

Sperm Morfolojisinin Sınıflandırılmasında Evrimsel Sinir Ağ ve Görü Dönüştürücü Modellerin Performans Analizi

The Performance Analysis of Convolutional Neural Networks and Vision Transformers in the Classification of Sperm Morphology

Abdulsamet Aktas
Computer Engineering Department
Yıldız Technical University
İstanbul, Türkiye
abdulsamet.aktas@std.yildiz.edu.tr

Gorkem Serbes
Biomedical Engineering Department
Yıldız Technical University
İstanbul, Türkiye
gserbes@yildiz.edu.tr

Hamza Osman İlhan
Computer Engineering Department
Yıldız Technical University
İstanbul, Türkiye
hoilhan@yildiz.edu.tr

Öz—Dünya genelinde kısırlık problemi gün geçtikçe daha büyük bir sorun haline gelmektedir ve problemin görülme sıklığının yaklaşık yarısında erkek faktörü etkilidir. Son zamanlarda, sperm kalitesini ölçmek için geliştirilen bilgisayar destekli sperm analiz sistemleri ile birlikte yapay zeka uygulamaları da kullanılmaya başlanmıştır. Özellikle evrimsel sinir ağları (ESA), sperm hücrelerinin sınıflandırılmasında kullanılan en yaygın yöntemlerden biridir. Son yıllarda popüler hale gelen ve görüntü sınıflandırma problemlerinde kullanılmaya başlanan bir diğer yöntem ise görü dönüştürücü (GD) (vision transformer (ViT)) modelleridir. Önerilen çalışma kapsamında erişime açık olan insan sperm hücre görüntüleri veri setlerinden HuSHeM, SMIDS ve SCIAN veri setleri üzerinde ESA ve GD mimarilerinin karşılaştırmalı olarak performans analizleri yapılmıştır. Geleneksel 5 farklı ESA mimarisi ve 3 farklı varyantta GD modeli olmak üzere toplam 8 adet model belirtilen veri setlerinin sınıflandırma performansları üzerinden kıyaslanmıştır. 5 kat çapraz doğrulama ve veri artırımı yapılarak modeller eğitilmiş ve elde edilen sonuçlar sunulmuştur. Sonuçlar t-test yöntemi ile kıyaslanarak performans analizi yapılmıştır. Her bir model 3 farklı veri seti üzerinde 7 farklı model ile kıyaslanarak toplam 21 karşılaştırma işlemi yapılmıştır. Yapılan karşılaştırmalar sonucunda ViT-L16 modeli ile 12 Kazanma, 9 Beraberlik ve 0 yenilgi alınmıştır. Kazanma oranları karşılaştırıldığında en yakın modelden yaklaşık %38 daha fazla kazanma oranına sahiptir.

Anahtar Sözcükler—Görü Dönüştürücü, Sperm Görüntü Sınıflandırma, Evrimsel Sinir Ağları

Abstract—The problem of infertility around the world is becoming more severe day by day. About half of the problem incidence is male-related. Computer Aided Sperm Analysis systems, accompanied by the usage of artificial intelligence, have been recently developed to measure the sperm quality. In particular, convolutional neural networks (CNNs) are one of the most widely used method for classifying sperm cells. Another method that has become popular in recent years and has started to be used in image classification problems is the vision transformer (ViT) models. In the proposed study, a comparative performance analysis of CNN and ViT architectures on open source HuSHeM, SMIDS and SCIAN human sperm cell image datasets was

performed. The results obtained from 8 models, including 5 different traditional CNN architectures and 3 different variants of ViT models, were compared with each other. After 5-fold cross validation and data augmentation, the models were trained and the results were obtained. The results obtained were validated with t-test and the performance analysis was performed. Each model was compared with 7 different models on 3 different datasets for a total of 21 comparisons. As a result of the comparisons, 12 wins, 9 draws and 0 defeats were obtained with the ViT-L16 model. When comparing win rates, it has about %38 more win rates than its closest model.

Keywords—Vision Transformer, Sperm Image Classification, Convolutional Neural Network

I. GİRİŞ

Üreme, tarih boyunca insanlık için en önemli konulardan biri olmuştur. Nesiller boyunca insanoğlu bilgi ve birikimini bir sonraki kuşaklara aktararak günümüze kadar gelmeyi başarmıştır. Fakat bazı çiftler fizyolojik, psikolojik ve/veya sağlıksal nedenlerden dolayı çocuk sahibi olamamaktadır. Dünya sağlık örgütü, en az bir yıl boyunca herhangi bir korunma yöntemi uygulamadan birliktelik olmasına rağmen çocuk sahibi olamama durumunu kısırlık olarak adlandırmaktadır [1]. Dünya genelinde yaklaşık %15 çiftin kısırlık problemi yaşadığı tahmin edilmektedir. Problemin %50'sinde erkek kaynaklı faktörlerin etkili olduğu düşünülmektedir [2].

Erkeklerde sperm kalitesi, kısırlık durumunu belirleyen en önemli faktörler arasında bulunmaktadır. Sperm kalitesini ölçmek için geliştirilen sistemler genel olarak Bilgisayar Destekli Sperm Analizi (BDSA) olarak adlandırılmaktadır. Çok etkili ve önemli olan bu sistemler klinik uygulamalarda oldukça yaygın bir şekilde kullanılmaktadır. Özellikle geleneksel yöntemlere göre çok daha başarılı sonuçlar vermektedir. Geleneksel yöntemlerde androloji uzmanları mikroskoba

yerleştirilmiş sperm örneğini çıplak gözle muayene ederek sperm kalitesini ölçmektedir. Bu yöntemde uzmanların tecrübesi oldukça önemlidir. Uzmana göre sonuçlar değişmekte ve uzmanlar arası tutarsızlıklar ortaya çıkmaktadır. BDSA sistemlerinde ise sonuçlar geleneksel yöntemlere göre daha tutarlı ve objektiftir. Fakat bu yöntem oldukça pahalıdır. Bu nedenle yaygın olarak geniş bir coğrafyada kullanılamamaktadır.

Derin öğrenme yöntemlerindeki gelişmeler ile çok farklı alanlarda görüntü işleme uygulamaları geliştirilmiştir. Bu yöntemler, sperm hücrelerinin sınıflandırılması, tespit ve takip edilmesi gibi alanlarda oldukça yaygın bir şekilde kullanılmaktadır. Literatürde bulunan çalışmalar incelendiğinde sperm sınıflandırma alanında genellikle geleneksel derin öğrenme yöntemlerinin kullanıldığı gözlemlenmiştir. VGG-16, VGG-19 [3], DenseNet [4], ResNet [5] gibi evrimsel sinir ağı (ESA) mimarileri bu yöntemlerin başında gelmektedir. Literatürde bulunan çalışmalar incelendiğinde özellikle son yıllarda popüler olan Görü Dönüştürücü (GD) [6] mimarilerinin, sperm sınıflandırma çalışmalarında kullanımı ise yenidir. Bu açığı kapatmayı, ve ESA mimarileri ile GD mimarilerinin sperm sınıflandırmadaki performanslarını kıyaslamayı amaçlayan çalışmamızda; literatürde bulunan erişime açık veri setleri HuSHeM [7], SCIANMorpho [8] ve SMIDS [9] üzerinde transfer öğrenme yöntemi kullanılarak geleneksel ESA ve GD mimarilerinin performans analizi yapılmıştır. Çalışmanın ilerleyen kısımlarında; literatürde bulunan çalışmalar Bölüm II'de, kullanılan yöntemler ve veri setleri Bölüm III'te, çalışmaya ait sonuçlar ise Bölüm IV'te sunulmuştur.

II. İLGİLİ ÇALIŞMALAR

Jason Riordon ve arkadaşları [10] yaptıkları çalışmada insan sperm hücrelerine ait görüntüler üzerinde sınıflandırma işlemi gerçekleştirmişlerdir. Veri seti olarak SCIAN ve HuSHeM kullanılarak, derin öğrenme yöntemi olarak ise VGG16 ağı seçilmiştir. 100 epoch eğitim işleminden sonra model ince-ayar yapılarak toplam 200 epoch daha eğitilmiştir. 3 seriden oluşan 5 kat çapraz doğrulama yapılmıştır. Her bir seride kullanılan örnekler diğer serilerle karıştırılmamış ve bu sayede 5 kat çapraz doğrulamanın her bir serisi için alınan sonuçların aynı örneklerden oluşması sağlanmıştır. Yapılan çalışma sonucunda HuSHeM veri seti için %94,1 ve SCIAN veri seti için %62 doğru pozitif oranı elde edilmiştir.

Lindsay Spencer ve arkadaşları [11] yaptıkları çalışmada sperm hücresi baş kısımlarını şekilsel olarak sınıflandırmak için topluluk derin öğrenme yöntemlerini kullanmışlardır. Veri seti olarak HuSHeM ve SCIAN kullanılırken, derin öğrenme modelleri olarak ise VGG16, VGG19, modifiye edilmiş ResNet-34 ve DenseNet-161 mimarilerinden faydalanılmıştır. Bu ana modellerden elde edilen öznetelik vektörleri birleştirilerek meta-sınıflandırıcıya verilmiştir. Meta sınıflandırıcının sonucuna göre nihai tahmin yapılmıştır. Eğitim işlemi sonucunda test verileri üzerinde yapılan deneylerde HuSHeM, SCIAN ve SCIAN-full veri setleri için

sırasıyla %98,2, %63,3 ve %81,2 F1-skor değerleri elde edilmiştir.

Omer Lutfu Tortumlu ve arkadaşları [12] yaptıkları çalışmada sperm morfoloji sınıflandırmasında mobil platform tabanlı ESA mimarilerini analiz etmişlerdir. Veri seti olarak HuSHeM, SMIDS ve SCIAN-Morpho veri setleri kullanılmıştır. ESA mimarisi olarak MobileNet V1 ve MobileNet V2 modelleri kullanılmıştır. Farklı veri çoğullama miktarı, dönem(epoch) sayısı ve eğitim zamanı gibi değerlere göre modeller birbirleri ile kıyaslanmıştır. HuSHeM veri seti için en başarılı sonuç MobileNet-V2 mimarisi ile %77, SMIDS ve SCIAN-Morpho veri setleri için en başarılı sonuç MobileNet-V1 mimarisi ile sırasıyla %88 ve %67 doğruluk olarak elde edilmiştir.

Mecit Yüzkat ve arkadaşları [13] yaptıkları çalışmada 6 farklı ESA mimarisi oluşturularak topluluk öğrenme yöntemi ile sperm sınıflandırma işlemi gerçekleştirmişlerdir. Çoğunluğa dayalı (hard) ve olasılığa dayalı (soft) oylama yöntemleri kullanılarak modeller birleştirilmiştir. Veri seti olarak HuSHeM, SCIAN ve SMIDS kullanılmıştır. Veri setleri 5 kat çapraz doğrulama ile ayrılmıştır. Deneysel çalışmalar sonucunda SMIDS, HuSHeM ve SCIAN-Morpho veri setleri için sırasıyla %90,73, %85,18 ve %71,91 doğruluk değerleri elde edilmiştir.

Hamza Osman Ilhan ve arkadaşları [14] yaptıkları çalışmada derin öğrenme yöntemlerini kullanarak sperm sınıflandırma işlemi gerçekleştirmişlerdir. Önerdikleri yöntemle modeller iki aşamalı olarak ince ayar yapılarak birleştirilmiştir. Derin öğrenme modeli olarak VGG16 ve GoogleNet mimarileri kullanılmıştır. Modellere öncelikle SMIDS veri seti kullanılarak ince ayar yapılmıştır. Ardından aynı model HuSHeM ve SCIAN-Morpho veri setleri ile ayrı ayrı eğitilerek performans artışı gözlemlenmiştir. Elde edilen model ile SMIDS veri setinde %90,87 doğruluk değeri elde edilmiştir. İki aşamalı ince ayar yöntemi uygulanmadan önce HuSHeM ve SCIAN-Morpho veri setleri için sırasıyla %88,89 ve %72,08 doğruluk değerleri elde edilirken, ilgili çalışmada önerilen iki aşamalı yöntemle doğruluk sırasıyla %92,1 ve %73,2 değerlerine çıkarılmıştır.

Imran Iqbal ve arkadaşları yaptıkları çalışmada insan sperm hücresi baş kısımlarını ESA ile sınıflandırmışlardır. Veri seti olarak SCIAN ve HuSHeM veri setleri, veri artırımı yöntemleri uygulanarak kullanılmıştır. ESA olarak toplam 53 evrim katmanından oluşan mimari kullanılmıştır. SCIAN veri seti için önerilen yöntemle %77 doğruluk, %88 hassasiyet ve %64 kesinlik değerleri elde edilirken, HuSHeM veri seti için ise %95,7 doğruluk, %95,5 hassasiyet ve %96,1 kesinlik değerlerine ulaşılmıştır.

Ahmad A. Mashaal ve arkadaşları [15] yaptıkları çalışmada ResNet-50 mimarisi kullanarak insan sperm hücrelerini sınıflandırmışlardır. Veri seti olarak McCallum [16] tarafından oluşturulmuş açık kaynak olarak kullanılabilen veri seti seçilmiştir. Veri seti, androloji uzmanları tarafından normal ve anormal olarak etiketlenmiş toplam 1200 adet görüntüden

oluşmaktadır. Medyan filtre kullanılarak görüntülerdeki tuz ve biber (İng. salt and pepper) gürültüsü giderilmiştir. Derin öğrenme modeli olarak ResNet-50 mimarisi kullanılmıştır. Önerilen modellerle test verileri üzerinde %96,66 doğruluk, %97,06 hassaslık, %98,21 kesinlik ve %97,63 F1-skor başarı değerleri elde edilmiştir.

Ao Chen ve arkadaşları [17] yaptıkları çalışmada gürültülü sperm görüntülerini sınıflandırmayı amaçlamışlardır. Veri seti olarak açık kaynak olarak kullanılan SVIA [18] veri setinin alt sınıfı olan subset-C kullanılmıştır. Subset-C, 121401 sperm görüntüsü ve 4479 gürültülü görüntüden oluşmaktadır. Sınıf bazında veri setinin dengeli olması için 4479 gürültülü görüntüden rastgele olarak seçilmiş 5058 tek sperm görüntüsü kullanılmıştır. Seçilen görüntüler rastgele bir şekilde %60 eğitim, %20 test ve %20 doğrulama olarak ayrılmıştır. Ardından bir çok derin öğrenme modeli eğitilerek performans karşılaştırılması yapılmıştır. Deneysel çalışmalar sonucunda ESA modellerinin GD modellerine göre çok daha iyi başarı metriklerine sahip olduğu gözlemlenmiştir. En stabil modelin VGG-16 modeli olduğu, doğrulama etkisinin en yüksek Inception-V3 modelinde olduğu, en hızlı çalışma süresinin AlexNet modeline ait olduğu ve en küçük modelin ShuffleNet-V2 modeli olduğu gözlemlenmiştir. T2T-t-ViT-24 modelinin en düşük doğruluk değeri ve en uzun çalışma süresine sahip olduğu raporlanmıştır.

TABLO I
MODELLERE AIT PARAMETRE SAYILARI VE EĞİTİM SÜRELERİ

Model	Parametre Sayısı	HuSHeM-40x	
		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

*İlgili veri setinde 1 epoch eğitim için harcanan süre

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

TABLO II
MODELLERE AIT PARAMETRE SAYILARI VE EĞİTİM SÜRELERİ (DEVAM)

Model	SCIAN-10x		SMIDS-10x	
	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

Literatürde bulunan çalışmalar incelendiğinde modellerin birbirleri ile kıyaslanmasında HuSHeM, SCIAN ve SMIDS veri setlerinin yaygın olarak kullanıldığı gözlemlenmiştir. Sınıflandırma çalışmalarında yaygın olarak kullanılan derin

öğrenme modellerinin ise VGG16, VGG19, ResNet50 ve DenseNet modelleri olduğu belirlenmiştir. Derin öğrenme ve GD modelleri sperm morfoloji analizi kapsamında Ao Chen ve arkadaşları tarafından kıyaslanmıştır [17]. Fakat bu kıyaslama yalnızca tek bir veri setinin alt seti olan subset-C üzerinden yapılmış ve yalnızca sperm ve gürültülü veri olmak üzere 2 sınıflı bir problem ekseninde değerlendirilmiştir. Mimarilerin birbirlerine olan üstünlükleri, sperm morfoloji analizi kapsamında kıyaslanmamıştır. Bu nedenle önerilen bu çalışmada sperm morfoloji analizi temelinde oluşturulmuş 3 farklı veri seti üzerinde farklı ESA ve GD mimarileri eğitilerek birbirlerine olan üstünlükleri tartışılmıştır.

III. MATERYAL VE YÖNTEM

A. Veri setleri

Literatürde bulunan çalışmalar incelendiğinde özellikle HuSHeM, SCIAN-Morpho ve SMIDS erişime açık veri setlerinin kıyaslama çalışmalarında oldukça fazla kullanıldığı gözlemlenmiştir. Bu nedenle ESA ve GD modellerinin sınıflandırma problemi kapsamında performans kıyaslanması bu üç veri seti üzerinde yapılmıştır.

1) *HuSHeM*: F. Shaker ve arkadaşları [7] tarafından oluşturulan veri seti 216 adet insan sperm baş görüntüsünden oluşmaktadır. Normal (54), konik (53), piriform (57) ve amorf (52) olmak üzere sağlıklı ve 3 adet morfolojik bozuntu çeşidine sahip 4 adet sınıf mevcuttur. Görüntüler RGB formatında ve 131×131 çözünürlüğündedir. Veri seti gürültü, çoklu sperm kafası ve karışmış sperm kuyrukları içermektedir.

2) *SCIAN*: V. Chang ve arkadaşları [8] tarafından oluşturulan veri setinde toplam 1132 adet insan sperm görüntüsü bulunmaktadır. Veri setinde normal (100), konik (228), piriform (76), amorf (656) ve ufak baş (72) olmak üzere sağlıklı ve 4 adet morfolojik bozuntu içeren toplam 5 adet sınıf bulunmaktadır. Görüntüler gri ölçekli olarak 35×35 piksel çözünürlüğünde elde kaydedilmiştir. Örnek-Sınıf dağılımının dengesiz ve görüntü çözünürlüğünün çok düşük olması literatürdeki çalışmalarda bu veri seti üzerinde düşük başarımların elde edilmesine neden olmuştur.

3) *SMIDS*: H. O. Ilhan ve arkadaşları [9] tarafından oluşturulan veri setinde toplam 3000 adet insan sperm hücrelerine ait görüntü bulunmaktadır. Veri seti, normal (1021), abnormal(1005) ve non-sperm(974) olmak üzere toplam 3 sınıftan oluşmaktadır. Sınıflar dengeli bir şekilde dağılmış ve görüntüler RGB formatında çeşitli çözünürlük değerlerine sahiptir.

B. Veri Ön İşleme

Veri setleri incelendiğinde özellikle HuSHeM veri setinde çok az sayıda örnek olduğu görülmektedir. Eğitim işlemi sırasında modellerin veriyi ezberlemesinin önüne geçmek için çeşitli veri artırımı yöntemleri uygulanmıştır. Döndürme, yatay ve dikey çevirme gibi yöntemler uygulanarak veri artırımı sağlanmıştır. Veri artırımı yapılmadan önce veri setleri 5 katlı çapraz doğrulama yapılarak eğitim ve test

örnekleri olarak ayrılmıştır. Veri artırımı işlemi sadece eğitim örneklerinde yapılarak test örneklerinde herhangi bir değişiklik yapılmamıştır. Veri artırımı işlemi sonrasında HuSHeM veri setinde bulunan eğitim örnekleri 40, SMIDS ve SCIAN veri setlerinde bulunan eğitim örnekleri 10 katına çıkarılmıştır.

C. Modeller

ESA mimarileri ilk kullanımından itibaren özellikle görüntü sınıflandırma alanında oldukça popüler hale gelmiştir. Ekran kartı alanında yapılan teknolojik gelişmeler sayesinde çok daha karmaşık ve hesap gücü gerektiren modeller tasarlanarak sinir ağı mimarilerinde daha detaylı tasarımlar gerçekleştirilebilmiştir. Imagenet [19] veri seti kullanılarak yapılan yarışmalarda ESA mimarilerine ilgi daha çok artarak, ESA mimarileri oldukça farklı alanlarda kullanılmaya başlanmıştır. ImageNet üzerinde eğitilen modellere ait ağırlık değerleri farklı veri setleri eğitiminde başlangıç ağırlık değerleri olarak kullanılabilir. Transfer öğrenme olarak adlandırılan bu yöntem sayesinde eğitim süresi dramatik bir şekilde kısalmaktadır [20]. Çalışma kapsamında, ImageNet üzerinde ön-öğrenme yapılmış ESA ve GD modelleri alınarak bu modellere 3 veri seti için ayrı ayrı ince-ayar işlemi yapılmış ve sınıflandırma sonuçları sunulmuştur.

ESA mimarilerinin yanı sıra GD mimarileri, doğal dil işlemedeki dikkat tabanlı modellerden [21] esinlenilerek geliştirilen, görüntü sınıflandırma alanında kullanılan yeni bir derin öğrenme yaklaşımıdır. Özellikle çok boyutlu veri setlerinde ESA modellerine göre daha iyi sonuçlar verdiği gözlemlenmiştir [6]. Fakat sperm veri setleri üzerinde genel performans değerlendirilmesi literatürde yer almamaktadır. ESA mimarilerinde olduğu gibi GD mimarilerinde de ImageNet üzerinde eğitilen modellere ait ağırlıklar kullanılarak sperm veri setleri ile ince-ayar işlemi yapılmış ve sınıflandırma gerçekleştirilmiştir. Çalışmada kullanılan ESA ve GD modelleri aşağıdaki başlıklarda açıklanmıştır.

1) *VGG16*: VGG16 ağı [3] 13 evrişim ve 3 tam bağlı katman olmak üzere 16 katmandan oluşan ESA mimarisidir. 3×3 boyutunda küçük evrişim filtreleri içeren bir mimariye sahip olan VGG16 mimarisi kendi zamanındaki en derin yapay sinir ağlarından biri olmayı başarmıştır.

2) *VGG19*: VGG19 ağı [3] VGG16 ağı ile benzer yapıdadır. Aralarındaki fark katman sayılarının farklı olmasıdır.

3) *DenseNet-121*: DenseNet mimarisinde her bir katman diğer bütün katmanlar ile ileri beslemeli bir şekilde birbirine bağlıdır. Geleneksel ESA mimarilerinde L adet katman L adet bağlantıya sahipken DenseNet modelinde $L(L+1)/2$ direkt bağlantı mevcuttur. Bu sayede öznitelikler ağ sonuna kadar verimli bir şekilde kullanılarak kaybolan gradyan problemi hafifletilmiştir.

4) *ResNet50*: ESA mimarilerindeki artan katman sayısına rağmen test verilerindeki başarı değeri düşmekteydi. Bunun önüne geçmek için her iki önceki katmanda elde edilen vektörler ileriye doğru birikimli şekilde verilerek öznitelik

kayı azaltılarak ResNet mimarileri oluşturuldu. Bu sayede elde edilen çok katmanlı kalıntılı ağlar, aynı sayıdaki düz ağlara göre çok daha iyi sonuçlar vermektedir [5].

TABLE III
T-TEST SONUCU ELDE EDİLEN SKOR DEĞERLERİ

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

5) *Efficient-B7*: ESA mimarilerindeki katmanların genişlik, derinlik ve görüntü çözünürlüğü özelliklerini daha verimli bir şekilde kullanmak için geliştirilmiş bir ağıdır. Genişlik, derinlik ve çözünürlük gibi ağ boyutlarını keyfi bir şekilde değiştiren geleneksel yöntemlerin aksine, her boyutu sabit bir kat sayı ile eşit bir şekilde ölçeklendirmektedir [22].

6) *ViT-L16*: ViT-L16 modeli direkt olarak BERT [23] modelinden türetilmiştir. İsmindeki L16, giriş yaması 16×16 boyutuna sahip "Large" varyantı anlamına gelmektedir. Giriş 16×16 yamalara (patches) bölünerek düzleştirilmektedir. Ardından eğitilebilir doğrusal projeksiyona verilerek yamalar ile konum bilgileri birleştirilmektedir. Toplam 24 katmandan oluşan ViT-L16 modeli 1024 gizli boyut ve 16 dikkat kafasından oluşmaktadır. 4096 çok katmanlı algılayıcı boyutuna sahip olan modelin toplam parametre sayısı 304 milyondur [24].

7) *ViT-S16*: ViT-S16 modeli diğer modellere göre daha küçük modeldir. Toplam 8 katmandan oluşmaktadır. 1000 gizli boyutu bulunan modelin 10 adet dikkat kafası bulunmaktadır. Çok katmanlı algılayıcı boyutu 1024, toplam parametre sayısı 21 milyondur [24].

8) *ViT-b16*: ViT-b16 modeli diğer GD modelleriyle benzer şekilde çalışmaktadır. Toplam 12 katmandan oluşmakta ve 768 gizli boyutu bulunan modelin 12 adet dikkat kafası bulunmaktadır. Çok katmanlı algılayıcı boyutu 3072, toplam parametre sayısı 86 milyondur [24].

TABLE IV
HUSHEM-40X VERİ SETİNE AİT DOĞRULUK DEĞERLERİ

	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

5 katlı çapraz doğrulama yapılan veri setleri üzerinde modeller ayrı ayrı eğitilmiştir. Eğitim sürecine ait işlem sürelerini ve parametre sayıları Tablo II'de belirtilmiştir. HuSHeM-40x veri seti için 25, SMIDS-10x ve SCIAN-10x veri setleri için toplam 50 epoch eğitim gerçekleştirilmiştir. Tabloda belirtilen toplam süreler bu değerlere göre hesaplanmıştır. Yapılan bütün eğitim işlemleri NVIDIA 4070-Ti ekran kartı üzerinde gerçekleştirilmiştir.

TABLE V
SMIDS-10x VERİ SETİNE AİT DOĞRULUK DEĞERLERİ

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

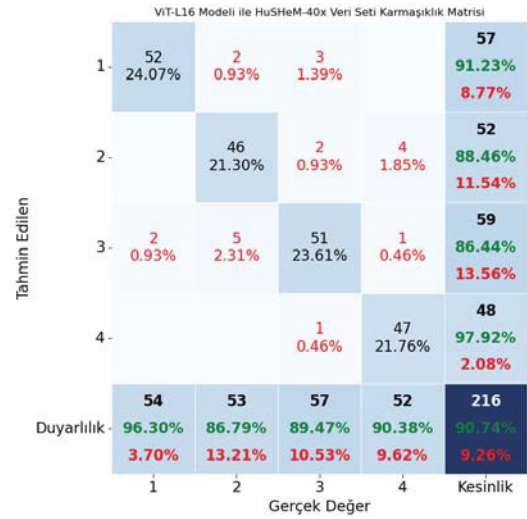
TABLE VI
SCIAN-10x VERİ SETİNE AİT DOĞRULUK DEĞERLERİ

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

IV. SONUÇLAR

Model eğitim süreci sonunda elde edilen modeller ile test verileri üzerinde sınıflandırma işlemi gerçekleştirilmiştir. Veri seti bazında her bir modele ait kat değerleri 0.9 güven değerine sahip iki kuruklu t-test analizi kullanılarak aralarında istatistiksel bir fark olup olmadığı tespit edilmiştir. Aralarında istatistiksel olarak fark olan modellerde daha başarılı olup olmadığı durumuna göre Kazanma, Kaybetme ve Berabere skorları hesaplanmıştır. Bütün modeller birbirleri ile kıyaslanarak elde edilen skor değerleri Tablo III'de sunulmuştur. T-test sırasında kullanılan HuSHeM, SMIDS ve SCIAN doğruluk değerlerini sırasıyla Tablo IV, Tablo V ve Tablo VI'de görmek mümkündür.

Tablo III'de ViT-L16 modeli ile Vgg16 modelinin kesişimi ele alındığında 2/1/0 veya 0/1/2 ifadesini görmek mümkündür. 2/1/0 hücresi ele alındığında deneysel çalışmaların yapıldığı 3 veri setinin 2'sinde ViT-L16 modelinin Vgg16 modeline göre daha iyi sonuçlar vererek kazandığını, 1 veri setinde iki model arasında anlamlı bir fark olmadığı için berabere kaldıklarını ve ViT-L16 modelinin VGG-16 modeline karşı herhangi bir yenilgi almadığı için 0 kez yenildiğini görmek mümkündür. Tablo III'de bulunan genel sonuçlar incelendiğinde ViT-L16



Şekil 1. ViT-L16 Modeli HuSHeM-40x Veri Seti Karmaşıklık Matrisi

modelinin 12 Kazanma(K), 9 Beraberlik(B) ve 0 yenilgi(Kyp) ile HuSHeM verisetinde en iyi ESA mimarisinden yaklaşık %6, SMIDS verisetinde ise yaklaşık %2 daha başarılı olduğu gözlemlenmiştir. SCIAN veri setinde ise Tablo VI incelendiğinde VGG-16'nın en yüksek ortalama doğruluğa sahip olduğu görülmektedir. Fakat t-test analizi yapıldığında ViT-L16 modeli ile VGG-16 modeli arasında istatistiksel bir fark olmadığı için iki model 'Berabere' olarak kabul edilmiştir. En başarılı model olan ViT-L16'nın her bir katta elde edilen modelleri ile kat bazındaki test değerleri üzerinde karmaşıklık matrisleri hesaplanarak birleştirilmiştir. HuSHeM, SMIDS ve SCIAN veri setlerine ait test verileri kullanılarak birleştirilmiş karmaşıklık matrisleri sırasıyla Şekil 1, Şekil 2 ve Şekil 3'te sunulmuştur. Karmaşıklık matrisinin en sağında bulunan sütunda sınıf bazında elde edilmiş kesinlik değerlerini, en alt satırda sınıf bazında hassaslık değerlerini en alt en sağ hücrede ise genel doğruluk değerlerini görmek mümkündür.

Yapılan deneysel çalışmalar sonucunda GD modellerinin özellikle çözünürlük değeri daha yüksek olan görüntü veri setlerinde ESA mimarilerine göre daha başarılı sonuçlar verdiği gözlemlenmiştir. Görüntü çözünürlüğünün 35x35 olduğu SCIAN verisetinde ViT-L16 modelinin diğer ESA mimarileri ile benzer sonuçlar verdiği görülmektedir. Çalışma zamanı olarak incelendiğinde GD modellerinin eğitimi için ESA mimarilerine göre daha fazla kaynak gereksinimi duydukları gözlemlenmiştir. Gelecek çalışmalarda GD modelleri kolektif öğrenme metodları ile birleştirilerek özellikle bu çalışmada da kullanılan insan sperm veri setleri üzerinde geleneksel ESA mimarilerine göre daha başarılı sonuçlar elde edilip edilmeyeceği araştırılacaktır.

BİLGİLENDİRME

Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) ARDEB 1001 Destek No 122E164 projesi tarafından desteklenmiştir.

Tahmin Edilen	Gerçek Değer			Kesinlik
	1	2	3	
1-	865 28.83%	53 1.77%	95 3.17%	1013 85.39%
2-	36 1.20%	907 30.23%	15 0.50%	958 94.68%
3-	104 3.47%	14 0.47%	911 30.37%	1029 88.53%
Duyarlılık	1005 86.07%	974 93.12%	1021 89.23%	3000 89.43%
	13.93%	6.88%	10.77%	10.57%

Şekil 2. ViT-L16 Modeli SMIDS-10x Veri Seti Karmaşıklık Matrisi

Tahmin Edilen	Gerçek Değer					Kesinlik
	1	2	3	4	5	
1-	37 3.27%	6 0.53%	1 0.09%	6 0.53%	23 2.03%	73 50.68%
2-		125 11.04%	11 0.97%		44 3.89%	180 69.44%
3-	3 0.27%	5 0.44%	17 1.50%	1 0.09%	15 1.33%	41 41.46%
4-	2 0.18%	2 0.18%		10 0.88%	19 1.68%	33 30.30%
5-	58 5.12%	90 7.95%	47 4.15%	55 4.86%	555 49.03%	805 68.94%
Duyarlılık	100 37.00%	228 54.82%	76 22.37%	72 13.89%	656 84.60%	1132 85.73%
	63.00%	45.18%	77.63%	86.11%	15.40%	34.28%

Şekil 3. ViT-L16 Modeli SCIAN-10x Veri Seti Karmaşıklık Matrisi

KAYNAKÇA

- [1] W. H. Organization *et al.*, *WHO laboratory manual for the examination and processing of human semen.* World Health Organization, 2021.
- [2] A. Agarwal, N. Parekh, M. K. P. Selvam, R. Henkel, R. Shah, S. T. Homa, R. Ramasamy, E. Ko, K. Tremellen, S. Esteves *et al.*, "Male oxidative stress infertility (mosi): proposed terminology and clinical practice guidelines for management of idiopathic male infertility," *The world journal of men's health*, vol. 37, no. 3, pp. 296–312, 2019.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition. arxiv 2015," *arXiv preprint arXiv:1512.03385*, vol. 14, 2015.
- [6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

- [7] F. Shaker, S. A. Monadjemi, J. Alirezaie, and A. R. Naghsh-Nilchi, "A dictionary learning approach for human sperm heads classification," *Computers in biology and medicine*, vol. 91, pp. 181–190, 2017.
- [8] V. Chang, A. Garcia, N. Hitschfeld, and S. Härtel, "Gold-standard for computer-assisted morphological sperm analysis," *Computers in biology and medicine*, vol. 83, pp. 143–150, 2017.
- [9] H. O. Ilhan, G. Serbes, and N. Aydin, "Automated sperm morphology analysis approach using a directional masking technique," *Computers in Biology and Medicine*, vol. 122, p. 103845, 2020.
- [10] J. Riordon, C. McCallum, and D. Sinton, "Deep learning for the classification of human sperm," *Computers in biology and medicine*, vol. 111, p. 103342, 2019.
- [11] L. Spencer, J. Fernando, F. Akbaridoust, K. Ackermann, and R. Nosrati, "Ensembled deep learning for the classification of human sperm head morphology," *Advanced Intelligent Systems*, vol. 4, no. 10, p. 2200111, 2022.
- [12] O. L. Tortumlu and H. O. Ilhan, "The analysis of mobile platform based cnn networks in the classification of sperm morphology," in *2020 Medical Technologies Congress (TIPTEKNO)*. IEEE, 2020, pp. 1–4.
- [13] M. Yüzkat, H. O. Ilhan, and N. Aydin, "Multi-model cnn fusion for sperm morphology analysis," *Computers in Biology and Medicine*, vol. 137, p. 104790, 2021.
- [14] H. O. Ilhan and G. Serbes, "Sperm morphology analysis by using the fusion of two-stage fine-tuned deep networks," *Biomedical Signal Processing and Control*, vol. 71, p. 103246, 2022.
- [15] A. A. Mashaal, M. A. Eldosoky, L. N. Mahdy, and A. E. Kadry, "Classification of human sperms using resnet-50 deep neural network," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023.
- [16] C. McCallum, J. Riordon, Y. Wang, T. Kong, J. B. You, S. Sanner, A. Lagunov, T. G. Hannam, K. Jarvi, and D. Sinton, "Deep learning-based selection of human sperm with high dna integrity," *Communications biology*, vol. 2, no. 1, p. 250, 2019.
- [17] A. Chen, C. Li, M. M. Rahaman, Y. Yao, H. Chen, H. Yang, P. Zhao, W. Hu, W. Liu, S. Zou *et al.*, "A comprehensive comparative study of deep learning methods for noisy sperm image classification: from convolutional neural network to visual transformer," *Intelligent Medicine*, 2023.
- [18] A. Chen, C. Li, S. Zou, M. M. Rahaman, Y. Yao, H. Chen, H. Yang, P. Zhao, W. Hu, W. Liu *et al.*, "Svia dataset: a new dataset of microscopic videos and images for computer-aided sperm analysis," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 204–214, 2022.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [20] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *CoRR*, vol. abs/1808.01974, 2018. [Online]. Available: <http://arxiv.org/abs/1808.01974>
- [21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, E. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [22] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [24] A. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit, and L. Beyer, "How to train your vit? data, augmentation, and regularization in vision transformers," *arXiv preprint arXiv:2106.10270*, 2021.