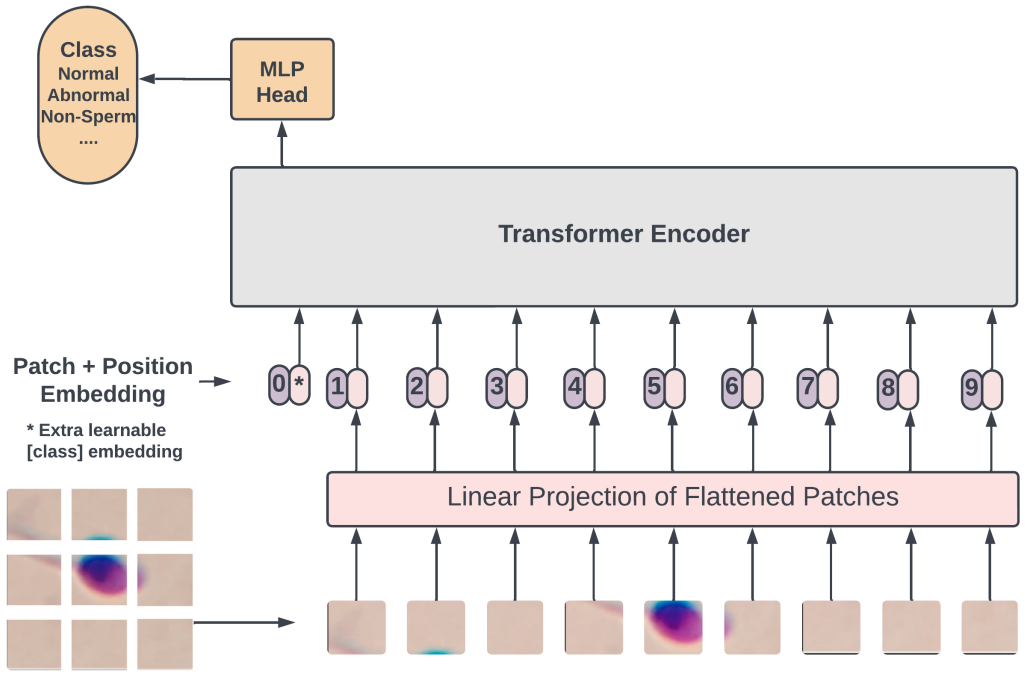


Şekil 3.5 Önerilen Transformer tabanlı sperm morfolojisi sınıflandırma yaklaşımının akış şeması

Yukarıda belirtilen kapsam doğrultusunda, HuSHeM ve SMIDS gibi halka açık veri setlerinde sekiz farklı Transformer modeli kullanarak sperm morfoloji sınıflandırma analizi yapılmıştır. Sunulan çalışmanın akış diyagramını Şekil 3.5'de görmek mümkündür. Transfer öğrenme ile kullanılan modeller veri artırma ölçeği, optimizasyon yöntemi ve öğrenme oranları bakımından detaylı bir şekilde analiz edilmiştir, böylece optimal model ve hiperparametre ilişkisi belirlenmiştir. Imagenet üzerinde önceden eğitim yapılarak elde edilen tüm ağırlıklar, her iki veri seti için ayrı ayrı eğitim örnekleriyle ince ayar edilmiştir. Bu çalışmada kullanılan modeller ve çalışma prensipleri aşağıdaki alt başlıklarda sunulmaktadır;

– **ViT Mimarisi:**
  Dosovitskiy ve arkadaşları tarafından önerilen ViT modeli [74], NLP alanında kullanılan Transformer'ların temel mimarisini görsel sınıflandırma problemleri için uyarlamaktadır. ViT modelinde, ham görseller bir dizi küçük yamalara (patch) bölünmektedir. Bu yamaların uzamsal bilgileri, uzamsal kodlama yöntemi kullanılarak elde edilmektedir. Bu uzamsal gömüler (positional embeddings),

yamalara eşlenerek Transformer Encoder beslenmektedir. Multi-head self attention (MHSA), Transformer'ın yamalar ve ek öğrenilebilir sınıf belirteci ile eğitilmesiyle hesaplanmaktadır. BERT'te kullanılan orijinal Dönüştürücü mimarisine yakın bir mimari sağlamak için görüntü temsili olarak kullanılan ekstra bir öğrenilebilir sınıf belirteci eklenmektedir. Bu belirteçlerin çıktısı, daha sonra tek gizli katmanlı ve doğrusal olmayan tanh fonksiyonu ile çalışan MLP kullanılarak sınıf tahminine dönüştürülmektedir. Eğitim sonrasında Transformer, yamalara ait öğrenilmiş gömme (embedding) değerlerini çıkartmaktadır. ViT modelinin çıktısındaki sınıf belirteçleri, görsellerin temsili olarak kullanılmaktadır. Son aşamada, ViT ile elde edilen bu temsiller, MLP ile sınıflandırılmaktadır. Sperm morfoloji görüntülerinin sınıflandırılması için uyarlanan modelin akış şeması Şekil 3.6'da görülmektedir. ViT'nin



Şekil 3.6 ViT akış şeması [74]

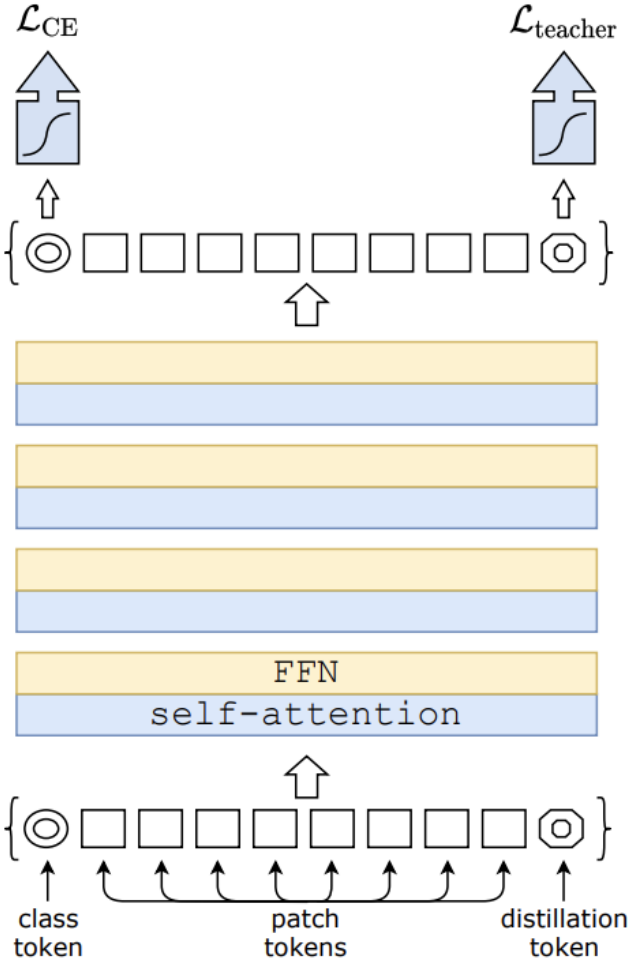
omurgalarında kullanılan parametre sayısına göre farklı isimlere sahip farklı varyasyonları bulunmaktadır. ViT-L16, ViT-B-16 ve ViT-S16 olarak adlandırılan mimariler sırasıyla 304M, 86.4M ve 21M parametreye sahiptir. Bu çalışmada, bu üç modelin her biri farklı optimizasyon yöntemleri ve öğrenme oranları ile HuSHeM ve SMIDS üzerinde eğitilmiştir.

– **The Data efficient Image Transformer (DEiT) Mimarisi:**
ViT modelleri, 300 milyon görüntüden oluşan ve özel olarak

etiketlenmiş en büyük veri kümelerinden biri olan JFT-300M veri kümesinde sınıflandırma performansında önemli bir iyileşme sağlamıştır [100]. Literatürde, ViT modellerinin iyi bir genelleme yapabilmesinin ancak yeterli miktarda veri ile eğitildiklerinde mümkün olduğuna dikkat çekilmiştir [74]. Ancak, büyük veri kümeleri üzerinde uygulanan eğitim süreci önemli miktarda hesaplama gücü gerektirmektedir. Bu sorunu çözmek için Touvron ve arkadaşları, farklı bir eğitim stratejisi ve yeni bir damıtma tekniği kullanan Data-efficient Image Transformer (DEiT) mimarisini tanıtmıştır [83]. DEiT mimarisinin temel amacı, ViT mimarisini temel alarak daha az miktarda eğitim verisi kullanarak daha iyi öğrenme performansı elde etmektir. DEiT mimarisinde yapılan gelişmeler ile temel ViT modelinin veri bağımlılığı azaltılmıştır. DEiT mimarisi, bilgi damıtma yöntemini kullanarak iki ağ arasında bilgi alışverişini kolaylaştırmaktadır. Daha büyük, daha güçlü ve önceden eğitilmiş olan ilk ağ öğretmen, rastgele başlatılan ağırlıklara sahip ikinci ağ ise öğrenci ağı olarak kullanılmaktadır.

Bilgi damıtma sürecinde, algoritmanın odak noktası öğretmen modelinin tahminlerinden öğrenme yapmaktır. ViT mimarisinden farklı olarak DEiT mimarisi, bir sınıf (CLS) belirtecine benzer şekilde işlev gören rastgele başlatılmış bir damıtma belirteci kullanmaktadır. Bu belirtecin amacı, öğretmen modeli tarafından tahmin edilen sabit etiketi çoğaltmaktır. Öğretmen modelinin ürettiği sonuçlar, iki farklı damıtma yöntemi kullanılarak çıkarılmaktadır: yumuşak (soft) ve sert etiket (hard-label) damıtma. Yumuşak damıtma yönteminde etiket değeri, öğretmen ve öğrenci modellerinin softmax sonuçları arasındaki Kullback-Leibler farklılığı [101, 102] en aza indirilerek elde edilmektedir. Sert etiketli damıtmada ise softmax ile elde edilen olasılık değerleri yerine öğretmen modelinin nihai (sert) kararı gerçek etiket olarak kullanılmaktadır. Damıtma belirteçleri, bu yöntemler kullanılarak elde edilmektedir. Elde edilen tüm belirteçler, model çıktılarını üretmek için üç adet kodlayıcı bloğundan geçirilmektedir. DEiT mimarisi parametre sayısına göre farklı varyasyonlarda kullanılabilmektedir. Bu çalışmada sırasıyla 5M, 22M ve 86M parametreye sahip DEiT-Tiny, DEiT-Small ve DEiT-Base mimarileri kullanılmıştır. DEiT mimarisinin daha iyi anlaşılması için Şekil 3.7'de akış diyagramı sunulmuştur.

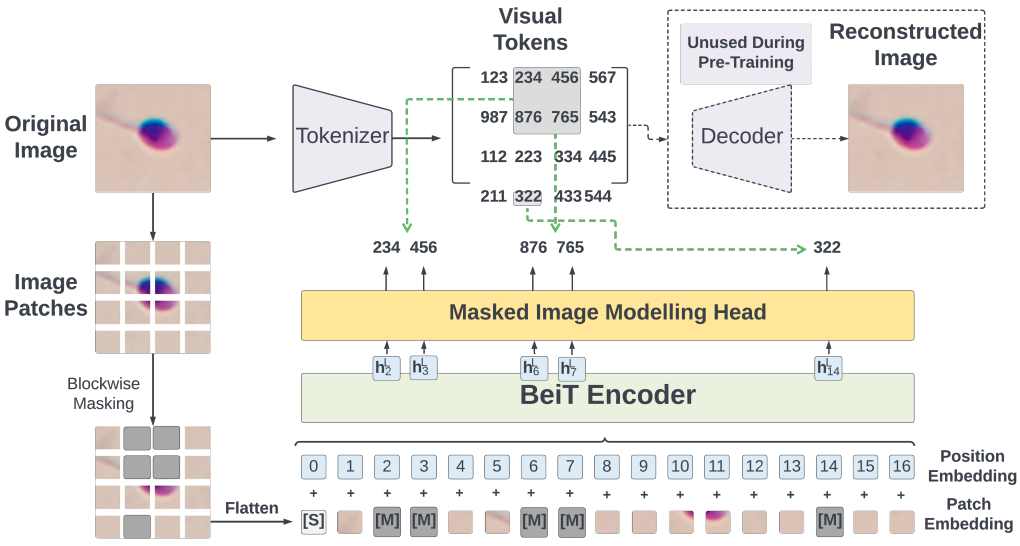
– **The BERT Pre-training of Image Transformer (BEiT) Mimarisi:**
Devlin ve arkadaşları tarafından geliştirilen BERT, NLP alanında



Şekil 3.7 DEiT akış şeması [83]

olağanüstü başarılar elde etmiştir [81]. BERT yönteminin temelini oluşturan Maskeli Dil Modeli (Masked Language Modeling - MLM), önce bir cümlenin rastgele bir kısmını maskelemektedir. Ardından, bu bozulmuş cümle kullanılarak Transformer modeli beslenmekte ve maskelenmiş belirteçler, kodlama vektörlerine göre geri yüklenmektedir. Bu yöntem temel alınarak, Bao ve arkadaşları tarafından ViT modellerini önceden eğitmek amacıyla BERT Pre-training of Image Transformer (BEiT) geliştirilmiştir [103]. Ancak, BERT mimarisinin doğrudan görsel verilere uyarlanması, görüntü parçaları arasında kelimelerin cümleleri oluşturma sırasındaki gibi bir ilişki bulunmaması nedeniyle zorlu bir görevdir. Bu sorunu çözmek için, NLP yöntemindeki maskelenmiş kelimelerin tahmin süreci, maskelenmiş parçalardan ham piksellerin tahmin edilmesi problemine dönüştürülmüştür. Ancak, bu piksel düzeyinde yapılan iyileştirme, modelin kapasitesini israf etme eğilimindedir. BERT'ten

ilham alan Bao ve arkadaşları, ön eğitim görevini gerçekleştirmek için Maskelenmiş Görüntü Modeli (Masked Image Modeling - MIM) yöntemini sunmuşlardır [103]. Bu yöntemde, ViT modelinin giriş kısmına uyacak şekilde, görüntü ızgarayla bölünmektedir. Rastgele seçilen parçalar, Blockwise masking yöntemi kullanılarak maskelenmektedir. Maskelenmiş parçalar, orijinal parçalar ve konumsal gömme (positional embeddings) vektörleri hep birlikte dönüştürücüye gönderilmektedir. Buna paralel olarak, ham görüntülerden, gizli değişken oto kodlayıcıya ait (Variational Auto Encoder - VAE) [104] gizli kodlar kullanılarak, ayrık görsel belirteçler üretilmektedir. Sonuç olarak, orijinal görüntü iki farklı şekilde temsil edilmektedir. Kod çözücü, görsel belirteçleri kullanarak orijinal görüntüyü yeniden oluşturmayı öğrenmekte ve maskelenmiş belirteçler MIM kullanılarak geri yüklenmektedir. Bu kendi kendine denetimli yaklaşım, dönüştürücülerin daha az veri bağımlılığı ile ön-eğitilmesini sağlamaktadır. Diğer dönüştürücü mimarileri gibi, BEiT mimarisinin de farklı varyasyonları bulunmaktadır. Bu çalışmada, sırasıyla 86M ve 307M parametreye sahip BEiT-Base ve BEiT-Large mimarileri kullanılmıştır. SMIDS'ten sperm morfoloji görüntüsüyle uyarlanan BEiT mimarisi Şekil 3.8'de görülmektedir.



Şekil 3.8 BEiT akış şeması [103]

Üç farklı dönüştürücü mimarisinin farklı parametre ayarlarına sahip alt varyasyonları kullanılarak sperm morfoloji sınıflandırma verisetleri üzerinde detaylı bir çalışma gerçekleştirilmiştir. Tablo 3.7 bu çalışmada kullanılan tüm

Tablo 3.7 Kullanılan transformer mimarilerin ImageNet top-1 doğruluk değerleri

Model	ImageNet top-1 Acc.	Parametre
ViT-S16[74]	80.4	21M
ViT-B16[74]	84.0	86M
ViT-L16[74]	85.7	307M
BEiT-Base[103]	83.2	86M
BEiT-Large[103]	85.2	307M
DEiT-Tiny[83]	72.2	5M
DEiT-Small[83]	79.9	22M
DEiT-Base[83]	81.8	86M

mimarilerin parametre sayılarını ve ImageNet [78] üzerindeki ilk-1 (top-1) doğruluklarını göstermektedir.

- **Performans Metrikleri:**
  Karmaşıklık matrisi, derin öğrenme modellerinin sınıflandırma performansını değerlendirmek amacıyla kullanılan temel bir araçtır. Örnek bir karmaşıklık matrisi, Şekil 3.9'da sunulmaktadır. Bu matris, modelin tahmin ettiği sınıflarla gerçek sınıflar arasındaki ilişkiyi özetleyerek, doğru ve yanlış sınıflandırmaların ayrıntılı bir analizini sağlamaktadır. Sınıflandırma problemlerinde yaygın olarak kullanılan karmaşıklık matrisi, sınıf bazlı doğruluk ve hata oranlarının değerlendirilmesi açısından kritik bir rol oynamaktadır.

  Matris, modelin performansını dört temel bileşenle ifade etmektedir: gerçek pozitif (True Positive, TP) ve gerçek negatif (True Negative, TN) değerleri doğru tahminleri temsil ederken, yanlış pozitif (False Positive, FP) ve yanlış negatif (False Negative, FN) değerleri hatalı tahminleri göstermektedir. Bu bileşenlerden hareketle, modelin genel doğruluk değeri ile birlikte sınıf bazlı Kesinlik (Precision, P), Duyarlılık (Recall, R) ve F1-skoru (F1) gibi önemli performans metrikleri hesaplanabilmektedir.

  Modellerin test setlerinde elde ettikleri doğruluk değerleri, ViT mimarilerinin başarısını değerlendirmek için ana performans ölçütü olarak kullanılmıştır. Eşitlik 3.1'de gösterildiği gibi doğruluk, doğru sınıflandırılan pozitif (TP) ve negatif (TN) örneklerin toplamının, veri kümesindeki tüm örneklerin toplamına bölünmesiyle hesaplanmaktadır. Başka bir ifadeyle, doğruluk değeri, modelin genel sınıflandırma performansını ve tüm veriseti üzerindeki

		GERÇEK DEĞERLER		
		Pozitif	Negatif	Toplam
TAHMİN DEĞERLERİ	Pozitif	TP	FP	Tahmin Pozitif Sayısı
	Negatif	FN	TN	Tahmin Negatif Sayısı
	Toplam	Gerçek Pozitif Sayısı	Gerçek Negatif Sayısı	

Şekil 3.9 Örnek karmaşıklık matrisi

doğruluğunu ifade etmektedir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Ancak, doğruluk metriği, özellikle dengesiz veri kümeleri üzerinde test edilen modeller için tek başına yeterli olmamaktadır. Bu nedenle, kesinlik, duyarlılık ve F1-skoru gibi ek metrikler, modellerin performansı hakkında daha ayrıntılı bilgi sağlamak amacıyla destekleyici ölçütler olarak kullanılmaktadır. Eşitlik 3.2'de tanımlandığı gibi kesinlik, modelin tahmin ettiği pozitif örnekler arasında gerçekten pozitif olanların oranını ifade etmektedir. Kesinlik, doğru pozitif (TP) örneklerin, tahmin edilen toplam pozitif örneklere bölünmesiyle elde edilmekte ve modelin pozitif tahminlerindeki güvenilirliği göstermektedir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \tag{3.2}$$

Duyarlılık ise, Eşitlik 3.3'te gösterildiği gibi, gerçek pozitif örneklerin ne kadarının doğru bir şekilde tespit edildiğini ölçmede kullanılmaktadır. Duyarlılık, doğru pozitif (TP) örneklerin, gerçek pozitif (TP) ve yanlış negatif (FN) örneklerin toplamına oranı ile hesaplanmaktadır. Bu metrik, modelin kaçırdığı pozitif örnekleri değerlendirmede oldukça önemlidir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \tag{3.3}$$

Eşitlik 3.4'te tanımlandığı üzere F1-skoru, kesinlik ve duyarlılık değerlerini tek bir ölçüt altında değerlendirmek için kullanılmaktadır. Bu skor, kesinlik ve duyarlılığın harmonik ortalaması olarak hesaplanmakta ve dengesiz veri kümelerinde modelin performansını daha adil bir şekilde yansıtmaktadır. Ayrıca, F1-skoru, eğitilen modelin sağlamlığını ve genel

başarısını vurgulayan önemli bir ölçüttür.

$$\text{F1-Skor} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{3.4}$$

3.1.2.2 Deneysel Sonuçlar

Tez kapsamında, ViT mimarileri kullanılarak otomatik sperm morfoloji analizine olanak sağlayan bir sistem tasarlanması amaçlanmıştır. Bu amaçla, son zamanlarda görüntü sınıflandırma alanında kullanılan ViT, BEiT ve DEiT mimarilerinin farklı varyasyonlarından sekiz farklı ViT, iki farklı veri kümesi üzerinde sperm morfoloji sınıflandırmasını analiz etmek için kullanılmıştır. Transformer modellerinin eğitimi sırasında, en uygun optimizasyon yöntemi, öğrenme oranı ve veri artırma ölçeğini belirlemek için ayrıntılı bir hiperparametre analizi süreci yürütülmüştür. İlk adım olarak, sekiz farklı ViT mimarisi Adamax, SGD ve RMSProp optimizasyon yöntemleri kullanılarak orijinal veri kümeleri üzerinde eğitilmiştir. Her optimizasyon yöntemi için 10^{-3}, 10^{-4} ve 10^{-5} olmak üzere üç öğrenme oranı seçilmiştir. Her bir veri kümesi için toplam 72 farklı eğitim süreci tamamlanmıştır. Optimizasyon yöntemlerinin ve öğrenme oranlarının ayrıntılı analizinden sonra, her optimizasyon yöntemi için en iyi iki model ve ilgili öğrenme oranı, ayrıntılı bir veri artırma ölçeği analizi yapmak için seçilmiştir. Uygulanan veri artırım ölçek analizi ile sadece veri artırma ölçeklerinin ViT mimarilerinin performansı üzerindeki etkisinin değerlendirilmesi değil, aynı zamanda son teknoloji CNN mimarileri ile karşılaştırıldığında en iyi performansı gösteren ViT mimarilerinin seçilmesi de amaçlanmıştır. Hiperparametre analizinde kullanılan parametre aralıkları Tablo 3.8'de listelenmiştir.

Tablo 3.8 Test edilen hiperparametre listesi

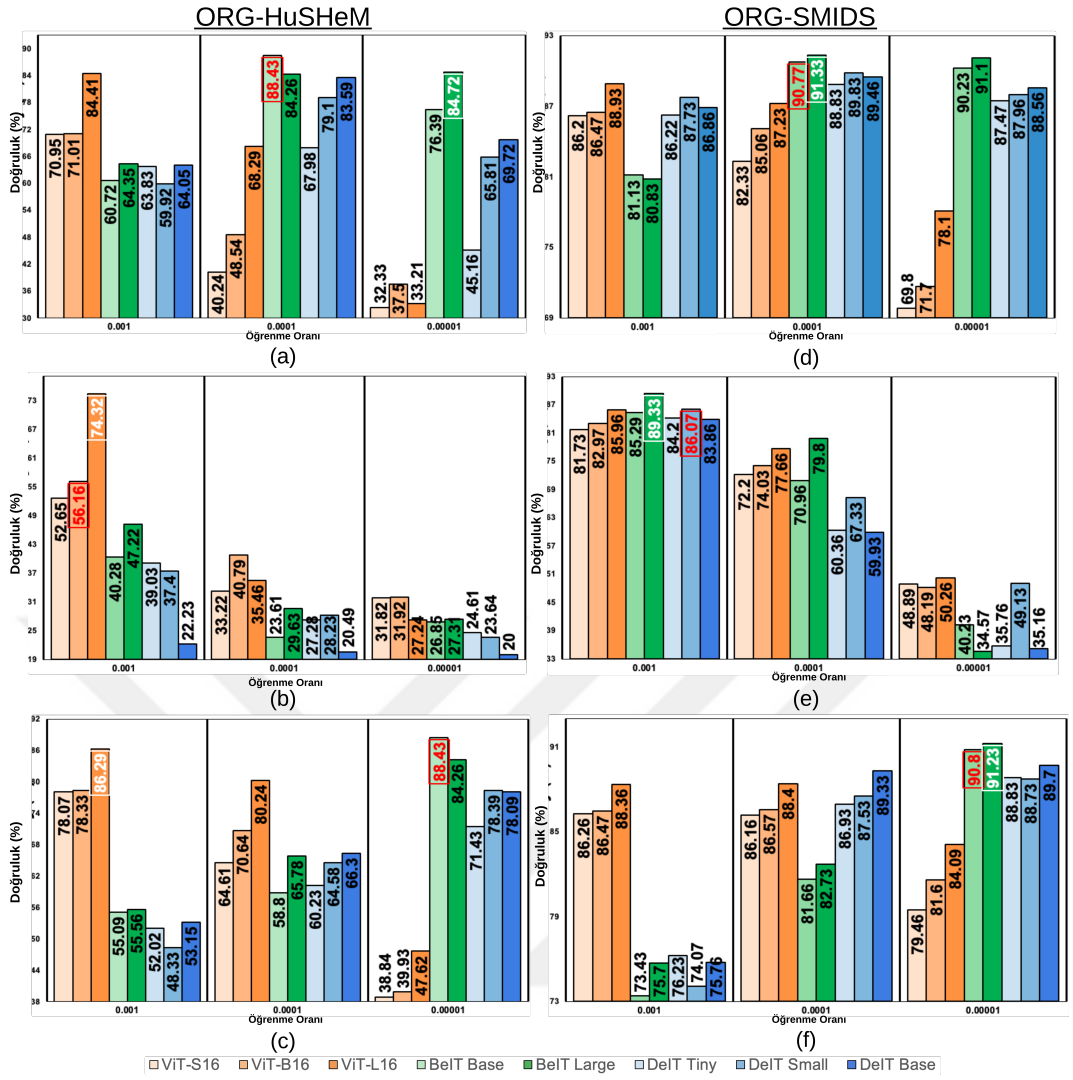
Parametre	Konfigürasyon	
	HuSHeM	SMIDS
Optimizasyon Yöntemi	Adamax, SGD, RMSProp	
Artırım Ölçeği	1× (Orijinal), 5×, 10×, 40×	1× (Orijinal), 3×, 5×, 10×
Öğrenme Oranı	10^{-3} , 10^{-4} , 10^{-5}	
Kayıp Fonksiyonu	Categorical Crossentropy	
Epoch	50	

Adil bir karşılaştırma yapmak için, çeşitli ViT mimariler k-kat çapraz doğrulama şeması kullanılarak değerlendirilmiştir. Veri kümesindeki örnekler rastgele *K* eşit parçaya bölünerek her katta aynı örnek sayısı dağılımına sahip bir yapılandırma elde edilmiştir. *K* değeri literatüre uygun olarak 5 olarak seçilmiştir [26–28, 50, 105]. Veri setinin %20'si test için kullanılırken, örneklerin kalan %80ínin modellerin eğitimi için kullanıldığı K-kat çapraz doğrulama tabanlı eğitim/test

seti ayrımı 5 kez tekrarlanmıştır. Bu çapraz doğrulama işlemi sonucunda, her tekrarda verilerin %20'lik farklı bir kısmı test için seçildiğinden, veri kümesindeki tüm örneklerin en az bir kez test kümesine dahil edilmesi sağlanmıştır. Her bir katın eğitiminde, herhangi bir önyargıdan kaçınmak için dönüştürücüler rastgele ağırlıklarla başlatılmıştır. Ayrıca, farklı mimarilerin aynı verilerle eğitilmesini sağlamak için, eğitim-test bölümleme işlemi başlangıçta bir kez gerçekleştirilmiş ve her tekrar için ilgili K-kat yapıları kaydedilmiştir. Bu yaklaşım, her modelin aynı kattaki aynı görüntüler üzerinde eğitilmesini ve test edilmesini sağlayarak dönüştürücülerin hem kendi aralarında hem de literatürdeki diğer çalışmalarla objektif ve anlamlı bir şekilde karşılaştırılmasına olanak sağlamaktadır.

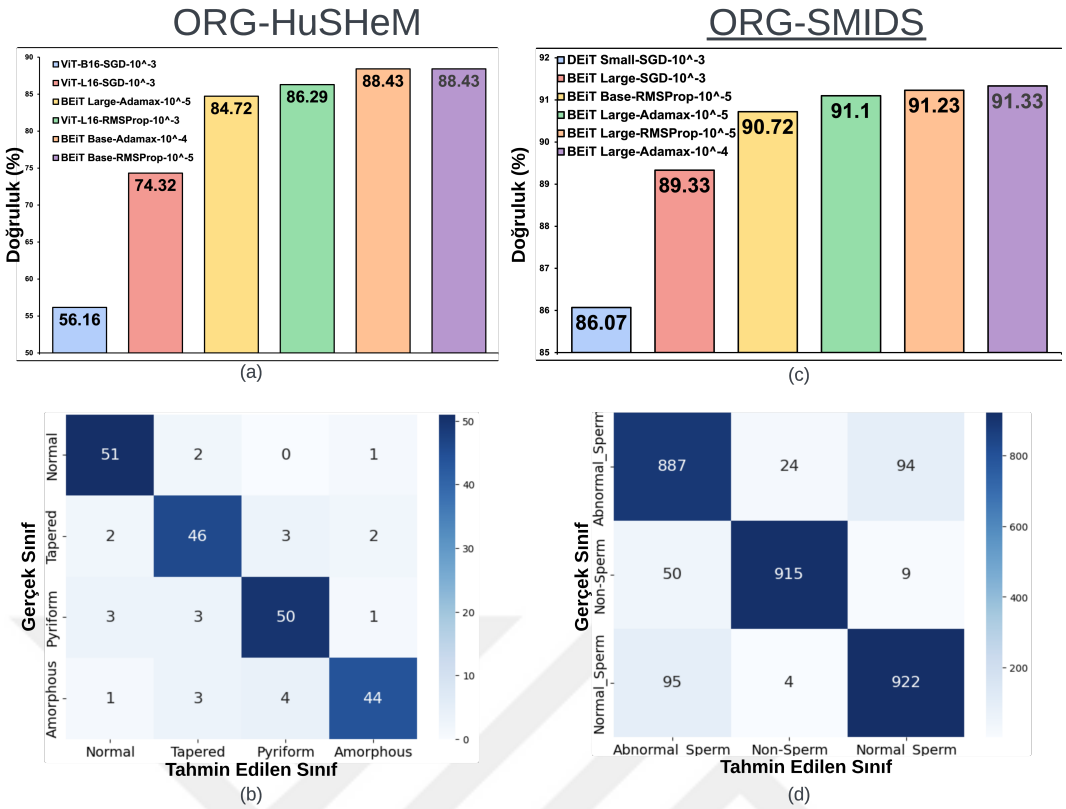
Deneysel çalışmalar, sperm morfoloji sınıflandırması için en iyi performansa sahip dönüştürücü mimarisini bulmak amacıyla her iki veri kümesi üzerinde modellerin eğitilmesiyle başlamıştır. Optimum hiperparametre seçim prosedürü, hangi ViT mimarisinin hangi optimizasyon yöntemi ve öğrenme oranı ile en iyi sonuçları verdiğinin belirlenmesi ile başlamıştır. Bu seçim süreciyle elde edilen her bir optimizasyon yöntemi için en başarılı iki dönüştürücü, veri artırma ölçeğinin sınıflandırma performansı üzerindeki etkisini incelemek için kullanılmıştır.

SMIDS için, sekiz farklı Transformer mimarisi ile her bir optimizasyon yöntemi için ayrı ayrı 24 farklı eğitim işlemi gerçekleştirilerek, toplamda 72 eğitim süreci gerçekleştirilmiştir. SMIDS üzerinde gerçekleştirilen eğitim süreçlerine ait test doğruluk grafikleri Şekil 3.10'un sağ sütununda görülmektedir. Eğitim sırasında üç optimizasyon yöntemi kullanılmıştır: Adamax, SGD ve RMSProp. Her bir optimizasyon yönteminde üç farklı öğrenme oranı kullanılarak elde edilen sonuçlar analiz edilmiştir. Adamax optimizasyon sonuçları incelendiğinde en yüksek doğruluk değeri %91,33 olarak BEiT Large mimarisi ile 10^{-4} öğrenme oranı kullanılarak elde edilmiştir. Aynı mimari 10^{-5} öğrenme oranı ile %91,1 ile ikinci en iyi sonucu elde etmiştir. SGD optimizasyon yöntemi ile elde edilen sonuçların Adamax kadar iyi olmadığı görülmektedir. SGD için en başarılı iki sonuç 10^{-3} öğrenme oranı ile elde edilmiştir. En başarılı iki model olan BEiT Large ve DEiT Small mimarileri sırasıyla %89,33 ve %86,07 doğruluk oranı elde edilmiştir. RMSProp sonuçlarına göre, en başarılı iki model 10^{-5} öğrenme oranı kullanılarak elde edilmiştir. BEiT Large mimarisi, Adamax yöntemine yakın sonuçlar elde eden RMSProp ile %91,23 doğrulukla en başarılı mimari olmuştur. BEiT Base modeli %90,80 doğruluk oranıyla ikinci en iyi model olmuştur. SMIDS üzerinde üç farklı optimizasyon yöntemi kullanılarak elde edilen en başarılı altı mimarinin sonuçları Şekil 3.11'in sağ tarafında ayrıntılı olarak gösterilmiştir. Şekil 3.11'de mimariler sonuçları elde ettikleri en uygun hiperparametrelere göre adlandırılmıştır.



Şekil 3.10 Transformer mimarilerinin Adamax (a,d), SGD (b,e) ve RMSProp (c,f) optimizasyon yöntemleri ve çeşitli öğrenme oranları kullanılarak orijinal (ORG) HuSHeM(a,b,c) ve orijinal (ORG) SMIDS(d,e,f) veri kümeleri üzerindeki doğrulukları

SMIDS veri setinde gerçekleştirilen işlemler, HuSHeM veri setinde de uygulanmıştır. HuSHeM'in test doğrulukları Şekil 3.10'un sol sütununda görülmektedir. Adamax sonuçlarına göre en yüksek doğruluk değeri olan %88,43, 10^{-4} öğrenme oranı kullanan BEiT Base mimarisi ile elde edilmiştir. Bunu 10^{-5} öğrenme oranı kullanarak %84,72 doğrulukla BEiT Large modeli takip etmektedir. SMIDS veri setinde olduğu gibi, SGD optimizasyon yöntemi HuSHeM veri setinde de çok başarılı olamamıştır. SGD için en başarılı sonuç, 10^{-3} öğrenme oranı kullanılarak %74,32 doğrulukla ViT-L16 mimarisi tarafından elde edilmiştir. Aynı öğrenme oranını kullanan ViT-B16 mimarisi ise %56,16 doğruluk değeri ile en başarılı ikinci model olmuştur. SGD ile elde edilen sonuçlar çok tatmin edici olmasa da, veri artırma ölçeği ile ilişkilerini incelemek için sonraki süreçlerde kullanılmıştır. RMSProp ile elde edilen sonuçlara göre BEiT Base mimarisi 10^{-5}



Şekil 3.11 a, c)Her üç optimizasyon yöntemindeki en iyi iki model için optimum hiperparametre yapılandırmaları ve doğruluk değerleri, b) ORG-HuSHeM üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi, d) ORG-SMIDS üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi

öğrenme oranı kullanılarak elde edilen %88,43 doğruluk ile en başarılı yöntem olmuştur. ViT-L16, 10^{-3} öğrenme oranıyla %86,29 ile ikinci en yüksek doğruluğu elde etmiştir. HuSHeM veri seti için en başarılı altı mimarinin sonuçları Şekil 3.11'in sol tarafında görülmektedir.

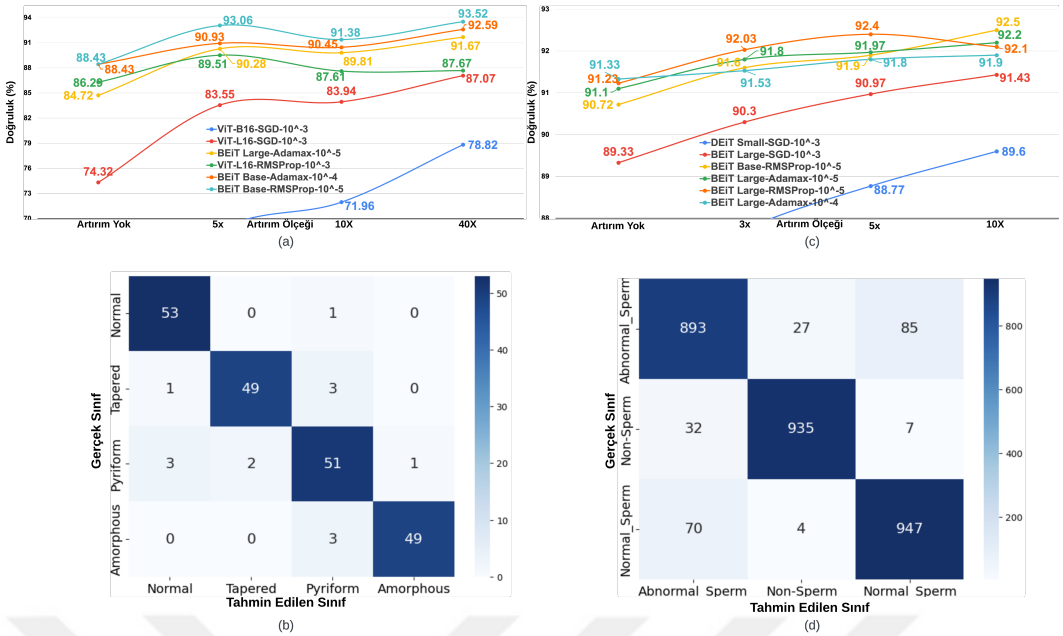
Çeşitli ölçekler kullanılarak veri artırımının uygulanmasının sınıflandırma performansı üzerindeki etkisi, orijinal veri kümeleri üzerinde gerçekleştirilen optimum parametre seçim sürecinden elde edilen hiperparametreler ve mimariler kullanılarak analiz edilmiştir. SMIDS, orijinal (ORG) eğitim veri kümesinin 3 katını temsil eden $3\times$, 5 katını temsil eden $5\times$ ve 10 katını temsil eden $10\times$ kadar artırılmıştır. SMIDS için seçilen altı mimari, Şekil 3.11'nin sağ tarafında gösterildiği gibi, en uygun optimizasyon yöntemi ve öğrenme oranıyla her bir büyütme ölçeğinde ayrı ayrı eğitilmiştir. Orijinal verilerin eğitimi sırasında kullanılan aynı K-kat çapraz doğrulama şeması bu eğitim süreçleri için de kullanılmıştır. SMIDS için veri artırma sürecinin sonunda, orijinal veri kümesine ek olarak üç farklı artırılmış eğitim veri kümesi elde edilmiştir. Seçilen altı öğrenme mimarisi bu üç artırılmış veri kümesi üzerinde

Tablo 3.9 BEiT Base-RMSProp-10^{-5}'in Şekil 3.11 ve 3.12'den elde edilen karmaşıklık matrisi tabanlı performans ölçütleri bakımından değerlendirilmesi

		HuSHeM				SMIDS		
		Normal	Tapered	Pyriform	Amorphous	Normal	Abnormal	Non-Sperm
Orijinal Veriseti	Kesinlik	0,89	0,85	0,87	0,92	0,86	0,97	0,9
	Duyarlılık	0,94	0,87	0,88	0,85	0,88	0,95	0,9
	F1 Skor	0,92	0,86	0,88	0,88	0,87	0,95	0,9
	Doğruluk	88,43%				90,80%		
Artırılmış Veriseti (40x-10x)	Kesinlik	0,91	0,96	0,88	0,98	0,90	0,97	0,91
	Duyarlılık	0,98	0,92	0,89	0,94	0,89	0,96	0,93
	F1 Skor	0,94	0,94	0,88	0,96	0,89	0,96	0,92
	Doğruluk	93,52%				92,5%		

en uygun parametre ayarları kullanılarak eğitilmiş ve 18 farklı eğitim süreci gerçekleştirilmiştir. Veri artırma ölçeğine göre SMIDS için elde edilen sonuçlar Şekil 3.12'ın sağ tarafında görülmektedir. Adil karşılaştırmalar yapabilmek için her bir öğrenme modeli, daha önce orijinal veri kümesinin eğitim sürecinde ayarlanmış olan aynı optimizasyon yöntemi ve öğrenme oranı kullanılarak eşit sayıda epoch kadar (50) eğitilmiştir. En yüksek doğruluk artışı DEiT Small-SGD-10^{-3} mimarisinde %86,07'den %89,6'ya sıçrama olarak gözlemlenirken, %92,5'lik en yüksek doğruluk değeri BEiT Base-RMSProp-10^{-5} mimarisi tarafından elde edilmiştir. Neredeyse tüm modellerde, veri artırma ölçeği arttıkça mimarilerin başarısı da artmıştır. Bu eğilimin tek istisnası, BEiT Large-RMSProp-10^{-5} modelinde, artırma ölçeği 5×'dan 10×'a değiştiğinde %0,3'lük ihmal edilebilir bir düşüş olarak gözlemlenmiştir. En iyi performansla ilgili olarak, 5× veri artırma ölçeğinde en yüksek değer olan %92,4 doğruluğa ulaşan BEiT Large-RMSProp-10^{-5} mimarisi hariç, neredeyse tüm öğrenme modellerinde 10× veri artırma ölçeğinde en yüksek doğruluk değerleri elde edilmiştir. En başarılı iki öğrenme modelinin performansı karşılaştırıldığında, BEiT Large-RMSProp-10^{-5} mimarisi ile Beit Base-RMSProp-10^{-5} mimarisi arasında sadece %0,1'lik bir fark olduğu görülmüştür. Ancak, model karmaşıklığı karşılaştırıldığında, BEiT Large-RMSProp-10^{-5} mimarisi 307M parametreye sahipken, BEiT Base-RMSProp-10^{-5} mimarisi sadece 86M parametreye sahiptir. Bu nedenle, sperm morfoloji sınıflandırması için Beit Base-RMSProp-10^{-5} mimarisini seçmek daha uygun olacaktır.

Her iki veri kümesi üzerinde elde edilen sonuçlar incelendiğinde BEiT Base-RMSProp-10^{-5} mimarisinin sperm morfoloji sınıflandırması için en başarılı model olduğu görülmektedir. Tablo 3.9, Şekil 3.11 ve Şekil 3.12'de verilen karmaşıklık matrisleri kullanılarak oluşturulmuştur. Tablo 3.9'da BEiT Base-RMSProp-10^{-5} mimarisi ile eğitilen modellerin hem orijinal veri kümesi üzerinde hem de 10× veri ile artırılmış SMIDS ve 40× veri ile artırılmış HuSHeM üzerinde P, R ve F1 değerleri sunulmuştur. Pozitif tahmin değeri olarak da



Şekil 3.12 a, c) Optimum hiperparametre seçim süreciyle elde edilen altı mimarinin veri artırma ölçeklerine göre sınıflandırma doğruluğu değerleri, b) 40×-HuSHeM üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi, d) 10×-SMIDS üzerinde eğitilen BEiT Base-RMSProp-10^{-5} modelinin karmaşıklık matrisi

adlandırılan P skorları karşılaştırıldığında, en uygun optimizasyon yöntemi ve öğrenme oranı ile eğitilen Beit Base mimarisi her iki veri kümesinde de iyi performans göstermiştir.

HuSHeM'de en yüksek artış %11 ile Tapered anormallik sınıfında elde edilirken, SMIDS'de artırılmış veriler kullanıldığında en yüksek artış %4 ile Normal sınıfında elde edilmiştir. HuSHeM'deki Amorf anormallik sınıfında da kesinlik skorunda umut verici bir iyileşme elde edilmiştir. Orijinal ve artırılmış testler arasındaki duyarlılık skorlarındaki fark incelendiğinde, Normal sınıfı için yanlış negatif değerleri önemli ölçüde azaltılarak HuSHeM için en yüksek duyarlılık değerine ulaşılmıştır. Öte yandan, duyarlılık değerinde en yüksek artış %9 ile Amorf sınıfında elde edilmiştir. SMIDS ile ilgili olarak, tüm sınıflarda duyarlılık değerlerinde bir artış gözlenmiştir. F1-skorları incelendiğinde, HuSHeM'de Pyriform sınıfı hariç tüm sınıflarda anlamlı bir artış elde edilmiştir. SMIDS veri setinde ise tüm sınıflarda F1-skor artışı sağlanmıştır. Genel doğruluk değerleri, artırılmış veri kullanılarak HuSHeM için %5,09 ve SMIDS için %1,7 artış göstermektedir. Elde edilen sonuçlar ışığında, veri artırım ölçeğinin model başarısı üzerinde olumlu bir etkiye sahip olduğu görülmektedir.

3.1.2.3 Tartışma ve Sonuç

Sperm morfoloji analizi, kısırlık tanısında kritik öneme sahiptir. Kliniklerde iki değerlendirme yöntemi kullanılmaktadır: GD ve BDSA değerlendirmeleri. Geleneksel yöntem olarak kabul edilen GD başarısı büyük ölçüde uzmanların deneyimine bağlıdır ve bu da potansiyel öznelliklere ve yanlışlıklara yol açmaktadır. BDSA tabanlı değerlendirmeler görsel değerlendirmeye kıyasla daha tutarlı ve tarafsız analizler sağlayabilmektedir. Bununla birlikte, bu sistemlerin çoğu sperm morfoloji analizi için tam otomatik bir çözüme sahip değildir. Bu sorunları ele almak için, son yıllarda sperm morfoloji analizi alanında önemli araştırmalar yapılmaktadır. Bu kapsamda, geleneksel görsel değerlendirme yöntemlerine ve geleneksel derin öğrenme yaklaşımlarına alternatif olarak daha doğru sınıflandırma performansına sahip ViT tabanlı bir analiz sistemi önerilmektedir.

Geleneksel makine öğrenmesi ve derin öğrenme yöntemleri tıbbi görüntü sınıflandırma alanında yaygın olarak kullanılmaktadır. Ancak son yıllarda dikkat mekanizması tabanlı ViT mimarileri, geleneksel CNN mimarilerine alternatif çözümler sunarak görüntü sınıflandırma alanında popülerlik kazanmaya başlamıştır. Sperm morfoloji analizinde, çeşitli anormallik türlerini ayırt etmek için spermlerin dokularından ziyade şekilleri daha önemlidir. ViT mimarileri, CNN mimarileri ile karşılaştıran deneysel çalışmalar, ViT mimarilerin görüntü sınıflandırmada dokuya bağlı olmaktan çok şekle bağlı olduğunu ortaya koymaktadır [89]. Bu nedenle, bu tez çalışmasında sperm morfoloji sınıflandırması için ViT mimarilerin performansları detaylı bir şekilde analiz edilmiştir. ViT mimariler, CNN mimarileri gibi veriye aç mimarilerdir ve örüntü öğrenme sürecinde önemli miktarda veri sağlanması gerekmektedir. Ayrıca, modellerin eğitimi sırasında seçilen hiperparametreler modelin başarısını önemli ölçüde etkileyebilmektedir. Bu bağlamda, ViT mimarilerinin eğitim sürecindeki en uygun hiperparametreler belirlenmiş ve veri artırım ölçeği ile performansları arasındaki ilişki detaylı bir şekilde analiz edilmiştir.

Sperm morfoloji analiz sistemlerinin performansını etkileyen en önemli faktörlerden biri modelin genelleme yeteneğidir. Modellerin farklı veri kümeleri üzerindeki yüksek doğruluğu, modelin güçlü genelleme kapasitesini göstermektedir. Sunulan çalışmanın deneylerinde literatürde yaygın olarak kullanılan sperm morfoloji veri setleri olan HuSHem ve SMIDS kullanılmıştır. Deneyler sonucunda, RMSProp optimizer ve 10^{-5} öğrenme oranına sahip Beit Base mimarisi, HuSHeM ve SMIDS üzerinde sırasıyla %93,52 ve %92,5 doğruluk değerleri elde ederek literatürdeki en gelişmiş makine öğrenmesi ve derin öğrenme yöntemlerinden daha iyi performans göstermiştir. Literatürdeki çalışmaların

Tablo 3.10 Önerilen yaklaşım ile literatürdeki yaklaşımlar arasında sperm morfoloji sınıflandırmasının karşılaştırılması

Veriseti Adı	Veriseti Bilgileri	Döndürme ve Kesme	Yayın	Yöntem	Doğruluk (%)
HuSHeM	216 Sperm Görüntüsü 131 × 131 Çözünürlük Üst üste bindirilmiş Dört Sperm Sınıfı	Manuel	Shaker ve ark. 2017 [7]	Manuel olarak Sperm Başı Döndürme ve 50 × 76 piksele kırpma Renk Uzayı Dönüştürme Sözlük Öğrenme	92.2
		Otomatik	Ilhan ve ark. 2019 [106]	Otomatik Yönlü Maskeleme k-NN sınıflandırması	57,4
		Manuel	Riordon ve ark. 2019 [25]	Veri Artırma VGG16 (ImageNet) kullanarak Transfer Öğrenimi	94
		Otomatik	Ilhan ve ark. 2020 [105]	Uyarlanabilir Gürültü Giderme Örtüşen Grup Küçültme Görüntü Gradyanı Yönlü Maskeleme MSER tanımlayıcısı ile Özellik Çıkarma SVM Sınıflandırma (doğrusal olmayan Çekirdek)	86,6
		Uygulanmadı	Yuzkat ve ark. 2021 [28]	Veri Artırma CNN Mimarileri Kullanarak Transfer Öğrenme Karar Seviyesi Füzyonu	85,18
		Uygulanmadı	Ilhan ve ark. 2021 [26]	Veri Artırma VGG16 ve GoogleNet (ImageNet + SMIDS) kullanarak Transfer Öğrenimi Karar Seviyesi Füzyonu (VGG16 + GoogleNet)	92,13
		Yarı-Manuel	Spencer ve ark. 2022[55]	Veri Artırma Transfer Öğrenme VGG16, VGG19, Resnet-34 ve DenseNet-161 Meta sınıflandırıcı	98,2
		Uygulanmadı	Mahali ve ark. 2023 [43]	Uygunsuz Veri Artırımı Transfer Öğrenmesi SwinTransformer, MobileNetV3 ve Otomatik kodlayıcı Füzyon modelleri (SwinMobile)	97,6
		Uygulanmadı	Aktas ve ark. 2023 [27]	Veri Artırma Transfer Öğrenme CNN'ler ve Transformerlar	90,85
		Uygulanmadı	Önerilen Tez Çalışması 2024	Veri Artırımı Transfer Öğrenme ViT, DEiT ve BEiT varyasyonları	93,52
SMIDS	536 Sperm Yaması 190 x 170 Çözünürlük Üst üste bindirilmiş Üç Sınıf	Uygulanmadı	Ilhan ve ark. 2018 [50]	Çift Ağaç Karmaşık Dalgacık Trans. SVM sınıflandırması	82,3
			Ilhan ve ark. 2018 [107]	Uyarlanabilir Gürültü Giderme SURF Tanımlayıcısı ile Özellik Çıkarma SVM sınıflandırması	83,4
	3000 Sperm Yaması 190 x 170 Çözünürlük Üst üste bindirilmiş Üç Sınıf	Uygulanmadı	Ilhan ve ark. 2020 [37]	Örtüşen Grup Küçültme Bulanık-C Araçları Kümeleme Veri Artırma MobileNet ile Sınıflandırma	87
		Otomatik	Ilhan ve ark. 2020 [105]	Uyarlanabilir Gürültü Giderme Örtüşen Grup Küçültme Görüntü Gradyanı Yönlü Maskeleme MSER tanımlayıcısı ile Özellik Çıkarma DVM Sınıflandırması (doğrusal olmayan Çekirdek)	85,7
		Uygulanmadı	Yuzkat ve ark. 2021 [28]	Veri Artırma CNN Mimarileri Kullanarak Transfer Öğrenme Karar Seviyesi Füzyonu	90,73
		Uygulanmadı	Ilhan ve ark. 2021 [26]	Veri Artırma VGG16 ve GoogleNet (ImageNet) kullanarak Transfer Öğrenimi Karar Seviyesi Füzyonu (VGG16 + GoogleNet)	90,87
		Uygulanmadı	Mahali ve ark. 2023 [43]	Uygunsuz Veri Artırımı Transfer Öğrenmesi SwinTransformer, MobileNetV3 ve Otomatik kodlayıcı Füzyon modelleri (SwinMobile)	91,7
		Uygulanmadı	Aktaş ve ark. 2023 [27]	Veri Artırma Transfer Öğrenme CNN ve Transformerlar	89,43
		Uygulanmadı	Önerilen Tez Çalışması 2024	Veri Artırımı Transfer Öğrenme ViT, DEiT ve BEiT varyasyonları	92,5

sonuçları ile önerilen çalışma ile elde edilen sonuçların detaylı bir karşılaştırması Tablo 3.10'da sunulmuştur.

Sözlük öğrenme yöntemi kullanılarak %92,2 doğruluk elde edilen HuSHeM ile ilgili ilk çalışma 2017 yılında Shaker ve arkadaşları tarafından gerçekleştirilmiştir [7]. Bir başka çalışmada Riordon ve arkadaşları transfer öğrenme yöntemiyle VGG-16 modelini kullanarak %94 doğruluk elde etmişlerdir [25]. Ancak, her iki çalışmada da HuSHeM veri setine manuel döndürme ve kırpma ön işleme adımı uygulanmış ve bu da onları tam otomatik sperm analiz sistemlerinin geliştirilmesi için uygunsuz hale getirmiştir. İlhan ve ark. [105] bu sorunu çözmek için yönlü maskeleme tekniği ile yeni yaklaşımlar kullanarak %86,6 doğruluk elde etmiştir. Bu yaklaşım manuel müdahale gerekliliğini ortadan kaldırsa da, veriler üzerinde rotasyon gibi ön işlemlerin uygulanması hala devam etmektedir. 2021 yılında İlhan ve arkadaşları VGG16 ve GoogleNet mimarilerini karar seviyesinde birleştirerek yeni bir yaklaşım ortaya koymuştur [26] . Bu yöntemde, modelleri sperm görüntülerini sınıflandırmaya daha uygun hale getirmek için iki aşamalı bir ince ayar işlemi gerçekleştirilmiştir. Tam otomatik ve ön işleme gerektirmeyen sistemler arasında, HuSHeM üzerinde %92,13 doğrulukla en modern yaklaşımı elde etmişlerdir. Spencer ve arkadaşları [55] geleneksel CNN mimarilerini birleştirerek sperm morfoloji sınıflandırması yapmış ve %98,2 doğruluk elde etmiştir, ancak görüntüler [7, 25] çalışmalarında olduğu gibi manuel olarak kırpılmış ve döndürülmüştür. Diğer çalışmalardan farklı olarak, sperm başları sola dönük olduğundan, veri büyütme için yalnızca dikey çevirme kullanılmıştır. Bu nedenle tam otomatik olarak kullanılamamaktadır. Mahali ve arkadaşları CNN mimarileri ile ViT mimarilerini birleştirerek SwinMobile yöntemini önermişlerdir [43]. HuSHeM üzerinde %97,6 doğruluk elde etmelerine rağmen, çalışmalarında veri artırma işleminin test verileri üzerinde de uygulandığı belirtilmiştir. Test verisi üzerinde veri artırma işleminin uygulanması öğrenme sürecindeki yanlılığı artırmakta ve modelin gerçekçi ve güvenilir olmayan yüksek doğruluk değerlerine ulaşmasını sağlamaktadır. Aktaş ve arkadaşları sperm morfoloji analizinde CNN ve ViT mimarilerinin performanslarını karşılaştırmışlardır. ViT-L16 modeli ile HuSHeM üzerinde %90,85 doğruluk değerine ulaşılmıştır. Önceki çalışmalarda HuSHeM veri setine uygulanan manuel ön-işlem adımları ve uygunsuz veri artırma yöntemlerinin aksine, bu çalışmada 10^{-5} öğrenme oranı konfigürasyonuna sahip RMSProp ile optimize edilmiş Beit Base mimarisi kullanılarak, son teknoloji yaklaşımlara kıyasla %93,52 doğruluk oranıyla en yüksek performans elde edilmiştir. Böylece, yanlılık içermeyen ve tam otomatik bir sperm morfoloji analiz sistemi önerilmiştir.

SMIDS ile ilgili çalışmalar incelendiğinde, SMIDS-small veri kümesi üzerinde

ilk çalışma 2018 yılında İlhan ve arkadaşları tarafından yapıldığı görülmektedir [50]. SMIDS verisetinin genişletilmiş bir versiyonu ilk olarak [37]'de tanıtılmış ve MobileNet kullanılarak %87 doğruluk elde edilmiştir. Geleneksel makine öğrenimi yöntemlerine alternatif olarak, İlhan ve arkadaşları [26] SMIDS üzerinde karar düzeyinde füzyon gerçekleştirmiş ve %90,87 gibi daha yüksek bir doğruluk değeri elde etmişlerdir. Benzer şekilde, Yüzkat ve arkadaşları çoklu CNN mimarilerini olasılık (soft) ve çoğunluk (hard) oylama topluluk yöntemleriyle birleştirmiş ve %90,73 doğruluk elde etmişlerdir [28]. HuSHeM'de görülen önyargıya benzer şekilde, Mahali ve arkadaşları [43] SMIDS üzerinde de uygunsuz veri artırımı gerçekleştirmiş ve %91,7 doğruluk elde etmişlerdir. Aktaş ve arkadaşları SMIDS üzerinde ViT-L16 dönüştürücü mimarisi ile %89,43 doğruluk elde etmiş, bu da literatürde daha önce sunulan son teknoloji CNN tabanlı çözümlerden daha iyi performans gösterememiştir. Bu çalışmada, 10^{-5} öğrenme oranı ayarlarına sahip RMSProp tarafından optimize edilen BEiT Base modeli, literatürdeki en yüksek doğruluğa sahip yöntemden yaklaşık %2 daha iyi olan %92,5 doğruluk elde etmiştir. Sperm morfoloji veri kümelerinde sperm, sperm başının şekline göre farklı kategorilerde sınıflandırılmaktadır. Sınıflandırmada spermin dokusundan ziyade şekli daha ayırt edici bir özelliktir. ViT mimarileri yapısal olarak görüntülerdeki nesnelerin dokularından ziyade şekillerine daha çok bağımlıdır. Bu da önerilen yaklaşımı SMIDS üzerinde çalışılan sperm morfoloji analiz problemi için en iyi performanslı tam otomatik çözüm (herhangi bir manuel ön işleme adımı olmadan) haline getirmektedir.

3.1.2.4 Sınırlamalar ve Gelecek Çalışmalar

Optimizasyon yöntemleri ve öğrenme oranları incelendiğinde RMSProp ve Adamax yöntemlerinin SGD yöntemine göre daha başarılı olduğu gözlemlenmiştir. Adil bir karşılaştırma sağlamak amacıyla tüm yöntemlerde aynı epoch sayısı kullanılmıştır. Ancak, SGD optimizasyon yöntemi, her üç öğrenme oranında da belirlenen epoch süresi boyunca diğer yöntemlere kıyasla anlamlı bir yakınsama gösterememiştir. RMSProp ve Adamax optimizasyon yöntemleri değerlendirildiğinde, her iki yöntemin de en başarılı sonuçları sırasıyla 10^{-5} ve 10^{-4} öğrenme oranlarında elde ettiği gözlemlenmiştir. Bu bulgu, farklı optimizasyon yöntemlerinin ve öğrenme oranlarının model başarısı üzerinde önemli bir etkiye sahip olduğunu ortaya koymaktadır. Bu doğrultuda, gelecekteki çalışmalarda daha fazla optimizasyon yöntemiyle kapsamlı bir deneysel süreç tasarlanarak, optimizasyon stratejilerinin ViT mimarilerinin öğrenme süreçlerine etkisi daha ayrıntılı bir şekilde incelenecektir. Deneysel bulgular, ViT mimarilerinin artan veri artırma ölçeği ile pozitif bir ilişkiye sahip olduğunu göstermektedir. Mevcut

veri kümelerinin artırılmasıyla, daha yüksek performans sergileyen modellerin oluşturulması hedeflenmektedir. Geleneksel veri artırma teknikleri, veri çeşitliliğini belirli bir noktaya kadar artırabilmekte olup, özellikle HuSHeM gibi sınırlı sayıda örnek içeren veri kümelerinde, döndürme ve yansıtma gibi yöntemler belirli bir eşiğin ötesinde çeşitliliği önemli ölçüde artıramamaktadır. Bu nedenle, veri kalitesini iyileştirmek ve daha fazla sayıda örnek içeren yeni veri kümeleri oluşturmak amacıyla Çekişmeli Üretici Ağlar (Generative Adversarial Networks - GAN) [108] gibi ileri düzey üretici modellerin kullanılması önerilmektedir. Modellerin daha çeşitli veri kümeleriyle eğitilmesi, farklı senaryolar için daha sağlam ve genellenebilir çözümler sunan modellerin geliştirilmesine olanak tanıyacaktır. Ayrıca, ViT mimarilerinin kendi aralarında ve CNN mimarileri ile birleştirilerek farklı veri kümeleri üzerindeki performanslarının karşılaştırılması hedeflenmektedir. Bu yaklaşım ile CNN mimarisinin genelleme yeteneği ve ViT mimarilerinin şekil bağımlı öğrenme kapasitesinin bir araya getirilmesiyle daha yüksek doğruluk oranlarına ulaşılabileceği öngörülmektedir. Ek olarak, İlhan ve arkadaşları tarafından önerilen iki aşamalı ince ayar sürecinin [26] ViT mimarilerine uygulanarak model başarısı üzerindeki etkilerinin detaylı bir şekilde analiz edilmesi planlanmaktadır.

3.2 Hi-LabSpermMorpho Verisetinin Sınıflandırılması

Sperm morfoloji analiz alanında önemli ilerlemeler kaydedilmiş olmasına rağmen, her türde sperm anormalliğini tam otomatik olarak analiz edebilen kapsamlı bir sistem hala geliştirilememiştir [26, 27, 29, 34, 38, 69, 106, 107, 109]. Temel sorun, yüksek genelleme kabiliyetine sahip modellerin eğitilmesinin önünü açabilecek yeterli veri setinin henüz literatürde mevcut olmamasıdır. Mevcut çalışmalar var olan veri setleri üzerinde başarılı sonuçlar verse de, var olan veri setlerindeki etiketli sınıf sayısının kısıtlı olması nedeniyle farklı anormallik türlerini tespit etme konusunda sınırlı kalmaktadır. Literatürdeki çoğu veriseti yalnızca sperm baş anormalliklerine odaklanmaktadır [7, 19, 29]. Diğer anormallik türleri (orta parça ve kuyruk) için ayrıntılı etiketlenmiş veri kümeleri mevcut değildir. Ancak, Rousso ve arkadaşları pyriform başlı spermler ile bükülmüş boyunlu ve kıvrılmış kuyruklu spermler arasında anlamlı bir pozitif korelasyon olduğunu çalışmalarında vurgulamaktadır [110]. Ayrıca, Menkveld ve arkadaşları, kısırlık tedavi yöntemlerinin anormallik türüne bağlı olarak farklılık gösterebileceğini belirtmiştir [11]. Bu doğrultuda, geniş kapsamlı bir anormallik sınıfı içeren veri setlerinin oluşturulması, tam otomatik sperm morfoloji analiz sistemlerinin geliştirilmesi açısından büyük önem taşımaktadır.

Mevcut veri setleri ya tek tip boyama ya da boyama yapılmayan yöntemler kullanılarak elde edilmiştir [7, 21, 29, 37]. Katz ve arkadaşları, Papanicolaou ve ıslak preparat boyama yöntemlerinin sperm morfolojisi üzerindeki etkilerini manuel değerlendirme yoluyla karşılaştırmıştır. Papanicolaou boyamasının ıslak preparasyon konfigürasyonlarına kıyasla sperm başının küçülmesine neden olduğu gözlemlenmiştir [111]. Benzer şekilde, Aksoy ve arkadaşları, manuel değerlendirme sürecinde farklı boyama yöntemlerinin spermin çeşitli bölgelerini daha belirgin hale getirdiğini göstermiştir [22]. Ancak çalışmalarında, derin öğrenme modellerinin eğitimi için ayrıntılı ve yeterli bir veri seti oluşturmamışlardır. Bu tez çalışmasının motivasyonlarından biri, derin öğrenmeye uygun farklı boyama teknikleri kullanılarak oluşturulan bir veri setinin, literatürdeki çalışmaların bulgularını doğrulamak ve genişletmek adına değerli bir kaynak sağlayacak olmasıdır [11, 22, 110, 111].

Mevcut veri setlerinin örneklem büyüklüğü, etiketlenmiş sperm anormallik türlerinin çeşitliliği ve tek tip boyama tekniğine dayanması gibi sınırlamalar nedeniyle, daha kapsamlı bir veri seti oluşturmaya önemli bir ihtiyaç vardır. Bu tez kapsamında, bu ihtiyaçları karşılamak için "HiLab-SpermMorpho" veri seti oluşturulmuştur [42]. Bu yeni veri seti, daha fazla sayıda sperm örneği sunarak, geniş bir yelpazede sperm morfoloji anormallik türlerini kapsamakta ve birden fazla boyama tekniği kullanarak mevcut veri setlerinde görülen eksiklikleri gidermeyi hedeflemektedir. Bu veri seti, sperm morfoloji sınıflandırması için tam otomatik sistemler oluşturmayı amaçlayan araştırmacılar için değerli bir kaynak görevi görecektir. HiLab-SpermMorpho, sperm morfoloji analizi alanında kullanılmak üzere DSÖ standartlarına göre uzmanlar tarafından etiketlenmiştir. Üç farklı Diff-Quick tabanlı boyama yöntemi (BesLab, Histoplus ve GBL) kullanılarak oluşturulmuş ve dört ana kategoride 18 sınıfa ayrılmış toplam 49.345 görüntüden oluşmaktadır: baş, orta parça (boyun), kuyruk ve normal. Baş kategorisinde amorphous, tapered, double, pyriform, pin, vacuolated, narrow acrosome area ve round yer almaktadır. Boyun kategorisi thin, thick, twisted ve asymmetrical olanları kapsamaktadır. Kuyruk kategorisi double, curly, long, short ve twisted sınıflarından oluşmaktadır. Dördüncü kategori normal sperm hücrelerinden oluşmaktadır. Veri setinin kullanılabilirliğini göstermek amacıyla 35 farklı derin öğrenme mimarisi kullanılarak sınıflandırma çalışmaları gerçekleştirilmiş ve temel performans sonuçları sunulmuştur.

3.2.1 Materyal ve Yöntem

Tez kapsamında, Hi-LabSpermMorpho veri seti için sınıflandırma performansı temel performans oluşturmak amacıyla 35 farklı derin öğrenme mimarisi

değerlendirilmiştir. Bu modeller, literatürde yaygın olarak kullanılmaları, yüksek performans göstermeleri, hesaplama verimlilikleri ve sperm morfoloji sınıflandırma çalışmalarındaki kullanım durumlarına dayanarak seçilmiştir. Hi-LabSpermMorpho veri setinin detaylı anlatımı bir önceki kısım olan Bölüm 2 Materyaller kısmında sunulmuştur.

3.2.1.1 Kullanılan Modeller

AlexNet, 2012 ImageNet yarışmasını kazanan ve derin öğrenme alanında devrim yaratan öncü mimarilerden biridir [87]. Modern mimarilere göre daha az katmana sahip olmasına rağmen, daha derin mimarilerle karşılaştırma yapmak için başlangıç noktası olarak tercih edilmiştir. AlexNet'ten daha derin olan VGG16 ve VGG19, sadelik ile performans arasında bir denge sunmaktadır [75]. DenseNet mimarisi ise, her bir katmanın önceki katmanlardan gelen özellik haritalarına erişmesine olanak tanıyarak gradyan akışını iyileştirmekte ve özelliklerin yeniden kullanımını sağlamaktadır [76]. DenseNet121 ve DenseNet201 gibi varyantlarla temsil edilen bu yapı, tıbbi görüntüleme ve biyoinformatik gibi alanlarda etkili ve hassas sonuçları nedeniyle yaygın olarak kullanılmaktadır.

ResNet mimarileri, ResNet18, ResNet50, ResNet101 ve ResNet152 gibi varyantlarıyla, çok derin ağlarda karşılaşılan gradyan kaybolması gibi sorunları çözmek için artık öğrenme (residual learning) kavramını tanıtmıştır [112]. Daha derin katmanlara sahip olmalarına rağmen, bu mimariler artık bağlantılar (residual connections) aracılığıyla verimli öğrenmeyi sağlamaktadır. ResNeXt50 ve ResNeXt101 mimarileri, özellik çıkarımını ve genel performansı artırmak amacıyla mevcut mimarilere paralel yolların sayısını belirleyen kardinalite adlı bir hiperparametrenin eklenmesiyle oluşturulmuştur [113]. Benzer şekilde, WideResNet50 ve WideResNet101 mimarilerinde katmanlar genişletilerek parametre sayısı ve özellik çıkarım kapasitesi artırılmıştır [114].

GoogLeNet (InceptionV1) [115] ve InceptionV3 [116], daha derin ve geniş ağların performansını değerlendirmek amacıyla seçilmiştir. İnception modülleri ile karakterize edilen bu modeller, farklı ölçeklerde özellik çıkarımı yaparak çok sınıflı veri setlerinde başarılı sonuçlar elde etmektedir. Verimlilik ve performans arasında denge sağlamak için EfficientNet modelleri de değerlendirmeye dahil edilmiştir. EfficientNetB0-B7 [45] ve EfficientNet-V2 Small, Medium ve Large [117] varyantları, farklı hesaplama kısıtları altında optimize edilmiş performans sunan ölçeklenebilir yapılarıyla bilinmektedir.

ShuffleNetV2 varyantları (ShuffleNetV2, ShuffleNetV2 x1.5 ve ShuffleNetV2

x2), hafif ve hızlı performansları sayesinde mobil ve düşük güç gereksinimi olan uygulamalar için ideal bir seçim olarak değerlendirilmiştir [118]. Farklı genişlik faktörleri ile esneklik ve performans sunan bu modeller, mobil cihazlar ve gömülü sistemler gibi kaynak kısıtlı ortamlar için optimize edilmiştir. Benzer şekilde, MobileNet varyantları, MobileNetV3 Small, MobileNetV3 Large [119] ve MobileNetV2 [120], mobil ve gömülü cihazlar için optimize edilmiş yapıları nedeniyle çalışmaya dahil edilmiştir. Ayrıca, AlexNet seviyesinde doğruluk elde ederken 50 kat daha az parametre kullanan SqueezeNet modeli de değerlendirmeye alınmıştır [121].

ViT-Base ve ViT-Large modelleri [74], özellikle tıbbi görüntülemede karmaşık veri setlerindeki küresel bağımlılıkları yakalamadaki üstün performansları nedeniyle çalışmaya dahil edilmiştir. ViT-Base, doğruluk ve verimlilik arasında bir denge sağlayarak daha hızlı çıkarım için uygun bir seçenek sunarken, ViT-Large, daha geniş model kapasitesiyle doğruluk oranını artırmaktadır. Bu modeller, sınıflandırma görevinde ileri düzey derin öğrenme tekniklerinin sağladığı avantajları keşfetmek amacıyla çalışmaya dahil edilmiştir.

Bu çalışma, çeşitli kullanım alanları ve özelliklere sahip geniş bir derin öğrenme mimarisi yelpazesini kapsamaktadır. AlexNet, VGG16 ve VGG19 gibi klasik modeller; ResNet ve DenseNet gibi derin ve karmaşık mimariler; MobileNet, ShuffleNet ve SqueezeNet gibi mobil ve gömülü cihazlar için optimize edilmiş hafif modeller; ve EfficientNet serisi gibi yüksek verimli mimariler çalışmaya dahil edilmiştir. Bu mimari çeşitlilik, önerilen veri setinin farklı özelliklerini ve zorluklarını etkili bir şekilde yakalamayı ve farklı CNN ve ViT mimarilerin temel performansını değerlendirmeyi amaçlamaktadır. Uygulanan her bir mimari, mobil cihazlar, büyük veri setleri ve yüksek çözünürlüklü görüntüler gibi belirli bir alanda optimal performans sunmaktadır. Bu geniş mimari yelpazesi, genel sınıflandırma performansını değerlendirmek ve farklı uygulama ihtiyaçlarına en uygun modeli belirlemek açısından kritik bir öneme sahiptir.

3.2.1.2 Uygulama Detayları

Deneylerimizde, üç farklı boyama tekniği kullanılarak oluşturulan DB3_v1 veri setlerinde, 35 farklı derin öğrenme mimarisi eğitilerek 18 sınıf için doğruluk değerleri elde edilmiştir. Modellerin eğitimi için en uygun optimizasyon yöntemi ve öğrenme oranını belirlemek amacıyla detaylı bir hiperparametre analizi gerçekleştirilmiştir. Bu bağlamda, Adamax, SGD ve RMSProp algoritmaları seçilmiştir. Her bir optimizasyon yöntemi için 10^{-3}, 10^{-4} ve 10^{-5} olmak üzere üç farklı öğrenme oranı test edilmiştir. Her bir DB3_v1 veriseti (BesLab, Histoplus,

GBL) için 297 farklı eğitim süreci tamamlanmış ve çalışma kapsamında toplamda 891 eğitim süreci gerçekleştirilmiştir. Optimizasyon yöntemleri ve öğrenme oranlarına yönelik detaylı analizler sonucunda, DB3_v2 veri setinde sınıflandırma için her üç boyama tekniğinde en iyi sonuçları sağlayan modeller ve hiperparametre konfigürasyonları belirlenmiştir.

3.2.2 Deneysel Çalışmalar

Adil bir karşılaştırma sağlamak amacıyla, veri setleri K-katlı çapraz doğrulama yöntemi kullanılarak eğitim ve test setlerine ayrılmıştır. Veri setindeki örnekler, her birinin eşit sayıda örnek içerdiği K katmana rastgele bölünmüştür. Sperm morfoloji analizine yönelik çalışmaların literatür incelemesinde K değerinin genellikle 5 olarak belirlendiği görülmektedir [26–28, 50, 105]. Bu nedenle, çalışmamızda da K değeri 5 olarak belirlenmiştir. K-katlı çapraz doğrulama sürecinde, veri setinin %20'si test için, kalan %80'i ise eğitim için kullanılmıştır. Bu süreç, her bir iterasyonda farklı bir %20'lik kısmın test için seçilmesi ile beş kez tekrarlanmıştır. Böylece, veri setindeki her bir örnek en az bir kez test setine dahil edilmiştir. Her katın eğitimi sırasında modeller, yanlılığı önlemek için rastgele ağırlıklarla başlatılmıştır. Ayrıca, her modelin aynı örnekler üzerinde eğitildiğinden emin olmak için, eğitim-test setlerine ayırma işlemi başlangıçta bir kez gerçekleştirilmiş ve ilgili K-kat ismi ile kaydedilmiştir. Bu yaklaşım, her modelin her katmanda aynı görüntü seti üzerinde eğitilmesini ve test edilmesini sağlayarak modeller arasında objektif ve anlamlı bir karşılaştırmayı mümkün kılmıştır.

3.2.2.1 Model Performansları

DB3_v1 üzerinde uygulanan 35 farklı derin öğrenme mimarisinin detaylı performans karşılaştırması, doğruluk (Acc), kesinlik (P), duyarlılık (R) ve f1-skora (F1) ilişkin değerler açısından, farklı optimizasyon yöntemleri ve öğrenme oranları dikkate alınarak yapılmıştır. Elde edilen sonuçlar, DB3_v1_BesLab, DB3_v1_HistoPlus ve DB3_v1_GBL verisetleri için sırasıyla Tablo 3.11, Tablo 3.12 ve Tablo 3.13'te sunulmuştur.

BesLab veri seti üzerinde gerçekleştirilen deneylerden elde edilen bulgulara göre, EfficientNet ailesinin üstün bir performans sergilediği görülmektedir (Şekil 3.13). EfficientNet V2 medium modeli, %65,05 ile en yüksek doğruluk oranını elde etmiş, bunu sırasıyla %64'ün üzerinde doğruluk değerleriyle EfficientNet V2 small ve V2 large modelleri takip etmiştir. EfficientNet modellerinin tutarlı şekilde yüksek performans sergilemesi, BesLab boyama yöntemiyle elde edilen morfolojik varyasyonları etkili bir şekilde işleyebilme kabiliyetlerini vurgulamaktadır. Benzer

Tablo 3.11 DB3_v1_BesLab veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması

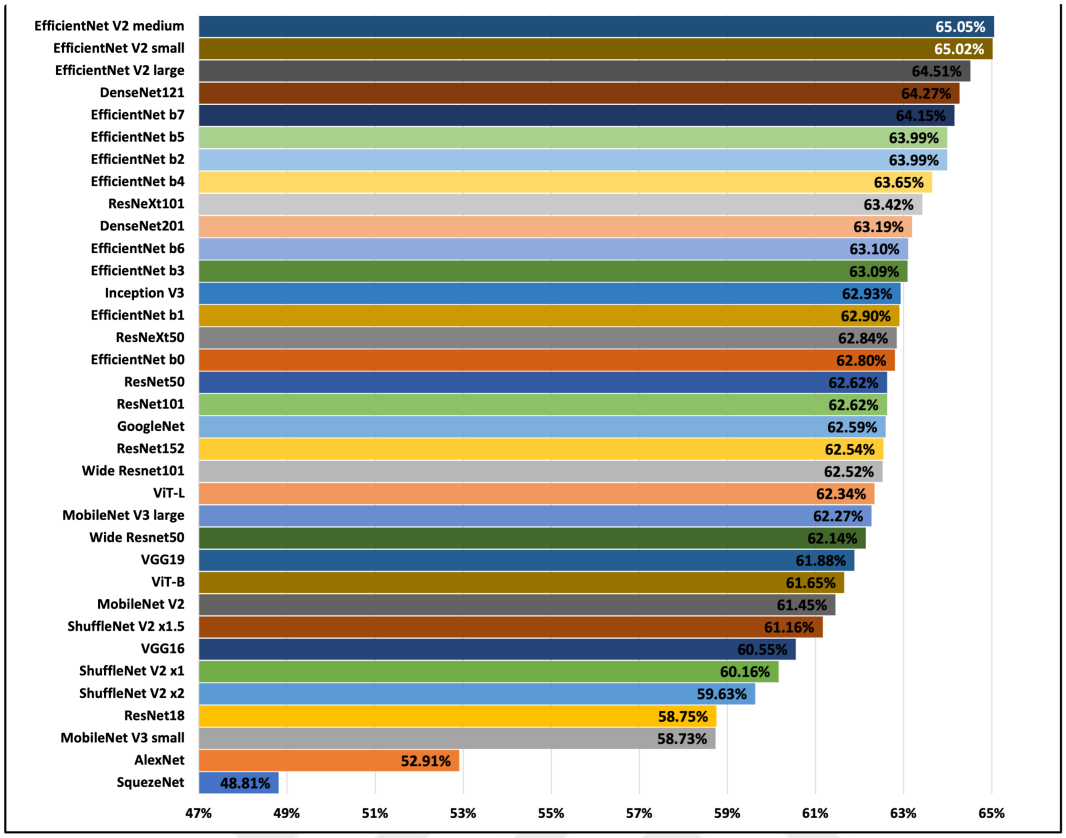
	BesLab					
Model Adı	**Opt.**	**Lr.**	**P**	**R**	**F1**	**Acc.**
AlexNet	Adamax	10^{-4}	MA: 46,00 WA: 51,50	MA: 42,11 WA: 52,91	MA: 42,53 WA: 51,75	52,91
ResNet18	RMSProp	10^{-5}	MA: 52,67 WA: 58,32	MA: 51,16 WA: 58,75	MA: 51,35 WA: 58,37	58,75
ResNet50	SGD	10^{-3}	MA: 55,31 WA: 61,91	MA: 51,42 WA: 62,62	MA: 52,24 WA: 62,06	62,62
ResNet101	SGD	10^{-3}	MA: 61,39 WA: 62,36	MA: 53,45 WA: 62,62	MA: 54,62 WA: 62,29	62,62
ResNet152	SGD	10^{-3}	MA: 57,85 WA: 62,24	MA: 53,29 WA: 62,54	MA: 54,57 WA: 62,24	62,54
DenseNet121	Adamax	10^{-3}	MA: 61,46 WA: 63,98	MA: 58,73 WA: 64,27	MA: 59,68 WA: 64,05	64,27
DenseNet201	RMSProp	10^{-5}	MA: 60,18 WA: 62,94	MA: 54,86 WA: 63,19	MA: 56,53 WA: 62,89	63,19
ShuffleNet V2 x1	RMSProp	10^{-4}	MA: 57,14 WA: 59,14	MA: 50,73 WA: 60,16	MA: 51,34 WA: 59,32	60,16
ShuffleNet V2 x1.5	Adamax	10^{-3}	MA: 57,69 WA: 60,64	MA: 52,94 WA: 61,06	MA: 54,36 WA: 60,71	61,06
ShuffleNet V2 x2	SGD	10^{-3}	MA: 54,69 WA: 58,61	MA: 48,57 WA: 59,63	MA: 49,12 WA: 58,61	59,63
MobileNet V3 S	Adamax	10^{-3}	MA: 54,66 WA: 57,95	MA: 49,42 WA: 58,73	MA: 50,60 WA: 58,11	58,73
MobileNet V3 L	RMSProp	10^{-4}	MA: 58,93 WA: 61,44	MA: 54,66 WA: 62,27	MA: 55,90 WA: 61,65	62,27
MobileNet V2	SGD	10^{-3}	MA: 57,11 WA: 60,35	MA: 50,04 WA: 61,45	MA: 50,68 WA: 60,59	61,45
VGG16	SGD	10^{-3}	MA: 54,11 WA: 59,57	MA: 50,20 WA: 60,55	MA: 51,12 WA: 59,91	60,55
VGG19	SGD	10^{-3}	MA: 55,08 WA: 61,01	MA: 51,17 WA: 61,88	MA: 52,09 WA: 61,25	61,88
ResNext50	RMSProp	10^{-5}	MA: 60,94 WA: 62,34	MA: 54,66 WA: 62,84	MA: 56,48 WA: 62,37	62,84
ResNext101	Adamax	10^{-4}	MA: 60,14 WA: 63,06	MA: 55,26 WA: 63,42	MA: 56,73 WA: 63,07	63,42
Wide Resnet50	SGD	10^{-3}	MA: 56,83 WA: 61,79	MA: 51,63 WA: 62,14	MA: 52,60 WA: 61,74	62,14
Wide Resnet101	SGD	10^{-3}	MA: 53,52 WA: 61,98	MA: 51,85 WA: 62,52	MA: 52,42 WA: 62,14	62,52
SqueezeNet	Adamax	10^{-4}	MA: 38,46 WA: 46,95	MA: 34,99 WA: 48,81	MA: 34,69 WA: 46,98	48,81
EfficientNet b0	RMSProp	10^{-4}	MA: 58,04 WA: 61,77	MA: 53,45 WA: 62,79	MA: 54,36 WA: 62,05	62,80
EfficientNet b1	RMSProp	10^{-4}	MA: 58,87 WA: 62,18	MA: 55,00 WA: 62,90	MA: 55,96 WA: 62,39	62,90
EfficientNet b2	RMSProp	10^{-4}	MA: 58,57 WA: 63,07	MA: 54,74 WA: 63,99	MA: 55,58 WA: 63,33	63,99
EfficientNet b3	Adamax	10^{-3}	MA: 59,71 WA: 62,44	MA: 55,72 WA: 63,09	MA: 56,79 WA: 62,59	63,09
EfficientNet b4	RMSProp	10^{-4}	MA: 60,98 WA: 63,22	MA: 56,35 WA: 63,65	MA: 57,84 WA: 63,33	63,65
EfficientNet b5	RMSProp	10^{-4}	MA: 57,33 WA: 63,27	MA: 54,56 WA: 63,99	MA: 55,36 WA: 63,54	63,99
EfficientNet b6	RMSProp	10^{-4}	MA: 59,33 WA: 62,56	MA: 55,76 WA: 63,10	MA: 56,90 WA: 62,71	63,10
EfficientNet b7	RMSProp	10^{-4}	MA: 56,12 WA: 63,15	MA: 53,36 WA: 64,15	MA: 54,03 WA: 63,41	64,15
EfficientNet V2 S	**SGD**	10^{-3}	**MA: 56,15 WA: 63,85**	**MA: 53,34 WA: 65,02**	**MA: 53,59 WA: 64,06**	**65,02**
EfficientNet V2 M	**SGD**	10^{-3}	**MA: 61,98 WA: 64,35**	**MA: 55,30 WA: 65,05**	**MA: 56,52 WA: 64,39**	**65,05**
EfficientNet V2 L	**RMSProp**	10^{-5}	**MA: 63,90 WA: 63,93**	**MA: 54,49 WA: 64,51**	**MA: 54,76 WA: 63,88**	**64,51**
GoogleNet	Adamax	10^{-3}	MA: 59,86 WA: 62,15	MA: 52,89 WA: 62,59	MA: 54,65 WA: 62,08	62,59
Inception	Adamax	10^{-3}	MA: 59,81 WA: 62,38	MA: 55,26 WA: 62,93	MA: 56,57 WA: 62,48	62,93
ViT-Base	Adamax	10^{-4}	MA: 58,43 WA: 61,34	MA: 54,69 WA: 61,65	MA: 56,02 WA: 61,37	61,65
ViT-Large	Adamax	10^{-4}	MA: 60,40 WA: 61,67	MA: 53,35 WA: 62,34	MA: 55,09 WA: 61,74	62,34

Tablo 3.12 DB3_v1_Histoplus veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması

	Histoplus					
Model Adı	**Opt.**	**Lr.**	**P**	**R**	**F1**	**Acc.**
AlexNet	Adamax	10^{-4}	MA: 49,73 WA: 56,37	MA: 44,50 WA: 57,57	MA: 45,64 WA: 56,43	57,57
ResNet18	SGD	10^{-3}	MA: 54,66 WA: 60,92	MA: 49,67 WA: 61,68	MA: 51,42 WA: 61,07	61,68
ResNet50	SGD	10^{-3}	MA: 58,25 WA: 63,93	MA: 53,19 WA: 64,36	MA: 55,08 WA: 63,97	64,36
ResNet101	SGD	10^{-3}	MA: 56,76 WA: 63,92	MA: 53,77 WA: 64,38	MA: 55,05 WA: 64,06	64,38
ResNet152	SGD	10^{-3}	MA: 57,95 WA: 64,48	MA: 54,22 WA: 64,83	MA: 55,74 WA: 64,54	64,83
DenseNet121	Adamax	10^{-3}	MA: 59,33 WA: 66,27	MA: 56,62 WA: 66,85	MA: 57,46 WA: 66,43	66,85
DenseNet201	RMSProp	10^{-5}	MA: 64,53 WA: 65,63	MA: 56,18 WA: 66,13	MA: 58,52 WA: 65,67	66,13
ShuffleNet V2 x1	RMSProp	10^{-4}	MA: 55,28 WA: 62,06	MA: 50,78 WA: 63,29	MA: 51,59 WA: 62,26	63,29
ShuffleNet V2 x1,5	Adamax	10^{-3}	MA: 59,63 WA: 63,17	MA: 53,86 WA: 63,57	MA: 55,42 WA: 63,22	63,57
ShuffleNet V2 x2	Adamax	10^{-3}	MA: 57,23 WA: 61,45	MA: 52,84 WA: 63,18	MA: 54,95 WA: 62,31	63,18
MobileNet V3 S	RMSProp	10^{-4}	MA: 59,55 WA: 62,98	MA: 54,56 WA: 63,53	MA: 56,02 WA: 63,11	63,53
MobileNet V3 L	Adamax	10^{-3}	MA: 60,22 WA: 64,35	MA: 55,78 WA: 64,65	MA: 57,04 WA: 64,38	64,65
MobileNet V2	Adamax	10^{-3}	MA: 59,46 WA: 63,95	MA: 55,10 WA: 64,44	MA: 56,33 WA: 64,07	64,44
VGG16	SGD	10^{-3}	MA: 55,06 WA: 62,22	MA: 51,27 WA: 63,12	MA: 52,66 WA: 62,50	63,12
VGG19	SGD	10^{-3}	MA: 55,64 WA: 62,75	MA: 52,96 WA: 63,51	MA: 53,82 WA: 62,98	63,51
ResNext50	RMSProp	10^{-5}	MA: 63,86 WA: 65,39	MA: 56,33 WA: 65,89	MA: 58,53 WA: 65,42	65,89
ResNext101	SGD	10^{-3}	MA: 62,43 WA: 65,58	MA: 57,23 WA: 66,20	MA: 59,23 WA: 65,72	66,20
Wide Resnet50	SGD	10^{-3}	MA: 57,13 WA: 64,16	MA: 53,97 WA: 64,56	MA: 55,31 WA: 64,26	64,56
Wide Resnet101	Adamax	10^{-4}	MA: 57,81 WA: 64,18	MA: 53,99 WA: 64,59	MA: 55,55 WA: 64,28	64,59
SqueezeNet	SGD	10^{-3}	MA: 52,75 WA: 58,76	MA: 46,41 WA: 59,81	MA: 48,02 WA: 58,79	59,81
EfficientNet b0	Adamax	10^{-3}	MA: 62,56 WA: 65,26	MA: 56,34 WA: 65,71	MA: 58,19 WA: 65,34	65,71
EfficientNet b1	RMSProp	10^{-4}	MA: 61,30 WA: 65,09	MA: 56,55 WA: 65,41	MA: 58,05 WA: 65,13	65,41
EfficientNet b2	Adamax	10^{-3}	MA: 62,37 WA: 65,58	MA: 57,52 WA: 65,98	MA: 59,05 WA: 65,69	65,98
EfficientNet b3	Adamax	10^{-3}	MA: 61,92 WA: 65,62	MA: 57,33 WA: 65,97	MA: 58,86 WA: 65,71	65,97
EfficientNet b4	RMSProp	10^{-4}	MA: 62,74 WA: 65,74	MA: 57,20 WA: 66,08	MA: 59,00 WA: 65,80	66,08
EfficientNet b5	Adamax	10^{-3}	MA: 61,55 WA: 65,68	MA: 56,94 WA: 66,09	MA: 58,40 WA: 65,79	66,09
EfficientNet b6	Adamax	10^{-3}	MA: 61,09 WA: 65,28	MA: 56,78 WA: 65,69	MA: 58,20 WA: 65,40	65,69
EfficientNet b7	Adamax	10^{-3}	MA: 62,80 WA: 65,93	MA: 58,29 WA: 66,26	MA: 59,96 WA: 66,00	66,26
EfficientNet V2 S	**SGD**	10^{-3}	**MA: 63,81 WA: 66,54**	**MA: 57,22 WA: 67,00**	**MA: 59,25 WA: 66,61**	**67,00**
EfficientNet V2 M	**RMSProp**	10^{-5}	**MA: 63,77 WA: 67,00**	**MA: 59,51 WA: 67,42**	**MA: 61,11 WA: 67,13**	**67,42**
EfficientNet V2 L	**RMSProp**	10^{-5}	**MA: 64,77 WA: 66,69**	**MA: 59,83 WA: 67,07**	**MA: 61,68 WA: 66,78**	**67,07**
GoogleNet	Adamax	10^{-3}	MA: 64,24 WA: 64,83	MA: 54,68 WA: 65,09	MA: 56,50 WA: 64,72	65,09
Inception	RMSProp	10^{-5}	MA: 62,62 WA: 64,09	MA: 54,47 WA: 64,58	MA: 56,95 WA: 64,14	64,58
ViT-Base	Adamax	10^{-4}	MA: 61,20 WA: 65,51	MA: 55,59 WA: 65,77	MA: 57,43 WA: 65,51	65,77
ViT-Large	Adamax	10^{-4}	MA: 62,85 WA: 65,49	MA: 55,72 WA: 65,95	MA: 57,85 WA: 65,53	65,95

Tablo 3.13 DB3_v1_GBL veri seti kullanılarak farklı sınıflandırıcıların görüntü sınıflandırma sonuçlarının karşılaştırılması

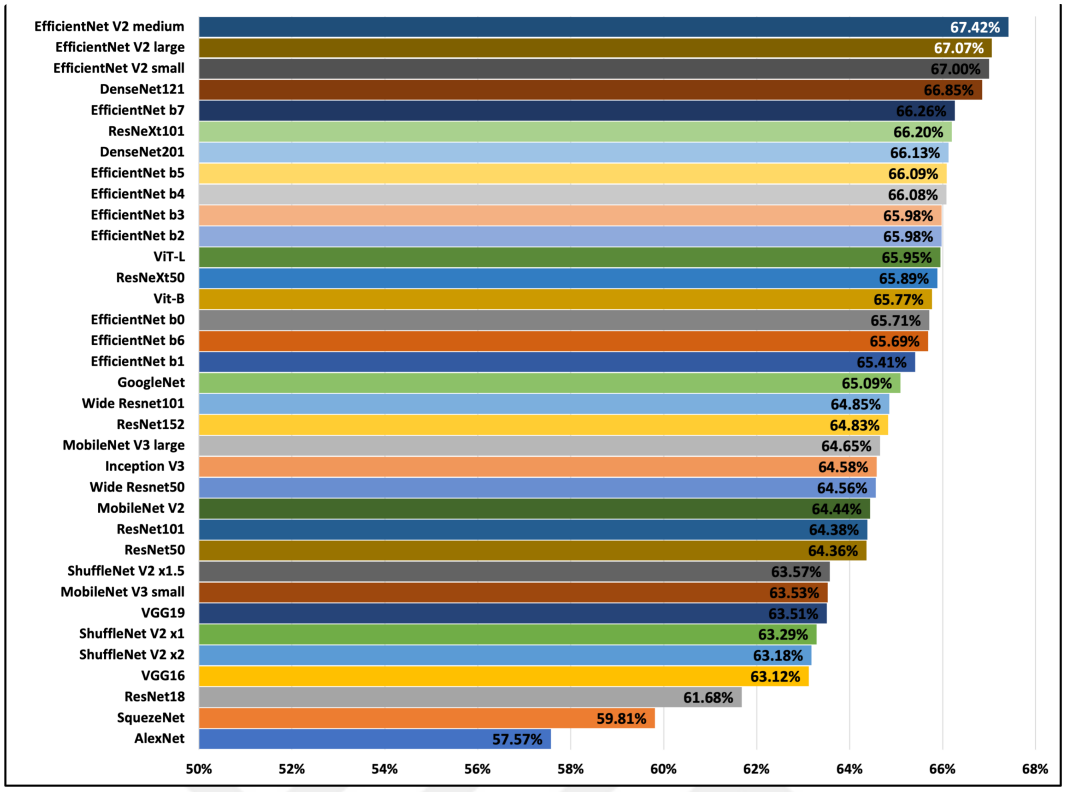
	GBL					
Model Adı	**Opt.**	**Lr.**	**P**	**R**	**F1**	**Acc.**
AlexNet	Adamax	10^{-4}	MA: 42,77 WA: 51,21	MA: 36,98 WA: 53,51	MA: 37,70 WA: 51,78	53,51
ResNet18	Adamax	10^{-3}	MA: 53,83 WA: 56,17	MA: 41,02 WA: 57,28	MA: 42,77 WA: 56,06	57,28
ResNet50	RMSProp	10^{-5}	MA: 56,45 WA: 59,40	MA: 45,24 WA: 59,75	MA: 48,10 WA: 59,13	59,75
ResNet101	SGD	10^{-3}	MA: 55,67 WA: 59,93	MA: 46,82 WA: 60,51	MA: 49,58 WA: 59,90	60,51
ResNet152	SGD	10^{-3}	MA: 55,16 WA: 60,33	MA: 47,12 WA: 60,81	MA: 49,72 WA: 60,28	60,81
DenseNet121	Adamax	10^{-3}	MA: 55,91 WA: 61,36	MA: 47,86 WA: 62,29	MA: 49,65 WA: 61,47	62,29
DenseNet201	RMSProp	10^{-5}	MA: 59,39 WA: 62,80	MA: 54,50 WA: 63,39	MA: 56,46 WA: 62,96	63,39
ShuffleNet V2 x1	RMSProp	10^{-4}	MA: 50,13 WA: 55,64	MA: 41,80 WA: 57,93	MA: 42,03 WA: 56,01	57,93
ShuffleNet V2 x1,5	RMSProp	10^{-4}	MA: 53,37 WA: 58,26	MA: 47,06 WA: 58,83	MA: 49,31 WA: 58,33	58,83
ShuffleNet V2 x2	Adamax	10^{-3}	MA: 55,53 WA: 58,74	MA: 48,53 WA: 59,24	MA: 50,88 WA: 58,80	59,24
MobileNet V3 S	RMSProp	10^{-4}	MA: 53,23 WA: 58,56	MA: 48,75 WA: 59,37	MA: 50,22 WA: 58,82	59,37
MobileNet V3 L	Adamax	10^{-3}	MA: 56,02 WA: 59,14	MA: 50,08 WA: 59,42	MA: 52,18 WA: 59,15	59,42
MobileNet V2	Adamax	10^{-3}	MA: 55,94 WA: 59,71	MA: 48,23 WA: 60,37	MA: 50,48 WA: 59,80	60,37
VGG16	SGD	10^{-3}	MA: 51,70 WA: 57,86	MA: 46,18 WA: 59,23	MA: 47,61 WA: 58,24	59,23
VGG19	SGD	10^{-3}	MA: 52,93 WA: 58,90	MA: 47,99 WA: 60,06	MA: 49,26 WA: 59,27	60,06
ResNext50	RMSProp	10^{-5}	MA: 60,08 WA: 61,42	MA: 50,48 WA: 62,10	MA: 53,71 WA: 61,39	62,10
ResNext101	SGD	10^{-3}	MA: 58,37 WA: 61,10	MA: 49,24 WA: 61,98	MA: 52,33 WA: 61,14	61,98
Wide Resnet50	SGD	10^{-3}	MA: 54,81 WA: 59,79	MA: 45,47 WA: 60,22	MA: 47,75 WA: 59,64	60,22
Wide Resnet101	SGD	10^{-3}	MA: 55,98 WA: 59,77	MA: 45,79 WA: 60,22	MA: 48,51 WA: 59,60	60,22
SqueezeNet	SGD	10^{-3}	MA: 46,31 WA: 53,48	MA: 40,04 WA: 55,54	MA: 41,22 WA: 53,98	55,54
EfficientNet b0	Adamax	10^{-3}	MA: 58,14 WA: 61,13	MA: 51,47 WA: 61,78	MA: 53,73 WA: 61,25	61,78
EfficientNet b1	RMSProp	10^{-4}	MA: 55,98 WA: 60,51	MA: 51,24 WA: 61,12	MA: 53,02 WA: 60,69	61,12
EfficientNet b2	RMSProp	10^{-4}	MA: 56,96 WA: 61,52	MA: 52,21 WA: 62,32	MA: 53,74 WA: 61,75	62,32
EfficientNet b3	Adamax	10^{-3}	MA: 58,66 WA: 61,48	MA: 52,92 WA: 61,99	MA: 55,10 WA: 61,59	61,99
EfficientNet b4	RMSProp	10^{-4}	MA: 59,32 WA: 62,59	MA: 53,31 WA: 62,76	MA: 55,61 WA: 62,56	62,76
EfficientNet b5	Adamax	10^{-3}	MA: 60,25 WA: 62,20	MA: 52,85 WA: 62,67	MA: 55,38 WA: 62,26	62,67
EfficientNet b6	Adamax	10^{-3}	MA: 58,45 WA: 61,67	MA: 52,72 WA: 62,11	MA: 54,92 WA: 61,74	62,11
EfficientNet b7	Adamax	10^{-3}	MA: 59,19 WA: 62,36	MA: 53,65 WA: 62,91	MA: 55,81 WA: 62,47	62,91
EfficientNet V2 S	**SGD**	10^{-3}	**MA: 61,08 WA: 63,03**	**MA: 54,55 WA: 63,58**	**MA: 57,03 WA: 63,11**	**63,58**
EfficientNet V2 M	**RMSProp**	10^{-5}	**MA: 60,70 WA: 63,03**	**MA: 54,81 WA: 63,46**	**MA: 57,14 WA: 63,10**	**63,46**
EfficientNet V2 L	**RMSProp**	10^{-5}	**MA: 59,25 WA: 62,91**	**MA: 53,29 WA: 63,54**	**MA: 55,58 WA: 63,06**	**63,54**
GoogleNet	Adamax	10^{-3}	MA: 57,20 WA: 60,52	MA: 48,37 WA: 60,94	MA: 51,27 WA: 60,41	60,94
Inception	Adamax	10^{-3}	MA: 57,70 WA: 60,93	MA: 50,30 WA: 61,34	MA: 52,96 WA: 60,91	61,34
ViT-Base	Adamax	10^{-4}	MA: 58,71 WA: 62,22	MA: 52,51 WA: 62,49	MA: 54,85 WA: 62,21	62,49
ViT-Large	Adamax	10^{-4}	MA: 57,53 WA: 61,62	MA: 50,79 WA: 62,33	MA: 53,17 WA: 61,76	62,33



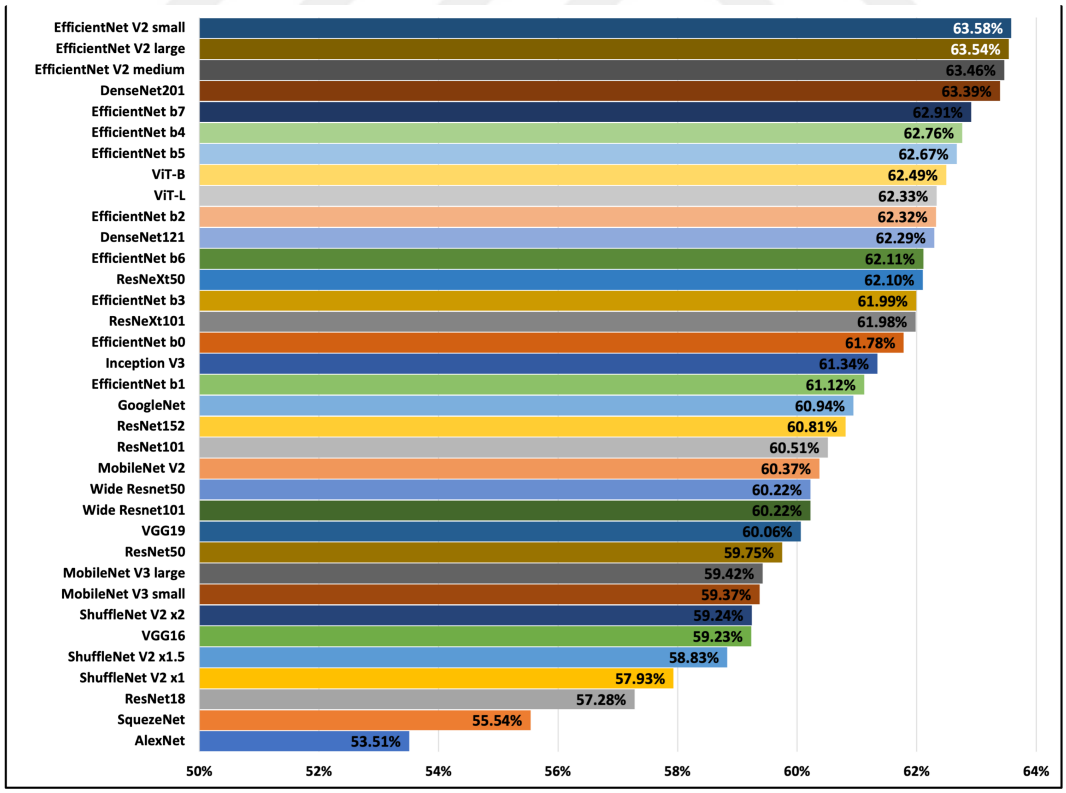
Şekil 3.13 DB3_v1_BesLab üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu

şekilde, Histoplus veri seti sonuçları (Şekil 3.14) EfficientNet V2 medium, V2 large ve EfficientNet V2 small modellerinin sırasıyla %67,42, %67,07 ve %67,00 ile en yüksek doğruluk oranlarını elde ettiğini göstermektedir. Bu sonuçlar, Histoplus boyama yöntemiyle yakalanan ayrıntılı özelliklerin, daha derin ve karmaşık EfficientNet modelleri tarafından etkin bir şekilde kullanıldığını ortaya koymaktadır. Ayrıca, %66'nın üzerinde doğruluk değerleri elde eden DenseNet121, DenseNet201 ve ResNext101 modelleri de üstün performans göstermiştir. Bu modellerin yoğun bağlantıları ve artık öğrenme (residual learning) yetenekleri, gelişmiş özellik çıkarımı ve model eğitimi verimliliği ile ilişkilendirilebilmektedir.

Şekil 3.15'te görüldüğü gibi, GBL veri seti için en yüksek doğruluk oranı, yaklaşık %63,58 ile EfficientNet V2 Small ve V2 Large modelleri tarafından elde edilmiştir. Bu sonuçlar, diğer boyama tekniklerinden elde edilen bulgularla tutarlı olup, EfficientNet modellerinin farklı boyama yöntemleri karşısında gösterdiği çok yönlülük ve sağlamlığı ortaya koymaktadır. DenseNet ve ResNext ailesi de genel olarak %61'in üzerinde doğruluk değerleriyle başarılı bir performans göstermiştir. EfficientNet ve DenseNet mimarilerinin varyantlarında gözlemlenen bu performans tutarlılığı, farklı boyama teknikleri ve morfolojik özelliklerle etkili bir şekilde başa çıkabilme yeteneklerini ortaya koymaktadır. Cep telefonları, IoT cihazları ve

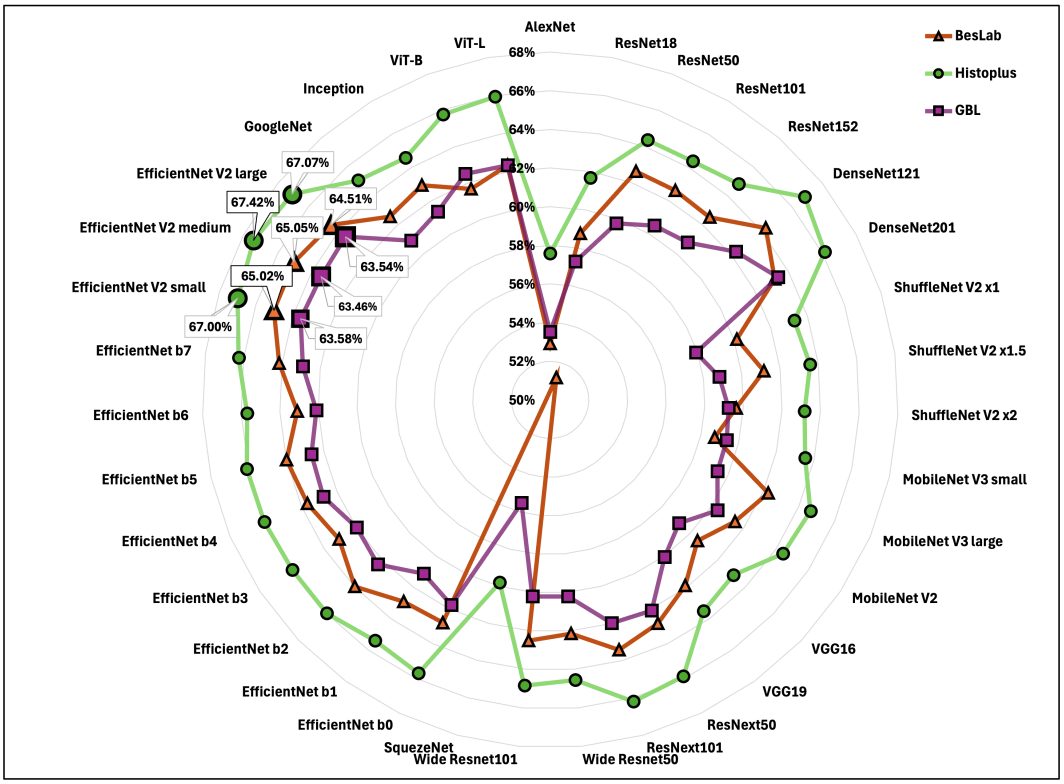


Şekil 3.14 DB3_v1_Histoplus üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu



Şekil 3.15 DB3_v1_GBL üzerinde derin öğrenme mimarilerinin sınıflandırma doğruluğu

tabletler gibi uç cihazlarda işlem yapılmasını sağlayan hafif modeller arasında yer alan MobileNet ve ShuffleNet mimarilerinin sonuçları incelendiğinde, MobileNet V3 Large modelinin BesLab ve Histoplus boyama teknikleri için sırasıyla %62,27 ve %64,65 doğruluk ile en yüksek performansa ulaştığı gözlemlenmiştir. GBL boyama tekniği için ise MobileNet V2 modeli %60,37 doğruluk oranı elde etmiştir. Deneysel çalışmalar, daha az sayıda parametreye sahip olmalarına rağmen hafif modellerin, diğer modellere kıyasla etkileyici sonuçlar sunduğunu ortaya koymuştur.

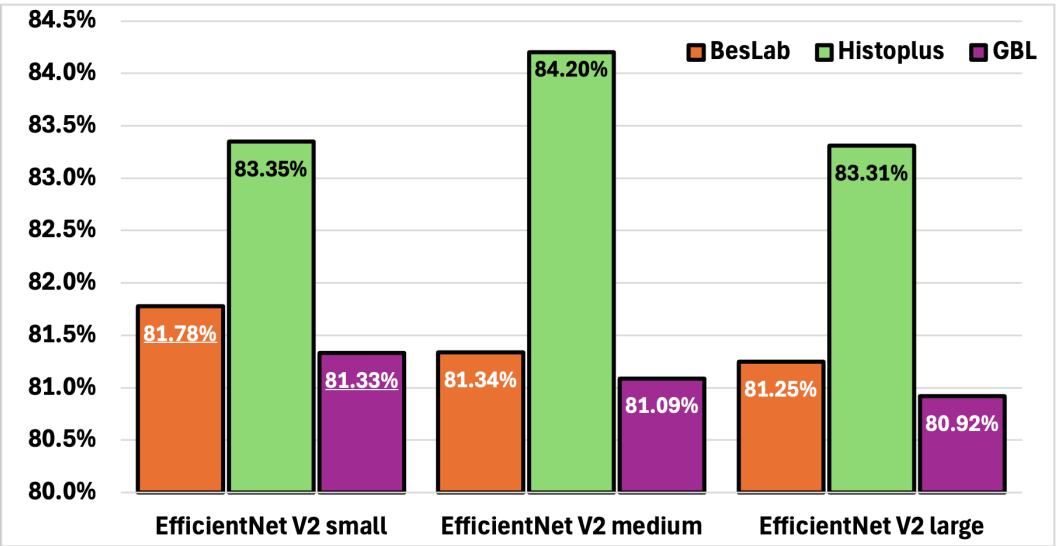


Şekil 3.16 Üç farklı boyama yöntemiyle elde edilen DB3_v1 veri setleri için sınıflandırma doğruluğunun radar grafiği

Tablo 3.11, 3.12 ve 3.13'te sunulan sonuçlar incelendiğinde, optimizasyon yöntemi ve öğrenme oranı seçiminin model performansı üzerinde önemli bir etkisi olduğu görülmektedir. Örneğin, Adamax ve RMSProp, sırasıyla 10^{-4} ve 10^{-5} öğrenme oranları ile farklı mimarilerde genellikle SGD optimizatörüne kıyasla daha etkili olmuştur.

3.2.2.2 Boyama Tekniklerinin Karşılaştırmalı Analizi

Şekil 3.16 olarak verilen radar grafiği, doğruluk performans ölçütü olarak kullanıldığında, 35 farklı derin öğrenme mimarisinin üç farklı boyama tekniğindeki (BesLab, Histoplus, GBL) performansını karşılaştırarak ilginç bir bulguyu vurgulamaktadır. 35 derin öğrenme modelinin tümü için Histoplus boyama tekniği

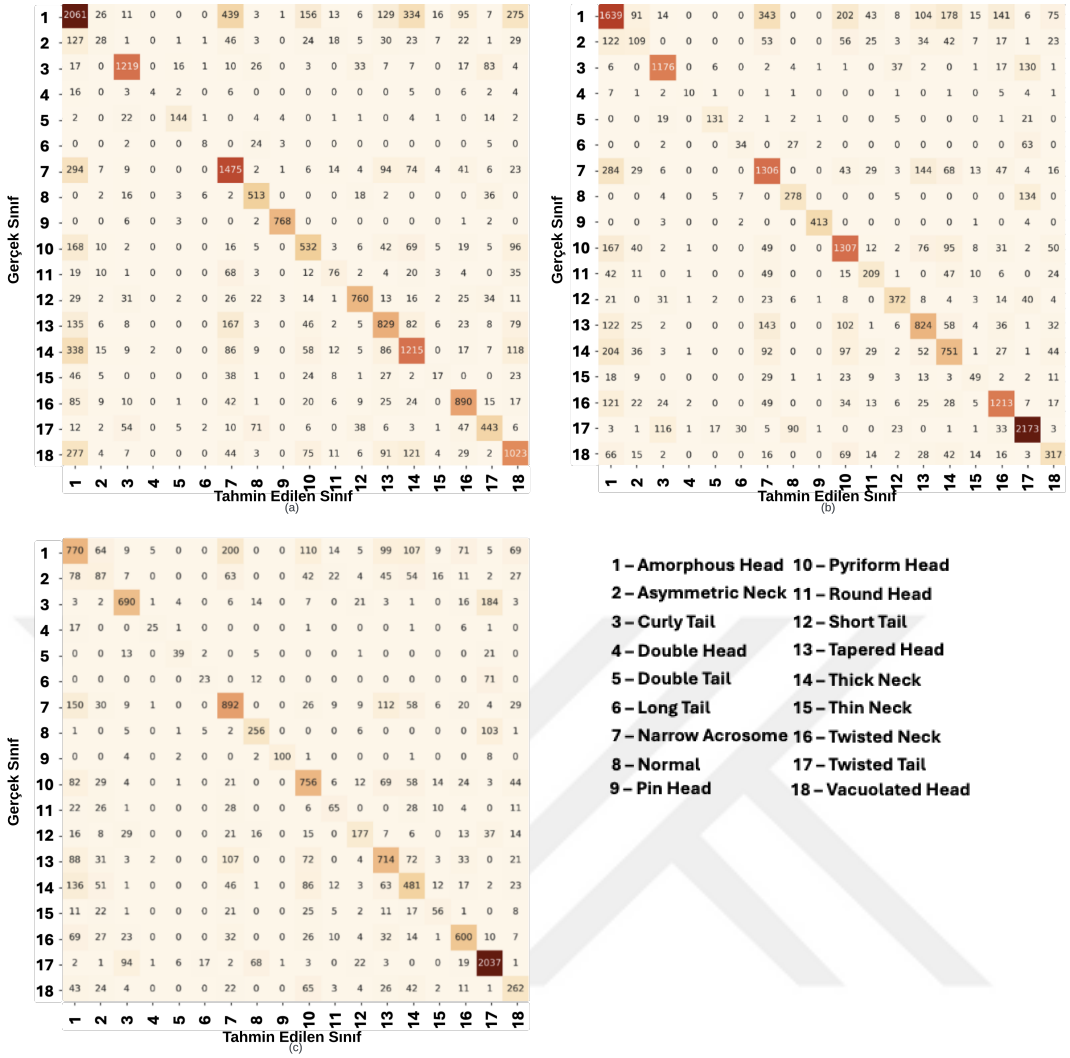


Şekil 3.17 DB3_v2 veri seti için sınıflandırma doğrulukları

en yüksek doğrulukla sonuçlanmaktadır. Bu tutarlı üstün performans, söz konusu mimarilerin sperm örneklerinin morfolojik özelliklerine ilişkin bilgileri öğrenmeleri için histoplus boyama tekniğinin daha ayırt edilebilir özellikler sağladığını göstermektedir. Öte yandan, hem BesLab hem de GBL boyama teknikleri makul sınıflandırma performansı sergilemiş olsa da, BesLab bazı modellerde GBL'yi geride bırakırken Histoplus boyama tekniğine göre daha düşük bir performans sergilemiştir.

3.2.2.3 Baş, Boyun ve Kuyruk Grupları (DB3v2) Analiz Sonuçları

DB3_v1 veri kümesi üzerinde yapılan deneysel çalışmalar sonucunda en başarılı modellerin EfficientNet V2 ailesinin small, medium ve large mimarileri olduğu görülmüştür. DB3_v1'deki başarılarından dolayı bu modeller DB3_v2 veri kümesindeki ana kategorileri sınıflandırmak için kullanılmıştır. Sonuçlar incelendiğinde, DB3_v1 veri kümesine benzer şekilde DB3_v2 veri kümesinde de en yüksek doğruluk değerlerinin Histoplus boyama tekniği ile elde edildiği görülmüştür. BesLab ve GBL boyama tekniklerinin sonuçları karşılaştırıldığında, bu iki yöntem arasındaki performans farkı DB3_v2'de DB3_v1'e kıyasla %1'in altına düşmüştür. Ayrıca, sınıflandırma doğrulukları DB3_v1'e kıyasla %60'tan %80'in üzerine çıkmıştır. BesLab ve GBL boyama teknikleri için en yüksek doğruluk sırasıyla %81,78 ve %81,33 ile EfficientNet V2 small modeli kullanılarak elde edilmiştir. Histoplus boyama tekniği için en yüksek doğruluk değeri ise %84,20 doğrulukla EfficientNet V2 medium modeli kullanılarak elde edilmiştir. DB3_v2 veri kümesi için ayrıntılı deneysel sonuçlar Şekil 3.17'de sunulmuştur.



Şekil 3.18 18 sınıflı DB3_v1_BesLab (a) ve DB3_v1_Histoplus (b) üzerinde eğitilen EfficientNet v2 medium'un karmaşıklık matrisleri, c) 18 sınıflı DB3_v1_GBL üzerinde eğitilen EfficientNet V2 small'un karmaşıklık matrisi

3.2.2.4 Yanlış Sınıflandırma Analizi

Tablo 3.11, 3.12 ve 3.13'teki sonuçlara göre, EfficientNet V2 medium DB3_v1_BesLab ve DB3_v1_Histoplus veri kümeleri için en iyi doğruluğu elde ederken, EfficientNet V2 small DB3_v1_GBL veri kümesi için diğerlerinden daha iyi performans göstermiştir. Bu modellerin karmaşıklık matrisleri Şekil 3.18'de sunulmuştur. Ayrıca, bu karmaşıklık matrislerinden elde edilen sınıf bazında kesinlik, duyarlılık ve F1-skor değerleri Tablo 3.14'te ayrıntılı olarak sunulmuştur. Şekil 3.18'de gösterilen karmaşıklık matrisleri incelendiğinde, AmorfHead sınıfının her üç boyama yönteminde de sıklıkla NarrowAcrosome, PyriformHead, TaperedHead, ThickNeck, TwistedNeck ve VacuolatedHead sınıfları ile yanlış sınıflandırıldığı tespit edilmiştir. Bu yanlış sınıflandırma, amorf sperm hücrelerinin diğer sınıflara kıyasla belirli bir morfolojik bozulma derecesiyle bütünleşmemesi ve çeşitli morfolojik anormalliklerin amorf olarak tanımlanmasına

yol açması nedeniyle meydana gelmektedir (Şekil 1.1). Sonuç olarak, amorf sperm hücreleri benzer uzamsal oranları paylaşan bu sınıflarla ortak özellikler sergilemektedir. Benzer şekilde, curly tail, short tail ve normal sınıfları, her üç boyama yönteminde de ağırlıklı olarak twisted tail sınıfı olarak yanlış sınıflandırılmıştır. Ayrıca, boyun ile ilişkili sınıfların, muhtemelen sperm başına yakın uzamsal yakınlıkları nedeniyle, kendi aralarında değil, baş anormallikleri ile daha sık yanlış sınıflandırıldığı gözlemlenmiştir. Sperm görüntüleri elde etmek için kullanılan boyama yöntemleri de yanlış sınıflandırma modellerini etkilemede önemli bir rol oynamıştır. GBL boyama yöntemi, daha sert geçişler ve daha koyu görüntülerle sonuçlanmış, bu da morfolojik özellikleri genellikle gizleyerek sınıflar arasında karışıklığın artmasına yol açmıştır. Buna karşılık, BesLab ve Histoplus boyamaları, daha net görselleştirmeler sunmuş, bu da genellikle daha iyi sınıflandırma performansı ile sonuçlanmıştır. Ancak, doğal biyolojik benzerlikler nedeniyle bazı çakışmalar devam etmiştir.

Tablo 3.14 Şekil 3.18'deki karmaşıklık matrislerinin sınıf bazında ayrıntılı performans ölçütleri

	BesLab				Histoplus				GBL			
Class Name	**P (%)**	**R (%)**	**F1 (%)**	**Destek**	**P (%)**	**R (%)**	**F1 (%)**	**Destek**	**P (%)**	**R (%)**	**F1 (%)**	**Destek**
AmorfHead	56,84	57,70	57,27	3572	58,08	57,33	57,70	2859	51,75	50,1	50,91	1537
AsymmetricNeck	22,22	07,65	11,38	366	28,02	22,15	24,74	492	21,64	19,00	20,23	458
CurlyTail	86,39	84,48	85,42	1443	83,64	84,97	84,3	1384	76,92	72,25	74,51	955
DoubleHead	66,67	08,33	14,81	48	58,82	28,57	38,46	35	71,43	48,08	57,47	52
DoubleTail	81,36	72,00	76,39	200	80,86	71,58	75,94	183	72,22	48,15	57,78	81
LongTail	42,11	19,05	26,23	42	45,33	26,56	33,5	128	48,94	21,70	30,07	106
NarrowAcrosome	59,6	71,81	65,14	2054	60,43	65,56	62,89	1992	60,97	65,83	63,31	1355
Normal	73,81	85,79	79,35	598	67,97	64,20	66,03	433	68,45	67,37	67,90	380
PinHead	98,46	98,21	98,34	782	98,33	97,64	97,98	423	99,01	84,75	91,32	117
PyriformHead	54,51	54,40	54,45	978	66,79	70,96	68,81	1842	60,92	67,32	63,96	1123
RoundHead	46,06	29,57	36,02	257	54,43	50,36	52,32	415	44,52	32,34	37,46	201
ShortTail	84,54	76,69	80,42	991	77,66	69,14	73,16	538	64,60	49,30	55,92	359
TaperedHead	59,86	59,26	59,55	1399	62,90	60,77	61,82	1356	60,30	62,09	61,18	1150
ThickNeck	60,78	61,46	61,12	1977	56,98	56,04	56,51	1340	51,17	51,50	51,33	934
ThinNeck	25,76	08,85	13,18	192	37,12	28,32	32,13	173	43,41	31,11	36,25	180
TwistedNeck	72,01	77,12	74,48	1154	75,53	77,46	76,48	1566	70,92	70,18	70,55	855
TwistedTail	66,12	62,75	64,39	706	83,71	86,99	85,32	2498	81,84	89,46	85,48	2277
VacuolatedHead	58,62	60,28	59,44	1697	51,29	52,48	51,88	604	50,38	51,47	50,92	509

P: Precision R: Recall F1: F1-Score

Şekil 3.19 a, c) 4 sınıflı DB3_v2_BesLab ve DB3_v2_GBL üzerinde eğitilen EfficientNet v2 small'ın karmaşıklık matrisleri, b) 4 sınıflı DB3_v2_Histoplus üzerinde eğitilen EfficientNet V2 Medium'un karmaşıklık matrisi

Tablo 3.15 Şekil 3.19'daki karmaşıklık matrislerinin detaylı performans metrikleri

	BesLab				Histoplus				GBL			
Sınıf İsimleri	**P (%)**	**R (%)**	**F1 (%)**	**Destek**	**P (%)**	**R (%)**	**F1 (%)**	**Destek**	**P (%)**	**R (%)**	**F1 (%)**	**Destek**
HeadAbnormalities	85,68	88,18	86,91	10789	83,88	91,65	87,59	9529	80,15	89,03	84,35	6045
NeckAbnormalities	64,74	58,74	61,60	3690	71,45	54,20	61,64	3571	64,41	47,51	54,68	2427
Normal	69,94	82,11	75,54	598	68,51	54,27	60,57	433	68,05	56,05	61,47	381
TailAbnormalities	88,47	86,46	87,45	3384	93,21	94,57	93,88	4736	92,44	93,28	92,86	3778

P: Precision R: Recall F1: F1-Score

Tablo 3.14'e göre, diğer anormalliklere kıyasla daha belirgin karakteristik özellikler sergileyen PinHead, ShortTail, CurlyTail, DoubleTail, TwistedNeck ve TwistedTail

gibi anormallikler tipik olarak %70 veya üzerinde F1 skorlarına sahiptir. Sonuçlar umut verici olsa da, sınıf dağılımındaki dengesizlikler DoubleHead, LongTail ve ThinNeck gibi belirli sınıflar için doğruluğun azalmasına neden olmuştur. Bu sorunu hafifletmek için 4 sınıflı DB3_v2 veri kümeleri oluşturulmuş ve en iyi performans gösteren EfficientNet V2 küçük, orta ve büyük modelleri DB3_v2 veri kümeleri kullanılarak eğitilmiştir. En başarılı modeller için karmaşıklık matrisleri BesLab, Histoplus ve GBL boyama teknikleri için sırasıyla Şekil 3.19.a., 3.19.b. ve 3.19.c.'de gösterilmiştir. Sınıf bazında kesinlik (P), duyarlılık (R) ve F1-skor (F1) değerleri karmaşıklık matrislerinden elde edilmiştir ve Tablo 3.15'te kapsamlı bir şekilde detaylandırılmıştır. Karmaşıklık matrisleri incelendiğinde, DB3_v1 veri kümelerine benzer şekilde, DB3_v2 veri kümelerinde de HeadAbnormalities ve NeckAbnormalities sınıflarının birbirleriyle yanlış sınıflandırıldığı görülmektedir. Ayrıca, Normal sınıfına ait örneklerin genellikle TailAbnormalities sınıfı ile yanlış sınıflandırıldığı tespit edilmiştir.

3.2.2.5 ANOVA ve Tukey HSD ile Boyama Yöntemlerinin Sınıf Bazında İstatistiksel Olarak Karşılaştırılması

Tablo 3.16, 3.17 ve 3.18 BesLab, Histoplus ve GBL boyama yöntemleri arasındaki sırasıyla Kesinlik, Duyarlılık ve F1-Skoru performans ölçütleri üzerinde Tukey HSD testleri ile ANOVA kullanılarak yapılan istatistiksel karşılaştırmaları temsil etmektedir. İlk olarak, üç boyama yöntemi arasında istatistiksel olarak anlamlı farklılıklar olup olmadığını belirlemek için ANOVA kullanılmıştır. Daha sonra, hangi spesifik boyama yöntemlerinin performansta önemli farklılıklar sergilediğini belirlemek için Tukey HSD testi yapılmıştır. Her bir anormallik sınıfı ve her bir boyama yöntemi çifti için tablolarda şu bilgiler verilmiştir: i) P değeri: Gözlenen farklılıkların şans eseri meydana gelme olasılığı, ii) Reject: Sıfır hipotezinin (boyama yöntemleri arasında fark yoktur) $\alpha = 0.05$ anlamlılık düzeyinde reddedilip reddedilmediğini (True) veya (False) göstermektedir. Toplam-Sınıf satırları, o kategorideki toplam karşılaştırmalardan anlamlı farklılıkların (True) sayısını özetlemekte ve Toplam satırları ise tüm sınıflar için genel bir özet sunmaktadır. Tablolara ilişkin yorumlar ve detaylı bilgiler, bir sonraki alt başlık olan 'Tartışma ve Sonuç' bölümünde sunulacaktır.

3.2.3 Tartışma ve Sonuç

Sperm morfoloji analizi, kısırlık tanı ve tedavisinde kritik bir öneme sahiptir. Kliniklerde uzmanlar tarafından kullanılan iki temel değerlendirme yöntemi, BDSA ve görsel değerlendirme, kısırlık tanısında yaygın olarak uygulanmaktadır. Görsel

Tablo 3.16 Kesinlik (Precision) değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları

Kesinlik

		BesLab-HistoPlus		BesLab-GBL		Histoplus-GBL		
	Class	P value	Reject	P value	Reject	P value	Reject	TOTAL
HEAD	AmorfHead	0,0046	True	0	True	0	True	3\3
	DoubleHead	0,9942	False	0,0003	True	0,0002	True	2\3
	NarrowAcrosome	0,8661	False	0,0058	True	0,0012	True	2\3
	PinHead	0,992	False	0	True	0	True	2\3
	PyriformHead	0	True	0	True	0,0001	True	3\3
	RoundHead	0	True	0,6888	False	0	True	2\3
	TaperedHead	0,1785	False	0,9987	False	0,1955	False	0\3
	VacuolatedHead	0	True	0	True	0,0002	True	3\3
	Toplam-Sınıf	-	4\8	-	6\8	-	7\8	17\24
NECK	AsymetricNeck	0	True	0	True	0	True	3\3
	ThickNeck	0,027	True	0	True	0	True	3\3
	ThinNeck	0	True	0,0001	True	0,0087	True	3\3
	TwistedNeck	0	True	0,0011	True	0	True	3\3
	Toplam-Sınıf	-	4\4	-	4\4	-	4\4	12\12
TAIL	CurlyTail	0,0703	False	0	True	0	True	2\3
	DoubleTail	0,9912	False	0	True	0	True	2\3
	LongTail	0,0191	True	0,9004	False	0,0054	True	2\3
	ShortTail	0	True	0	True	0	True	3\3
	TwistedTail	0	True	0	True	0,9991	False	2\3
	Toplam-Sınıf	-	3\5	-	4\5	-	4\5	11\15
Normal	Normal	0	True	0	True	0,531	False	2\3
	Toplam	-	12\18	-	15\18	-	15\18	42\54

değerlendirme yönteminin başarısı büyük ölçüde uzmanların deneyimine bağlıdır. İnsan faktörünün önemli bir rol oynadığı bu yöntem, subjektif teşhislere ve yüksek oranda yanlışlığa yol açabilmektedir. BDSA sistemleri görsel değerlendirmeye kıyasla daha tutarlı ve tarafsız analizler sağlayabilmektedir. Bununla birlikte, bu sistemlerin çoğu sperm morfoloji analizinden ziyade sperm hareketlilik tahmini için kullanılmaktadır. Bu sorunları ele almak için son yıllarda sperm morfoloji analizi alanında birçok çalışma yürütülmektedir. Bilgisayar destekli tam otomatik sperm morfoloji analiz sistemleri tasarlamak için araştırmacıların ihtiyaç duyduğu en önemli kaynak iyi hazırlanmış veri kümeleridir. Sperm morfoloji analiz alanındaki veri setlerinin yaygınlaşması, bu alana olan ilgiyi önemli ölçüde artırmıştır. Ancak, literatürde analiz edilen veri kümeleri arasında, gelişmiş derin öğrenme modellerini yeterince destekleyen ve farklı sperm bölgelerindeki anormallikler için oldukça genelleştirilebilir sınıflandırıcıların geliştirilmesine olanak tanıyan herhangi bir veri setine rastlanmamıştır. Bu tez kapsamında, çok daha fazla sayıda anormallik sınıfına ve üç farklı boyama tekniğiyle elde edilen örnek görüntülere sahip olan Hi-LabSpermMorpho veri kümesi, kapsamlı, iyi etiketlenmiş ve çeşitli bir yapı sağlayarak mevcut sınırlamaları ortadan kaldırmak amacıyla sunulmuştur. Bu veri setinin oluşturulması, yüksek kaliteli ve tutarlı veriler sağlamak için sistematik görüntü toplama ve ön işleme adımlarının ardından titiz hazırlık ve

Tablo 3.17 Duyarlılık değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları

		Duyarlılık						
		BesLab-HistoPlus		BesLab-GBL		Histoplus-GBL		
	Class	P value	Reject	P value	Reject	P value	Reject	TOTAL
HEAD	AmorfHead	1	False	0	True	0	True	2\3
	DoubleHead	0,8201	False	0	True	0	True	2\3
	NarrowAcrosome	0,2618	False	0,1904	False	0,0036	True	1\3
	PinHead	0,316	False	0	True	0	True	2\3
	PyriformHead	0	True	0	True	0,0125	True	3\3
	RoundHead	0	True	0	True	0	True	3\3
	TaperedHead	0,4996	False	0,0037	True	0,0837	False	1\3
	VacuolatedHead	0	True	0	True	0,0053	True	3\3
	Toplam-Sınıf	-	3\8	-	7\8	-	7\8	17\24
NECK	AsymetricNeck	0	True	0	True	0,9999	False	2\3
	ThickNeck	0	True	0	True	0	True	3\3
	ThinNeck	0	True	0	True	0,9991	False	2\3
	TwistedNeck	0,1615	False	0	True	0	True	2\3
	Toplam-Sınıf	-	3\4	-	4\4	-	2\4	9\12
TAIL	CurlyTail	0,002	True	0	True	0	True	3\3
	DoubleTail	0,1489	False	0	True	0	True	2\3
	LongTail	0	True	0,1999	False	0	True	2\3
	ShortTail	0	True	0	True	0	True	3\3
	TwistedTail	0	True	0	True	0,6954	False	2\3
	Toplam-Sınıf	-	4\5	-	4\5	-	4\5	12\15
Normal	Normal	0	True	0	True	0,8987	False	2\3
	Toplam	-	11\18	-	16\18	-	13\18	40\54

boyama süreçlerini içermektedir. Oluşturma aşamasında, çeşitli anormallik türleri için farklılık gösterebilecek çok çeşitli morfolojik ayrıntıları yakalamak amacıyla üç farklı Diff-Quick boyama seti (BesLab, Histoplus ve GBL) kullanılmıştır. Farklı boyamaların sperm anormalliklerinin görünürlüğünü ve sınıflandırılmasını etkilediği gözlemlenmiştir. Sınıflandırma sonuçlarından Histoplus boyasının en ayırt edilebilir özellikleri sağladığı ve tüm mimarilerde daha yüksek sınıflandırma doğrulukları elde edildiği gözlemlenmiştir.

Sperm örneği kayıt aşamasında, ham görüntüleri yakalamak için gelişmiş parlak alan mikroskopları ve laboratuvarımız tarafından tasarlanan yenilikçi bir görüntü toplama sistemi kullanılmıştır. Görüntü elde etmek için cep telefonu kullanımı, sperm morfoloji analizini daha erişilebilir ve ölçeklenebilir hale getirerek, önerilen sistemin önemli bir avantajını oluşturmuştur. Bu sayede, standart mikroskoplar ve özel kamerası olmayan cihazlarla bile sperm örneklerinin yüksek kaliteli görüntüleri elde edilip, otomatik sperm morfoloji analizi yapılabilir. Ayrıca, çoğu merkezde yetersiz uzman istihdamı nedeniyle kısırlık tedavisinde karşılaşılan uzmanlık eksikliğinin, önerdiğimiz otomatik sperm morfoloji analiz çerçevesiyle aşılabileceğine inanmaktayız.

Tablo 3.18 F-Skor değerlerine dayalı boyama yöntemleri arasında istatistiksel karşılaştırma: Tukey HSD ile ANOVA sonuçları

F-Score

		BesLab-HistoPlus		BesLab-GBL		Histoplus-GBL		
	Class	P value	Reject	P value	Reject	P value	Reject	TOTAL
HEAD	AmorfHead	0,1184	False	0	True	0	True	2\3
	DoubleHead	0,7848	False	0	True	0	True	2\3
	NarrowAcrosome	0,3725	False	0,0012	True	0	True	2\3
	PinHead	0,5003	False	0	True	0	True	2\3
	PyriformHead	0	True	0	True	0,0009	True	3\3
	RoundHead	0	True	0,0006	True	0	True	3\3
	TaperedHead	0,2485	False	0,1155	False	0,9201	False	0\3
	VacuolatedHead	0	True	0	True	0,0008	True	3\3
	Toplam-Sınıf	-	3\8	-	7\8	-	7\8	17\24
NECK	AsymetricNeck	0	True	0	True	0,271	False	2\3
	ThickNeck	0	True	0	True	0	True	3\3
	ThinNeck	0	True	0	True	0,5453	False	2\3
	TwistedNeck	0,0005	True	0	True	0	True	3\3
	Toplam-Sınıf	-	4\4	-	4\4	-	2\4	10\12
TAIL	CurlyTail	0,0092	True	0	True	0	True	3\3
	DoubleTail	0,4599	False	0	True	0	True	2\3
	LongTail	0	True	0,0748	False	0	True	2\3
	ShortTail	0	True	0	True	0	True	3\3
	TwistedTail	0	True	0	True	0,9339	False	2\3
	Toplam-Sınıf	-	4\5	-	4\5	-	4\5	12\15
Normal	Normal	0	True	0	True	0,745	False	2\3
	Toplam	-	12\18	-	16\18	-	13\18	41\54

Önerilen çalışmanın model eğitimi aşamasında, 35 farklı derin öğrenme mimarisi kullanılarak ayrıntılı hiperparametre analizi ile sınıflandırma temel performansı oluşturulmuştur. Bu 35 modelin seçimi, sperm morfoloji sınıflandırma literatüründe yaygın olarak benimsenmiş olmaları, performansları, hesaplama verimlilikleri ve kullanımları gibi kriterlere dayanmaktadır. Literatürdeki önceki çözümlere bakıldığında, tıbbi görüntü sınıflandırmasında geleneksel özellik çıkarma yöntemlerinin yaygın olarak kullanıldığı görülmektedir. KAZE, SURF, MSER gibi tanımlayıcı tabanlı özellikler ve DWT, DTCWT gibi doku özellikleri manuel olarak elde edilmekte ve geleneksel makine öğrenmesi yöntemlerinde çeşitli sınıflandırma problemleri için kullanılmaktadır [50, 106]. Ancak, bu yöntemlerle geliştirilen sistemler manuel ön işleme gerektirdiği için tam otomatik bir şekilde kullanılamamaktadır. Geleneksel öznitelik çıkarma yöntemlerine alternatif olan derin öğrenme yöntemlerinde ise öznitelik çıkarma için herhangi bir ön işleme gerekliliği bulunmamaktadır. Derin öğrenme yöntemleri veriye oldukça aç mimarilerdir ve yüksek genelleme kapasitesine sahip modellerin eğitilebilmesi için örüntü öğrenme sürecinde önemli miktarda verinin sağlanması gerekmektedir. Ayrıca, eğitim sürecinde seçilen hiperparametre konfigürasyonu modelin eğitimini önemli ölçüde etkilemektedir. Bu tezde, derin öğrenme mimarilerinin eğitim sürecinde en uygun hiperparametre yapılandırmasını belirlemek için çeşitli

optimizasyon yöntemleri ve öğrenme oranları kullanılmıştır.

Sperm morfoloji analiz sistemlerinin performansını etkileyen en önemli faktörlerden biri modellerin genelleme kapasitesidir. Yüksek genelleme kapasitesine sahip modeller, eğitim verilerindeki performanslarına benzer şekilde, gerçek dünya verileri üzerinde aşırı uyum göstermeden iyi performans göstermektedir. Veri kümesindeki örnek sayısı ve sınıf çeşitliliği, modellerin genelleme kapasitesini etkileyen faktörler arasındadır. Sperm morfoloji alanında yaygın olarak kullanılan veri kümelerine ilişkin bilgiler Tablo 1.1'de ayrıntılı olarak sunulmuştur.

Bu alandaki [19] adlı çalışmada, Yi ve arkadaşları 640×480 piksel çözünürlükte yaklaşık 300 sperm başı görüntüsü içeren bir veri kümesi oluşturmuşlardır. Bu veri kümesi toplam 4 sınıf içermektedir: Normal, Small, Elongated ve Megalo olmak üzere toplam 4 sınıf içeren bu veri kümesi, görüntüler arasındaki ortalama kare hatasına göre sınıflandırılmıştır. Benzer bir çalışmada Ramos ve arkadaşları 512×512 piksel çözünürlükte toplam 590 sperm başı görüntüsü içeren bir veri kümesi oluşturmuşlardır [20]. Sperm başı görüntüleri bilgisayar destekli bir karyometrik görüntü analiz programı kullanılarak analiz edilmiş ve normal ve anormal olmak üzere iki kategoride sınıflandırılmıştır. Bu iki çalışmadan farklı olarak Abbiramy ve arkadaşları [21] çalışmalarında üç farklı yapay sinir ağı tabanlı sınıflandırıcı kullanarak küçük bir veri kümesi üzerinde sınıflandırma yapmıştır. DSÖ laboratuvar kılavuzundaki görüntüler kullanılarak oluşturulan veri kümesi, 86×100 piksel çözünürlükte sadece 91 sperm başı görüntüsünden oluşmaktadır. Normal ve anormal olarak kategorize edilen etiketli görüntüler ikili sınıflandırma problemi olarak ele alınmıştır. Bu çalışmalarda kullanılan her üç veri kümesi de yalnızca sperm başı görüntülerinden oluşmaktadır. Ayrıca, bu veri kümelerindeki görüntü sayısı otomatik sperm morfoloji analiz sistemleri geliştirmek için yeterli değildir. Bu çalışmaların aksine, [22] ve [23] araştırmacıları görsel değerlendirme ve BDSA sistemlerini kullanmışlardır. Aksoy ve ark. [22] farklı boyama tekniklerinin sperm morfolojisi üzerindeki etkilerini araştırmak için 67 hastadan alınan örnekleri kullanarak bir veri seti oluşturmuştur. Görsel değerlendirme yönteminin kullanıldığı çalışma, sperm uzunluklarının kullanılan boyama tekniğine göre değişebileceğini gösterse de, veri seti tam otomatik sperm morfoloji analiz sistemleri oluşturmak için uygun değildir. Lammers ve arkadaşları, geleneksel manuel değerlendirme ile BDSA sistemlerini karşılaştırmak için 250 hastanın örneklerini kullanarak bir veri seti oluşturmuştur [23]. Ancak bu veri kümesi de tam otomatik sistemler geliştirmek için uygun değildir.

Görüntüleme teknolojisindeki gelişmeler, sperm morfolojisindeki veri setlerini

hem kalitatif hem de kantitatif iyileştirmelerle zenginleştirmiştir. Ghasemian ve arkadaşları, 235 hastadan elde edilen 576×764 piksel çözünürlükte toplam 1457 sperm görüntüsü içeren HSMA-DS veri setini oluşturmuştur [24]. SMA sistemi ile bazı ön işlemlerden sonra, spermin baş ve kuyruk gibi farklı kısımları tespit edilmiş ve normal ve anormal olarak sınıflandırılmıştır. Javadi ve arkadaşları aynı veri kümesini değiştirerek MHSMA veri kümesini oluşturmuşlardır [34]. HSMA-DS [24] veri setinden farklı olarak, sperm başları manuel olarak işaretlenmiş ve görüntüler sperm başı tam merkezde olacak şekilde kırpılmıştır. 128×128 piksel çözünürlükte 1540 görüntü içeren veri kümesi, sperm başı, kuyruk-boyun, akrozom ve vakuol kategorileri altında normal ve anormal olmak üzere iki sınıf içermektedir. Bu iki veri kümesi biraz daha fazla görüntü içermesine rağmen detaylı bir anormallik sınıflandırması içermemektedir. Özellikle spermlerin manuel olarak belirli bir formata getirildiği MHSMA veri kümesi, gerçek dünya problemlerine uygulanabilecek tam otomatik bir sperm analiz sistemini eğitmek için uygun değildir.

Literatürde mevcut veri kümeleri incelendiğinde, en fazla beş farklı anormallik sınıfının etiketlendiği ve bu sınıfların sadece kafa anormalliklerine ait olduğu görülmüştür [69]. Veri kümelerindeki örnek sayısı ile ilgili olarak, daha az sınıfa sahip veri kümeleri nispeten daha fazla örnek içerirken [20, 21, 24, 34, 35, 38, 44, 105], üçten fazla sınıfa sahip veri kümelerinin genellikle az sayıda örnek içerdiği tespit edilmiştir [7, 19, 29]. Nispeten çok sayıda numune içeren veri kümelerinde, ayrıntılı anormallik sınıflarının etiketlenmediği ve sperm numunelerinin normal ve anormal olmak üzere yalnızca iki ana sınıfa ayrıldığı sıklıkla gözlemlenmiştir. Boyama tekniklerinin sperm morfolojileri üzerindeki etkilerini inceleyen görsel değerlendirme yöntemlerine rağmen, özellikle otomatik sperm analiz sistemleri geliştirmek için hiçbir veri seti oluşturulmamıştır [22].

Bu tez kapsamında, önerilen veri toplama yöntemi kullanılarak Hi-Lab-SpermMorho adı verilen kapsamlı bir sperm morfoloji analiz veri seti oluşturulmuştur. Elde edilen görüntüler, iki uzman tarafından normal sınıf da dahil olmak üzere toplam 18 sınıf olarak kapsamlı bir şekilde etiketlenmiştir. Sperm başı kategorisi toplam 8 sınıf içermektedir: amorphous, tapered, double, pyriform, pin, vacuolated, narrow acrosome area ve round. Boyun kategorisi thin, thick, twisted ve asymmetrical olmak üzere 4 sınıftan, kuyruk kategorisi ise double, curly, long, short ve twisted olmak üzere 5 sınıftan oluşmaktadır. Sonuç olarak, Hi-Lab-SpermMorpho veri kümesi normal sınıf da dahil olmak üzere toplam 18 farklı sınıf içermektedir. Önerilen veri kümesi, literatürdeki diğer veri kümelerine kıyasla benzersiz ve çeşitlilik açısından zengindir. Boyama yöntemlerinin sperm morfolojisi üzerindeki etkileri incelenerek, Histoplus Diff-Quick boyama yöntemi

ile elde edilen görüntülerin tam otomatik bir sperm morfoloji analiz sistemi geliştirilmesinde daha etkili olduğu gösterilmiştir. Literatürde derin öğrenme yöntemleri için kullanılabilecek farklı boyama yöntemleriyle elde edilmiş tek veri kümesidir. Ayrıca, kamuya açık Hi-LabSpermMorpho_v1 veri setindeki görüntüler ana başlıklar altında kategorize edilerek çeşitli senaryolarda kullanılmak üzere 4 sınıflı Hi-LabSpermMorpho_v2 veri seti oluşturulmuştur. Geleneksel makine öğrenimi yöntemlerinin ve geleneksel özellik çıkarma tekniklerinin tam otomatik sperm morfoloji analiz sistemleri geliştirmedeki sınırlamaları göz önünde bulundurularak, 35 farklı derin öğrenme yöntemi kullanılarak çeşitli senaryolara uygun temel sınıflandırma performansları oluşturulmuştur. Önerilen Hi-LabSpermMorpho_v1 ve Hi-LabSpermMorpho_v2 veri kümelerinin kullanılabilirliği ayrıntılı olarak sunulmuştur.

Uygulanan tüm derin öğrenme mimarilerinin Kesinlik, Duyarlılık ve F1-Skoru gibi performans metrikleri ile yapılan istatistiksel analiz, boyama yöntemi seçiminin sperm anormalliği sınıflandırma sistemlerinin performansını önemli ölçüde etkilediğini göstermektedir. Elde edilen istatistiksel analiz sonuçları Tablo 3.16, 3.17 ve 3.18'de sunulmaktadır. Özellikle, BesLab ile GBL ve Histoplus ile GBL arasındaki karşılaştırmalar, Tukey HSD sonuçlarıyla ANOVA'nın Toplam satırlarında gösterildiği gibi, tüm ölçümler ve anormallik sınıfları arasında en yüksek sayıda anlamlı farklılık göstermektedir. Bu, GBL boyama yönteminin sınıflandırma performansı üzerindeki etkisinin, hem BesLab hem de Histoplus'tan belirgin şekilde farklı olduğunu ve GBL'nin genellikle daha düşük performans gösterdiğini ortaya koymaktadır. Toplam-Sınıf satırlarına odaklanıldığında, boyama yönteminin etkisinin farklı anormallik kategorilerine göre değiştiği görülmektedir. Boyun anormallikleri için, boyama yöntemleri arasındaki tüm karşılaştırmalar, Kesinlik ve F1-Skoru açısından anlamlıdır. Bu durum, boyama yöntemlerinin bu kategorideki sınıflandırma performansını önemli ölçüde etkilediğini vurgulamaktadır. Benzer şekilde, kuyruk anormallikleri için yapılan karşılaştırmaların çoğu, özellikle GBL'yi içerenler, önemli farklılıklar göstermektedir. Bu durum, boyama yönteminin kuyruk anormalliklerinin tespitini ve sınıflandırılmasını önemli ölçüde etkilediğini göstermektedir. Bununla birlikte, baş anormallikleri için BesLab ve Histoplus arasındaki karşılaştırmalarda daha az önemli farklılıklar gözlenmiştir. Bu durum, Baş anormalliklerinin BesLab ve Histoplus arasındaki seçimden daha az etkilendiğini, ancak GBL boyama yönteminden olumsuz olarak etkilendikleri göstermektedir. Baş anormallikleri içinde ise Tapered Head sınıfıyla ilgili ilginç bir gözlem ortaya çıkmaktadır. Diğer sınıfların aksine, Tapered Head sınıfı istatistiksel testlerde yüksek sayıda "Yanlış" sonuç göstermiştir, bu da boyama yöntemleri arasındaki performans metriklerindeki

farklılıkların bu sınıf için istatistiksel olarak anlamlı olmadığını göstermektedir. Bu da Tapered Head anormalliği için sınıflandırma performansının her üç boyama yönteminde de nispeten tutarlı olduğunu göstermektedir. Önemli farklılıkların olmaması, kullanılan boyama tekniğinden bağımsız olarak tutarlı bir şekilde yakalanan Tapered Head spermin farklı morfolojik özelliklerinden kaynaklanıyor olabilir. Bununla birlikte, tutarlı performans ölçütlerine rağmen, Tapered Head sınıfı kayda değer yanlış sınıflandırma oranları sergilemiştir. Örnekler sıklıkla Amorf Head, Narrow Acrosome, Pyriform Head ve Thin Neck olarak yanlış sınıflandırılmıştır. Bu durum, Tapered Head anormalliklerinin bu sınıflarla morfolojik benzerlikler paylaştığını ve sınıflandırma sürecinde karışıklığa yol açtığını göstermektedir. Bu anormallikler arasındaki şekil ve yapı benzerlikleri, uygulanan boyama yönteminden bağımsız olarak derin öğrenme modellerinin bunları doğru bir şekilde ayırt etmekte zorlanmasına neden olmuş olabilir. Tapered Head sınıfı için yüksek sayıda "Yanlış" sonuç, boyama yönteminin bu özel anormallik için sınıflandırma performansını iyileştirmek için tek başına yeterli olmayabileceğini göstermektedir. Bunun yerine, modelin Tapered Head ve benzer sınıflar arasında ayrım yapma kabiliyetini artırmak için ek stratejiler yapmak gerekmektedir. Bu stratejiler arasında, Tapered Head ve yanlış sınıflandırıldığı anormallik türleri için eğitim verilerinin miktarını ve çeşitliliğini artırmak veya ince morfolojik farklılıkları yakalayabilen daha gelişmiş modelleme teknikleri kullanmak yer alabilir. Genel olarak bulgular, GBL'in önemli farklılıkları ve genellikle daha zayıf performans ölçütleri nedeniyle derin öğrenme tabanlı otomatik sperm morfoloji analiz sistemleri için daha az uygun olduğunu göstermektedir. GBL ile ilgili görüntü kalitesi sorunları (sperm başlarındaki aşırı emilim nedeniyle aşırı kararma ve gürültüye neden olan ve kritik morfolojik özellikleri gizleyen daha sert geçişler gibi), derin öğrenme modelinin sperm anormalliklerini doğru bir şekilde sınıflandırma yeteneğini olumsuz yönde etkilemektedir. Buna karşılık, BesLab ve Histoplus otomatik analiz için daha etkili boyama yöntemleridir. Her iki yöntem de GBL boyama yöntemine kıyasla daha yüksek genel doğruluk sağlamaktadır. Histoplus, özellikle boyun ve kuyruk anormallikleri gibi yapıların görünürlüğünü artırmada, gelişmiş boya emilimi nedeniyle bazı sınıflarda hafif avantajlar sunabilmektedir. Ancak, bu avantajlar tüm sınıflarda tutarlı bir şekilde istatistiksel olarak anlamlı değildir. BesLab ve Histoplus arasındaki seçim özel ihtiyaçlara bağlı olabilmektedir. Histoplus, gelişmiş leke emiliminin özellik görünürlüğünü artırdığı sınıflarda marjinal faydalar sunabilmekte, ancak sedimantasyon nedeniyle arka plan karmaşıklığı ortaya çıkarmaktadır. BesLab, daha az arka plan gürültüsü ile daha pürüzsüz görüntüler sağlamaktadır, ancak yine de Histoplus ile karşılaştırılabilir performans sunmasına rağmen bazı sperm bölümlerini az boyayabilmektedir. Özel kullanım durumları boyama yönteminin

seçimini etkileyebilmektedir. Boyun ve kuyruk anormalliklerinin tespit edilmesine odaklanılıyorsa, Histoplus gelişmiş boya emilimi nedeniyle biraz daha avantajlı olabilmektedir. Genel amaçlar için veya arka plan karmaşıklığı bir endişe kaynağı olduğunda, BesLab daha pürüzsüz görüntüleri ve daha az arka plan gürültüsü nedeniyle tercih edilebilir.

Özetle, istatistiksel analiz, GBL boyama yönteminin çoğu sperm anormallik sınıfı ve performans metriği açısından BesLab ve Histoplus boyama yöntemlerinden anlamlı derecede farklı olduğunu ve GBL'nin sınıflandırma performansı üzerinde daha düşük bir etki sağladığını göstermektedir. BesLab ve Histoplus, daha az anlamlı farkla birbirlerine daha çok benzemektedir.

3.2.4 Gelecek Çalışmalar

Gelecek çalışmalarda CNN mimarilerine alternatif olarak kullanılan ViT [74] mimarileri ile Hi-LabSpermMorpho veri kümesi üzerinde sınıflandırma çalışmaları yapılması hedeflenmektedir. Bir sonraki çalışmada, yüksek doğrulukta bir sperm morfolojisi analiz sistemi tasarlamak amacıyla CNN ve ViT mimarileri, çeşitli topluluk öğrenme yöntemleriyle birleştirilecektir. Ayrıca, veri artırma tekniklerinin veri miktarını artırarak sistem performansını olumlu yönde etkilediği, daha önceki çalışmalarda kanıtlanmıştır [26, 27]. Bu nedenle, döndürme ve yansıtma gibi geleneksel veri artırma yöntemleri özellikle normal sınıf görüntülerine uygulanacak ve sistem performansına etkileri analiz edilecektir. Geleneksel veri artırma yöntemlerinin yetersiz kaldığı durumlarda, görüntülerin hem sayısını hem de kalitesini artırmak için Generative Adversarial Networks (GAN) [108] gibi gelişmiş modeller kullanılacaktır. Ayrıca Kontrast Sınırlı Adaptif Histogram Eşitleme (CLAHE), Parçacık Sürü Optimizasyonu (PSO), Yapay Arı Kolonisi (ABC), Genetik Algoritma (GA), Diferansiyel Evrim (DE) gibi görüntü iyileştirme algoritmaları kullanılacaktır, Simulated Annealing (SA), Bacteria Foraging Algorithm (BFA), Cat Swarm Optimization (CSO), Dragonfly Algorithm (DA) ve Black Hole'un görüntü iyileştirmede model performansına etkilerini incelemek için kullanılması planlanmaktadır [122]. Özellikle, RMSProp ve Adamax optimizasyon yöntemleri analiz edildiğinde, modellerin genel olarak her üç boyama setinde de 10^{-3} ve 10^{-5} öğrenme oranları ile en başarılı sonuçları elde ettiği gözlemlenmiştir. Bu gözlem, farklı optimizasyon yöntemlerinin ve öğrenme oranlarının model performansı üzerinde önemli bir etkisi olduğunu ortaya koymaktadır. Bu nedenle, gelecek çalışmalarda optimizasyon yöntemlerinin CNN ve ViT mimarilerinin eğitim süreci üzerindeki etkilerini gözlemlemek için daha kapsamlı bir süreç tasarlanacaktır. Ayrıca önerilen çalışmamızda Hi-LabSpermMorpho veri kümesi ile sperm morfoloji analizi alanında ImageNet

benzeri bir [123] veri kümesi oluşturulmuştur. Hi-LabSpermMorpho veri kümesini kullanarak, İlhan ve arkadaşları tarafından [26]'de önerilen iki aşamalı ince ayar yaklaşımını, performans iyileştirmesi için mevcut diğer sperm morfoloji analizi veri kümelerine uygulamak amaçlanmaktadır. Bu yöntemle, ImageNet kullanılarak oluşturulan ön eğitimli modeller gibi sperm morfoloji alanına özgü ön eğitimli modeller üretilecektir. Bu ön-eğitim modellerinin mevcut ve yeni sperm morfoloji veri kümeleri üzerindeki etkileri karşılaştırmalı olarak analiz edilecektir. Toplam 35 farklı derin öğrenme mimarisi ile oluşturulan sınıflandırma temel performansı, veri setinin çeşitli çalışma alanlarında uygulanabilirliğini göstermektedir. Deneysel çalışmalar incelendiğinde MobileNet V2 ve ShuffleNet gibi hafif ağlar ile elde edilen sonuçların umut verici olduğu görülmektedir. Bu bulgulardan yola çıkarak TensorFlow Lite, ONNX Runtime, PyTorch Mobile ve ML Kit for Firebase gibi kütüphaneler kullanılarak uç cihazlarda çalışabilecek modellerin elde edilmesi hedeflenmektedir. Bu modeller ile cep telefonu, tablet veya IoT cihazları gibi uç cihazlar üzerinde tam otomatik sperm morfoloji analizi yapabilen sistemlerin geliştirilmesi hedeflenmektedir.

4
DERİN ÖĞRENME YÖNTEMLERİ İLE SPERM YÖRÜNGE ANALİZİ

4.1 Hi-LabSpermTracking Veri Seti ile Sperm Yörünge Analizi

Sperm analizinde, sperm morfolojisi, konsantrasyonu, DNA bütünlüğü ve hareketliliği gibi parametreler, kısırlık tanısında önemli bir rol oynamaktadır. Ancak, klinik değerlendirmelerde erkek kısırlık teşhisi genellikle sperm konsantrasyonu ve hareketliliği üzerinde yoğunlaşmaktadır. Sperm konsantrasyonu, ejakülatın birim hacmi başına düşen sperm hücrelerinin sayısını ifade etmektedir. DSÖ tarafından laboratuvarlardaki test yöntemlerini standartlaştırmak amacıyla yayınlanan el kitabına göre, mililitre başına 15 milyon sperm bulunması normal kabul edilmektedir [2]. Düşük sperm konsantrasyonu (oligospermi), genetik anomaliler [124], spermatogenezdeki fonksiyonel bozukluklar [125], inmemiş testis [126] veya azospermi [127] gibi durumların bir göstergesi olabilmektedir.

Sperm konsantrasyonu, potansiyel patolojilerin altında yatan sebepler hakkında önemli bilgiler sağlasa da, spermlerin işlevsel kapasitesini tam anlamıyla yansıtmamaktadır. Erkek fertilitesine ilişkin daha kapsamlı bir değerlendirme yapabilmek için sperm hareketliliğinin incelenmesi büyük önem taşımaktadır. Sperm hareketliliği, sperm hücrelerinin hareket ve yörüngelerine dayalı olarak değerlendirilmektedir. Bu bağlamda, sperm hareketliliği dört kategoriye ayrılmaktadır: düzgün doğrusal hareketli (Grade A), hareketli (Grade B), doğrusal olmayan hareketli (Grade C) ve hareketsiz (Grade D) [2]. DSÖ'nün yayınladığı el kitabına göre, toplam sperm sayısının en az %40'ının hareketli olması, ayrıca hareketli sperm hücrelerinin en az %4'ünün düzgün doğrusal hareketli (Grade A+B) olması gerekmektedir. Ayrıca, düşük sperm hareketliliği (astenozoospermi), testis hasarı [16], sperm mitokondriyal disfonksiyonu [17] veya varikosel [18] gibi durumların bir göstergesi olabilmektedir.

Günümüzde sperm analizi iki farklı teknik kullanılarak yapılmaktadır. Bunlardan

ilki olan görsel değerlendirme, emek yoğun ve verimliliği düşük bir yöntemdir. Sperm hücreleri, Makler odacığı ve hemositometre gibi tıbbi malzemeler yardımıyla mikroskop altında incelenerek sayım ve hareketlilik analizi işlemleri gerçekleştirilmektedir. Ancak bu yöntemle elde edilen sonuçlar önemli ölçüde gözlemcinin deneyimine ve uzmanlığına bağlıdır. Bu süreci otomatikleştirmek için geliştirilen BDSA sistemleri, görsel değerlendirmeye kıyasla daha doğru, tutarlı ve objektif analizler gerçekleştirme olanağı sağlamaktadır. Bu sistemler, semen örneklerinin mikroskobik görüntüleri ve videolarında sperm hücrelerini tanıyıp lokalize ederek, sperm konsantrasyonu ve hareketliliğini hesaplamak için birden fazla spermin izlenmesine olanak tanımaktadır. Ancak, yüksek kaliteli mikroskoplara entegre edilmiş yüksek performanslı bilgisayarların kullanılması, veri işleme ve depolama gereksinimlerini önemli ölçüde artırmakta ve pahalı ekipmanlar, profesyonel teknisyenlerin istihdamını zorunlu kılmaktadır. Bu sebeplerle, BDSA sistemlerinin küresel ölçekte yaygın bir şekilde uygulanması oldukça zordur [73].

Yukarıda belirtilen kısıtlamalar nedeniyle, erkek kısırlığının ekonomik, erişilebilir, doğru ve tarafsız bir şekilde teşhis edilmesine duyulan ihtiyaç dünya çapında devam etmektedir [3]. Bu ihtiyaca yanıt olarak, son yıllarda sperm morfoloji analizi alanında çeşitli çalışmalar yapılmaktadır [26–28, 55, 99, 105]. Ancak, kısırlık tanısının kritik bir bileşeni olan sperm hareketlilik analizine odaklanan çalışmalar hala sınırlı kalmaktadır. Bu boşluğun temel nedenlerinden biri, sperm hücrelerinin hareketlilik ve izleme özelliklerini içeren geniş, çeşitli ve iyi etiketlenmiş açık kaynak veri setlerinin eksikliğidir. Bu tür veri kümelerinin oluşturulması, manuel etiketleme süreci ve kontrollü laboratuvar ortamları gereksinimi nedeniyle önemli zorluklar doğurmakta ve süreci hem zaman alıcı hem de maliyetli hale getirmektedir. Sperm hareketliliğini ölçmek için görsel değerlendirme ve BDSA gibi geleneksel yaklaşımlar uzun süredir kullanılmakta olsa da, bu yöntemler tek başlarına güvenilir ve ölçeklenebilir çözümlere yönelik artan talebi tam anlamıyla karşılayamamaktadır. Sonuç olarak, sperm hareketlilik analizi alanında etkili, hızlı, doğru, uygun maliyetli ve objektif değerlendirmeleri sağlayacak daha gelişmiş otomatik teknolojilerin geliştirilmesine küresel çapta ihtiyaç bulunmaktadır. Geleneksel yöntemlerle karşılaştırıldığında, en yeni derin öğrenme yaklaşımları sperm morfoloji analizi için uygun maliyetli ve tarafsız çözümler sunmaktadır [26–28, 55, 99, 105]. Benzer şekilde, sperm hareketlilik ölçümü için derin öğrenme tabanlı gelişmiş tespit ve izleme algoritmalarının kullanılması, bu mevcut eksikliklerin üstesinden gelmek için umut verici bir strateji olarak kabul edilmektedir [39, 41, 62, 64, 65, 67].

Tez kapsamında, semen örneklerinde sperm hücrelerinin tespiti ve takibi için

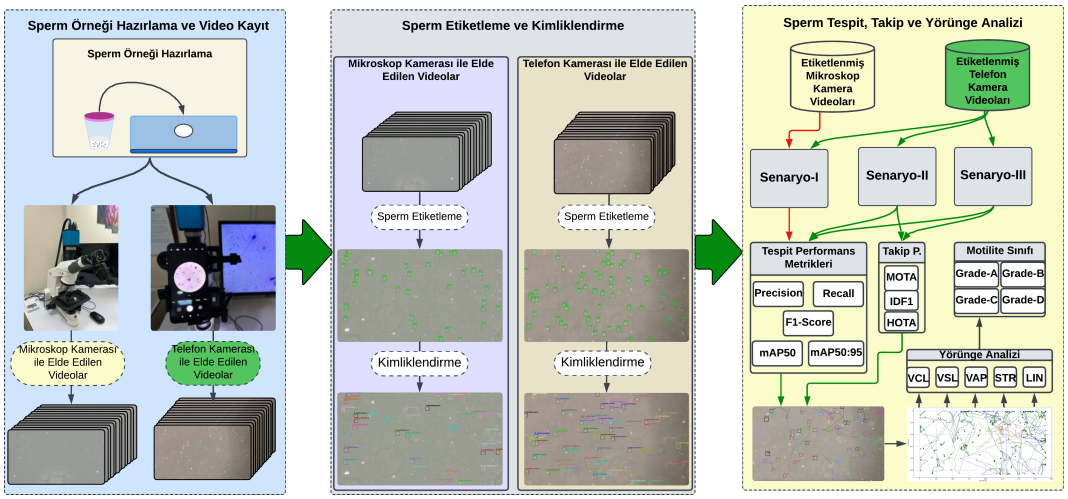
yeni bir yaklaşım sunulmaktadır. Geliştirdiğimiz Hi-LabSpermTracking veri kümesi kullanılarak, CNN tabanlı You Only Look Once v8 (YOLOv8) mimarisi ile dönüştürücü tabanlı Real-Time Detection Transformer (RT-DETR)[128] mimarisinin sonuçları birleştirilerek teorik ve pratik kullanımı içeren üç farklı senaryoda değerlendirilmiştir. Bu çalışmanın amacı, YOLOv8'in hızlı özellik çıkarma kabiliyetini RT-DETR'nin bağlamsal analiz gücü ile birleştirerek sahne içindeki nesne ilişkilerinin daha kapsamlı bir şekilde incelenmesini sağlayan bir sistem tasarlamaktır. YOLO hızlı ve güvenilir birincil tespit sağlarken, RT-DETR tarafından sunulan derinlemesine analiz, sistemin kapsamlılığını artırmaktadır. Ayrıca tez kapsamında, sperm hücrelerini izlemek için DeepSORT izleme algoritması kullanılarak tespit ve izleme metriklerinin analizi yapılmıştır. Bu kombinasyon, tespit ve izleme performansını artırarak kısırlık teşhisi ile ilgili uygulamalar için etkili bir çözüm sunmaktadır.

4.1.1 Materyal ve Yöntem

Bu tez kapsamında oluşturulan Hi-LabSpermTracking veri seti üzerinde, literatürde yaygın olarak kullanılan teorik veri seti bölütleme yöntemlerinin yanı sıra, gerçek dünya problemlerini simüle eden pratik veri seti bölütleme yöntemleri uygulanarak, üç farklı senaryo altında sperm tespit ve takip çalışmaları gerçekleştirilmiştir. Önerilen sisteme ait akış diyagramı Şekil 4.1'de sunulmaktadır. Sperm örneği hazırlama ve video kaydı aşamasında, hastalardan alınan sperm örnekleri mikroskop altında incelenmiş ve iki farklı görüntü elde etme yöntemi kullanılarak video kayıtları oluşturulmuştur. Elde edilen videolar, sperm etiketleme ve kimliklendirme adımında işlenerek etiketlenmiştir. Hi-LabSpermTracking veri setinin oluşturulma süreci, Bölüm 2 Materyaller başlığı altında ayrıntılı bir şekilde açıklanmıştır. Oluşturulan Hi-LabSpermTracking veri seti, üç farklı senaryo çerçevesinde sperm tespit ve takip problemleri üzerinde değerlendirilmiştir. Bu senaryolarda kullanılan yöntemler, aşağıdaki başlıklar altında detaylı bir şekilde ele alınmıştır.

4.1.1.1 Sperm Tespit Modülü

Bu çalışmada, Hi-LabSpermTracking verisetinde sperm tespiti gerçekleştirmek amacıyla son teknoloji derin öğrenme yöntemleri olan YOLOv8 ve RT-DETR [128] kullanılmıştır. Bu modeller, sahip oldukları tamamlayıcı yetenekler ve birbirleriyle yüksek uyumlulukları dikkate alınarak seçilmiştir. YOLOv8, nesne tespitinde hızlı ve hafif bir yapı sunarken, RT-DETR, dikkat mekanizmalarına dayalı daha derin bağlamsal analiz yetenekleriyle öne çıkmaktadır. Bu bölümde, her



Şekil 4.1 Önerilen yönteme ait akış diyagramı

iki modelin mimari tasarımlarının detaylı bir incelemesi yapılmakta, modellerin doğal avantajları açıklanmakta ve bu araştırma kapsamında benimsenmelerinin arkasındaki bilimsel ve teknik gerekçeler analiz edilmektedir.

YOLOv8, 'You Only Look Once' ailesinin en yeni sürümlerinden biridir ve nesne algılama, sınıflandırma ve segmentasyon gibi görevler için optimize edilmiş bir derin öğrenme mimarisi sunmaktadır. Model, görüntülerdeki çok ölçekli özellik haritalarını etkili bir şekilde kullanmak için Özellik Piramit Ağı (FPN) ve Yol Toplama Ağı (PANet) gibi gelişmiş mimarileri kullanmaktadır. Bu sayede hem küçük hem de büyük nesnelerin tespitinde yüksek performans göstermektedir. YOLOv8, nesne konumlarını ve sınıf bilgilerini çıkarmak için girdiyi tek bir geçişte işleyen tek aşamalı bir algılayıcıdır. İki aşamalı algılayıcıların aksine, daha basit bir mimari sağlayan, hesaplama maliyetlerini azaltan ve önemli ölçüde daha hızlı çalışmayı mümkün kılan Anchor-free bir tasarım benimsemektedir. Ek avantajları arasında hızlı bir eğitim süreci, düşük gecikme süresi ve çeşitli veri kümelerinde güçlü genelleme yeteneği bulunmaktadır. Bununla birlikte, yoğun ve karmaşık sahnelerde veya sınıf dengesizliği olan veri kümelerinde performansı sınırlı kalabilmektedir. Örneğin, YOLOv8 trafik senaryolarında küçük ve tıkalı nesneleri tespit etmekte zorlanmakta, bu da yoğun nüfuslu sahnelerdeki doğruluğu etkilemektedir. Ayrıca, aydınlatma koşullarındaki değişiklikler de algılama performansını akış boyunca sürdürme becerisini zorlayabilmektedir [129]. Bu zorluklara rağmen, YOLOv8 hız açısından önemli avantajlar sunarak, özellikle hızın çok önemli olduğu senaryolarda gerçek zamanlı nesne algılama görevleri için oldukça verimli bir seçim olacaktır [130].

YOLOv8, farklı ölçekler ve uygulama gereksinimleri için optimize edilmiş çeşitli varyantlar sunmaktadır. Bu varyantlar beş kategoriye ayrılmıştır: Nano (n), Küçük

(s), Orta (m), Büyük (l) ve Ekstra Büyük (xl). Modeller arasındaki temel farklar kullanılan parametre sayısı ve mimarilerinin karmaşıklık düzeyidir. Nano model çok düşük parametre sayısı ile öne çıkmakta, hızlı çalışması ve düşük donanım gereksinimi nedeniyle kaynak kısıtlı ortamlar ve gerçek zamanlı uygulamalar için uygun bir seçenek sunmaktadır. Öte yandan, Ekstra Büyük varyant, yüksek parametre sayısı ve artan mimari karmaşıklığı ile daha zorlu görevlerde yüksek doğruluk sağlamakta fakat daha yüksek donanım kapasitesi gerektirmektedir. Bu bağlamda, YOLOv8 ailesinin Nano varyantı, kaynak verimliliği ve gerçek zamanlı performans avantajları nedeniyle tercih edilmiştir.

Bu çalışmada kullanılan bir diğer nesne algılama yöntemi de dönüştürücü mimarisine dayanan RT-DETR [128] modelidir. Bu model doğruluk ve hız arasında dengeli bir performans sağlayarak özellikle gerçek zamanlı uygulamalar için etkili olmaktadır. RT-DETR, dikkat mekanizmalarına dayalı bir mimari kullanmakta ve çok ölçekli özellikleri işlemek için Denoising Training yöntemini uygulamaktadır. Bu yaklaşım, modelin gürültülü veya yanlış etiketlenmiş veri setleri ve karmaşık sahneler içeren videolar karşısında dayanıklılığını artırmaktadır. RT-DETR mimarisi, farklı uygulama gereksinimlerine uygun çeşitli mimari boyutları sunmaktadır. Bu çalışmada, nesneler arasındaki bağlamsal farkındalığı daha yüksek düzeyde sağlayan RT-DETR-L modeli tercih edilmiştir.

Bu çalışmada kullanılan nesne tespit yöntemleri incelendiğinde, YOLOv8-n modelinin hızlı nesne tespiti gerçekleştirdiği, RT-DETR-L modelinin ise dikkat mekanizması ile sahne ve nesneler arasındaki bağlamsal farkındalığı artırdığı görülmektedir. Bu bağlamda, her iki modelin çıktılarının birleştirilmesi, güçlü yönlerinin birbirlerinin sınırlamalarını telafi etmesine ve böylece genel algılama performansının optimize edilmesine olanak sağlamaktadır.

4.1.1.2 Sperm Takip Modülü

Nesne takibi, özellikle biyomedikal görüntüleme alanında, hareketli nesnelerin zaman içindeki uzamsal konumlarının belirlenmesini ve yörüngelerinin sürekli olarak izlenmesini içerdiğinden kritik öneme sahiptir. Simple Online and Realtime Tracking with a Deep Association Metric (DeepSORT) [131], nesne izleme algoritmaları arasında yaygın olarak benimsenmiş ve oldukça verimli bir yaklaşımdır. Bu algoritma, gerçek zamanlı izleme yeteneklerini derin öğrenme tabanlı özellik çıkarma ile birleştirerek nesne izleme performansını optimize etmeyi amaçlamaktadır.

DeepSORT, SORT (Simple Online and Realtime Tracking) algoritmasının

geliştirilmiş bir versiyonudur. SORT algoritması temel olarak Kalman filtresi [132] ve Macar algoritması [133] kullanarak nesne izleme görevini yerine getirmektedir. Bununla birlikte, SORT algoritması izleme için yalnızca nesne algılama yöntemlerinden elde edilen geometrik özelliklere (örneğin, sınırlayıcı kutu koordinatları) dayanmaktadır. Bu yaklaşım, benzer görünümlere sahip veya üst üste binen nesnelerle çalışırken eşleştirme hatalarına yol açabilmektedir. Bu sınırlamanın üstesinden gelmek için DeepSORT yöntemi, görsel benzerliklere dayalı bir derin öğrenme yaklaşımının SORT algoritmasına entegre edilmesiyle geliştirilmiştir [131].

DeepSORT algoritmasının çalışma prensibi tespit ve takip ilişkilendirme mekanizmasına dayanmaktadır. Her karede, tespit edilen nesneler bir önceki karedeki izleme bilgileriyle eşleştirilerek nesneler takip edilmektedir. Bu eşleştirme işlemi iki ana faktöre dayanmaktadır: görsel özellikler ve hareket tahmini. Görsel özellik çıkarımında, nesnelerin gömülü özelliklerini çıkarmak için derin öğrenme tabanlı yöntemler kullanılmaktadır. Bu özellikler nesnelerin benzersiz kimliklerini temsil etmekte ve nesnelerin görsel benzerliğine dayalı olarak kareden kareye doğru eşleştirme yapılmasını mümkün kılmaktadır. Öte yandan hareket tahmini, nesnelerin uzamsal konumlarını tahmin etmek için Kalman filtresi [132] kullanılarak gerçekleştirilmektedir. Eşleştirme stratejisinin bir parçası olarak, algılama ve izleme bilgilerini ilişkilendirmek için bir maliyet matrisi oluşturulmaktadır. Bu matris, nesnelerin görsel benzerliklerine ve geometrik uzaklıklarına göre hesaplanmaktadır. Macar algoritması [133] bu matristeki en uygun eşleşmeleri belirlemekte ve nesnelerin takibi gerçekleştirilmektedir. Buna ek olarak, DeepSORT'un görsel özellik tabanlı yaklaşımı, nesnelerin geçici olarak kaybolması veya çerçevede yeniden görünmesi durumunda kimlik bilgilerinin (ID'ler) korunmasına olanak tanımaktadır. Bu mekanizma, izleme doğruluğunu artırarak uzun süreli nesne takibini daha güvenilir hale getirmektedir.

DeepSORT, görünüm tabanlı eşleştirme mekanizması sayesinde, sperm video analizinde yaygın olarak karşılaşılan karmaşık hareket modellerini ve görsel olarak benzer sperm hücrelerini izlemede olağanüstü performans göstermektedir. Algoritmanın gerçek zamanlı çalışma kabiliyeti, mikroskobik videolarda hızlı hareket eden sperm hücrelerinin etkili bir şekilde izlenmesini mümkün kılmaktadır. Derin öğrenme tabanlı görsel özellik çıkarımı ve hareket tahmini gibi özellikleriyle DeepSORT, biyomedikal görüntü işlemede yüksek doğruluk ve güvenilirlik sağlayarak sperm hücrelerinin takibine önemli bir katkıda bulunmaktadır. Bu yetenekler, algoritmayı bu alanda gelecekteki araştırmalar için referans olabilecek sağlam bir yöntem haline getirmektedir.

4.1.1.3 Sperm Yörünge Analiz Modülü

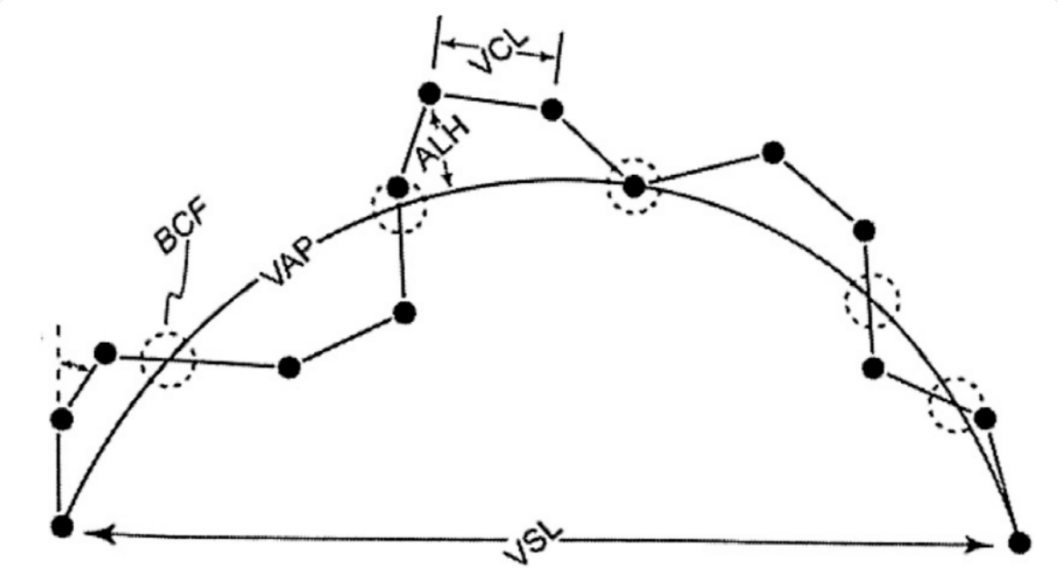
Sperm tespit ve takip modülleri aracılığıyla videolardaki her bir sperme ait yörüngeler elde edilmektedir. Bu yörüngeler üzerinde hareket analizi gerçekleştirebilmek için, DSÖ tarafından belirlenen bazı özelliklerin çıkarılması gerekmektedir. DSÖ tarafından sperm yörüngelerinin değerlendirilmesi amacıyla tanımlanan parametreler Şekil 4.2'de sunulmaktadır. Bu şekil incelendiğinde, VAP (Average Path Velocity), VCL (Curvilinear Velocity) ve VSL (Straight-Line Velocity) parametrelerinin temel bileşenleri oluşturduğu görülmektedir. Diğer hareket parametreleri ise bu üç ana parametre kullanılarak hesaplanmaktadır. Yörünge analizi kapsamında sperm hareketlilik türünün belirlenmesinde kullanılan tanımlanmış hareketlilik özellikleri aşağıda sıralanmaktadır:

- $Curve\ Linear\ Velocity\ (VCL)$ = Spermin katetmiş olduğu toplam yol kullanılarak elde edilen ortalama hız parametresidir (μm/s).
- $Average\ Path\ Velocity\ (VAP)$ = Spermin izlediği kavisli yolun yumuşatılması sonucu elde edilen yol kullanılarak hesaplanan ortalama hız parametresidir (μm/s).
- $Straight\ Line\ Velocity\ (VSL)$ = Spermin başlangıç ve bitiş noktaları arasında kalan doğrusal yol kullanılarak hesaplanan hız parametresidir (μm/s).
- $Beat\ Cross\ Frequency\ (BCF)$ = Sperm başının, VAP değerini hesaplamak için kullanılan yumuşatılmış yol ile arasındaki kesişim frekansını belirtmektedir. Kaç adet kesişim yapıldığı ile ölçülmektedir.
- $Straightness\ (STR) = \frac{VSL}{VAP}$ (%) ile hesaplanmaktadır. Yumuşatılmış yol kullanılarak spermin doğrusallığı ölçülmektedir.
- $Linearity\ (LIN) = \frac{VSL}{VCL}$ (%) ile hesaplanmaktadır. Kavisli yol kullanılarak spermin doğrusallığı ölçülmektedir.
- $Wobble\ (WOB) = \frac{VAP}{VCL}$ (%) ile hesaplanmaktadır. Spermin salınımını ölçmek için kullanılmaktadır.

VCL, spermin izlediği gerçek yola ait ortalama hızı belirlemekte olup, doğrusal olmayan hareketleri de kapsayarak spermin tüm hareketlerini içermektedir. Bu parametre, spermin hareket türünü belirleme açısından önemli bir ölçüt olarak kabul edilmektedir. VSL, spermin düz bir çizgi boyunca hareket etme hızını ölçmekte olup, hareket karakteristiğinin değerlendirilmesinde kritik bir rol oynamaktadır. Bu

parametre aracılığıyla spermin doğrudan ilerleme yeteneği analiz edilmektedir. Bir diğer önemli parametre olan VAP, spermin kavisli yol boyunca ortalama hızını hesaplamaktadır. Bu parametreler kullanılarak, spermin hareket türünü ölçmek amacıyla LIN ve STR parametreleri elde edilmektedir. LIN ve STR, spermin ne kadar düzgün ve verimli hareket ettiğini belirlemek için kullanılmaktadır. Ayrıca, spermin hareket yönündeki ani veya küçük değişiklikler Wobble olarak tanımlanmaktadır. Bu değer, VCL ve VAP oranı ile hesaplanmakta olup, 1'e ne kadar yakınsa bu iki değerin birbirine o kadar uyumlu olduğunu göstermektedir

Tüm video boyunca her bir spermin yörüngesi çıkarılarak ilgili hareketlilik parametreleri elde edilmektedir. Elde edilen parametreler, belirli referans değerler doğrultusunda sperm hareketlilik türlerini belirlemek amacıyla sınıflandırılmaktadır. Bu sınıflandırma, Witkowski ve arkadaşları tarafından belirlenen referans değerlerine dayanmakta olup, ilgili değerler Tablo 4.1'de sunulmaktadır [134].



Şekil 4.2 DSÖ tarafından tanımlanmış sperm hareketlilik değerlendirme parametreleri [3]

Tablo 4.1 Witkowski sperm hareketlilik parametreleri referans değerleri [134]

	VCL ($\mu m/s$)	VSL ($\mu m/s$)	VAP ($\mu m/s$)	STR (%)	LIN (%)
Referans Grade A			>25		>75
Referans Grade B			<25		>75
Referans Grade C					<75
Referans Grade D			<5		

4.1.1.4 Performans Metrikleri

1. **Tespit Performans Metrikleri**

Sperm tespit ve takip problemlerinde modellerin performansı çeşitli metrikler ve yaklaşımlar kullanılarak değerlendirilmektedir. Tespit görevlerinde modellerin doğruluğunu ve etkinliğini ölçmek için Kesinlik (P), Duyarlılık (R) ve bu iki metriğin harmonik ortalamasını temsil eden F1-Skoru gibi metrikler kullanılmaktadır. Ayrıca, modelin genel tespit performansının daha kapsamlı bir değerlendirmesini sağlayan Ortalama Hassasiyet (mAP) metriği de bu çalışmaya dahil edilmiştir. Kesinlik, modelin doğru tespitleri yapma oranını ifade etmekte ve Eşitlik 4.1 kullanılarak hesaplanmaktadır. Duyarlılık, doğru tespit edilen pozitif örneklerin toplam gerçek pozitif örnek sayısına oranını temsil etmekte ve modelin gerçek pozitifleri ne kadar iyi yakaladığını göstermektedir. Duyarlılık, Eşitlik 4.2'de tanımlandığı şekilde hesaplanır. Bu iki metriği dengeleyerek modelin genel performansını değerlendiren F1-Skoru, Eşitlik 4.3'te gösterildiği gibi her iki metriğin harmonik ortalaması alınarak hesaplanmaktadır.

$$P = \frac{TP}{TP + FP} \tag{4.1}$$

$$R = \frac{TP}{TP + FN} \tag{4.2}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{4.3}$$

$$mAP_{0.5} = \frac{1}{n} \sum_{i=1}^{n} AP_i^{IoU=0.5} \tag{4.4}$$

$$mAP_{0.5:0.95} = \frac{1}{n \times k} \sum_{i=1}^{n} \sum_{j=1}^{k} AP_i^{IoU=0.5+0.05(j-1)} \tag{4.5}$$

mAP, bir modelin genel tespit performansını değerlendirmek için kullanılan bir metriktir. Bu metrik, farklı Intersection over Union (IoU) eşik seviyelerinde hesaplanan Ortalama Hassasiyet (AP) değerlerinin ortalaması alınarak elde edilmektedir. Bu çalışmada, literatürde yaygın olarak kullanılan mAP50 ve mAP50:95 metrikleri kullanılmıştır. Eşitlik 4.4'te mAP50, IoU değeri 0,5 veya daha yüksek olan tespitleri doğru olarak tanımlamakta ve bu eşiğin altındaki tahminleri yanlış olarak sınıflandırmaktadır. Eşitlik 4.5'te verilen mAP50:95 metriği, IoU eşiğinin 0,5 ile 0,95 arasında 0,05'lik adımlarla değiştirilmesiyle elde edilen AP değerlerinin ortalamasını

hesaplamaktadır. Eşitlik 4.4 ve 4.5'te n sınıf sayısını, AP_i ise i'inci sınıf için ortalama hassasiyeti göstermektedir. Eşitlik 4.5'teki k, IoU eşik değerlerinin sayısını ifade etmektedir.

Senaryo-I'de, iki YOLOv8n modeli, cep telefonu ve mikroskop kamerasına dayalı videolar kullanılarak bağımsız olarak eğitilmiş ve tespit performansı metrikleri elde edilmiştir. Eğitilen bu modellerin tespit performansları arasında istatistiksel olarak anlamlı bir fark olup olmadığını değerlendirmek için bir ANOVA (Varyans Analizi) analizi yapılmıştır [135]. ANOVA'nın ardından, her iki modelle elde edilen 30 farklı videodan oluşan test setleri için P, R, F1, mAP50 ve mAP50:95 metriklerinde ikili karşılaştırmalar yapılmıştır. Bu karşılaştırmalar Tukey-HSD testi [136] kullanılarak yapılan bir post-hoc analiziyle desteklenmiştir. Elde edilen sonuçlar Deneysel Sonuçlar bölümünde ayrıntılı olarak sunulmuştur.

2. **Takip Performans Metrikleri**

   Nesne izleme senaryolarında, modellerin izleme performansını değerlendirmek için tespit metriklerine ek olarak Multiple Object Tracking Accuracy (MOTA), Higher Order Tracking Accuracy (HOTA) ve Identity F1-Score (IDF1) metrikleri kullanılmıştır. MOTA, MOT (Multiple Object Tracking) görevlerinde algılama ve izleme algoritmalarının performansını ölçmek için yaygın olarak kullanılan bir metriktir. Bu metrik, kaçırılan tespitler, yanlış pozitifler ve nesne takibi sırasında meydana gelen kimlik değişimleri gibi faktörleri göz önünde bulundurarak performansı değerlendirmektedir. MOTA metriği Eşitlik 4.6 kullanılarak hesaplanmaktadır. Bu eşitlikte, Yanlış Pozitif (FP) ve Yanlış Negatif (FN) değerleri, 0,5'lik bir IoU eşiği kullanılarak hesaplanmaktadır. Eşitlikte t çerçeve indekslerini, FN_t kaçırılan tespit sayısını, FP_t yanlış tespit sayısını, $\text{identity switches (IDSW)}_t$ kimlik değiştirme sayısını ve GT_t ise t-inci çerçevedeki temel gerçek nesne sayısını ifade etmektedir. Bu değerler her çerçeve için ayrı ayrı hesaplanarak tüm çerçeveler için toplam hata değeri elde edilmektedir. Hesaplanan toplam hata değeri, toplam doğruluk nesnesi sayısına bölünerek normalleştirilmekte ve bu değer 1'den çıkarılarak nihai MOTA metriği elde edilmektedir. MOTA'ya ek olarak, Multiple Object Detection Accuracy (MODA) metriği, kaçırılan tespitlere ve yanlış pozitiflere odaklanarak nesne tespitinin doğruluğunu ölçerken, Multiple Object Tracking Precision (MOTP), tespit edilen ve gerçek referans değer (ground truth) nesneleri arasındaki ortalama mesafeyi dikkate alarak tespit

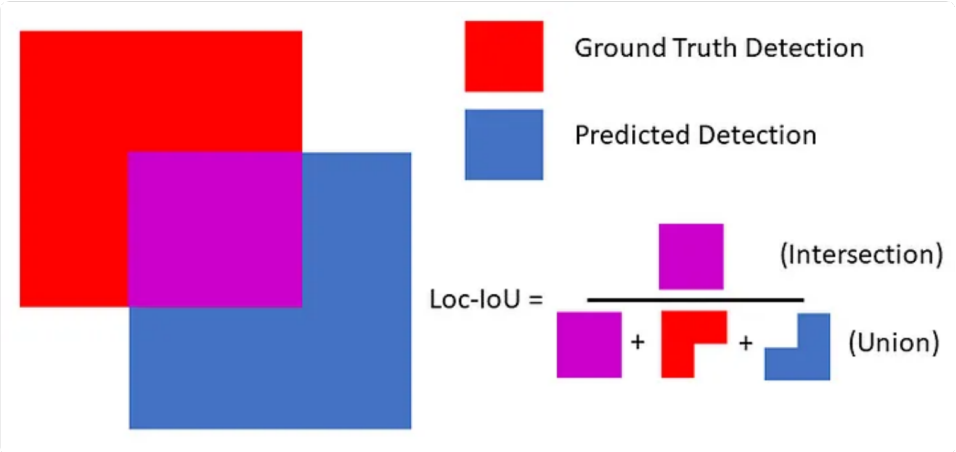
edilen nesnelerin yerelleştirme hassasiyetini değerlendirmektedir.

$$\text{MOTA} = 1 - \frac{\sum_t (\text{FN}_t + \text{FP}_t + \text{IDSW}_t)}{\sum_t \text{GT}_t} \tag{4.6}$$

Bir başka MOT ölçütü ise izleme algoritmalarının hem doğru eşleştirme hem de kimlik tutarlılığını değerlendirmek için kullanılan IDF1'dir. IDF1, yanlış tespitler ve gözden kaçan nesneler gibi faktörlerden etkilenen takip edilen nesnelerin kimlik doğruluğunu değerlendirmektedir. Bu metrik, özellikle kimlik değişimlerine karşı duyarlıdır ve bu tür hataların algoritmanın performansı üzerindeki etkisini vurgulamaktadır.

$$\text{IDF1} = 2 \times \frac{\text{IDP} \times \text{IDR}}{\text{IDP} + \text{IDR}} \tag{4.7}$$

Eşitlik 4.7'de tanımlandığı gibi, IDF1 metriği Identity Precision (IDP) ve Identity Recall (IDR) değerlerinin harmonik ortalaması alınarak hesaplanmaktadır. Identity Precision (IDP) doğru eşleşmelerin takip edilen toplam nesne sayısına oranını temsil ederken, Identity Recall (IDR) doğru eşleşmelerin doğru eşleşmesi gereken toplam nesne sayısına oranını temsil etmektedir. IDF1 metriği yalnızca nesne tespitinin başarısını değerlendirmekle kalmaz, aynı zamanda bu nesnelerin kimliklerini doğru bir şekilde izleme performansını da değerlendirir. Bu özellikleri nedeniyle izleme algoritmalarının değerlendirilmesinde önemli bir metrik olarak öne çıkmaktadır. MOTA ve IDF1, MOT çalışmalarında yaygın olarak kullanılan



Şekil 4.3 Localization Intersection over Union (Loc-IoU) metrik hesaplamasının görselleştirilmesi [137]

ölçütler olsa da bazı sınırlamaları bulunmaktadır. MOTA öncelikle nesne

tespitinin doğruluğuna dayalı bir değerlendirme sağlarken, IDF1 nesne kimliklerinin doğru ilişkilendirilmesine odaklanarak daha ayrıntılı bilgi sunmaktadır. Bu iki metriği daha dengeli bir şekilde temsil etmek için J. Luiten ve arkadaşları HOTA metriğini geliştirmişlerdir [137]. HOTA, üç farklı IoU skorunun bir kombinasyonu olarak hesaplanmakta ve izleme performansını üç alt görevde değerlendiren kapsamlı bir yaklaşım sunmaktadır: nesne algılama, kimlik ilişkilendirme ve lokalizasyon. Bu görevlerin her biri için IoU değerleri hesaplanmakta ve nihai HOTA değerini elde etmek için birleştirilmektedir. Böylece HOTA metriği, izleme algoritmalarının daha dengeli ve ayrıntılı bir şekilde değerlendirilmesi için bir fırsat sunmaktadır.

Lokalizasyon adımında, tahmin edilen bir tespit ile gerçek referans değer(grount truth) tespit arasındaki uzamsal uyum değerlendirilmektedir. Şekil 4.3, Loc-IoU değerini hesaplama yöntemini görselleştirmektedir. Eşitlik 4.8'de belirtildiği gibi, LocA metriği, veri kümesindeki tüm eşleşen gerçek referans değer ve tahmin edilen tespit çiftlerinin Loc-IoU değerlerinin ortalaması alınarak hesaplanmaktadır. Bu denklemde, $|TP|$ toplam doğru pozitif eşleşme sayısını temsil ederken, $c \in TP$ doğru eşleşen her bir nesneyi temsil etmektedir. LocA, izleme algoritmalarının uzamsal doğruluğunu değerlendirmek için çok önemli bir ölçüt olarak kullanılmaktadır.

Tespit adımında, tüm gerçek referans değer tespitler kümesi ile tüm tahmin edilen tespitler kümesi arasındaki hizalama ölçülür. Eşitlik 4.9'da tanımlandığı gibi Detection IoU (Det-IoU), detection accuracy (DetA)'yı değerlendirmek için de kullanılmaktadır. DetA hesaplanırken, gerçek referans değer ve tahmin edilen tespitler arasındaki kesişme oranını belirlemek için belirli bir Loc-IoU eşik değeri kullanılmaktadır. Ancak bu yaklaşımda, bir tahmin birden fazla gerçek referans değerle kesişebilmekte veya bunun tam tersi gerçekleşerek potansiyel belirsizliklere yol açabilmektedir. Bu tür durumların üstesinden gelmek için, iki set arasında bire bir eşleşme sağlamak için Macar Algoritması uygulanmaktadır.

Elde edilen eşleşmeler Eşitlik 4.9'da TP (Doğru Pozitifler) sembolü ile temsil edilmektedir. Eşleşmeyen tahminler $|FP|$ (Yanlış Pozitifler) olarak tanımlanırken, eşleşmeyen gerçek referans değerler $|FN|$ (Yanlış Negatifler) olarak adlandırılmaktadır. Veri setindeki Doğru Pozitiflerin (TPs), Yanlış Pozitiflerin (FPs) ve Yanlış Negatiflerin (FNs) toplam sayısına göre hesaplanan Det-IoU, genel DetA değerini vermektedir. Bu sayede algoritmanın tespit adımındaki performansı kapsamlı bir şekilde değerlendirilmiş olmaktadır.

Son olarak, ilişkilendirme adımında, sistemin zaman içinde aynı tespitlere doğru kimlikler atama yeteneği, veri setinde sunulan gerçek kimlik referansları (ground-truth ID) ve gerçek yörünge referansları (ground-truth track) kullanılarak değerlendirilmektedir. Daha önce belirtildiği gibi, bu ölçüm Macar algoritması ile tahmin edilen yörüngeler ile gerçek referans yörüngelerinin eşleştirilmesi yoluyla gerçekleştirilmekte ve tahmin edilen tespitlerin tüm yörüngeleri ile gerçek referans tespitlerin yörüngeleri arasındaki hizalama hesaplanmaktadır. Eşitlik 4.10'de tanımlandığı üzere, True Positive Associations (TPA), tahmin edilen yörüngeler (predicted tracks) ile gerçek yörünge referansları (ground-truth tracks) arasındaki doğru eşleşmeleri ifade etmektedir. Öte yandan, tahmin edilen yörüngelerin başka bir gerçek yörünge referansı ile hatalı eşleşmesi veya herhangi bir yörünge ile eşleşmemesi durumu False Positive Associations (FPA) olarak adlandırılmaktadır. Benzer şekilde, herhangi bir doğru tahminle eşleşmeyen tüm gerçek yörüngeler ve tespitler False Negative Associations (FNA) olarak tanımlanmaktadır. TPA, FNA ve FPA kavramlarının görsel temsilleri Şekil 4.4'de sunulmaktadır. Şekilde sol alt köşede yer alan kalın kesik çizgili büyük mavi ve açık mavi daireler, gerçek referans tespitlerini ve yörüngelerini göstermektedir. İnce kesik çizgilere sahip küçük siyah ve gri daireler ise tahmin edilen tespitleri ve yörüngeleri temsil etmektedir. Her farklı renkteki daire, ayrı bir nesneyi simgelemektedir. Kırmızı kare, eşleşmiş tahmini ve gerçek referans tespit çiftini gösterirken, iki tespit arasındaki ilişkinin niteliğini değerlendirmek amacıyla yörüngeler boyunca kesişen tespitler (yeşil kareler) TPA'yı, eşleşmeyen tespitler ise sırasıyla FPA (sarı kareler) ve FNA (kahverengi kareler) olarak gösterilmektedir. Eşitlik 4.11'de belirtildiği üzere, ortalama Ass-IoU değeri, veri seti genelinde gerçek referans ve tahmin edilen tespitler arasındaki eşleşmeler kullanılarak hesaplanmakta ve bu değer, Association Accuracy (AssA) metriğinin belirlenmesinde kullanılmaktadır.

$$\text{LocA} = \frac{1}{|TP|} \sum_{c \in TP} \text{Loc-IoU}(c) \tag{4.8}$$

$$\text{Det-A} = \text{Det-IoU} = \frac{|TP|}{|TP| + |FP| + |FN|} \tag{4.9}$$

$$\text{Ass-IoU} = \frac{|TPA|}{|TPA| + |FPA| + |FNA|} \tag{4.10}$$

$$\text{AssA} = \frac{1}{|TP|} \sum_{c \in TP} \text{Ass-IoU}(c) \tag{4.11}$$

$$= \frac{1}{|TP|} \sum_{c \in TP} \frac{|TPA(c)|}{|TPA(c)| + |FNA(c)| + |FPA(c)|} \tag{4.12}$$

$$\text{HOTA}_\alpha = \sqrt{\text{DetA}_\alpha \cdot \text{AssA}_\alpha} \tag{4.13}$$

$$= \sqrt{\frac{\sum_{c \in \text{TP}_\alpha} \text{Ass-IoU}_\alpha(c)}{|\text{TP}_\alpha| + |\text{FN}_\alpha| + |\text{FP}_\alpha|}} \tag{4.14}$$

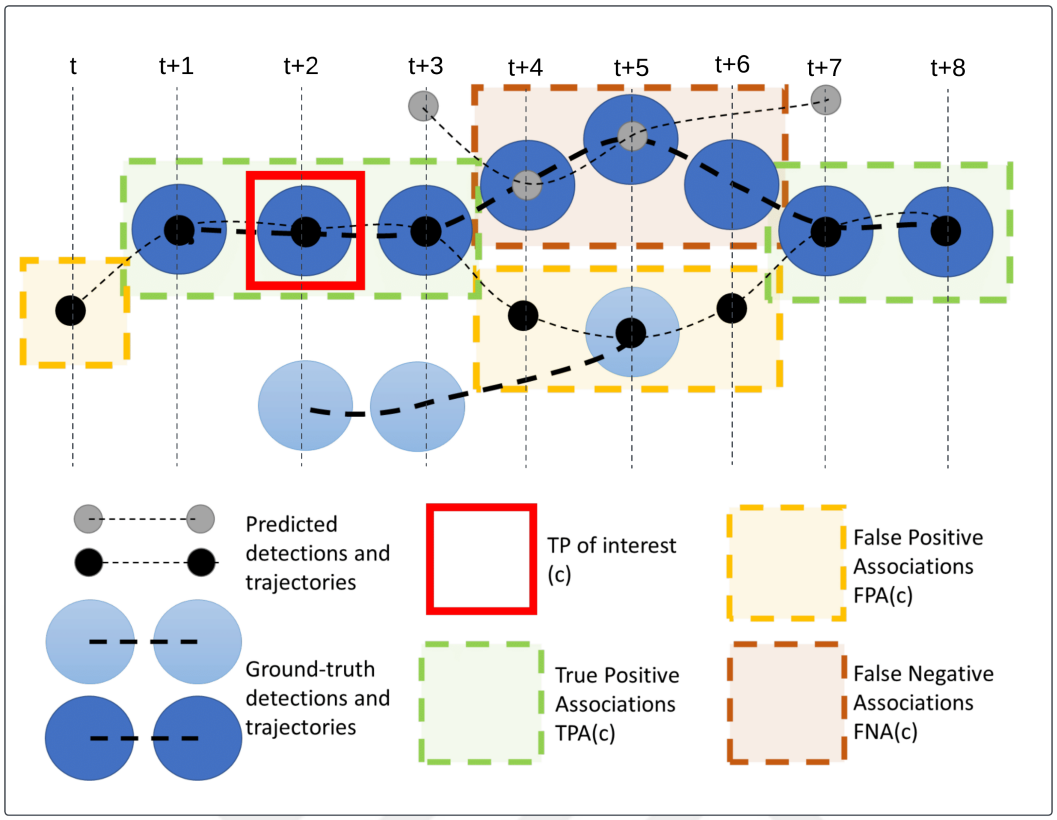
$$\text{HOTA} = \int_{0<\alpha\leq 1} \text{HOTA}_\alpha \tag{4.15}$$

$$\approx \frac{1}{19} \sum_{\alpha=0.05}^{0.95} \text{HOTA}_\alpha \tag{4.16}$$

Yukarıda bahsedilen lokalizasyon, tespit ve ilişkilendirme metrikleri MOT problemi bağlamında oldukça önemlidir. Bu üç metriğin birleşimi, bir sistemin performansının kapsamlı bir değerlendirmesini sağlayan HOTA'yı oluşturmakta ve sistemin performansını üç farklı açıdan tek bir metrik değeriyle ölçmek için yeterli bir araç sunmaktadır. Eşitlik 4.13'te gösterildiği gibi HOTA, DetA ve AssA değerlerinin geometrik ortalaması alınarak hesaplanmakta ve her iki metriğin eşit olarak ağırlıklandırılması sağlanmaktadır. DetA ve AssA'nın hesaplanması sırasında, atama işlemi için Macar algoritması kullanılmış ve belirli bir Loc-IoU eşiği (α) uygulanmıştır. Ancak, HOTA metriği hesaplanırken bu (α) değeri 0,05 ila 0,95 arasında değiştirilerek, belirli bir eşikle sınırlandırılmadan daha genel bir performans değerlendirmesi yapılmasına olanak tanınmıştır. Bu yaklaşım, LocA'yı nihai değerlendirmenin bir parçası haline getirmekte ve HOTA skorunun, tespit, ilişkilendirme ve tespitlerin uzamsal doğruluğu açısından modelin performansını kapsamlı bir şekilde değerlendirmesini sağlamaktadır.

4.1.1.5 Deneysel Senaryolar

Literatürdeki çalışmalar incelendiğinde, farklı senaryolarda sperm tespiti ve takibi üzerine çeşitli çalışmalar olduğu görülmektedir. Sperm tespitine odaklanan çalışmalarda [39, 62, 66, 67], veri kümelerinin geleneksel veri bölütleme yöntemleri kullanılarak hazırlandığı, tüm görüntülerin karıştırıldığı, belirli bir yüzdenin test kümesi olarak ayrıldığı, kalan kısmın ise eğitim ve doğrulama kümeleri için kullanıldığı görülmektedir. Öte yandan, sperm takibine odaklanan çalışmalarda [41,



Şekil 4.4 AssA metriğinin hesaplanmasında kullanılan kavramların görsel gösterimi [137]

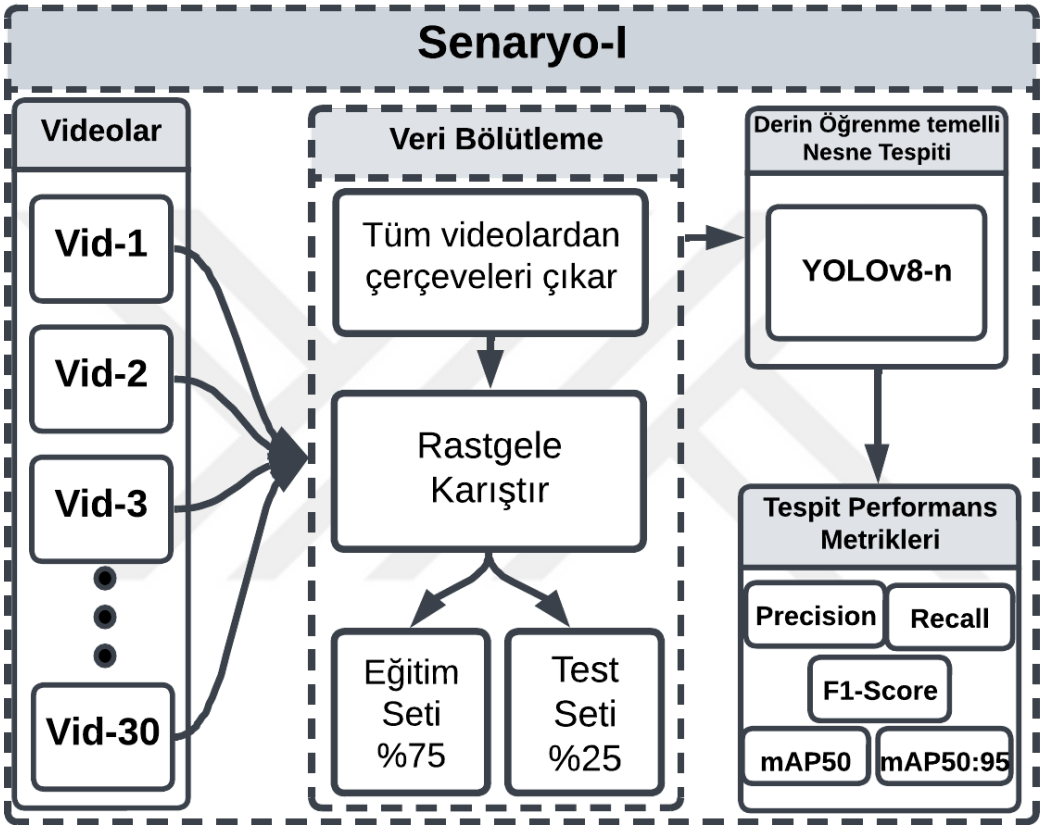
67], yukarıda bahsedilen yaklaşıma ek olarak, gerçek dünya kullanım senaryoları doğrultusunda eğitim ve test setlerinin tamamen farklı videolardan elde edildiği yöntemlerin de uygulandığı görülmektedir.

Bu çalışmada, cep telefonu tabanlı veri toplama yöntemiyle elde edilen videolar kullanılarak deneyler gerçekleştirilmiştir. Videoların etkinliğini değerlendirmek amacıyla, YOLOv8n modeli hem mikroskop kamerasıyla hem de cep telefonu kamerasıyla kaydedilen videolar üzerinde ayrı ayrı eğitilmiş ve elde edilen sonuçlar istatistiksel olarak karşılaştırılmıştır. Analizler, iki veri türü arasında anlamlı bir fark olmadığını göstermiştir. Bu nedenle, sonraki deneylerde cep telefonu videoları kullanılmıştır. Elde edilen veriler doğrultusunda, oluşturulan veri kümesi üç farklı senaryo altında çeşitli performans metrikleriyle değerlendirilmiştir.

1. **Senaryo-I**

   İlk senaryoda, literatürdeki birçok çalışmada olduğu gibi [39, 62, 66, 67], videolardan çıkarılan çerçeveler karıştırılmış ve veri kümesi %75 eğitim ve %25 test olarak ayrılmıştır. Bu senaryoda, veri kümesi üzerinde sperm tespiti gerçekleştirilmiş ve modellerin P, R, F1, mAP50 ve mAP50:95 değerleri sunulmuştur. Bu senaryoda eğitim ve test kümeleri rastgele

oluşturulmuştur. Test setindeki kareler rastgele seçildiği için bir video akışını temsil etmemektedir. Ancak sperm takibi yapabilmek ve HOTA, MOTA gibi metrikleri hesaplayabilmek için ardışık karelerdeki spermleri tespit etmek, bu spermlere ID atamak ve yörüngelerini belirlemek gerekmektedir. Bu gereksinimler doğrultusunda ilk senaryoya alternatif olarak ikinci bir senaryo geliştirilmiştir. Senaryo-I'e ait akış şeması Şekil 4.5'te sunulmuştur.

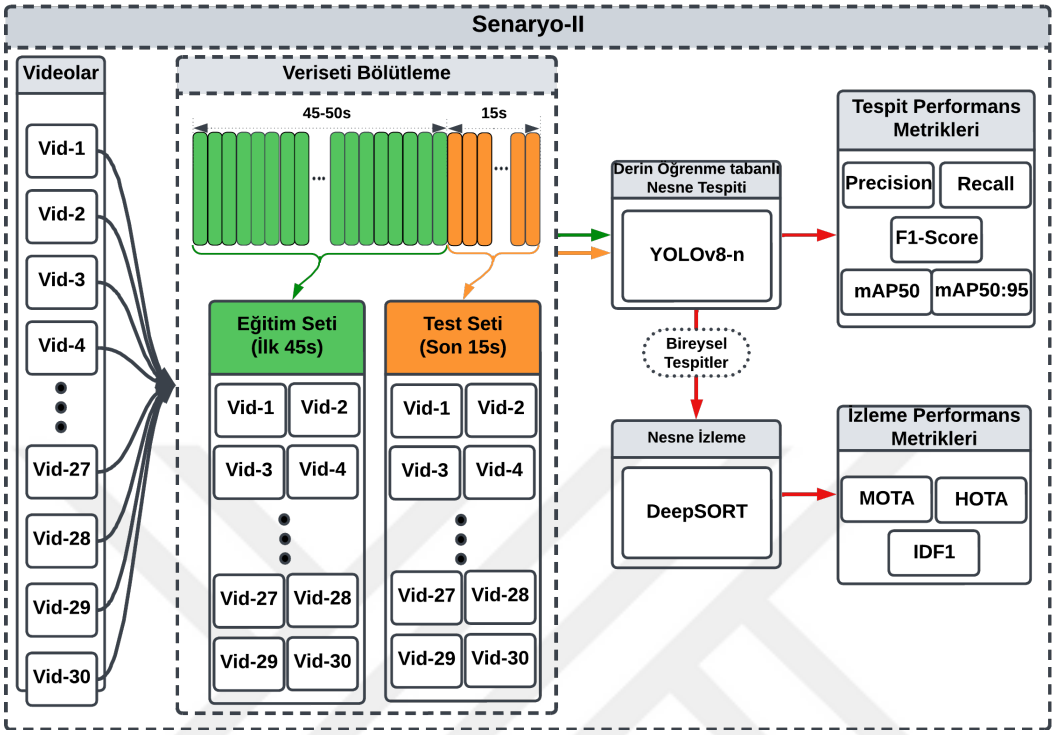


Şekil 4.5 Senaryo-I akış diyagramı

2. **Senaryo-II**

   İkinci senaryoda, videoların ilk 45 saniyesi eğitim seti, kalan 15 saniyesi ise test seti olarak kullanılmıştır. Bu yöntem her video için uygulanmış olup, tüm videoların yaklaşık %75'i eğitim, %25'i ise test kümesi olarak ayrılmıştır. Bu senaryo kapsamında eğitilen YOLOv8n modeli kullanılarak, her bir videonun son 15 saniyesi için P, R, F1, mAP50 ve mAP50:95 metrikleri hesaplanmış; ayrıca DeepSORT algoritmasıyla MOTA, HOTA ve IDSW gibi izleme metrikleri değerlendirilmiştir. İkinci senaryoda elde edilen sonuçlar, literatürde yaygın olarak kullanılan birinci senaryoya kıyasla daha güvenilir olsa da, gerçek dünyadaki kullanım senaryolarıyla tam olarak örtüşmemektedir. Örneğin, hastaneye ilk kez başvuran bir hastanın

videosunun analiz edilmesi bu senaryoda mümkün değildir. Senaryo-II için akış şeması Şekil 4.6'da sunulmuştur.



Şekil 4.6 Senaryo-II akış diyagramı

3. **Senaryo-III**

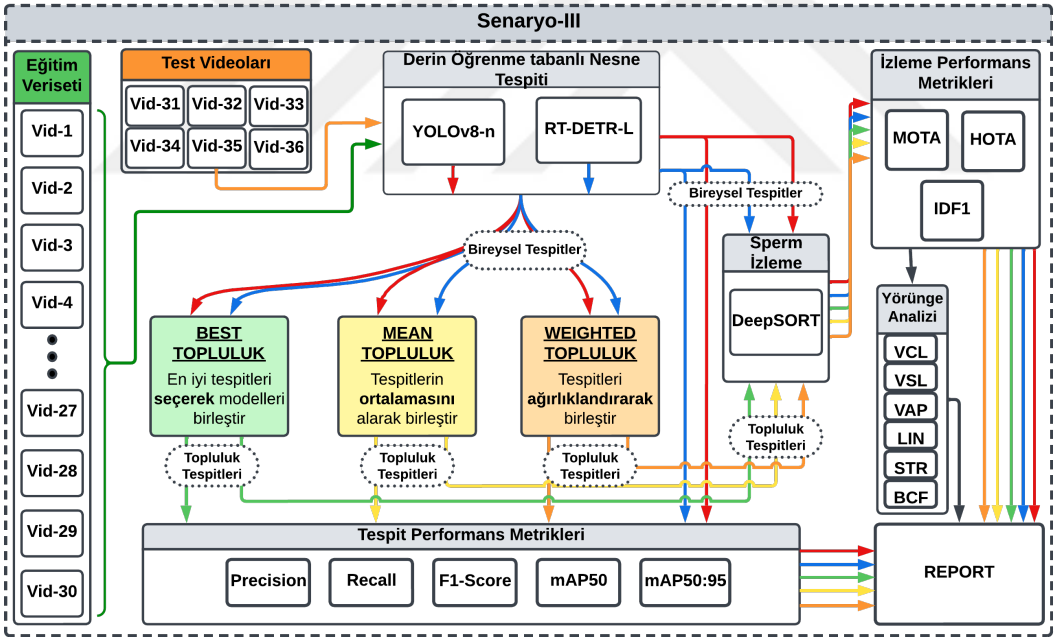
Yukarıda açıklanan senaryolar teorik olarak faydalı olsa da, gerçek hayatta pratik kullanımları açısından bazı sınırlamalar taşımaktadır. Bunun temel nedeni, eğitim ve test setlerinin aynı videonun farklı karelerinden oluşturulmasıdır. Ancak, gerçek dünya senaryolarında modelin eğitim seti, test aşamasında değerlendirilen hastalara ait herhangi bir kare içermemektedir. Bu nedenle, gerçek dünya uygulamalarıyla daha uyumlu bir yaklaşımı temsil eden üçüncü senaryo geliştirilmiştir.

Üçüncü senaryoda, önceki iki senaryoda kullanılan 30 video eğitim setine dahil edilirken, test aşaması için ek olarak 6 yeni video veri setine eklenmiştir. Bu genişletilmiş veri kümesi, hem YOLOv8n modeli hem de transformatör tabanlı bir mimari olan RT-DETR-L modelinin değerlendirilmesi için kullanılmıştır. Elde edilen sonuçlar incelendiğinde, YOLOv8n modelinin RT-DETR-L modeline kıyasla daha hızlı çalıştığı gözlemlenmiştir. Özellikle, çok sayıda sperm içeren videolarda, YOLOv8n modeli RT-DETR-L'ye kıyasla daha fazla tahmin üretmektedir. Ancak, kesinlik (precision) değeri açısından değerlendirildiğinde, YOLOv8n modelinin RT-DETR-L'ye kıyasla

daha düşük performans gösterdiği belirlenmiştir. Bunun temel sebebi, YOLOv8n modelinin sperm içermeyen nesneleri yanlışlıkla sperm olarak sınıflandırmaya eğilimli olması ve bunun sonucunda yanlış pozitiflerin (false positive) artmasıdır.

Öte yandan, RT-DETR-L modeli daha yavaş çalışmasına ve daha az tahmin üretmesine rağmen, daha doğru ve güvenilir sonuçlar vermektedir. Literatür incelendiğinde, CNN ve transformatör tabanlı modellerin özellikle karmaşık sahnelerde tespit doğruluğunu artırmak amacıyla sıklıkla birleştirildiği görülmektedir [138, 139]. Bu çalışmada, YOLOv8n ve RT-DETR-L modellerinin sonuçları üç farklı topluluk yöntemi kullanılarak birleştirilmiş ve sperm takibi için DeepSORT algoritması uygulanmıştır. Bu yöntemle, her iki modelin güçlü yönleri bir araya getirilerek daha doğru ve güvenilir tespitlerin elde edilmesi hedeflenmiştir.

Senaryonun sonunda, modellerin performans metrikleri kapsamlı bir şekilde analiz edilmiştir. Senaryo-III'e ait akış şeması Şekil 4.7'de gösterilmektedir.



Şekil 4.7 Senaryo-III akış diyagramı

4.1.2 Deneysel Sonuçlar

Bu bölümde, modeller Hi-LabSpermTracking veri kümesi kullanılarak belirtilen üç farklı senaryoda eğitilmiş ve kapsamlı bir şekilde analiz edilmiştir. Önerilen cep telefonu tabanlı görüntü elde etme yönteminin etkinliğini göstermek amacıyla, senaryolar öncesinde istatistiksel testler gerçekleştirilmiştir. Bu bağlamda, cep

telefonu videoları ve mikroskop kamera videoları, iki farklı YOLOv8n modeli ile ayrı ayrı eğitilmiş ve her iki modelin test verileri üzerinde P, R, F1, mAP50 ve mAP50:95 gibi performans metrikleri hesaplanmıştır. Elde edilen performans metrikleri kullanılarak, cep telefonu videoları ile mikroskop videoları arasında istatistiksel olarak anlamlı bir fark olup olmadığını tespit etmek amacıyla ANOVA testi ve sonrasında Tukey'in Dürüstçe Anlamlı Fark (HSD) post hoc analizi uygulanmıştır. Sonuçlar, Tablo 4.2'de özetlenmiştir.

Telefon Videoları ve Mikroskop Videolarının karşılaştırmalı analizi, Tablo 4.2'de sunulduğu gibi, çoğu metrikte (P, R, F1 ve mAP50 gibi) p-adj değerlerinin 0,05'ten büyük olduğu ve dolayısıyla önemli bir fark gözlenmediği sonucunu ortaya koymaktadır. Ancak, mAP50:95 için istatistiksel olarak anlamlı bir fark (p-adj = 0,007) tespit edilmiştir; cep telefonu ile kaydedilen videolar, mikroskop kamerası ile kaydedilen videolara (0,734) kıyasla daha yüksek bir ortalama değere (0,778) sahiptir. Pozitif ortalama fark (MeanDiff = 0,0441), cep telefonuyla kaydedilen videoların daha geniş bir tespit senaryosu yelpazesini yakalamada üstün performans gösterdiğini ortaya koymaktadır. İstatistiksel testlerde p-adj, düzeltilmiş p-değeri olup, 0,05'in altındaki değerler anlamlı bir farkı işaret etmektedir. "Reject", p-adj < 0,05 olduğunda boş hipotezin reddedildiğini belirtmektedir. Alt ve üst sınırlar ise güven aralığını ve gerçek ortalama farkın muhtemel olduğu aralığı (bu durumda 0,012 ila 0,076) göstermektedir. Bu bulgular, cep telefonuyla kaydedilen videoların etkinliğini doğrulamakta ve çalışma kapsamındaki çeşitli senaryolarda daha geniş uygulamalar için potansiyellerinin altını çizmektedir.

Tablo 4.2 Tukey'in Honestly Significant Difference (HSD) sonuçları: telefon ve mikroskop kamera tabanlı video performansının değerlendirilmesi

Metrikler	Telefon Videoları	Mikroskop Videoları	MeanDiff	p-adj	Lower	Upper	Reject
P	97,77	98,20	-0,0047	0,31	-0,014	0,004	False
R	98,14	97,90	0,0016	0,69	-0,0065	0,009	False
F1	97,90	98,00	-0,0016	0,71	-0,01	0,006	False
mAP50	98,90	98,90	0	0,99	-0,0045	0,004	False
50:95	77,80	73,40	0,0441	0,007	0,0122	0,075	True

4.1.2.1 Senaryo-I Sonuçları

Literatür incelendiğinde, sperm tespit problemini ele alan birçok çalışmanın veri kümesindeki tüm videolardan kareleri rastgele bir şekilde eğitim ve test kümelerine böldüğü ve önerdikleri sistemler için değerlendirme metriklerini raporladığı görülmektedir [39, 62, 65–67]. Bu bağlamda, Senaryo-I için veri kümesindeki 30 videonun tüm kareleri birleştirilmiş ve %75'i eğitim için, %25'i test için rastgele

bir şekilde ayrılmıştır. Senaryo-I için test sonuçları, Tablo 4.2'nin Telefon Videoları sütununda verilmiştir. Elde edilen sonuçlar incelendiğinde, modelin Senaryo-I'de %97,77 Kesinlik, %98,14 Duyarlılık, %97,9 F1-Skor, %98,9 mAP50 ve %77,80 mAP50:95 gibi yüksek değerler elde ettiği gözlemlenmektedir. Bu sonuçlar, Tablo 1.3'de sunulan sperm tespit çalışmalarına kıyasla önemli ölçüde daha yüksek performans sergilemektedir. Fakat, test seti görüntülerinin rastgele seçilmesi nedeniyle izleme metriklerini hesaplamak mümkün olmamıştır. Gerçek dünya senaryolarında izleme performansını değerlendirmek için kare sırasının korunması büyük bir önem taşımaktadır. Bu durum, modelin video kareleri arasındaki nesneleri doğru şekilde takip etme yeteneğinin etkin bir şekilde değerlendirilmesini mümkün kılmaktadır.

4.1.2.2 Senaryo-II Sonuçları

Senaryo-II kapsamında gerçekleştirilen deneylerin tespit ve takip sonuçları Tablo 4.3'de sunulmuştur. Bu senaryoda, her bir videonun son 15 saniyesi test seti olarak ayrılmış ve videolar için tespit metrikleri YOLOv8-n modeli kullanılarak elde edilmiştir. YOLOv8-n modeli tarafından tespit edilen spermler, DeepSORT algoritması ile takip edilmiştir. Senaryo-II sonuçları için tespit ve MOT metriklerinin ortalamaları Tablo 4.3'nin son satırında özetlenmiştir. Tablo 4.3'de yer alan GTDets ve GTIDs sütunları, her bir video için toplam nesne sayısını ve benzersiz kimlikleri göstermektedir. Elde edilen sonuçlara göre, 30 video için ortalama tespit metrikleri sırasıyla %92,81 P, %84,19 R, %87,63 F1, %83,88 mAP50 ve %63,98 mAP50:95 olarak bulunmuştur. Ayrıca, video sekanslarının son 15 saniyesi test seti olarak kullanıldığından, aynı videolar için MOT metrikleri de hesaplanmıştır. Senaryo-II kapsamında izleme performansı incelendiğinde, 30 video için ortalama değerlerin sırasıyla %72,27 HOTA, %77,88 MOTA ve %86,31 IDF1 olduğu görülmektedir. Elde edilen tespit ve MOT metrik performansları, Hi-LabSpermTracking veri setinin bu alandaki araştırmaları önemli ölçüde ilerletme potansiyeline sahip olduğunu göstermektedir. Ancak, bu veri bölümleme yaklaşımı teorik olarak geçerli olsa da, test seti olarak her videonun son bölümüne güvenmek, modelin daha önce karşılaşmadığı hasta videolarıyla ilk kez karşılaşacağı gerçek dünya senaryosunu tam olarak temsil etmemektedir.

4.1.2.3 Senaryo-III Bireysel Sonuçları

Senaryo I ve II'de gerçekleştirilen deneyler, Hi-LabSpermTracking veri setinin sperm tespiti ve takibi için teorik olarak son derece uygun olduğunu doğrulamıştır. Ancak, her iki senaryo da hem eğitim hem de test setlerinde, aynı hastalardan alınan,

Tablo 4.3 Senaryo II: her videonun son 15 saniyesi (test seti) için tespit ve MOT metrikleri

Video	Tespit Performansı				Takip Performansı					Video Özellikleri				
Adı	**P**	**R**	**F1**	**mAP50**	**50:95**	**HOTA**	**DetA**	**AssA**	**MOTA**	**MOTP**	**MODA**	**IDF1**	**Dets**	**GTDets**	**IDs**	**GTIDs**	**IDSW**
Vid-1	77,57	74,53	75,64	73,48	51,65	62,36	50,13	79,20	53,92	79,48	53,92	76,83	1791	1812	23	13	0
Vid-2	96,02	65,85	77,43	65,62	44,43	60,22	52,71	69,71	66,10	78,07	66,40	76,40	1220	1717	15	13	5
Vid-3	72,98	55,27	60,47	52,21	32,89	43,83	34,28	56,05	37,35	72,78	37,55	61,94	361	514	11	7	1
Vid-4	89,97	85,07	87,28	84,99	68,60	74,02	64,76	85,57	74,80	84,07	74,87	87,03	1434	1480	12	10	8
Vid-5	81,26	85,76	82,96	85,08	54,42	65,01	54,09	78,30	62,76	79,08	62,76	82,03	1601	1493	16	11	0
Vid-6	87,69	76,33	81,39	75,70	51,08	64,06	55,54	75,01	63,81	80,17	64,18	77,08	1934	2194	26	19	8
Vid-7	96,06	72,40	82,31	72,25	43,87	61,55	54,94	69,76	70,37	75,66	70,49	81,48	2023	2626	21	18	3
Vid-8	97,48	82,45	89,07	82,41	66,93	75,67	68,22	84,99	79,84	85,56	80,13	86,86	2040	2441	23	19	7
Vid-9	95,84	80,23	87,23	80,14	57,06	69,11	63,38	76,51	78,69	81,50	78,98	83,39	2666	3107	29	21	9
Vid-10	98,57	92,40	95,31	92,37	66,52	77,64	73,53	84,05	93,26	81,15	93,28	95,55	3473	3648	28	25	1
Vid-11	92,91	96,95	94,85	96,92	76,96	80,58	74,49	88,86	89,15	84,26	89,21	93,54	5909	5586	48	38	3
Vid-12	99,87	92,20	95,83	92,20	78,49	82,86	77,93	88,95	90,93	86,93	90,97	94,92	2672	2933	21	20	1
Vid-13	92,75	75,89	83,25	75,70	49,76	64,85	58,05	73,16	69,43	79,98	69,69	81,83	1882	2290	26	19	6
Vid-14	95,35	78,74	86,04	78,68	66,60	72,27	65,47	80,90	73,26	85,74	73,57	80,19	1603	1956	19	15	6
Vid-15	88,56	58,68	69,69	57,14	30,02	47,34	41,01	54,70	48,92	72,73	49,55	63,52	714	1106	21	11	7
Vid-16	92,74	89,26	90,85	89,02	70,62	74,89	69,46	83,71	82,92	82,83	83,24	86,56	2428	2482	26	18	8
Vid-17	98,75	96,76	97,72	96,76	81,27	86,05	81,24	92,12	95,63	87,12	95,80	96,96	4072	4144	33	28	7
Vid-18	98,86	89,58	93,95	89,57	81,13	84,17	79,35	89,99	88,71	89,58	88,94	91,27	3179	3491	29	24	8
Vid-19	97,40	99,20	98,26	99,19	75,17	81,60	78,27	86,16	97,12	82,24	97,12	98,57	2719	2676	24	20	0
Vid-20	91,41	84,64	87,68	84,40	59,72	70,06	64,21	78,63	77,75	79,64	77,99	86,87	2769	2881	33	21	7
Vid-21	85,06	94,57	89,44	94,41	70,60	73,84	66,27	83,50	76,13	82,68	76,41	85,85	2030	1793	22	13	5
Vid-22	98,98	92,61	95,59	92,61	80,92	85,64	81,05	91,52	94,13	88,03	94,13	96,98	2193	2315	17	16	0
Vid-23	99,78	97,33	98,31	97,33	83,68	86,07	83,46	90,04	98,65	85,83	98,65	99,32	442	444	3	3	0
Vid-24	99,44	97,33	98,08	97,33	86,75	85,99	83,87	88,49	98,34	87,00	98,34	99,17	418	421	3	3	0
Vid-25	96,79	86,90	90,75	86,38	52,80	66,29	63,92	69,26	84,29	77,14	84,57	90,38	630	700	7	5	2
Vid-26	99,41	91,50	95,18	91,50	74,95	81,80	76,07	89,21	92,48	84,55	92,48	96,10	1099	1184	8	8	0
Vid-27	87,53	90,00	88,58	89,85	77,31	77,48	69,29	87,65	77,03	86,94	77,10	85,64	1549	1480	13	10	1
Vid-28	89,96	90,87	90,25	90,72	66,59	73,30	69,04	78,45	80,01	82,07	80,36	87,41	1492	1446	14	10	5
Vid-29	90,46	87,67	88,95	87,53	72,92	76,41	67,63	86,95	78,55	84,42	78,55	89,22	1759	1776	16	12	0
Vid-30	94,82	64,86	76,63	64,79	45,65	63,18	52,61	76,83	61,97	80,91	62,26	76,47	719	1036	10	7	3
Ortalama & Birleştirme	**92,81**	**84,19**	**87,63**	**83,88**	**63,98**	**72,27**	**65,81**	**80,61**	**77,88**	**82,27**	**78,05**	**86,31**	**58821**	**63172**	**597**	**457**	**111**

yeni yönelim ve konumlardaki aynı sperm nesnelerini içeren kareler sunduğundan, bu senaryolar bir modelin tamamen yeni hasta verileriyle ilk kez karşılaştığı

gerçek dünya koşullarını tam olarak yansıtamamaktadır. Bu sınırlamayı ele almak amacıyla, Senaryo III gerçek dünya uygulama senaryosu olarak tasarlanmıştır. Bu kurulumda, veri kümesindeki 30 video yalnızca eğitim için kullanılmış ve eğitilen modelleri değerlendirmek için, daha önce görülmemiş 6 ek video (4 farklı hastadan kaydedilmiş) eklenmiştir. Tablo 4.4, her videodaki nesne sayısını ve benzersiz kimlikleri gösteren GTDets ve GTIDs sütunlarını sunmaktadır. Bu altı yeni videoda toplamda 166.666 nesne ve 248 benzersiz kimlik bulunmaktadır. Bu tasarım, sperm algılama ve izleme modellerinin ilk kez yeni hastalardan alınan veriler üzerinde test edilmesini sağlayarak, gerçek klinik uygulamaya daha yakın bir değerlendirme yapılmasını mümkün kılmaktadır.

Senaryo-III kapsamında eğitilen bireysel modellerin sonuçları Tablo 4.4'te ayrıntılı olarak sunulmuştur. Modeller Hi-LabSpermTracking veri kümesi kullanılarak 20 epoch boyunca eğitilmiştir. Eğitim sürecinin ardından modeller test veri kümesi olarak kullanılan videolar üzerinde çalıştırılmış ve her bir videoya özgü tespit ve takip performans metrikleri hesaplanmıştır. Bu metriklerin hesaplanması için güven skoru ve IoU eşiği 0,5 olarak alınmıştır. YOLOv8n modelinin performansı değerlendirildiğinde, en yüksek tespit başarısının, diğer videolara kıyasla daha az nesne (1.783) içeren Vid-35 videosunda elde edildiği gözlemlenmiştir. Bu videoda modelin tespit performansını yansıtan metrik değerleri sırasıyla %99,70 P, %98,57 R, %98,98 F1, %98,57 mAP50 ve %78,12 mAP50:95 olarak tespit edilmiştir. İzleme metrikleri incelendiğinde Vid-35 videosunun %83,25 HOTA, %99,38 MOTA ve %99,69 IDF1 elde ettiği görülmüştür. Bu videoda tespit edilen nesneler başarılı bir şekilde kimliklendirilmiş ve hiçbir IDSW meydana gelmemiştir. Buna karşılık, en düşük performans daha fazla sayıda nesne (45.998) içeren Vid-36 videosunda gözlemlenmiştir. Bu videoda YOLOv8n modeli %81,80 P, %65,04 R, %72,39 F1, %60,47 mAP50 ve %21,13 mAP50:95 tespit performansı metrikleri elde edilmiştir. İzleme metrikleri %49,14 HOTA, %53,16 MOTA ve %65,70 IDF1 olarak hesaplanmıştır. Vid-36 videosunda modelin yaklaşık 8.000 nesneyi tespit edemediği ve toplam 84 IDSW yaptığı görülmektedir. Modelin tüm videolardaki ortalama performansı incelendiğinde, tespit metrikleri %90,19 P, %83,23 R, %86,37 F1-skoru, %81,98 mAP50 ve %47,43 mAP50:95 olarak elde edilmiştir. İzleme problemleri için ortalama değerler %64,53 HOTA, %74,54 MOTA ve %82,31 IDF1 olmuştur. Toplam 106.157 nesne doğru bir şekilde tespit edilmiş ve 149 IDSW meydana gelmiştir.

Tablo 4.4'te sunulan RT-DETR-L modelinin sonuçları incelendiğinde, YOLOv8n modeline benzer şekilde tespit ve takip performans metriklerinin Vid-35 videosunda en iyi değerlerine ulaştığı görülmektedir. Vid-35 için elde edilen tespit performansı metrikleri %98,48 P, %96,47 R, %97,20 F1-skor, %96,47 mAP50 ve %67,56

Tablo 4.4 Senaryo-III: görünmeyen test videoları için tespit ve MOT metrikleri

Video Adı	Tespit Yöntemi	Tespit Performansı					Takip Performansı						Video Özellikleri				
		P	R	F1	mAP50	50-95	HOTA	DetA	AssA	MOTA	MOTP	MODA	IDF1	Dets	GTDets	IDs	GTIDs	IDSW
Vid-31	Ensemble	88,69	92,41	90,44	92,32	54,75	66,40	60,30	74,58	80,90	73,45	80,97	88,74	16,288	15,598	65	38	12
Vid-32	Ensemble	79,41	87,24	83,06	84,16	46,10	62,47	52,02	77,64	61,80	74,73	61,87	79,93	21,964	19,297	70	40	14
Vid-33	Ensemble	93,60	90,21	91,82	90,07	56,02	71,17	65,83	79,85	84,95	77,74	85,02	90,13	18,628	19,156	53	36	13
Vid-34	Ensemble	87,73	86,09	86,82	85,88	49,51	65,67	57,67	78,81	72,79	75,20	72,87	84,08	14,855	14,834	56	34	11
Vid-35	Ensemble	98,92	99,01	98,84	99,01	73,50	81,86	79,52	84,91	99,50	81,78	99,50	99,75	1,784	1,783	3	3	0
Vid-36	Ensemble	80,18	72,71	76,22	67,86	23,67	52,03	46,61	60,46	58,61	66,70	58,77	72,10	43,015	45,998	158	97	73
Vid-31	RT-DETR-L	92,34	79,82	85,51	79,48	40,57	62,32	55,13	72,11	74,66	72,98	74,73	84,01	13,845	15,598	62	38	12
Vid-32	RT-DETR-L	85,93	74,96	79,88	72,41	34,95	59,37	50,37	72,36	63,19	73,01	63,29	78,54	18,111	19,297	70	40	20
Vid-33	RT-DETR-L	94,93	77,38	85,16	77,19	46,18	64,15	57,42	72,70	75,65	75,87	75,73	84,08	15,992	19,156	47	36	14
Vid-34	RT-DETR-L	94,93	80,87	87,22	80,69	44,35	61,28	57,98	66,38	76,97	74,11	77,12	77,05	12,986	14,834	62	34	23
Vid-35	RT-DETR-L	98,48	96,47	97,20	96,47	67,56	77,67	74,84	81,64	97,98	78,69	97,98	98,99	1,769	1,783	3	3	0
Vid-36	RT-DETR-L	79,62	61,87	69,56	57,44	19,92	48,11	42,16	57,55	48,45	65,77	48,69	64,57	37,478	45,998	172	97	113
Vid-31	YOLOv8n	92,72	84,95	88,59	84,83	48,14	64,24	56,63	74,80	74,90	72,92	75,00	84,86	14,587	15,598	54	38	16
Vid-32	YOLOv8n	83,98	84,04	83,93	81,88	44,08	62,33	53,47	75,39	66,88	74,60	66,97	79,81	19,916	19,297	71	40	18
Vid-33	YOLOv8n	93,71	83,83	88,40	83,57	48,41	67,04	60,96	76,22	80,10	76,18	80,17	87,49	17,713	19,156	51	36	13
Vid-34	YOLOv8n	89,24	82,95	85,90	82,59	44,70	61,19	56,47	68,71	72,82	74,18	72,94	76,33	14,148	14,834	58	34	18
Vid-35	YOLOv8n	99,70	98,57	98,98	98,57	78,12	83,25	81,36	85,49	99,38	83,45	99,38	99,69	1,780	1,783	3	3	0
Vid-36	YOLOv8n	81,80	65,04	72,39	60,47	21,13	49,14	43,67	58,15	53,16	67,52	53,34	65,70	38,013	45,998	151	97	84
Ortalama & Birleştirme	Ensemble	88,09	87,94	87,87	86,55	50,59	66,60	60,32	76,04	76,42	74,93	76,50	85,79	116,534	116,666	405	248	123
	RT-DETR-L	91,04	78,56	84,09	77,28	42,25	62,15	56,32	70,46	72,82	73,41	72,92	81,21	100,181	116,666	416	248	182
	YOLO	90,19	83,23	86,37	81,98	47,43	64,53	58,76	73,13	74,54	74,81	74,64	82,31	106,157	116,666	388	248	149

mAP50:95 olarak kaydedilmiştir. MOT problemi için metrikler %77,67 HOTA, %97,98 MOTA ve %98,99 IDF1 değerleri ile oldukça başarılı bir performans göstermiştir. Tespit edilen nesne sayısının (1.769) olması gereken nesne sayısına (1.783) çok yakın olduğu ve video boyunca hiçbir IDSW meydana gelmediği görülmektedir. Ancak, YOLOv8n modelinde olduğu gibi, en düşük tespit ve takip performansı en fazla sayıda nesne içeren Vid-36 videosunda gözlemlenmiştir. Vid-36'nın analizinde %79,62 P, %61,87 R, %69,56 F1-skoru, %57,44 mAP50 ve %19,92 mAP50:95 tespit performansı elde edilmiştir. İzleme performansı açısından model %48,11 HOTA, %48,45 MOTA ve %64,57 IDF1 değerleri elde edilmiştir. Modelin Vid-36 boyunca yaklaşık 9.000 nesneyi tespit edemediği ve toplam 113 IDSW yaptığı gözlemlenmiştir. RT-DETR-L modelinin tüm test setindeki genel performansı incelendiğinde, tespit problemleri için ortalama değerler %91,04 P, %78,56 R, %84,09 F1, %77,28 mAP50 ve %42,24 mAP50:95 olarak elde edilmiştir. MOT problemi için ortalama değerler %62,15 HOTA, %72,82 MOTA ve %81,21 IDF1'dir. Tüm videolarda tespit edilen toplam nesne sayısının 100.181 olduğu ve toplam 182 IDSW meydana geldiği görülmüştür.

Modellerin sonuçları incelendiğinde (Tablo 4.4'te gösterildiği gibi), her iki modelin de tüm videolarda tutarlı bir performans sergilediği, ancak bir modelin diğerine göre avantajlı ve dezavantajlı olduğu bazı durumlar olduğu görülmüştür. Modellerin genel performansı değerlendirildiğinde, deneysel çalışmalar YOLOv8n modelinin RT-DETR-L modelinden daha iyi performans gösterdiğini ortaya koymaktadır. Ancak sonuçlar detaylı olarak incelendiğinde RT-DETR-L modelinin %91,04 P değeri ile YOLOv8n modeline kıyasla daha yüksek bir performans elde ettiği görülmüştür. Her iki modelin toplam tespit (DET) sayıları incelendiğinde, RT-DETR-L modelinin YOLOv8n modeline göre yaklaşık 6.000 daha az tespit gerçekleştirdiği, ancak yine de daha yüksek bir P değerine sahip olduğu görülmektedir. RT-DETR-L modelinin YOLOv8n modelinden daha yüksek bir P değeri elde ettiği Vid-32, Vid-33 ve Vid-34 videoları incelendiğinde, YOLOv8n modelinin tespit sayılarının RT-DETR-L modeline kıyasla GTDets'e çok daha yakın olduğu, hatta bazı durumlarda (örneğin Vid-32) daha fazla tespit gerçekleştirdiği görülmektedir. Bulgular doğrultusunda, RT-DETR-L modelinin YOLOv8n modeline kıyasla daha yüksek bir P oranına ulaştığı ve nesneleri (sperm hücreleri) doğru şekilde tespit etme konusunda daha iyi bir beceriye sahip olduğu sonucuna varılabilmektedir. Bununla birlikte, RT-DETR-L modeli R metriğinde yetersiz kalmaktadır, bu da YOLOv8n'e kıyasla karede bulunan nesnelerin daha küçük bir oranının tespit edildiği anlamı çıkarılmaktadır. Bu bağlamda, tespit performansında P ve R metriklerinin dengesini sağlamak amacıyla her iki modelin avantajları entegre edilmiş ve bu entegrasyonun izleme performansı üzerindeki

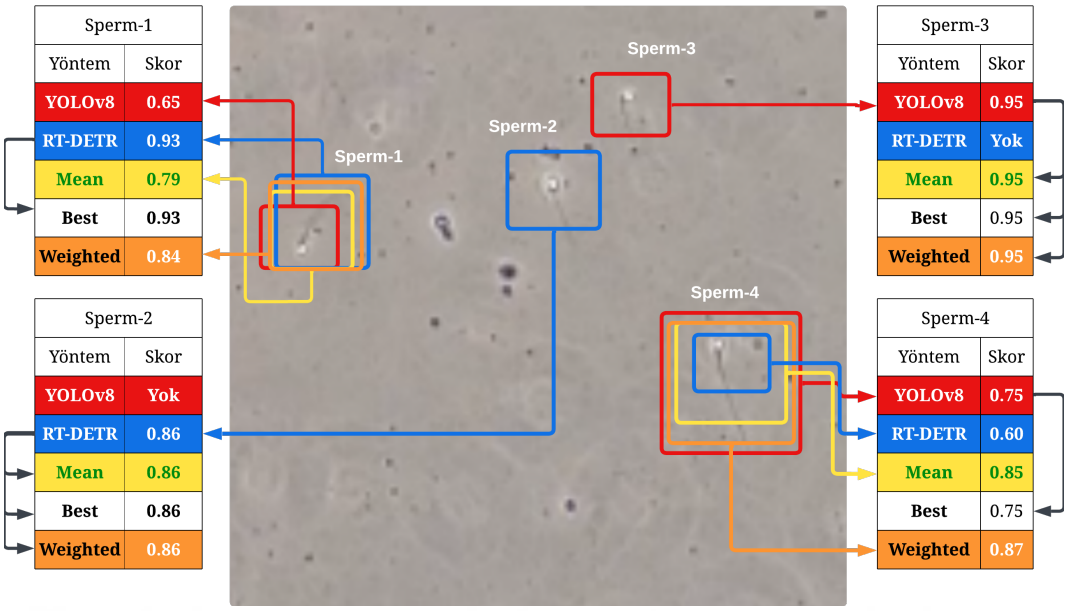
etkisi ayrıntılı bir şekilde incelenmiştir.

4.1.2.4 Senaryo-III Ensemble Modellerinin Sonuçları

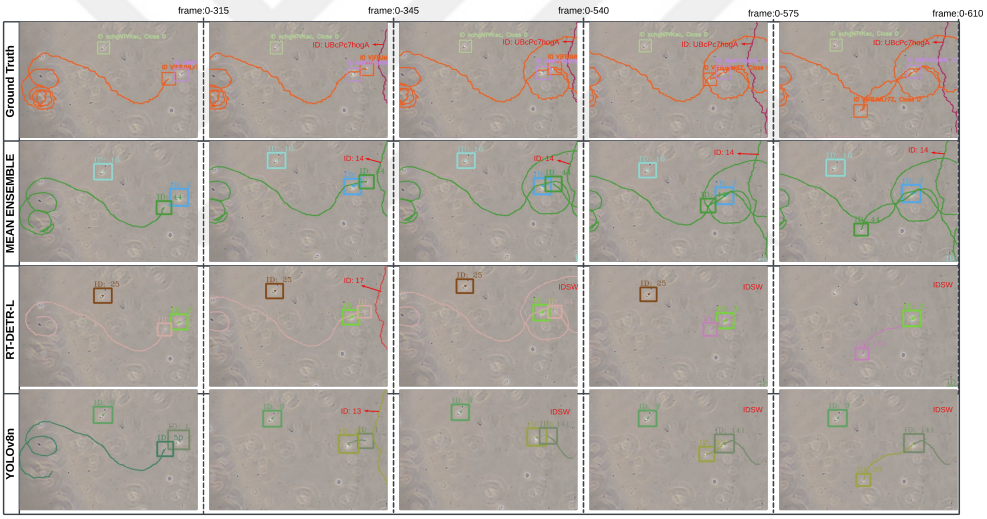
Bu senaryoda, hem YOLO hem de RT-DETR modellerinin güçlü yönlerinden yararlanarak, özellikle karmaşık ve çeşitli senaryolarda tespit doğruluğunu artırmak amacıyla topluluk yöntemleri uygulanmıştır. Genel performansı ve sağlamlığı iyileştirmeyi hedefleyen bu yaklaşımda, bireysel modellerden elde edilen tespit sonuçlarını birleştirmek için BEST, MEAN ve WEIGHTED olmak üzere üç farklı topluluk yöntemi önerilmiştir. Her bir yöntem, tespit sonuçlarını birleştirmek için farklı bir strateji benimsemektedir. Önerilen topluluk yöntemlerinin stratejilerini daha iyi açıklamak için, Şekil 4.8'de örnek bir çerçeve üzerinden her bir yönteme göre birleştirilmiş sperm tespitlerinin temsili gösterilmektedir. BEST yönteminde, her bir karede modeller tarafından tespit edilen nesneler arasında en yüksek güven skoruna sahip olan sınırlayıcı kutu doğru kabul edilmektedir. Örneğin, Şekil 4.8'de sunulan Sperm-1'de, nesne YOLOv8n modeli tarafından %65 güven skoru ile tespit edilirken, aynı nesne RT-DETR-L modeli tarafından %93 güven skoru ile tespit edilmiştir. BEST topluluk yöntemi uygulandığında, YOLOv8n modelinin tespiti yerine RT-DETR-L modelinin tespiti doğru olarak kabul edilerek performans metrikleri bu seçime göre hesaplanmaktadır.

MEAN yönteminde, her iki model tarafından yapılan tespit edilen sınırlayıcı kutuların koordinat değerlerinin ortalaması alınarak yeni bir sınırlayıcı kutu oluşturulmaktadır. Performans ölçütleri bu yeni sınırlayıcı kutu kullanılarak hesaplanmaktadır. Örneğin, Şekil 4.8'de gösterilen Sperm-4 tespitinde, modellere ait sınırlayıcı kutular mavi ve kırmızı ile gösterilirken, bu kutuların ortalaması alınarak oluşturulan yeni sınırlayıcı kutu sarı ile gösterilmektedir. Son yöntem olan WEIGHT yönteminde ise modellerin güven skorları sınırlayıcı kutuların koordinatlarını ağırlıklandırmak için kullanılmaktadır. Bu yöntemde model tespitinden elde edilen güven skorları normalize edilerek ağırlıkların toplamının 1'e eşit olması sağlanmaktadır. Normalize edilmiş ağırlıklar daha sonra modeller tarafından tespit edilen sınırlayıcı kutuların koordinat değerlerine uygulanarak ağırlıklandırılmış yeni bir sınırlayıcı kutu elde edilmektedir. Bu yaklaşım, daha başarılı modellerin nihai sınırlayıcı kutunun belirlenmesinde daha büyük bir etkiye sahip olmasını sağlamaktadır.

Sperm-2 ve Sperm-3'te olduğu gibi, sadece bir model tarafından tespit edilen nesneler için, başka birleştirilecek nesne bulunmadığından, bu tespitler herhangi bir ek işleme tabi tutulmadan doğrudan o karedeki tespit edilen nesneler arasına dahil edilmekte ve performans metrikleri buna göre hesaplanmaktadır. Modeller



Şekil 4.8 Önerilen topluluk yöntemlerinin derin öğrenme tabanlı tespit modelleri ile gösterilmesi



Şekil 4.9 Vid-34 üzerinde YOLO, RT-DETR ve MEAN topluluk yöntemi kullanılarak elde edilen tespit ve yörünge sonuçlarının karşılaştırılması

tarafından tespit edilen nesneleri eşleştirmek için IoU (Intersection over Union) yöntemi kullanılmış ve eşik değeri 0,5 olarak ayarlanmıştır. Birleştirme işleminden sonra, aynı nesne için birden fazla tespit sonucu varsa, en doğru olanı seçilmektedir. Ardından, birleştirilmiş tespitler, DeepSORT algoritmasına verilerek video boyunca nesnelerin takip edilmesi sağlanmıştır. Önerilen yöntemlerin tespit etme ve izleme performans metrikleri, Tablo 4.5'te sunulmaktadır.

Tablo 4.5'te yer alan topluluk yöntemlerinin sonuçları incelendiğinde, hem tespit hem de izleme problemleri açısından en başarılı yöntemin, tespitlerin ortalamasını alarak birleştirme işlemini gerçekleştiren MEAN yöntemi olduğu görülmektedir.

Tablo 4.5 Senaryo-III: Ensemble-metotların sonuçları

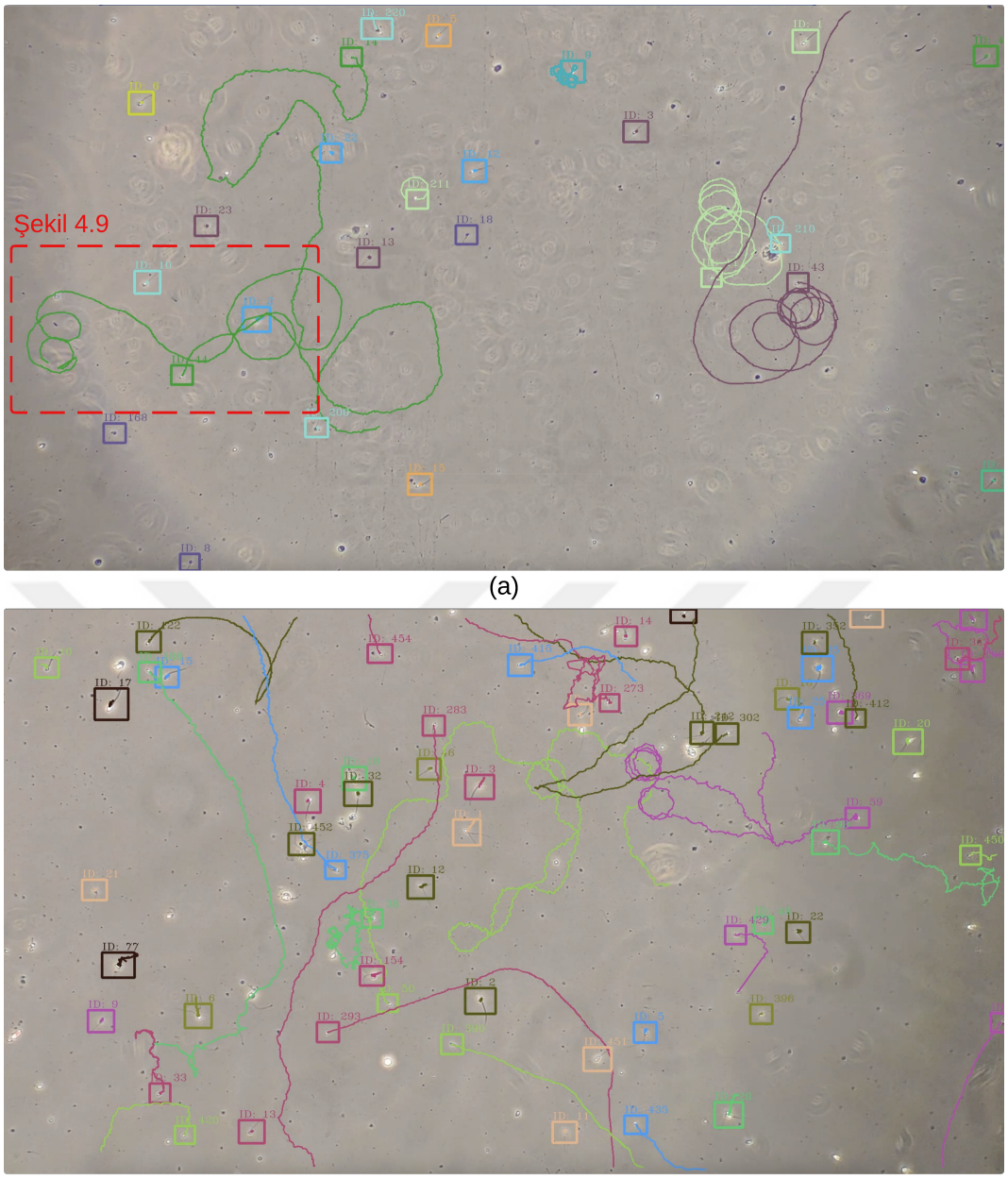
Video Adı	Topluluk Yöntemi	Tespit Performansı					Takip Performansı						Video Özellikleri				
		P	R	F1	mAP50	50-95	HOTA	DetA	AssA	MOTA	MOTP	MODA	IDF1	Dets	GTDets	IDs	GTIDs	IDSW
Vid-31	BEST	88,58	92,41	90,46	92,32	51,55	64,24	57,25	73,72	76,59	71,89	76,65	87,13	16,620	15,598	62	38	10
Vid-32	BEST	79,13	86,86	82,73	83,70	44,03	60,86	51,84	73,70	61,51	74,50	61,58	77,91	21,978	19,297	66	40	14
Vid-33	BEST	93,58	90,03	91,71	89,89	52,56	69,02	64,57	75,12	84,84	76,51	84,91	89,47	18,660	19,156	54	36	14
Vid-34	BEST	87,08	85,36	86,13	85,07	48,99	64,20	57,30	73,23	71,64	75,68	71,73	81,95	14,869	14,834	58	34	14
Vid-35	BEST	99,03	99,12	98,95	99,12	78,58	83,41	81,48	85,70	99,50	83,64	99,50	99,75	1,784	1,783	3	3	0
Vid-36	BEST	80,59	72,91	76,52	69,00	23,92	50,99	46,64	57,33	57,80	66,97	57,98	71,07	43,099	45,998	162	97	84
Vid-31	MEAN	88,69	92,41	90,44	92,32	54,75	66,40	60,30	74,58	80,90	73,45	80,97	88,74	16,288	15,598	65	38	12
Vid-32	MEAN	79,41	87,24	83,06	84,16	46,10	62,47	52,02	77,64	61,80	74,73	61,87	79,93	21,964	19,297	70	40	14
Vid-33	MEAN	93,60	90,21	91,82	90,07	56,02	71,17	65,83	79,85	84,95	77,74	85,02	90,13	18,628	19,156	53	36	13
Vid-34	MEAN	87,73	86,09	86,82	85,88	49,51	65,67	57,67	78,81	72,79	75,20	72,87	84,08	14,855	14,834	56	34	11
Vid-35	MEAN	98,92	99,01	98,84	99,01	73,50	81,86	79,52	84,91	99,50	81,78	99,50	99,75	1,784	1,783	3	3	0
Vid-36	MEAN	80,18	72,71	76,22	67,86	23,67	52,03	46,61	60,46	58,61	66,70	58,77	72,10	43,015	45,998	158	97	73
Vid-31	WEIGHTED	88,58	92,41	90,45	92,32	54,68	65,72	59,08	74,58	78,79	73,36	78,86	87,81	16,627	15,598	64	38	12
Vid-32	WEIGHTED	79,46	87,21	83,08	84,13	46,13	62,52	52,07	77,76	61,87	74,74	61,94	79,97	21,979	19,297	69	40	13
Vid-33	WEIGHTED	93,62	90,02	91,73	89,89	56,00	71,37	66,00	80,13	85,05	77,73	85,12	90,30	18,660	19,156	54	36	13
Vid-34	WEIGHTED	87,74	86,01	86,78	85,79	49,60	65,73	57,72	78,87	72,86	75,24	72,93	84,13	14,865	14,834	58	34	11
Vid-35	WEIGHTED	98,92	99,01	98,84	99,01	73,86	82,02	79,74	84,98	99,61	81,90	99,61	99,80	1,784	1,783	3	3	0
Vid-36	WEIGHTED	80,33	72,67	76,26	67,82	23,72	51,99	46,60	60,37	58,66	66,70	58,84	72,51	43,085	45,998	163	97	85
Ortalama & Birleştirme	BEST	88,00	87,78	87,75	86,52	49,94	65,45	59,85	73,13	75,31	74,86	75,39	84,55	117,010	116,666	405	248	136
	MEAN	88,09	87,94	87,87	86,55	50,59	66,60	60,32	76,04	76,42	74,93	76,50	85,79	116,534	116,666	405	248	123
	WEIGHTED	88,11	87,89	87,86	86,49	50,67	66,56	60,20	76,11	76,14	74,94	76,22	85,75	117,000	116,666	411	248	134

Özellikle tespit metrikleri açısından, topluluk yöntemleri arasında belirgin bir fark bulunmamaktadır. Bununla birlikte, elde edilen performans değerleri birbirine yakın olmasına rağmen, video boyunca spermlere atanan kimliklerin tutarlılığını yansıtan IDSW (Kimlik Değiştirme) değerinin MEAN yönteminde diğer yöntemlere kıyasla daha düşük olduğu, dolayısıyla bu yöntemin daha tutarlı bir kimliklendirme sağladığı görülmektedir.

MEAN yönteminin topluluk yöntemleri arasında en başarılısı olduğu göz önünde bulundurularak, bireysel modeller ile MEAN modeli arasındaki karşılaştırma Tablo 4.4'te sunulmuştur. Performans metrikleri incelendiğinde, MEAN modelinin, bireysel modellere kıyasla Kesinlik (%88,09) ve Duyarlılık (%87,94) arasındaki dengeyi koruyarak daha yüksek bir F1-skoru (%87,87) elde ettiği görülmektedir. Özellikle, temel nesne algılama ölçütleri olan mAP50 ve mAP50:95'te MEAN yöntemi, en iyi performans gösteren bireysel model olan YOLOv8n'e kıyasla sırasıyla yaklaşık %5 ve %3 oranında daha iyi performans sergilemektedir.

İzleme performansı açısından yapılan değerlendirmelerde, önerilen MEAN topluluk yöntemi, RT-DETR ve YOLOv8n modellerine kıyasla hem HOTA hem de MOTA metriklerinde yaklaşık %2 oranında daha iyi sonuçlar elde etmektedir. Ayrıca, tüm videolardaki toplam IDSW değerleri karşılaştırıldığında, MEAN topluluk modelinin RT-DETR-L modeline kıyasla 59, YOLOv8n modeline kıyasla ise 26 daha az kimlik değiştirme gerçekleştirdiği gözlemlenmektedir. Sonuç olarak, MEAN yöntemi ile, RT-DETR-L ve YOLOv8n modellerine kıyasla sırasıyla yaklaşık %5 ve %3 daha yüksek IDF1 değeri elde edilmiştir.

Modellerin Vid-34 üzerindeki tespit ve takip sonuçları Şekil 4.9'da sunulmuştur. Yöntemlerin performansı, özellikle sperm çarpışmaları sırasında meydana gelen İDSW' e bağlı olarak değerlendirilmiştir. Bu bağlamda, Vid-34'ten seçilen spermler kullanılarak ayrıntılı bir analiz yapılmıştır. Şekil 4.9'da, sperm hücrelerinin GT yörüngeleri ve kimlikleri, önerilen MEAN ensemble yöntemi ve bireysel modeller ile elde edilen yörüngeler ve kimliklerle birlikte sunulmaktadır. Şekil 4.9'un üst kısmında yer alan 'frame0:315' ifadesi başlangıçtan (frame 0) 315. kareye kadar olan video segmentini ifade etmektedir. 0-315 çerçeve aralığı incelendiğinde, MEAN ensemble yönteminin spermlere sırasıyla 16, 44 ve 2 kimlik atadığı görülmektedir. Diğer yöntemlerde ise RT-DETR-L 25, 64 ve 3, YOLOv8n ise 9, 50 ve 1 numaralı kimlikleri aynı spermlere atamıştır. Yöntemler karşılaştırıldığında, Vid-34'te 44, 64 ve 50 kimlikli spermlerin aynı spermi temsil ettiği görülmektedir. Ayrıca, bu spermin 0-315 kare aralığında her yöntemde farklı bir yörünge çizdiği görülmektedir. MEAN ensemble yönteminin oluşturduğu yörünge, diğer bireysel yöntemlere kıyasla önemli ölçüde daha uzundur. Bunun



Şekil 4.10 MEAN ensemble yöntemi kullanılarak (a) Vid-34 ve (b) Vid-36'da tespit ve izleme sonuçları (Çerçeve: 0-Son)

nedeni, bireysel yöntemlerin ilgili sperme ilk karelerde bir kimlik atadıktan sonra 315. frame'e gelene kadar kareler arasında daha fazla sayıda kimlik değişimi yapmasıdır.

GT yörüngesiyle karşılaştırıldığında, MEAN ensemble yöntemi diğer yöntemlere göre daha az IDSW gerçekleştirmiş ve spermin gerçek yörüngesine daha yakın bir sonuç elde etmiştir. Ayrıca, 315. kare 44, 64 ve 50 numaralı spermlerin 2, 3 ve 1 numaralı spermlerle çarpışmasından hemen önceki ana ait olan kareyi temsil etmektedir. Sperm çarpışmasından sonra IDSW meydana gelip gelmediğini

görmek için 0-345 kare aralığı incelenmiştir. 0-345 arası kareler incelendiğinde, MEAN ensemble ve RT-DETR-L yöntemlerinde IDSW gözlenmemiştir. Ancak, YOLOv8n yönteminde ID değeri 50 olan sperm ile ID değeri 1 olan sperm arasında IDSW meydana gelmiş ve ID değeri 50 olan sperm, takibe ID değeri 1 ile devam edilmiştir. Bu süreçte, YOLOv8n yönteminde önceden ID değeri 1 olan sperme yeni bir ID atanmıştır. Ayrıca, 345. karenin en sağında, aşağıdan yukarıya doğru hareket eden bir sperme ait yörünge her üç yöntem tarafından da tespit edilmiştir. Tespit edilen sperm, incelenen diğer spermlerden daha yukarıda konumlandırıldığı için kimlik bilgisi görünmemektedir. Bu kimlik bilgisini takip etmek amacıyla, spermlerin kimlikleri görsele manuel olarak eklenmiş ve sırasıyla 14, 17 ve 13 değerleri, genel video çerçevesine bakılarak atanmıştır.0-540 çerçeve aralığına ilerledikçe, MEAN ensemble yönteminde IDSW (Kimlik Değiştirme) oluşmazken, RT-DETR-L ve YOLOv8n modellerinin tespit ettiği nesnelerde IDSW olayları gözlemlenmiştir. RT-DETR-L yönteminde, ID 17 olan spermin yörüngesinin artık karede görünmediği; YOLOv8n modelinde ise aynı spermin ID'sinin değişerek 13'ten farklı bir değere sahip olduğu ve bu nedenle yörüngesinin artık çizilmediği dikkat çekmektedir. Ayrıca, YOLOv8n modelinde, bir önceki karede ID değeri 1 olan spermin IDSW nedeniyle ID'sinin 141 olarak değiştiği ve bu yeni ID ile takibe devam ettiği gözlemlenmiştir. Çarpışma sonrası, 0-575 kare aralığındaki sperm ID değerleri incelendiğinde, MEAN ensemble yönteminde çarpışmaya rağmen ID değerlerinde herhangi bir değişiklik meydana gelmediği ve ID değerlerinin tutarlı bir şekilde atandığı görülmüştür. Buna karşın, diğer bireysel yöntemlerde IDSW gözlemlenmiştir. RT-DETR-L modelinde, daha önceki karelerde ID 64 olarak atanan sperm artık ID 182 ile devam etmektedir. YOLOv8n yönteminde ise, ID 141 ve 83 arasında bir IDSW olayı meydana gelmiştir. Vid-34'ün son karesi olan 0-610. kareler incelendiğinde, MEAN ensemble yöntemi ile her bir spermin eksiksiz yörüngelerinin videonun başından sonuna kadar herhangi bir IDSW olmadan tutarlı bir şekilde elde edildiği görülmüştür. Buna karşın, bireysel modellerde hareketsiz spermler için bile IDSW gözlemlenmiş ve hareketli spermlerin yörüngeleri tutarlı bir şekilde elde edilememiştir.

En yüksek GTID ve GTDET sayılarına sahip Vid-36 ve Vid-34'ün sperm yörüngeleri Şekil 4.10'da sunulmuştur. Şekilde gösterilen tespitler ve izleme yörüngeleri MEAN Ensemble yöntemi kullanılarak elde edilmiştir. Şekil 4.10'da kesikli çizgiler halinde gösterilmiş olan Şekil 4.9, sperm IDSW olayını göstermek için kesilmiştir. Her bir spermin tüm video boyunca izlediği yörüngeler Şekil 4.10'da görülmektedir. Bu yörüngeler incelendiğinde, bazı spermlerin herhangi bir salınım olmaksızın doğrusal hareket ettiği (örn. Şekil 4.10b, ID:375), bazılarının ise dairesel hareket sergilediği (örn, Şekil 4.10a, ID:11) görülmektedir. Benzer şekilde

bazıları sabit kalırken (örneğin, Şekil 4.10b, ID:2), bazıları da diğerlerine kıyasla daha az yer değiştirmektedir (örneğin, Şekil 4.10b, ID:33). Hi-LabSpermTracking veri setini oluşturmak için geniş açılı videoların kullanılması, bu tür analizler sağlayarak spermin farklı hareketlilik yörüngeleri hakkında bilgi vermektedir.

4.1.2.5 Sperm Yörünge Analizi

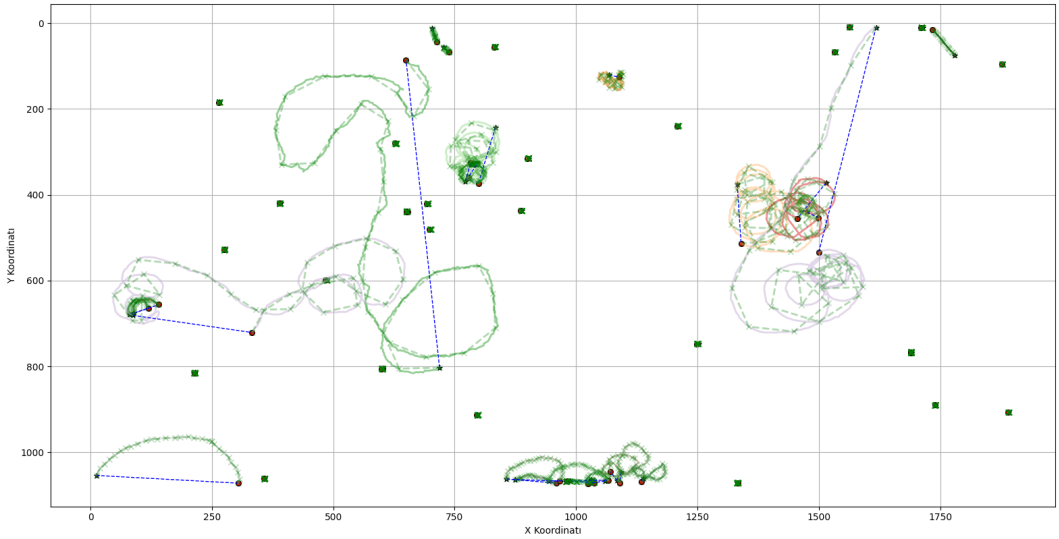
Tez kapsamında oluşturulan Hi-LabSpermTracking veri seti üzerinde, önerilen MEAN ensemble yöntemi kullanılarak sperm tespit ve takip süreçleri gerçekleştirilmiştir. Bu süreçlerin sonunda, test videolarında yer alan spermlere ait yörüngeler çıkarılmış ve DSÖ tarafından tanımlanan parametreler kullanılarak yörünge analizi gerçekleştirilmiştir.

Hi-LabSpermTracking veri setinde bulunan test videolarının her birinin gerçek referans dosyasında bulunan sperm koordinatları kullanılarak Tablo 4.1'de bulunan referans aralıklarına göre gerçek referans Grade seviyeleri elde edilmiştir. Her bir video için elde edilen Grade yüzdelerini Tablo 4.6'da görmek mümkündür. Aynı tabloda, benzer şekilde MEAN ensemble yöntemi kullanılarak videolarda bulunan spermlerin koordinat ve ID değerlerine göre parametre çıkarım işlemi yapılarak Grade yüzdeleri elde edilmiştir. Senaryo-III kapsamında, örnek görüntü olarak verilen Şekil 4.10'a benzer şekilde, aynı videolar (Vid-34 ve Vid-36) kullanılarak spermlere ait yörüngeler ve VAP, VCL, VSL hesaplamak için çizdirilen yollar, Şekil 4.11 ve 4.12'de örnek olarak sunulmuştur. Şekil 4.9'da yöntemler arasındaki IDSW'yi göstermek amacıyla sunulan görüntüde bulunan 44 numaralı sperme ait VAP, VCL, VSL, STR, LIN ve WOB değerleri ise Şekil 4.13'te yer almaktadır.

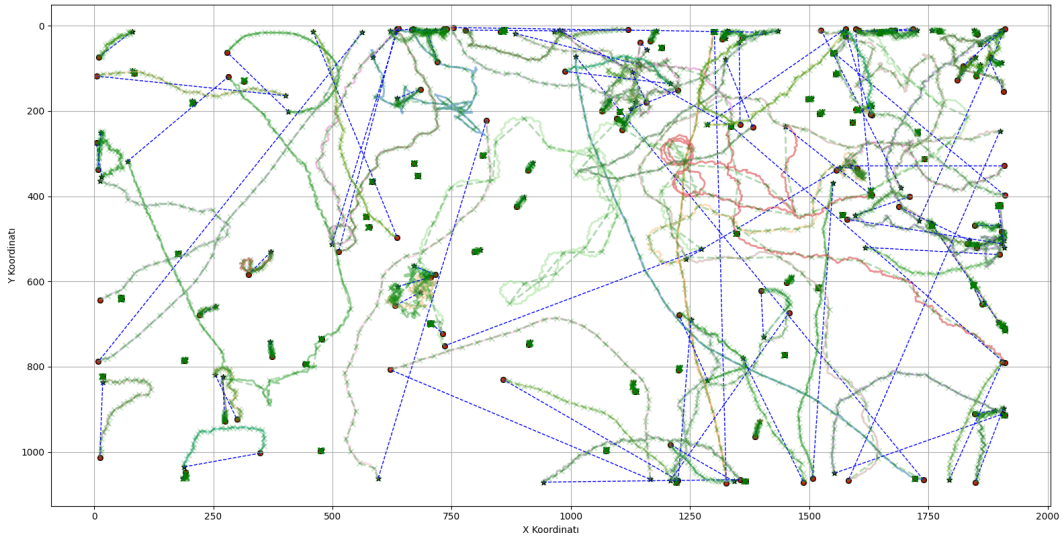
Tablo 4.6 Önerilen sperm yörünge analizi ile gerçek referans değerlerin karşılaştırılması

	Gerçek Referans (%)				**MEAN (%)**			
	Grade-A	**Grade-B**	**Grade-C**	**Grade-D**	**Grade-A**	**Grade-B**	**Grade-C**	**Grade-D**
Vid-31	0,0	13,2	18,4	68,4	0,0	9,4	10,9	79,7
Vid-32	0,0	12,5	20	67,5	0,0	11,4	12,9	75,7
Vid-33	0,0	19,4	19,4	61,1	0,0	15,1	11,3	73,6
Vid-34	0,0	14,7	23,5	61,8	0,0	6,9	15,5	77,6
Vid-35	0,0	0,0	33,3	66,7	0,0	0,0	33,3	66,7
Vid-36	0,0	25,8	14,4	59,8	0,0	24,4	12,8	62,8
Ortalama	0,0	14,3	21,5	64,32	0,0	11,2	16,1	72,7

Test setinde yer alan videoların gerçek referans koordinat ve kimlik bilgileri kullanılarak, gerçek referans hareketlilik seviyeleri elde edilmiştir. Benzer şekilde, bu çalışmada sunulan MEAN yöntemi ve DeepSORT yöntemleri ile elde edilen sperm yörüngeleri analiz edilerek, video bazlı sperm hareketlilik seviyeleri hesaplanmıştır. Her bir video için elde edilen hareketlilik seviyeleri Tablo 4.6'de sunulmuştur. Videoların bireysel olarak değerlendirildiği durumlarda,



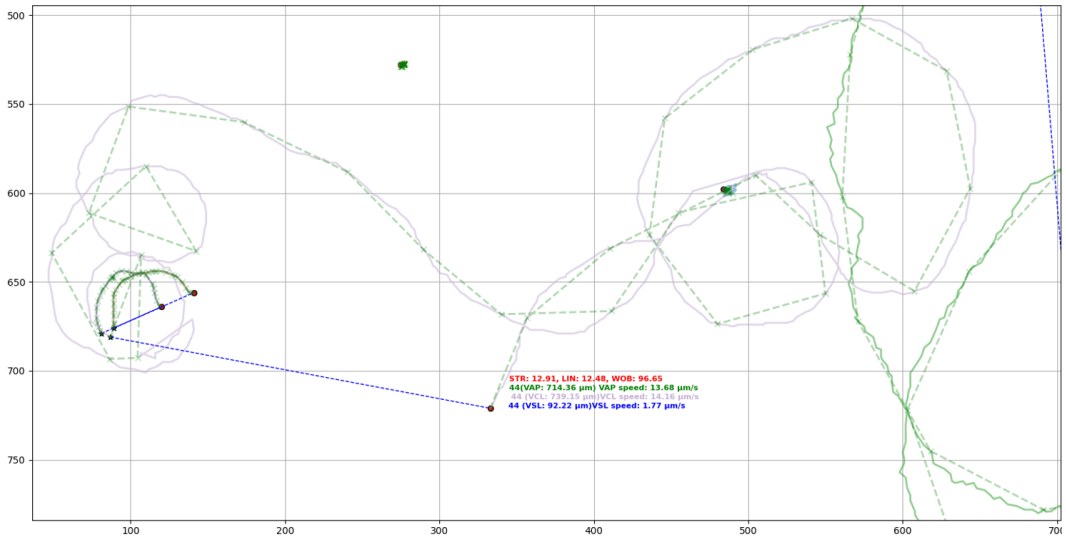
Şekil 4.11 Vid-34'e ait örnek sperm yörünge analizi



Şekil 4.12 Vid-36'ya ait örnek sperm yörünge analizi

en kalabalık ve karmaşık sperm hücrelerine sahip video olan Vid-36, tüm hareketlilik seviyelerinde gerçek referans değere oldukça yakın sonuçlar elde etmiştir. Diğer taraftan, sperm hücre sayısının en düşük olduğu Vid-35'te ise referans değerleriyle birebir örtüşen sonuçlar elde edilmiştir. Diğer videolarda ise bazı farklılıklar gözlemlenmekle birlikte, hareketlilik seviyeleri ve spermlerin yüzdesel dağılımı incelendiğinde, MEAN yönteminin gerçek referans değerlerle paralel sonuçlar ürettiği tespit edilmiştir. Tüm videoların hareketlilik seviyelerinin ortalamaları alındığında, gerçek referans değerleri ile MEAN yöntemiyle elde edilen değerler arasında yüksek bir tutarlılık olduğu ve bu modelin klinik çalışmalarda kullanılabilirliğinin sağlandığı anlaşılmaktadır.

Şekil 4.11 ve 4.12'de, her bir sperme ait hareketlilik özellikleri gösterilmektedir.



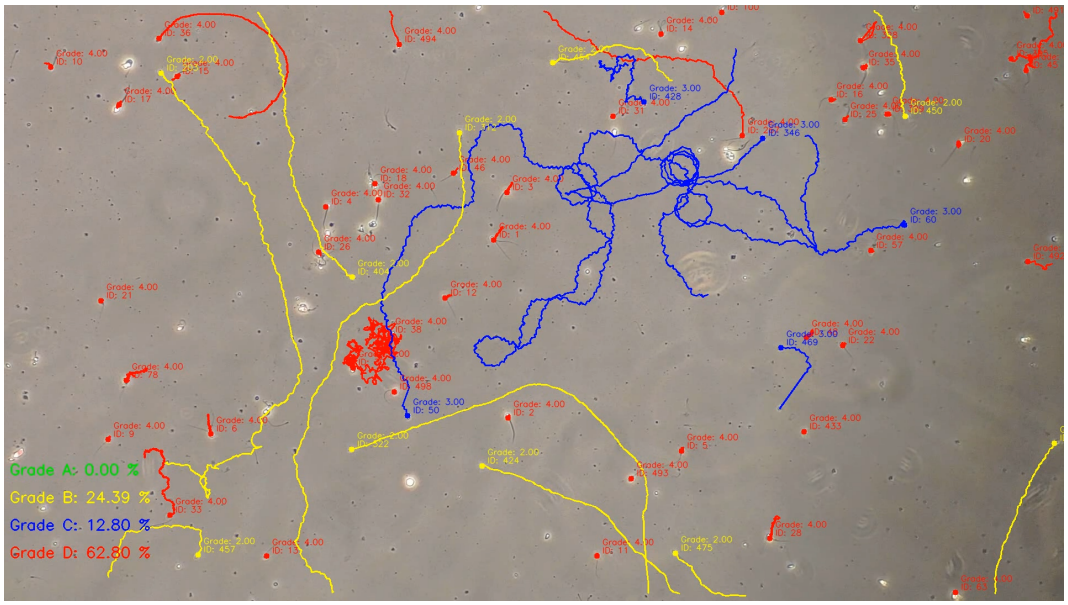
Şekil 4.13 Vid-34'de bulunan 44 ID değerine sahip sperme ait yörünge analizi

Bu veriler kullanılarak, videolarda bulunan spermlere ait yörüngeler, hareketlilik seviyelerine göre renklendirilmiştir. Videodaki spermlerin hareketlilik seviyelerine ilişkin yüzdesel değerler, videonun sol alt köşesinde yer alan yazılarla belirtilmiş ve böylece kullanıcıların video hakkında bilgi edinmesi sağlanmıştır. Örnek olarak, Vid-36'nın son karesi Şekil 4.14'de sunulmuştur.

Grade-A sperm hareketlilik seviyesine sahip sperm hücreleri yeşil, Grade-B sperm hücreleri sarı, Grade-C sperm hücreleri mavi ve Grade-D sperm hücreleri kırmızı renk ile gösterilmiştir. Sperm hücrelerinin yörüngeleri ve hareketlilik türleri incelendiğinde; hareketsiz veya çok yavaş hareket eden sperm hücrelerinin kırmızı, dairesel hareketler ile kendi eksenlerinde dönerek ilerleyen sperm hücrelerinin mavi, diğerlerine kıyasla daha düz bir şekilde ve belirli bir hızda ilerleyen sperm hücrelerinin ise sarı renkle sınıflandırıldığı gözlemlenmektedir. Sperm hücrelerinin video boyunca gerçekleştirdikleri yörünge ve hareketlilik sınıfları değerlendirildiğinde, tez kapsamında uygulanan sperm hareket analizinin tutarlı sonuçlar verdiği anlaşılmaktadır.

4.1.3 Tartışma

Bu tez kapsamında, CNN ve dönüştürücü tabanlı nesne tespit yöntemleri kullanılarak elde edilen tespitler, çeşitli topluluk öğrenme teknikleri ile birleştirilmiş ve DeepSORT nesne takip algoritması ile sperm tespit ve takip problemleri için performans değerlendirmesi yapılmıştır. Gerçekleştirilen deneyler, özellikle sperm tespit ve takip alanında literatüre önemli katkılar sağlaması beklenen yeni Hi-LabSpermTracking veri kümesi üzerinde yapılmıştır. Bu veri kümesi, 14 farklı hastaya ait toplam 66 video içermektedir ve çalışmanın bir



Şekil 4.14 Vid-36 son karesindeki sperm hücrelerinin yörüngeleri ve hareketlilik seviyelerine göre renklendirilmesi

parçası olarak kamuya açık hale getirilecektir. Tespit problemleri için bireysel ve topluluk modellerinin performans değerlendirmesi, P, R, F1, mAP50 ve mAP50:95 metrikleri kullanılarak yapılmıştır. MOT problemleri için ise modellerin performansı, HOTA, MOTA ve IDF1 metrikleri kullanılarak analiz edilmiştir.

Tablo 1.1, literatürde mevcut olan spermle ilgili veri setlerinin genel bir özetini sunmaktadır. Tablo incelendiğinde, mevcut veri kümelerinin çoğunun sperm morfoloji analizi için oluşturulduğu görülmektedir [7, 19–21, 24, 29, 35, 37, 38, 42]. Sperm tespit problemleri için geliştirilen veri kümeleri incelendiğinde, bu çalışmaların genellikle sperm hücrelerinin tespitine odaklandığı ve spermlerin yalnızca sınırlayıcı kutular (bbox) kullanılarak etiketlendiği gözlemlenmektedir [38, 39, 62, 66, 67]. Bununla birlikte, sperm izleme problemleri için, sınırlayıcı kutu etiketlerine ek olarak, her spermin tüm karelerde tutarlı bir şekilde benzersiz bir kimlikle ilişkilendirilmesi gerekmektedir. Bu ek gereklilik, sperm hücresi yörüngelerinin doğru bir şekilde analiz edilebilmesi için kritik öneme sahiptir. Literatürde, sperm izleme problemleri için gerekli olan benzersiz kimliklendirmeyi ve tüm karelerde tutarlı etiketlemeyi sunan tek veri seti, VISEM-Tracking veri setidir [41]. Bu veri seti, sperm takibi alanında önemli bir kaynak olmasına rağmen bazı kısıtlamalara sahiptir. İlk olarak, veri setindeki videolar yalnızca 30 saniye uzunluğundadır ve bu durum çalışmanın kapsamını önemli ölçüde sınırlandırmaktadır. Ayrıca, veri setindeki bazı videolar (örneğin, Video 14, 15, 21, 23, 35, 47, 54, 82) iki farklı videonun birleştirilmesiyle, bazılarının ise (örneğin, Video 29, 52, 60) üç farklı videonun birleştirilmesiyle oluşturulduğu tespit

edilmiştir. Bu durum, bir sperm hücresinin yörüngesini tüm video boyunca tutarlı bir şekilde izlemeyi imkansız hale getirmektedir. Ayrıca, VISEM-Tracking veri setindeki videolar 400× büyütme ile çekilmiş olup 640×480 çözünürlüğe sahiptir. Bu teknik özellikler, sperm hücrelerinin dar bir görüş açısıyla kaydedilmesine neden olmakta ve bu da yörüngelerinin ayrıntılı, geniş açılı bir şekilde analiz edilmesini sınırlamaktadır. Daha geniş bir görüş alanı, sperm yörüngelerinin daha geniş bir düzlemde izlenmesini sağlayarak daha kapsamlı bir perspektif sunmaktadır. Bu özellik, DSÖ'nün Grade A, B, C ve D sperm hareketlilik sınıflandırmaları için gerekli olan VCL (eğrisel hız), VAP (ortalama yol hızı) ve VSL (düz çizgi hızı) gibi sperm hareketlilik sınıflandırma metriklerinin güvenilir bir şekilde elde edilmesi için oldukça önemlidir [65, 134].

Bu çalışma kapsamında oluşturulan Hi-LabSpermTracking veri seti, herhangi bir kesinti veya atlama olmaksızın kaydedilen 60 saniye uzunluğundaki videolardan oluşmaktadır. Videolar, 100× büyütme oranına sahip bir mikroskoba entegre edilmiş bir cep telefonu kullanılarak kaydedilmiştir. Bu yaklaşım, sperm hücrelerinin hareket yörüngelerinin ayrıntılı analizine olanak tanıyan geniş bir görüş alanı sağlamaktadır. Hi-LabSpermTracking, kapsamlı etiketleme ve benzersiz kimlik bilgileri ile hem sperm tespiti hem de sperm takibi sorunlarında literatüre önemli katkılar sağlama potansiyeline sahiptir. Bu özellikleriyle veri seti, araştırmacılara sperm hücrelerinin davranışlarını daha ayrıntılı olarak inceleme ve yeni yöntemler geliştirme fırsatı sunmaktadır.

Hi-LabSpermTracking veri kümesi kullanılarak yapılan deneyler üç farklı senaryo altında gerçekleştirilmiştir. Senaryo-I'de, literatürde veri kümesi bölümleme için yaygın olarak kullanılan rastgele eğitim-test seti oluşturma yöntemi uygulanmıştır. Bu yaklaşımda, tüm videolardan alınan kareler birleştirilmiş ve rastgele karıştırılarak eğitim ve test kümeleri oluşturulmuştur [39, 62, 65–67]. Bu yaklaşım kullanılarak YOLOv8n ile %97,77 P, %98,14 R, %97,9 F1, %98,9 mAP50 ve %77,80 mAP50:95 gibi çok yüksek değerler elde edilmiştir. Ancak bu yüksek sonuçlara rağmen, bu yöntemin gerçek dünya problemleri için uygun olmadığı tespit edilmiştir. Bunun temel nedeni, test seti için çerçevelerin rastgele seçilmesi nedeniyle, spermin hareketlilik yörüngeleriyle ilgili herhangi bir metriğin hesaplanmasının mümkün olmamasıdır. Bu durum, bölümleme yönteminin, tespit problemleri için işlevsel olduğunu ancak izleme problemleri için yetersiz olduğunu göstermektedir. Bu sorunu ortadan kaldırmak ve daha gerçekçi bir senaryo sağlamak için Senaryo-II'de farklı bir veri bölümleme yöntemi uygulanmıştır. Bu yaklaşımla hem tespit hem de takip problemleri için daha kapsamlı değerlendirmeler yapılabilmesi amaçlanmıştır.

Senaryo-II'de tespit metriklerinin yanı sıra takip metriklerini de hesaplayabilmek için veri kümesindeki videolar son 15 saniyesi test kümesine, kalan kısmı ise eğitim kümesine ayrılacak şekilde bölünmüştür. Bu yaklaşım, literatürde yaygın olarak kullanılan rastgele bölme yönteminden farklı olarak test kümesinde bir video sekansı elde edilmesine olanak sağlamıştır. Video dizileri, algılama metriklerine ek olarak eğitilen model için izleme metriklerinin hesaplanmasını sağlamıştır.Bu senaryoda, YOLOv8n mimarisi kullanılarak elde edilen tespit metrikleri sırasıyla %92,81 Kesinlik, %84,19 Duyarlılık, %87,63 F1-Skoru, %83,88 mAP50 ve %63,98 mAP50:95 olarak belirlenmiştir. İzleme problemleri için ise HOTA, MOTA ve IDF1 metrikleri sırasıyla %72,27, %77,88 ve %86,31 olarak hesaplanmıştır. Elde edilen bu sonuçlar, modelin yalnızca tespit işlemleri için değil, aynı zamanda izleme problemleri için de etkili bir şekilde kullanılabileceğini göstermektedir.

Senaryo-I ve Senaryo-II'de elde edilen performans değerleri, Tablo 1.3'te listelenen sperm tespiti ve takibi üzerine yapılan çalışmalarla karşılaştırıldığında oldukça umut verici sonuçlar ortaya koymaktadır. VISEM veri kümesi videolarının küçük bir alt kümesini kullanan bir çalışmada, Valiuskaite ve arkadaşları Senaryo-I veri bölümleme yöntemini uygulamış ve %91,77 doğruluk elde etmiştir [62]. Benzer şekilde, aynı veri bölümleme yöntemini kullanan bir başka çalışma [66] YOLOv5s modelini kullanarak %87,5 mAP50 değeri elde etmiştir. Ayrıca, Yüzkat ve arkadaşları üç farklı derin öğrenme tabanlı tespit algoritmasını bir araya getirerek kendi oluşturdukları SeSVid veri kümesinde %85,2'lik bir mAP50 elde etmiştir [39]. Bu karşılaştırmalar, Hi-LabSpermTracking veri setinin, Senaryo-I ve Senaryo-II'de kullanılan yöntemlerin literatürdeki mevcut çalışmalarla rekabet edebildiğini ve bazı durumlarda daha iyi performans gösterdiğini ortaya koymaktadır.

Zhang ve arkadaşları, VISEM-Tracking veri setinde önerdikleri izleme algoritmasıyla birlikte modifiye edilmiş bir YOLOv8s modeli kullanmışlardır [67]. Çalışmalarında, veri kümesinden toplam 29.196 kare rastgele bölümlere ayrılmış ve hem eğitim hem de test kümelerine veri artırma teknikleri uygulanmıştır. Yazarlar, test aşamasında veri kümesinden hangi videoların kullanıldığını açıkça belirtmemelerine rağmen %74,303 HOTA ve %71,767 MOTA metrikleri elde etmişlerdir. Bununla birlikte, hem eğitim hem de test verilerine veri artırımı uygulamak, sonuçların yanlı olabileceği ve gerçek dünya sorunlarını temsil etmede sınırlamalara sahip olabileceği anlamına gelmektedir. Bu da modelin genellenebilirliği konusunda kayda değer bir endişe yaratmaktadır.

Gerçek dünyada modelin daha önce hiç karşılaşmadığı yeni sperm örnekleriyle karşılaşma ihtimali yüksektir. Bu nedenle Senaryo III'te eğitim ve test videolarının

birbirinden tamamen ayrıldığı bir veri bölümleme yöntemi benimsenmiştir. Bu senaryoda spermler önerilen topluluk öğrenme yöntemleri kullanılarak tespit edilmiş ve sperm takibi DeepSORT yöntemi ile gerçekleştirilmiştir. Bu yaklaşım, modelin gerçek dünyada uygulanabilirliğini artırmak ve daha güvenilir sonuçlar elde etmek için tasarlanmıştır.

Senaryo-III'te eğitilen bireysel modeller üç farklı topluluk yöntemi ile birleştirilerek sperm tespit ve takip performansları değerlendirilmiştir. Deneyler sonucunda, önerilen MEAN ensemble yöntemi Hi-LabSpermTracking veri kümesi üzerinde %87,87 F1-Score ve %86,55 mAP50 değerleri ile en başarılı tespit performansını elde etmiştir. Benzer şekilde, MOT problemlerinde, MEAN ensemble yöntemi %66,60 HOTA, %76,42 MOTA ve %85,79 IDF1 metrikleriyle en iyi izleme performansını göstermiştir. Senaryo-III'te kullanılan aynı veri bölümleme yöntemi Thambawita ve arkadaşları tarafından VISEM-Tracking veri kümesinde kullanılmıştır [41]. Bu çalışmada, veri kümesinin sperm tespit problemlerine uygunluğunu göstermek için YOLOv5 modelinin varyantları kullanılmıştır. YOLOv5m modeli ile yapılan deneyde en başarılı sonuçlar %57,12 P, %22,79 R ve %22,16 mAP50 metrikleri ile elde edilmiştir. Çalışmanın sonuçları incelendiğinde, VISEM-Tracking veri setinin sperm tespit problemlerinde çok düşük performans sergilediği görülmektedir [41]. Bu durum, kullanılan veri setinin sperm tespiti için yeterli çeşitlilik ve zenginliği sağlamadığını göstermektedir.

Deneysel bulgularımız, YOLOv8n ve RT-DETR-L gibi derin öğrenme tabanlı dedektörlerin Hi-LabSpermTracking veri setinde sperm tespiti ve takibi için güçlü performans elde edebileceğini göstermektedir. Ancak sonuçları yorumlamak için takip performansının farklı yönleri arasında ayrım yapmak gerekir: i) tespit (her bir sperm hücresinin ne kadar iyi bulunduğu ve lokalize edildiği) ve ii) ilişkilendirme (her bir sperm hücresinin kareler arasında ne kadar tutarlı bir şekilde takip edildiği). HOTA metriği, bu bileşenleri incelemek ve izleme algoritmalarının performansındaki değiş tokuşları anlamak için birleştirici bir çerçeve sunmaktadır. HOTA metriğinin temel öngörülerinden biri, izleme performansının hem tespit hem de ilişkilendirme hatalarını eşit olarak tartması gerekliliğidir. Bu tez bağlamında: Senaryo-II birçok videoda yüksek tespit doğruluğu (Kesinlik, Duyarlılık, mAP) ortaya koymuş ancak güçlü tespite rağmen kimlik tutarlılığı (yani düşük ID değişimleri) korunmazsa genel takip metriklerinin (MOTA ve IDF1 gibi HOTA benzeri ölçümler) yine de zarar görebileceğini göstermiştir. Bu, HOTA'nın mükemmel tespitin tek başına sağlam takip için yetersiz olduğu felsefesiyle uyumludur; güçlü ilişkilendirme de aynı derecede kritiktir. Eğitim ve test setlerini tamamen farklı videolara bölen (gerçek dünya kullanımını taklit eden) Senaryo-III, bu dengenin altını daha da çizmektedir. Hem YOLOv8n hem de

RT-DETR-L ayrı ayrı farklı derecelerde tespit başarısı göstermiştir - bazı videolar daha yüksek R ancak orta düzeyde P veya tam tersi elde etmiştir. Bununla birlikte, HOTA'da olduğu gibi, genel izleme başarısı da kareler arasında kimlik tutarlılığının korunmasına bağlıdır. Sonuçlarımızda, YOLOv8n genel olarak RT-DETR-L'den daha az cevapsız tespit (daha yüksek R) ve daha az toplam kimlik değişimi üretmiştir. Buna karşılık, RT-DETR-L genel olarak daha fazla kimlik değiştirmiş (182'ye karşı 149) ve ayrıca zorlu çerçevelerde daha fazla sperm kaçırarak daha düşük ortalama R'ye katkıda bulunmuştur. Bu etkileşim, HOTA'nın dar bir şekilde tespit (örn. MOTA) veya ilişkilendirmeye (örn. IDF1) odaklanan metriklerden kaçınma motivasyonunu yansıtmaktadır. Benzer şekilde, verilerimiz tek başına tespit veya ilişkilendirmeyi vurgulamanın gerçek izleme kalitesini yanlış yansıtabileceğini göstermektedir.

HOTA, kısa vadeli kimlik değişimlerinin ötesine bakmanın ve izlemede uzun vadeli tutarlılığı incelemenin önemini vurgulamaktadır. Sonuçlarımız, belirli videoların - özellikle yüksek sperm yoğunluğuna sahip olanların - çarpışma veya sıkışmalardan sonra artan IDSW sergilediğini göstermektedir. Daha basit videolarda (örneğin, daha az sperm hücresi ve minimum örtüşme olanlar), test edilen modellerimizin çoğu ideale yakın algılama göstermiş ve bir sperm hücresi tanımlandıktan sonra istikrarlı izleme göstererek, kimlik nadiren değişmiştir. Daha kalabalık videolarda (örn. Vid-36), örtüşen yörüngeler nedeniyle kimlik değişimleri daha sık hale gelmiştir. Bununla birlikte, çoğu spermi tespit etmede üstün olan (daha yüksek R) YOLOv8n, RT-DETR-L'ye (182) kıyasla toplamda daha az kimlik değişimi (149) ile sonuçlanmıştır. RT-DETR-L tespit etmede daha muhafazakar olmasına rağmen (daha düşük R), ID değişimlerini beklendiği gibi azaltmamıştır. Bu sonuç HOTA'nın temel önermesi olan sürekli, uzun vadeli ilişkilendirme tutarlılığının güçlü tespit kadar kritik olduğunun altını çizmektedir - tek başına tespitte artan kazançlar (veya muhafazakar olma girişimleri) uzun bir video dizisi boyunca mutlaka daha az kimlik hatası anlamına gelmemektedir.

Önerilen topluluk stratejilerimiz (BEST, MEAN, WEIGHTED) önemli bir ilkeyi yansıtmaktadır: tespit çıktılarının birleştirilmesi belirli hata türlerini azaltabilir, ancak yanlış pozitiflerin şişirilmesinden veya doğru ilişkilendirmelerin gözden kaçırılmasından sakınmak için dikkatli bir şekilde yapılmalıdır. HOTA'nın tespit ve ilişkilendirmeyi dengelediği gibi, topluluk yaklaşımımız da bir modelden gelen aşırı agresif sınırlayıcı kutu tahminlerinin diğerinin kesin sınırlayıcı kutularını gölgelememesini veya tam tersini sağlamayı amaçlamaktadır. En iyi performans gösteren topluluk (MEAN) ID değişimleri sürekli olarak düşürmüş ve mAP50:95 değerini artırmıştır. Bu sinerji, tutarlı izleme kimliklerini korurken hem yanlış tespitleri (P) hem de kaçırılan tespitleri (R) azaltmanın daha güçlü bir genel

performansa götürdüğü yönündeki HOTA bakış açısını yeniden doğrulamaktadır.

4.1.4 Sonuç ve Gelecek Çalışmalar

Spermiyogram testi, geleneksel görsel değerlendirme yöntemleri ve bilgisayar destekli sperm analiz sistemleri kullanılarak gerçekleştirilmektedir. Görsel değerlendirmenin uygulama açısından kolay ve maliyet etkin olduğu düşünülse de, bu alanda nitelikli uzmanların yetiştirilmesi uzun ve maliyetli bir süreçtir. Ayrıca bu yöntemle elde edilen değerlendirme sonuçları, uzmanların deneyimlerine göre büyük ölçüde değişkenlik göstermektedir. Bir diğer değerlendirme yöntemi olan BDSA sistemleri ise yüksek maliyetleri ve karmaşık yapıları nedeniyle yaygın olarak kullanılmamaktadır. Bu bağlamda mevcut yöntemlere alternatif olarak ucuz, kolay erişilebilir ve güvenilir sistemlerin geliştirilmesi önemli bir küresel ihtiyaç haline gelmiştir.

Bu tez kapsamında, sperm hücrelerinin otomatik olarak tespit ve takibini sağlayacak derin öğrenme tabanlı sistemlerin geliştirilmesi için gerekli olan etiketli Hi-LabSpermTracking veri seti tanıtılmakta ve derin öğrenme tabanlı sperm tespit ve takip algoritmaları ile elde edilen temel sonuçlar sunulmaktadır. Mikroskoba entegre cep telefonu tabanlı görüntü alma sistemi ile elde edilen videolardan oluşan Hi-LabSpermTracking veri seti, geliştirilen modellerin herhangi bir ek gelişmiş görüntüleme sistemine ihtiyaç duymadan dünya genelinde birçok laboratuvarda geleneksel bir mikroskop ve cep telefonu ile kullanılabilmesine olanak sağlamaktadır. Ayrıca çalışmada kullanılan YOLOv8n ve RT-DETR-L modelleri üç farklı topluluk öğrenme yöntemi ile birleştirilerek, elde edilen tespitler DeepSORT algoritması ile takip edilmiştir. Çalışma sonucunda, önerilen MEAN topluluk yöntemi ile tespit ve takipte bireysel modellere kıyasla daha başarılı sonuçlar elde edilmiştir.

Gelecek çalışmalarda, çalışma kapsamında eğitilen modeller kullanılarak, zaman açısından zahmetli ve maliyetli olan etiketleme işlemlerinin sözde (pseudo) etiketleme yöntemi ile otomatize edilmesi ve Hi-LabSpermTracking'de etiketlenen video sayısının artırılması hedeflenmektedir. Bu yöntem kullanılarak etiketlenmemiş diğer videoların etiketleme işlemleri daha etkin ve sistematik bir şekilde gerçekleştirilecektir. Bu sayede daha iyi genelleme kapasitesine sahip modellerin elde edilmesi hedeflenmektedir. Oluşturulacak veri kümesi üzerinde farklı sperm tespit ve takip algoritmaları ile eğitim yapılarak elde edilen sonuçlar sunulacak ve karşılaştırılacaktır. Çalışma kapsamında gerçekleştirilen sperm tespitlerini birleştirme yöntemlerine benzer şekilde ByteTrack, BotSort ve DeepOC-Sort gibi algoritmalar kullanılarak elde edilen sperm yörüngelerinin

birleştirilmesi planlanmaktadır [140–142]. Buna ek olarak, gelecekteki çalışmalarda, VCL, VAP ve VSL gibi temel hareketlilik metriklerini elde etmek amacıyla derin öğrenme tabanlı modeller aracılığıyla çıkarılan yörüngelerin kullanılmasının yanı sıra, hastalara ait kan değerleri, kalıtsal faktörler, çevresel etmenler gibi bilgilerin de incelenmesi hedeflenmektedir. Bu bilgilerin, kısırlık üzerindeki etkilerinin analiz edilmesi, daha kapsamlı ve etkileşimli bir değerlendirme yapmayı mümkün kılacaktır.

5
SONUÇ

Bu çalışmada, spermiyogram analizlerinde geleneksel yöntemlerin sınırlamaları göz önünde bulundurularak derin öğrenme tabanlı yeni yaklaşımlar geliştirilmiş ve sperm tespiti, takibi, yörünge analizi ile morfoloji analizi için iki farklı veri seti oluşturulmuştur. Geleneksel görsel değerlendirme yöntemlerinin uzman bağımlılığı, subjektif sonuçlar ve zaman alıcı olması gibi dezavantajlarının yanı sıra, BDSA sistemlerinin yüksek maliyetleri ve sınırlı kullanım alanları, bu alanda düşük maliyetli, erişilebilir ve güvenilir otomatik sistemlerin geliştirilmesini zorunlu kılmaktadır.

Sperm morfoloji analizi alanında, literatürde yeterli çeşitliliğe ve genelleme kapasitesine sahip veri setinin mevcut olmaması nedeniyle, bu eksikliği gidermek amacıyla Hi-LabSpermMorpho veri seti geliştirilmiştir. Bu çalışmada, sperm hücrelerinin baş, boyun ve kuyruk bölümlerine ait toplamda 18 farklı morfoloji sınıfı tanımlanmış ve bu sınıflar, üç farklı Diff-Quick boyama tekniği (BesLab, Histoplus ve GBL) kullanılarak detaylı bir şekilde etiketlenmiştir. Bu kapsamda, sperm morfoloji analizi için 35 farklı derin öğrenme modeli ile detaylı hiperparametre analizleri gerçekleştirilmiş ve literatürde yaygın olarak kullanılan geleneksel özellik çıkarma yöntemlerine kıyasla derin öğrenme modellerinin daha yüksek doğruluk sağladığı gözlemlenmiştir. Elde edilen sonuçlar, derin öğrenme tabanlı yaklaşımların sperm morfoloji analizi alanında tam otomatik ve yüksek doğruluklu sistemler geliştirilmesine önemli katkılar sunduğunu ortaya koymuştur. Deneysel çalışmalar sonucunda, farklı boyama yöntemlerinin derin öğrenme modellerinin başarısı üzerindeki etkileri, sınıf bazında detaylı bir şekilde incelenmiştir. Elde edilen bulgular, Histoplus boyama tekniğinin diğer tekniklere göre sperme ait daha belirgin özellikler sunduğunu ve bu sayede daha başarılı sınıflandırma sonuçlarının elde edildiğini göstermektedir.

Sperm morfoloji analizinin yanı sıra, sperm hareketlilik analizi de erkek kısırlığı teşhisinde önemli bir değerlendirme yöntemi olarak öne çıkmaktadır. Bu bağlamda, sperm hücrelerinin otomatik olarak tespitini ve takibini gerçekleştirebilecek derin

öğrenme tabanlı sistemlerin geliştirilmesine olanak sağlayan Hi-LabSpermTracking veri seti oluşturulmuş ve nesne tespit işlemleri YOLOv8n ile RT-DETR-L modelleri kullanılarak gerçekleştirilmiştir. Sperm hücreleri öncelikle her iki model ile tespit edilmiş, ardından üç farklı topluluk yöntemi kullanılarak modellerin çıktıları birleştirilmiş ve DeepSORT algoritması ile sperm yörüngeleri çıkarılmıştır. Yapılan deneysel çalışmalar, tez kapsamında önerilen topluluk yöntemlerinin, bireysel modellere kıyasla daha yüksek tespit ve izleme performansı sağladığını göstermiştir.

Gelecek çalışmalarda, sözde (pseudo) etiketleme teknikleri kullanılarak veri setlerinin genişletilmesi ve yeni sperm tespiti ile takip algoritmaları geliştirilerek doğruluk oranlarının artırılması hedeflenmektedir. Ayrıca, ByteTrack, BotSort ve DeepOC-SORT gibi gelişmiş takip algoritmalarının kullanılmasıyla sperm hücrelerinin yörünge analizlerinin birleştirilmesi ve farklı hareketlilik metrikleri (VCL, VAP, VSL) ile sperm hareketliliğinin daha detaylı bir şekilde incelenmesi planlanmaktadır. Bunlara ek olarak, sperm hücreleri ile hastalara ait klinik verilerin (kan değerleri, genetik ve çevresel faktörler) entegre edilerek kısırlık üzerindeki etkilerinin analiz edilmesi, geliştirilen sistemlerin klinik uygulamalarda daha etkin bir şekilde kullanılmasına katkı sağlayacaktır.

Sonuç olarak, bu çalışma, sperm tespiti, takibi ve morfoloji analizi için derin öğrenme tabanlı çözümler geliştirerek, geleneksel yöntemlerin sınırlamalarını aşan düşük maliyetli, erişilebilir ve güvenilir bir otomatik analiz sistemi sunmayı amaçlamaktadır. Geliştirilen Hi-LabSpermTracking ve Hi-LabSpermMorpho veri setleri, bu alandaki gelecekteki çalışmalar için önemli bir referans kaynağı oluşturmakta olup, tıbbi görüntüleme ve biyomedikal analizlerde yapay zekâ tabanlı yaklaşımların daha yaygın bir şekilde benimsenmesine katkı sağlayacaktır.

Bu tez kapsamında oluşturulan veri setleri ve geliştirilen analiz sistemleri, literatürde mevcut olan aşağıdaki boşlukları gidermeye yönelik katkılar sunmuştur;

- Literatürde ilk kez bu kadar çeşitli ViT modelleri, en başarılı öğrenme mimarisini belirlemek amacıyla halka açık iki farklı çok sınıflı RGB sperm morfoloji veri setinde (SMIDS, HuSHeM) uygulanmıştır.
- ViT modelleri için farklı optimizasyon yöntemlerinin, öğrenme oranlarının ve veri artırma oranlarının test edilmesini içeren ayrıntılı bir hiperparametre analiz prosedürü gerçekleştirilmiştir.
- ViT modellerinin CNN modellerine göre üstünlüğü, çok sınıflı sperm morfoloji analiz sistemleri için detaylı bir şekilde gösterilmiştir.

- İki uzman tarafından DSÖ standartlarına göre etiketlenmiş 18 farklı sperm morfoloji sınıfı içeren Hi-LabSpermMorpho veriseti oluşturulmuştur. Bu düzeyde ayrıntılı etiketlenmiş veri seti mevcut literatürde bulunmamaktadır.
- 49.345 etiketlenmiş RGB görüntüsü içeren Hi-LabSpermMorpho literatürdeki en büyük veri kümelerinden biri haline gelmiştir.
- Üç farklı Diff-Quick boyama tekniği kullanılarak elde edilen sperm görüntüleri, sperm morfolojisi üzerindeki boyama yöntemlerinin etkilerinin detaylı bir şekilde analiz edilmesine olanak sağlamaktadır.
- Hi-LabSpermMorpho veri setinin büyüklüğü ve çeşitliliği, yüksek genelleme kapasitesine sahip derin öğrenme modellerinin geliştirilmesini sağlamaktadır.
- Veri setleri ticari olmayan amaçlarla kamuya sunularak, araştırmacıların algoritmalarını doğrulamalarına ve karşılaştırmalarına olanak sağlamaktadır.
- 22.865 görüntü içeren 36 telefon kamerası tabanlı video (her biri ortalama 60 saniye) ve 18.943 görüntü içeren 30 mikroskop kamerası tabanlı video olmak üzere toplam 41.808 görüntü içeren Hi-LabSpermTracking veri seti oluşturulmuştur.
- Hi-LabSpermTracking veri seti, derin öğrenme yöntemleri ve çoklu nesne izleme (MOT) algoritmaları kullanılarak sperm tespiti ve takibi için gerekli altyapıyı sunmaktadır. Bu veri seti, sadece sınırlayıcı kutu etiketleriyle sınırlı kalmayıp, aynı zamanda her bir sperme ait kimlik bilgilerini de içermektedir.
- Kaydedilen videolar, herhangi bir kesinti veya atlama olmaksızın tek bir kesintisiz 60 saniyelik klip olarak çekilmiş ve bu süre zarfında sperm yörüngeleri hakkında kapsamlı bilgiler sunmuştur. Bu eşsiz özellik, veri setini diğerlerinden (örneğin, [41]) ayırarak, literatürde sürekli ve kapsamlı sperm izleme verileri sağlayan tek veri seti olmasını sağlamaktadır.
- CNN tabanlı YOLO, dönüştürücü tabanlı RT-DETR ve DeepSORT izleme yöntemlerini kullanarak yapılan tespit, izleme ve yörünge analizi, derin öğrenme için Hi-LabSpermTracking veri setinin uygunluğunu ortaya koymuştur. Bu yaklaşım, sperm tespiti, takibi ve yörünge analizi konusunda literatürde daha önce benzeri görülmemiş bir kapsamlılık düzeyinde sonuçlar elde edilmesini sağlamıştır.
- Tez kapsamında önerilen MEAN topluluk yöntemi, hem tespit hem de takip problemlerinde bireysel modellere oranla daha üstün sonuçlar elde edilmesini sağlamıştır.

- Önerilen sperm tespit ve takip yöntemleri kullanılarak sperm hücrelerine ait yörüngeler elde edilmiştir. Bu yörüngeler, detaylı bir şekilde analiz edilerek VCL, VSL, VAP, LIN, STR ve WOB gibi hareketlilik özellikleri tamamen otomatik bir şekilde çıkarılmış ve DSÖ standartlarına göre sperm hareketliliği sınıflandırması yapılmıştır.

- Cep telefonu tabanlı bir video kayıt yöntemi kullanılarak oluşturulan Hi-LabSpermTracking veri seti, üzerinde eğitilen modellerin farklı mikroskop türleriyle evrensel uyumluluk sağlamasına olanak tanımaktadır. Bu yaklaşım, mikroskoba entegre edilmiş bir akıllı telefon aracılığıyla sperm hareketliliği analizini basitleştirirken, ek bir kamera ihtiyacını ortadan kaldırmaktadır.

- Sperm tespit ve takip veri seti olan Hi-LabSpermTracking, ticari olmayan amaçlarla kamuya sunulmuş ve araştırmacıların algoritmalarını doğrulama ve karşılaştırma imkanı sağlamıştır. Bu sayede, literatüre önemli bir katkı yapılmıştır.

KAYNAKÇA

[1] A. B. Turp ve diğ., "Infertility and surrogacy first mentioned on a 4000-year-old Assyrian clay tablet of marriage contract in Turkey," *Gynecological Endocrinology*, c. 34, no. 1, ss. 25–27, 2018.

[2] W. H. Organization ve diğ., *WHO laboratory manual for the examination and processing of human semen*. World Health Organization, 2021.

[3] W. H. Organization, *Infertility prevalence estimates, 1990–2021*. World Health Organization, 2023.

[4] A. Agarwal ve diğ., "Male oxidative stress infertility (MOSI): proposed terminology and clinical practice guidelines for management of idiopathic male infertility," *The world journal of men's health*, c. 37, no. 3, ss. 296–312, 2019.

[5] M. Freund, "Standards for the rating of human sperm morphology. A cooperative study," *International journal of fertility*, c. 11, no. 1, ss. 97–180, 1966.

[6] P. Brennan A. Silman, "Statistical methods for assessing observer variability in clinical measures.," *BMJ: British Medical Journal*, c. 304, no. 6840, s. 1491, 1992.

[7] F. Shaker, S. A. Monadjemi, J. Alirezaie A. R. Naghsh-Nilchi, "A dictionary learning approach for human sperm heads classification," *Computers in biology and medicine*, c. 91, ss. 181–190, 2017.

[8] R. P. Amann D. Waberski, "Computer-assisted sperm analysis (CASA): Capabilities and potential developments," *Theriogenology*, c. 81, no. 1, ss. 5–17, 2014.

[9] M. J. Tomlinson A. Naeem, "CASA in the medical laboratory: CASA in diagnostic andrology and assisted conception," *Reproduction, Fertility and Development*, c. 30, no. 6, ss. 850–859, 2018.

[10] J. Talarczyk-Desole, A. Berger, G. Taszarek-Hauke, J. Hauke, L. Pawelczyk P. Jedrzejczak, "Manual vs. computer-assisted sperm analysis: can CASA replace manual assessment of human semen in clinical practice?" *Ginekologia Polska*, c. 88, no. 2, ss. 56–60, 2017.

[11] R. Menkveld, "Sperm morphology assessment using strict (tygerberg) criteria," *Spermatogenesis: Methods and Protocols*, ss. 39–50, 2013.

[12] M. De Braekeleer, M. H. Nguyen, F. Morel A. Perrin, "Genetic aspects of monomorphic teratozoospermia: a review," *Journal of assisted reproduction and genetics*, c. 32, ss. 615–623, 2015.

[13] J. Rubes ve diğ., "Episodic air pollution is associated with increased DNA fragmentation in human sperm without other changes in semen quality," *Human Reproduction*, c. 20, no. 10, ss. 2776–2783, 2005.

[14] J. Auger ve diğ., "Sperm morphological defects related to environment, lifestyle and medical history of 1001 male partners of pregnant women from four European cities," *Human reproduction*, c. 16, no. 12, ss. 2710–2717, 2001.

[15] G. Bellastella ve diğ., "Dimensions of human ejaculated spermatozoa in Papanicolaou-stained seminal and swim-up smears obtained from the Integrated Semen Analysis System (ISAS®)," *Asian journal of andrology*, c. 12, no. 6, s. 871, 2010.

[16] F. Gholami Jourabi, S. Yari, P. Amiri, A. Heidarianpour H. Hashemi, "The ameliorative effects of methylene blue on testicular damage induced by cisplatin in rats," *Andrologia*, c. 53, no. 1, e13850, 2021.

[17] D. Durairajanayagam, D. Singh, A. Agarwal R. Henkel, "Causes and consequences of sperm mitochondrial dysfunction," *Andrologia*, c. 53, no. 1, e13666, 2021.

[18] J. S. Su, N. J. Farber S. C. Vij, "Pathophysiology and treatment options of varicocele: an overview," *Andrologia*, c. 53, no. 1, e13576, 2021.

[19] W. Yi, K. Park J. Paick, "Parameterized characterization of elliptic sperm heads using Fourier representation and wavelet transform," *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, IEEE, c. 2, 1998, ss. 974–977.

[20] L. Ramos, J. C. Hendriks, P. Peelen, D. D. Braat A. M. Wetzels, "Use of computerized karyometric image analysis for evaluation of human spermatozoa," *Journal of andrology*, c. 23, no. 6, ss. 882–888, 2002.

[21] V. Abbiramy A. Tamilarasi, "A comparative study on human spermatozoa images classification with artificial neural network based on FOS, GLCM and morphological features," *International Conference on Digital Image Processing and Information Technology*, Springer, 2011, ss. 220–228.

[22] E. Aksoy, T. M. Aktan, S. Duman G. Cuce, "Assessment of spermatozoa morphology under light microscopy with different histologic stains and comparison of morphometric measurements," *Int J Morphol*, c. 30, no. 4, ss. 1544–1550, 2012.

[23] J. Lammers, C. Splingart, P. Barrière, M. Jean T. Fréour, "Double-blind prospective study comparing two automated sperm analyzers versus manual semen assessment," *Journal of assisted reproduction and genetics*, c. 31, ss. 35–43, 2014.

[24] F. Ghasemian, S. A. Mirroshandel, S. Monji-Azad, M. Azarnia Z. Zahiri, "An efficient method for automatic morphological abnormality detection from human sperm images," *Computer methods and programs in biomedicine*, c. 122, no. 3, ss. 409–420, 2015.

[25] J. Riordon, C. McCallum D. Sinton, “Deep learning for the classification of human sperm,” *Computers in biology and medicine*, c. 111, s. 103 342, 2019.

[26] H. O. Ilhan G. Serbes, “Sperm morphology analysis by using the fusion of two-stage fine-tuned deep networks,” *Biomedical Signal Processing and Control*, c. 71, s. 103 246, 2022.

[27] A. Aktas, G. Serbes H. O. Ilhan, “The Performance Analysis of Convolutional Neural Networks and Vision Transformers in the Classification of Sperm Morphology,” *2023 8th International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2023, ss. 330–335.

[28] M. Yüzkat, H. O. Ilhan N. Aydin, “Multi-model CNN fusion for sperm morphology analysis,” *Computers in Biology and Medicine*, c. 137, s. 104 790, 2021.

[29] V. Chang, A. Garcia, N. Hitschfeld S. Härtel, “Gold-standard for computer-assisted morphological sperm analysis,” *Computers in biology and medicine*, c. 83, ss. 143–150, 2017.

[30] C. Cortes V. Vapnik, “Support-vector networks,” *Machine learning*, c. 20, ss. 273–297, 1995.

[31] L. Breiman, J. H. Friedman, R. A. Olshen C. J. Stone, “Classification and Regression Trees,” 1984.

[32] G. H. John P. Langley, “Estimating continuous distributions in Bayesian classifiers,” *arXiv preprint arXiv:1302.4964*, 2013.

[33] T. Cover P. Hart, “Nearest neighbor pattern classification,” *IEEE transactions on information theory*, c. 13, no. 1, ss. 21–27, 1967.

[34] S. Javadi S. A. Mirroshandel, “A novel deep learning method for automatic assessment of human sperm images,” *Computers in biology and medicine*, c. 109, ss. 182–194, 2019.

[35] K. K. Mohammed, H. M. Afify, F. Fouda, A. E. Hassanien, S. Bhattacharyya S. Vaclav, “Classification of human sperm head in microscopic images using twin support vector machine and neural network,” *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019, Volume 1*, Springer, 2020, ss. 857–871.

[36] C. McCallum ve diğ., “Deep learning-based selection of human sperm with high DNA integrity,” *Communications biology*, c. 2, no. 1, s. 250, 2019.

[37] H. O. Ilhan, I. O. Sigirci, G. Serbes N. Aydin, “A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods,” *Medical & biological engineering & computing*, c. 58, ss. 1047–1068, 2020.

[38] A. Chen ve diğ., “SVIA dataset: A new dataset of microscopic videos and images for computer-aided sperm analysis,” *Biocybernetics and Biomedical Engineering*, c. 42, no. 1, ss. 204–214, 2022.

[39] M. Yuzkat, H. O. Ilhan N. Aydin, “Detection of sperm cells by single-stage and two-stage deep object detectors,” *Biomedical Signal Processing and Control*, c. 83, s. 104 630, 2023.

[40] T. B. Haugen ve diğ., "Visem: A multimodal video dataset of human spermatozoa," *Proceedings of the 10th ACM Multimedia Systems Conference*, 2019, ss. 261–266.

[41] V. Thambawita ve diğ., "VISEM-Tracking, a human spermatozoa tracking dataset," *Scientific Data*, c. 10, no. 1, s. 260, 2023.

[42] A. Aktas, G. Serbes, M. Huner Yigit, N. Aydin, H. Uzun H. Osman Ilhan, "Hi-LabSpermMorpho: A Novel Expert-Labeled Dataset With Extensive Abnormality Classes for Deep Learning-Based Sperm Morphology Analysis," *IEEE Access*, c. 12, ss. 196 070–196 091, 2024. doı: 10.1109/ACCESS.2024.3521643.

[43] M. I. Mahali ve diğ., "A Dual Architecture Fusion and AutoEncoder for Automatic Morphological Classification of Human Sperm," *Sensors*, c. 23, no. 14, s. 6613, 2023.

[44] H. Yang ve diğ., "Multidimensional Morphological Analysis of Live Sperm Based on Multiple-target Tracking," *Computational and Structural Biotechnology Journal*, 2024.

[45] M. Tan Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *International conference on machine learning*, PMLR, 2019, ss. 6105–6114.

[46] E. Alegre, M. Biehl, N. Petkov L. Sanchez, "Assessment of acrosome state in boar spermatozoa heads using n-contours descriptor and RLVQ," *Computer methods and programs in biomedicine*, c. 111, no. 3, ss. 525–536, 2013.

[47] E. Alegre, V. González-Castro, R. Alaiz-Rodrıguez M. T. Garcıa-Ordás, "Texture and moments-based classification of the acrosome integrity of boar spermatozoa images," *Computer Methods and Programs in Biomedicine*, c. 108, no. 2, ss. 873–881, 2012.

[48] M. Y. Khachane, R. R. Manza R. J. Ramteke, "Notice of Removal: Fuzzy rule based classification of human spermatozoa," *2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)*, 2015, ss. 1–5. doı: 10.1109/EESCO.2015.7253656.

[49] O. Garcıa-Olalla, E. Alegre, L. Fernández-Robles, P. Malm E. Bengtsson, "Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors," *Computer methods and programs in biomedicine*, c. 120, no. 1, ss. 49–64, 2015.

[50] H. O. Ilhan, G. Serbes N. Aydin, "Dual tree complex wavelet transform based sperm abnormality classification," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2018, ss. 1–5.

[51] P. F. Alcantarilla, A. Bartoli A. J. Davison, "KAZE features," *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12*, Springer, 2012, ss. 214–227.

[52] H. Bay, T. Tuytelaars L. Van Gool, "Surf: Speeded up robust features," *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*, Springer, 2006, ss. 404–417.

[53] J. Matas, O. Chum, M. Urban T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, c. 22, no. 10, ss. 761–767, 2004.

[54] O. L. Tortumlu H. O. Ilhan, "The analysis of mobile platform based CNN networks in the classification of sperm morphology," *2020 Medical Technologies Congress (TIPTEKNO)*, IEEE, 2020, ss. 1–4.

[55] L. Spencer, J. Fernando, F. Akbaridoust, K. Ackermann R. Nosrati, "Ensembled Deep Learning for the Classification of Human Sperm Head Morphology," *Advanced Intelligent Systems*, c. 4, no. 10, s. 2 200 111, 2022.

[56] V. Chang, L. Heutte, C. Petitjean, S. Härtel N. Hitschfeld, "Automatic classification of human sperm head morphology," *Computers in biology and medicine*, c. 84, ss. 205–216, 2017.

[57] H. Jabbari N. Bigdeli, "New conditional generative adversarial capsule network for imbalanced classification of human sperm head images," *Neural Computing and Applications*, c. 35, no. 27, ss. 19 919–19 934, 2023.

[58] K. Dewan, T. Rai Dastidar M. Ahmad, "Estimation of sperm concentration and total motility from microscopic videos of human semen samples," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, ss. 2299–2306.

[59] V. Thambawita, P. Halvorsen, H. Hammer, M. Riegler T. B. Haugen, "Extracting temporal features into a spatial domain using autoencoders for sperm video analysis," *arXiv preprint arXiv:1911.03100*, 2019.

[60] S. A. Hicks ve diğ., "Machine learning-based analysis of sperm videos and participant data for male fertility prediction," *Scientific reports*, c. 9, no. 1, s. 16 770, 2019.

[61] V. Thambawita, P. Halvorsen, H. Hammer, M. Riegler T. B. Haugen, "Stacked dense optical flows and dropout layers to predict sperm motility and morphology," *arXiv preprint arXiv:1911.03086*, 2019.

[62] V. Valiuškaitė, V. Raudonis, R. Maskeliūnas, R. Damaševičius T. Krilavičius, "Deep learning based evaluation of spermatozoid motility for artificial insemination," *Sensors*, c. 21, no. 1, s. 72, 2020.

[63] S. Ottl, S. Amiriparian, M. Gerczuk B. W. Schuller, "motilitAI: A machine learning framework for automatic prediction of human sperm motility," *iscience*, c. 25, no. 8, 2022.

[64] S. Adinugroho A. Nakazawa, "Deep learning-based sperm motility and morphology estimation on stacked color-coded MotionFlow," *Informatics in Medicine Unlocked*, c. 45, s. 101 459, 2024.

[65] H. O. Ilhan, M. Yuzkat N. Aydin, "Sperm Motility Analysis by using Recursive Kalman Filters with the smartphone based data acquisition and reporting approach," *Expert Systems with Applications*, c. 186, s. 115 774, 2021.

[66] R. Zhu ve diğ., "YOLOv5s-SA: light-weighted and improved YOLOv5S for sperm detection," *Diagnostics*, c. 13, no. 6, s. 1100, 2023.

[67] C. Zhang, Y. Zhang, Z. Chang C. Li, "Sperm YOLOv8E-TrackEVD: A Novel Approach for Sperm Detection and Tracking," *Sensors*, c. 24, no. 11, s. 3493, 2024.

[68] H. O. Ilhan N. Aydin, "Smartphone based sperm counting-an alternative way to the visual assessment technique in sperm concentration analysis," *Multimedia Tools and Applications*, ss. 1–27, 2019.

[69] V. Chang, J. M. Saavedra, V. Castañeda, L. Sarabia, N. Hitschfeld S. Härtel, "Gold-standard and improved framework for sperm head segmentation," *Computer methods and programs in biomedicine*, c. 117, no. 2, ss. 225–237, 2014.

[70] G. L. Monte, F. Murisier, I. Piva, M. Germond R. Marci, "Focus on intracytoplasmic morphologically selected sperm injection (IMSI): a mini-review," *Asian journal of andrology*, c. 15, no. 5, s. 608, 2013.

[71] W. H. Organisation, *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction*. Cambridge university press, 1999.

[72] D. Riddell, A. Pacey K. Whittington, "Lack of compliance by UK andrology laboratories with World Health Organization recommendations for sperm morphology assessment," *Human reproduction*, c. 20, no. 12, ss. 3441–3445, 2005.

[73] H. O. Ilhan N. Aydin, "A novel data acquisition and analyzing approach to spermiogram tests," *Biomedical Signal Processing and Control*, c. 41, ss. 129–139, 2018.

[74] A. Dosovitskiy ve diğ., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

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

[76] G. Huang, Z. Liu, L. Van Der Maaten K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 4700–4708, 2017.

[77] K. He, X. Zhang, S. Ren J. Sun, "Deep residual learning for image recognition. arXiv 2015," *arXiv preprint arXiv:1512.03385*, c. 14, 2015.

[78] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, ss. 248–255.

[79] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang C. Liu, "A survey on deep transfer learning," *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, Springer, 2018, ss. 270–279.

[80] A. Vaswani ve diğ., "Attention is all you need," *Advances in neural information processing systems*, c. 30, 2017.

[81] J. Devlin, M.-W. Chang, K. Lee K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[82] A. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit L. Beyer, "How to train your vit? data, augmentation, and regularization in vision transformers," *arXiv preprint arXiv:2106.10270*, 2021.

[83] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles H. Jégou, "Training data-efficient image transformers & distillation through attention," *International conference on machine learning*, PMLR, 2021, ss. 10 347–10 357.

[84] J. Li, G. Zhu, C. Hua ve diğ., "A Systematic Collection of Medical Image Datasets for Deep Learning. arXiv 2106.12864 [preprint] http://arxiv.org/abs/2106.12864. Posted June 24, 2021," *Accessed October*, 2021.

[85] C. Shorten T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, c. 6, no. 1, s. 60, 2019.

[86] Y. LeCun, L. Bottou, Y. Bengio P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, c. 86, no. 11, ss. 2278–2324, 1998.

[87] A. Krizhevsky, I. Sutskever G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, c. 25, ss. 1097–1105, 2012.

[88] D. L. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert J. J. DiCarlo, "Performance-optimized hierarchical models predict neural responses in higher visual cortex," *Proceedings of the national academy of sciences*, c. 111, no. 23, ss. 8619–8624, 2014.

[89] K. Hermann, T. Chen S. Kornblith, "The origins and prevalence of texture bias in convolutional neural networks," *Advances in Neural Information Processing Systems*, c. 33, ss. 19 000–19 015, 2020.

[90] C. Cao ve diğ., "Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks," *Proceedings of the IEEE international conference on computer vision*, 2015, ss. 2956–2964.

[91] L. Dong, S. Xu B. Xu, "Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition," *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2018, ss. 5884–5888.

[92] L. C. Vila, C. Escolano, J. A. Fonollosa M. R. Costa-Jussa, "End-to-End Speech Translation with the Transformer.," *IberSPEECH*, 2018, ss. 60–63.

[93] M. O. Topal, A. Bas I. van Heerden, "Exploring transformers in natural language generation: Gpt, bert, and xlnet," *arXiv preprint arXiv:2102.08036*, 2021.

[94] T. Brown ve diğ., "Language models are few-shot learners," *Advances in neural information processing systems*, c. 33, ss. 1877–1901, 2020.

[95] K. He ve diğ., "Transformers in medical image analysis," *Intelligent Medicine*, c. 3, no. 1, ss. 59–78, 2023.

[96] S. Tuli, I. Dasgupta, E. Grant T. L. Griffiths, "Are convolutional neural networks or transformers more like human vision?" *arXiv preprint arXiv:2105.07197*, 2021.

[97] C. Liu Q. Yin, "Automatic diagnosis of covid-19 using a tailored transformer-like network," *Journal of Physics: Conference Series*, IOP Publishing, c. 2010, 2021, s. 012 175.

[98] J. C. Than ve diğ., "Preliminary study on patch sizes in vision transformers (vit) for covid-19 and diseased lungs classification," *2021 IEEE National Biomedical Engineering Conference (NBEC)*, IEEE, 2021, ss. 146–150.

[99] A. Chen ve diğ., "A Comprehensive Comparative Study of Deep Learning Methods for Noisy Sperm Image Classification: from Convolutional Neural Network to Visual Transformer," *Intelligent Medicine*, 2023.

[100] C. Sun, A. Shrivastava, S. Singh A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," *Proceedings of the IEEE international conference on computer vision*, 2017, ss. 843–852.

[101] G. Hinton, O. Vinyals J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[102] L. Wei, A. Xiao, L. Xie, X. Zhang, X. Chen Q. Tian, "Circumventing outliers of autoaugment with knowledge distillation," *European Conference on Computer Vision*, Springer, 2020, ss. 608–625.

[103] H. Bao, L. Dong, S. Piao F. Wei, "Beit: Bert pre-training of image transformers," *arXiv preprint arXiv:2106.08254*, 2021.

[104] A. Ramesh ve diğ., "Zero-shot text-to-image generation," *International Conference on Machine Learning*, PMLR, 2021, ss. 8821–8831.

[105] H. O. Ilhan, G. Serbes N. Aydin, "Automated sperm morphology analysis approach using a directional masking technique," *Computers in Biology and Medicine*, s. 103 845, 2020.

[106] H. Ilhan, G. Serbes N. Aydin, "Automatic directional masking technique for better sperm morphology segmentation and classification analysis," *Electronics Letters*, c. 55, no. 5, ss. 256–258, 2019.

[107] H. O. Ilhan, I. O. Sigirci, G. Serbes N. Aydin, "The effect of nonlinear wavelet transform based de-noising in sperm abnormality classification," *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2018, ss. 658–661.

[108] I. Goodfellow ve diğ., "Generative adversarial networks," *Communications of the ACM*, c. 63, no. 11, ss. 139–144, 2020.

[109] A. Bijar, M. Mikaeili, R. Khayati ve diğ., "Fully automatic identification and discrimination of sperm's parts in microscopic images of stained human semen smear," *Journal of Biomedical Science and Engineering*, c. 5, no. 2012, 2012.

[110] D. Rousso, A. Kourtis, G. Mavromatidis, F. Gkoutzioulis, G. Makedos D. Panidis, "Pyriform head: a frequent but little-studied morphological abnormality of sperm," *Archives of andrology*, c. 48, no. 4, ss. 267–272, 2002.

[111] D. F. KATZ, J. W. OVERSTREET, S. J. SAMUELS, P. W. NISWANDER, T. D. BLOOM E. L. LEWIS, "Morphometric analysis of spermatozoa in the assessment of human male fertility," *Journal of Andrology*, c. 7, no. 4, ss. 203–210, 1986.

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

[113] S. Xie, R. Girshick, P. Dollár, Z. Tu K. He, "Aggregated residual transformations for deep neural networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 1492–1500, 2017.

[114] S. Zagoruyko N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[115] C. Szegedy ve diğ., "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 1–9, 2015.

[116] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 2818–2826, 2016.

[117] M. Tan Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *arXiv preprint arXiv:2104.00298*, 2021.

[118] N. Ma, X. Zhang, H.-T. Zheng J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," *Proceedings of the European Conference on Computer Vision (ECCV)*, ss. 116–131, 2018.

[119] A. Howard ve diğ., "Searching for mobilenetv3," *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, ss. 1314–1324.

[120] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, ss. 4510–4520, 2018.

[121] F. N. Iandola, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.

[122] H. İLHAN, A. Elbir, G. SERBES N. Aydin, "The evaluation of nature-inspired optimization techniques for contrast enhancement in images: A novel software tool," *Traitement du Signal*, c. 40, no. 4, 2023.

[123] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, ss. 248–255. doı: 10.1109/CVPR.2009.5206848.

[124] J. Karibe ve diğ., "Monozygotic adult twins with nonmosaic Klinefelter syndrome with different results of sperm retrieval," *Andrologia*, c. 54, no. 1, e14266, 2022.

[125] R. Cannarella, M. Gül, A. Rambhatla A. Agarwal, "Temporal decline of sperm concentration: role of endocrine disruptors," *Endocrine*, c. 79, no. 1, ss. 1–16, 2023.

[126] A. Sharma, S. Minhas, W. S. Dhillo C. N. Jayasena, "Male infertility due to testicular disorders," *The Journal of Clinical Endocrinology & Metabolism*, c. 106, no. 2, e442–e459, 2021.

[127] A. Agarwal ve diğ., "Male infertility," *The Lancet*, c. 397, no. 10271, ss. 319–333, 2021.

[128] Y. Zhao ve diğ., "Detrs beat yolos on real-time object detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, ss. 16 965–16 974.

[129] B. Li, Q. Meng, X. Li, Z. Wang, X. Liu S. Kong, "Enhancing YOLOv8's Performance in Complex Traffic Scenarios: Optimization Design for Handling Long-Distance Dependencies and Complex Feature Relationships," *Electronics*, c. 13, no. 22, s. 4411, 2024.

[130] V. Eddye, M. J. Iorpenda V. Willert, "Evaluation of YOLOv8n as a Suitable Tool for UAV Navigation," *2024 22nd International Conference on Research and Education in Mechatronics (REM)*, IEEE, 2024, ss. 262–267.

[131] N. Wojke, A. Bewley D. Paulus, "Simple online and realtime tracking with a deep association metric," *2017 IEEE international conference on image processing (ICIP)*, IEEE, 2017, ss. 3645–3649.

[132] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[133] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, c. 2, no. 1-2, ss. 83–97, 1955.

[134] Ł. Witkowski, "A computer system for a human semen quality assessment," *Biocybernetics and Biomedical Engineering*, c. 33, no. 3, ss. 179–186, 2013.

[135] L. St, S. Wold ve diğ., "Analysis of variance (ANOVA)," *Chemometrics and intelligent laboratory systems*, c. 6, no. 4, ss. 259–272, 1989.

[136] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017.

[137] J. Luiten ve diğ., "Hota: A higher order metric for evaluating multi-object tracking," *International journal of computer vision*, c. 129, ss. 548–578, 2021.

[138] W. Lu ve diğ., "A cnn-transformer hybrid model based on cswin transformer for uav image object detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, c. 16, ss. 1211–1231, 2023.

[139] W. F. Hendria, Q. T. Phan, F. Adzaka C. Jeong, "Combining transformer and CNN for object detection in UAV imagery," *ICT Express*, c. 9, no. 2, ss. 258–263, 2023.

[140] N. Aharon, R. Orfaig B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022.

[141] G. Maggiolino, A. Ahmad, J. Cao K. Kitani, "Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification," *2023 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2023, ss. 3025–3029.

[142] Y. Zhang ve diğ., “Bytetrack: Multi-object tracking by associating every detection box,” *European conference on computer vision*, Springer, 2022, ss. 1–21.

TEZDEN ÜRETİLMİŞ YAYINLAR

Makale

1. Aktas, A., Serbes, G., Yigit, M. H., Aydin, N., Uzun, H., & Ilhan, H. O. (2024). Hi-LabSpermMorpho: A Novel Expert-Labeled Dataset with Extensive Abnormality Classes for Deep Learning-Based Sperm Morphology Analysis. IEEE Access.

Konferans Bildirisi

1. Aktas, A., Serbes, G., & Ilhan, H. O. (2023, September). The Performance Analysis of Convolutional Neural Networks and Vision Transformers in the Classification of Sperm Morphology. In 2023 8th International Conference on Computer Science and Engineering (UBMK) (pp. 330-335). IEEE.

Proje

1. Sperm Morfoloji Analizi Için Derin Öğrenme Tabanlı Bilgisayar Destekli Sistem Tasarımı- Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) - ARDEB 1001- Proje No: 122E164

2. Cep Telefonu Tabanlı Sperm Hareket Analiz ve Raporlama Sistemi- Türkiye Sağlık Enstitüleri Başkanlığı (TÜSEB)- B Grubu AR-GE Projesi - Proje No: 27698