

Diyabetik Retinopatinin Tespitinde ve Derecelendirilmesinde Derin Öğrenme Teknolojilerinin Kullanılması

Esra ODABAŞ YILDIRIM¹, Esmâ Fazilet KARAGÜLLE^{2*}, Mustafa YILDIRIM³

Öz

Diyabetik Retinopati (DR), diyabet hastalığının retina damarlarını etkileyen ciddi ve yaygın bir komplikasyondur. Hastalığın ilerlemesiyle retina dokusunda mikro anevrizmalar, kanamalar ve ödem gelişmekte; tedavi edilmediği takdirde ileri evrelerde görme kaybı ve körlüğe yol açabilmektedir. Bu nedenle DR'nin erken evrede tespiti ve teşhisi klinik açıdan büyük önem taşımaktadır. Bu çalışmada, Kaggle APTOS 2019 Körlük Tespiti yarışmasından elde edilen retina görüntüleri kullanılarak farklı Evrişimli Sinir Ağı (ESA) ve Transformer (Dönüşüm) tabanlı derin öğrenme mimarilerinin karşılaştırmalı olarak performansı değerlendirilmiştir. Modellerin performansları, uygulanan gelişmiş ön işleme ve veri artırma stratejilerinin etkileri bağlamında detaylı olarak incelenmiştir. Deneylerde ResNet18, EfficientNetB4, VGG16 ve DenseNet121 gibi ESA tabanlı mimariler ile Swin Transformer gibi Transformer tabanlı modeller kullanılmıştır. Tüm modeller, 5-katlı çapraz doğrulama yöntemi ile eğitilmiş ve performansları karşılaştırılmıştır. Elde edilen bulgular, veri artırma tekniklerinin DR sınıflandırma başarısını istatistiksel olarak anlamlı düzeyde artırdığını ortaya koymuştur. Ayrıca, Swin Transformer modeli %85.00 doğruluk ve %91.37 QWK (Quadratic Weighted Kappa) ile en yüksek performansı sergilemiş ve Transformer tabanlı modellerin geleneksel ESA tabanlı modellere kıyasla DR sınıflandırmasında daha iyi performans sergilediğini göstermiştir.

Anahtar Kelimeler: Derin öğrenme, Diyabetik retinopati, Evrişimli sinir ağları, Görüntü işleme, Transformer tabanlı mimariler, Veri artırma.

Utilization of Deep Learning Technologies in The Detection and Grading of Diabetic Retinopathy

Abstract

Diabetic Retinopathy (DR) is a serious and common complication of diabetes affecting the retinal vessels. As the disease progresses, microaneurysms, hemorrhages, and edema develop in the retinal tissue. If left untreated, it can lead to vision loss and blindness in advanced stages. Therefore, early detection and diagnosis of DR is of great clinical importance. In this study, we evaluated the comparative performance of different Convolutional Neural Network (CCN) and Transformer-based deep learning architectures using retinal images obtained from the Kaggle APTOS 2019 Blindness Detection Competition. The performance of the models is examined in detail in the context of the effects of the applied advanced preprocessing and data augmentation strategies. ESA-based architectures such as ResNet18, EfficientNetB4, VGG16, and DenseNet121, and Transformer-based models such as Swin Transformer were used in the experiments. All models were trained using 5-fold cross-validation, and their performances were compared. The findings revealed that data augmentation techniques significantly increased the accuracy of DR classification. Furthermore, the Swin Transformer model exhibited the highest performance with 85.00% accuracy and 91.37% QWK (Quadratic Weighted Kappa), and Transformer-based models performed better in DR classification compared to traditional ESA-based models.

Keywords: Convolutional neural networks, Data augmentation, Deep learning, Diabetic retinopathy, Image processing, Transformer architectures.

¹Atatürk Üniversitesi, Yazılım Mühendisliği, Erzurum, Türkiye, esra.odabas@atauni.edu.tr

²Atatürk Üniversitesi, Yazılım Mühendisliği, Erzurum, Türkiye, esmafazilet.karagulle17@ogr.atauni.edu.tr

³Atatürk Üniversitesi, Tıp Fakültesi, Erzurum, Türkiye, mustafa.yildirim@atauni.edu.tr

*Sorumlu Yazar/Corresponding Author

1. Giriş

Halk arasında şeker hastalığı olarak bilinen Diabetes Mellitus (DM), insulin hormonunun yokluğu, yetersizliği veya metabolizmasında gerçekleşen bozulmalar sonucunda oluşmaktadır. Uluslararası Diyabet Federasyonu'nun (International Diabetes Federation, IDF) 2021 yılı raporuna göre dünya çapında 537 milyon kişi diyabet hastasıdır ve bu sayının 2045 yılına kadar 783 milyona ulaşması beklenmektedir (IDF 2021; Sun vd., 2022). DM, çok sayıda komplikasyonu da beraberinde getiren bir rahatsızlıktır. Bu komplikasyonlar arasında kardiyovasküler hastalıklar, böbrek yetmezliği, sinir hasarı ve göz hastalıkları önemli bir yer tutmaktadır (Zhang vd. 2010). DM'nin neden olduğu göz hastalıklarından biri görme kaybına neden olabilen DR olarak bilinmektedir. Uzun süreli yüksek kan şekeri seviyeleri retina damarlarında geçirgenlik artışı ve yapısal bozulmalara yol açarak DR gelişimine sebep olur (Zhang vd., 2010; Antonetti vd., 2012). Bu durum retina üzerinde kanamalar, eksüda birikimi, mikro anevrizmalar ve ilerleyen durumlarda retina dekolmanına sebep olmaktadır. Dünya Sağlık Örgütü 2023 raporuna göre DR, dünya genelinde önlenebilir körlüğün önde gelen nedenlerinden biridir (World Health Organization, WHO. 2023).

DR genellikle erken evrelerde belirgin bir semptom göstermediğinden hastaların çoğu ileri evrelere ulaşana kadar hastalığı fark etmemektedir. Bu durum, tedaviye geç başlanmasına ve kalıcı görme kaybı riskinin artmasına neden olmaktadır (Yau vd., 2012). Bu gecikme tedavi süreçlerinin de daha karmaşık hale gelmesine neden olmaktadır. Geleneksel tanı yöntemleri uzmanların manuel değerlendirmelerine dayanmakta ve bu durum süreci hem zaman alıcı hem de iş gücüne bağımlı hale getirmektedir. Ancak artan hasta sayısı ile sağlık uzmanlarının sınırlı olduğu ve kaynakların yetersiz olduğu bölgelerde bu yöntem yetersiz kalmaktadır. Özellikle düşük ve orta gelirli ülkelerde düzenli göz taramalarına erişimin sınırlı olması DR'nin geç teşhis edilmesine ve kalıcı görme kaybına yol açmasına neden olabilmektedir (WHO, 2020). Bu bağlamda, derin öğrenme temelli otomatik tarama sistemlerinin önemi giderek artmaktadır. (Gulshan vd., 2016) tarafından geliştirilen bir yapay zekâ sistemi retina görüntülerinden otomatik DR tespiti yaparak uzman eksikliğinin olduğu bölgelerde tarama hizmetlerini daha erişilebilir hale getirme potansiyelini ortaya koymuştur.

Günümüzde yapay zekâ ve özellikle derin öğrenme teknolojileri tıbbi görüntüleme ve tanı sistemlerinde önemli yenilikler sunmaktadır. Derin öğrenme, büyük veri kümelerinden karmaşık desenleri ve ilişkileri öğrenerek yüksek doğruluk oranları sağlayan bir yapay zekâ yaklaşımıdır (LeCun vd., 2015). Özellikle tıbbi görüntülerde kullanılan ESA'lar, DR gibi göz hastalıklarının fundus görüntülerinden lezyonları otomatik olarak tanıma yeteneği sayesinde hastalık teşhisinde önemli başarılar elde etmiştir (Gulshan vd., 2016).

Bununla birlikte son yıllarda bilgisayarla görü alanında geleneksel ESA mimarilerine alternatif olarak Transformer tabanlı modeller öne çıkmaktadır. Transformer mimarisi ilk olarak (Vaswani vd.,

2017) tarafından doğal dil işleme (natural language processing - NLP) alanında geliştirilmiş ve dil çevirisi, soru-cevap gibi görevlerde büyük başarılar elde etmiştir. Bu mimari, veriler arasındaki uzun mesafeli ilişkileri modellemedeki başarısı sayesinde araştırmacılar tarafından görüntü verilerine uygulanabilirliği araştırılmaya başlanmıştır. Görüntü verilerine ilk uygulamalardan biri olan Vision Transformer (Görsel Dönüşüm - ViT), görüntüleri sabit boyutlu yamalara (patch) bölerek bu yamaları kelime örnekleri gibi işlemiş ve öz dikkat (self-attention) mekanizmaları ile tüm görüntü alanındaki küresel ilişkileri doğrudan modellemiştir (Dosovitskiy vd., 2020). Bu yöntem, görüntülerdeki küresel bağlamı etkili biçimde modelleyebilmekte ancak yüksek miktarda veri gereksinimi nedeniyle verinin sınırlı olduğu durumlarda yetersiz kalabilmektedir.

Bu nedenle Transformer mimarisinin görüntü işleme görevlerine daha verimli şekilde uyarlanmasını hedefleyen Swin Transformer, dikkat hesaplamalarını yerel pencereler içinde gerçekleştirerek işlem verimliliğini artırmış ve hiyerarşik yapısıyla çok ölçekli özellik öğrenimini mümkün kılmıştır (Liu vd., 2021). Böylece ilk etapta dil tanıma problemleri için tasarlanmış olan Transformer mimarisi fundus görüntülerinin sınıflandırılması gibi yüksek çözünürlüklü ve detay odaklı tıbbi görüntüleme görevlerinde ESA'larla karşılaştırılabilir bir alternatif haline gelmiştir. Tüm bu gelişmeler yapay zekâ temelli modellerin tıbbi karar destek sistemlerinde etkin biçimde kullanılabileceğini göstermektedir.

Bu çalışma, fundus görüntüleri üzerinde hem DR tespiti hem de derecelendirilme görevlerini gerçekleştiren derin öğrenme tabanlı bir sistem önermektedir. Modelin başarısı farklı ESA ve Transformer mimarileri kullanılarak test edilmiş ve görüntüler üzerinde gelişmiş ön işleme teknikleri uygulanarak sınıflandırma performansı artırılmaya çalışılmıştır. Çalışmanın devamında sırasıyla güncel literatür çalışmaları, kullanılan veri seti ve mimariler, gerçekleştirilen deneysel analizler ve sonuçlar detaylı olarak sunulacak ve tartışılacaktır.

1.1. Literatür Araştırması

Yi ve ark. (2021) DR tespiti için RA-EfficientNet adını verdikleri bir model geliştirmişlerdir. Bu model, EfficientNet mimarisine artık dikkat (Residual Attention - RA) bloğu ekleyerek DR sınıflandırmasında iyileşme sağlamıştır. Model, APTOS 2019 veri setinde test etmiş ve çoklu sınıflandırmada %93.55 doğruluk oranı elde etmişlerdir (Yi vd., 2021). Rao ve ark. (2020) DR tespiti için ImageNet üzerinde önceden eğitilmiş ResNet50 tabanlı bir derin öğrenme modeli kullanmışlardır. Çalışmada APTOS 2019 veri seti kullanarak modelin ikili sınıflandırma performansını değerlendirmişlerdir. Yapılan deney sonucunda %96.59 doğruluk oranı elde etmişlerdir (Rao vd., 2020). Bodapati ve ark. (2021), DR evrelerini sınıflandırmak için Xception ve VGG16 ağlarını kapılı dikkat (gated attention) mekanizmasıyla birleştirerek bir derin öğrenme modeli

geliştirmişlerdir. Çalışmada, Xception ve VGG16 ağlarının özelliklerini birleştirerek DR'nin çeşitli evrelerini sınıflandırmada %82.5 doğruluk oranı elde etmişlerdir (Bodapati vd., 2021). Patel ve Chaware (2020), MobileNetV2 tabanlı bir model geliştirmiş ve modeli transfer öğrenme yöntemiyle DR tespiti için ince ayarlamışlardır. Çalışmada APTOS 2019 veri seti üzerinde veri artırma teknikleri ve farklı dönüşümler (dönme, yakınlaştırma vs.) uygulayarak modelin doğruluk oranını %81 olarak bildirmişlerdir (Patel vd., 2020). Minarno ve Bagaskara (2025), APTOS 2019 veri setinde InceptionV3 modeliyle yaptıkları denemede veri artırma uygulanmadığında modelde hızlıca aşırı uyum gözlemlendiğini uygun artırma teknikleriyle eğitim yapıldığında test doğruluğunun %82.7'ye ulaştığını raporlamışlardır. Çalışmada veri artırmanın yokluğunda eğitim ve test başarımı arasındaki farkın açıldığı yani modelin genelleme kabiliyetinin ciddi şekilde azaldığı gösterilmiştir (Minarno vd., 2025). Thota ve Reddy (2020), DR'nin şiddetini sınıflandırmak için transfer öğrenme yöntemini kullanmışlardır. Çalışmada önceden eğitilmiş bir VGG16 modeline ince ayar yapmışlardır. Çalışmada yüksek doğruluk seviyelerine ulaşmak için veri artırma, toplu normalleştirme ve öğrenme oranının ayarlanması gibi çeşitli teknikler kullanmışlardır. Çalışmada APTOS 2019 veri seti kullanılarak modelin performansı test edilmiş ve %74 doğruluk, %80 duyarlılık, %65 özgüllük ve %80 AUC elde etmişlerdir (Thota vd., 2020). Nguyen ve ark. (2020), derin öğrenme tekniklerini kullanarak DR tespitine yönelik çok sınıflı bir model geliştirmişlerdir. Çalışmada, ESA, VGG-16 ve VGG-19 gibi derin öğrenme modelleri kullanarak DR tespiti yapmışlardır. Çalışmada veri setini artırmak amacıyla yansıtma ve döndürme uygulamışlardır. Model EyePacs veri seti üzerinde test edilmiş ve %80 duyarlılık, %82 doğruluk, %82 özgüllük ve %90 AUC elde etmişlerdir (Nguyen vd., 2020). Lavanya ve ark. (2020), DR tespiti ve sınıflandırılması için Raspberry Pi kullanan bir sistem geliştirmişlerdir. Çalışmada DR tespiti için düşük maliyetli bir çözüm önerisi sunmayı amaçlamışlardır. Raspberry Pi cihazına entegre edilen ESA tabanlı derin öğrenme modeli ile retina görüntüleri analiz edilmiş ve %90 sınıflandırma doğruluğu ve %63 test doğruluğu elde etmişlerdir (Lavanya vd., 2020). Elzennary ve ark. (2020), DR'nin şiddetini belirlemek için transfer öğrenme ile DenseNet-121 sinir ağı mimarisini kullanmışlardır. Çalışmada APTOS 2019 veri setine Gauss bulanıklaştırma, kontrast artırma ve çeşitli görüntü işleme teknikleri uygulamışlardır. Çalışmada doktorların DR'yi tespit etmek için kullanabileceği arayüz oluşturmak için Flask adlı Python çerçevesinden yararlanmışlardır. Yapılan testler sonucunda %95.64 doğruluk oranı elde etmişlerdir (Elzennary vd., 2020). Pamadi ve ark. (2022), MobileNetV2'yi kullanarak DR tespiti ve derecelendirilmesi için ikili ve çoklu sınıflandırma yapan bir model geliştirdiler. APTOS 2019 veri setine gauss bulanıklaştırma gibi görüntü işleme ve geometrik döndürme gibi veri artırma tekniklerini uygulamışlardır. Çalışmada ikili sınıflandırma için %97, çoklu sınıflandırma için %78 doğruluk oranı elde etmişlerdir (Pamadi vd., 2022). Saranya ve ark. (2022), DR tespiti için DenseNet tabanlı bir derin öğrenme modeli önermişlerdir. APTOS 2019 veri seti kullanılarak modeli eğitmiş ve retina

görüntülerinde DR'nin tespitine yönelik ikili sınıflandırma yapmışlardır. Çalışmada, eğitime başlamadan önce görüntü işleme ve veri dengeleme adımlarını uygulayarak %83 doğruluk elde etmişlerdir (Saranya vd., 2022).

Tablo 1. DR görüntülerinin sınıflandırılması üzerine farklı çalışmaların özeti.

Çalışma	Veri Seti	Yöntem	İkili Sınıflandırma	Çoklu Sınıflandırma
Rao ve ark. (2020)	Aptos-19	ResNet50 (transfer öğrenme)	%96.59 Doğruluk	-
Patel ve Chaware (2020)	Aptos-19	İnce ayarlanmış MobileNetV2	-	%81 Doğruluk
Nguyen ve ark. (2020)	Aptos-15	VGG16	-	%82 Doğruluk
Lavanya ve ark. (2020)	EyePacs	ESA, Raspberry Pi arayüzü	-	%63 Doğruluk
Thota ve Reddy (2020)	EyePacs	VGG-16, Veri artırma	-	%74 Doğruluk, %80 Hassasiyet, %65 Özgüllük
Elzennary ve ark. (2020)	Aptos-19	Gauss, kontrast artırma, Transfer öğrenme, DenseNet-121, Flask Arayüzü	-	%95.64 Doğruluk
Taufiqurrahman ve ark. (2021)	Aptos-19	MobileNetV2 + SVM(10-kat doğrulama), Veri artırma	-	%85 Doğruluk, %92.5 Kappa
Bodapati ve ark. (2021)	Aptos-19	Xception & VGG16 hibrit + Kapalı Dikkat	-	%82.5 Doğruluk, %79.0 Kappa
Yi ve ark. (2021)	Aptos-19 & EyePacs	RA-EfficientNet-B4	%98.36 Doğruluk	%93.55 Doğruluk, %89.20 Doğruluk (EyePacs)
Majumder ve Kehtarnavaz (2021)	Aptos-19 & EyePacs	SE-DenseNet201, Veri artırma	-	%88 Doğruluk (EyePacs), %90 Kappa (Aptos-19)
Pamadi ve ark. (2022)	Aptos-19	Gauss, MobileNetV2	%97 Doğruluk	%78 Doğruluk
Saranya ve ark. (2022)	Aptos-19	DenseNet-121, Veri dengeleme	-	%83 Doğruluk
Batool ve ark. (2023)	EyePacs & DeepDRiD	EfficientNet-B0-B6 modelleri (ön eğitilmiş), Gauss + veri artırma ile iyileştirme	-	EyePACS- %84 F1; DeepDRiD- %87 F1
İncir ve Bozkurt (2024)	EyePacs	Yedi farklı önceden eğitilmiş ESA mimarisinin karşılaştırılması; en iyi performans EfficientNetV2-M (afine dönüşümlerle veri ön işleme ve artırma)	-	%97.65 Doğruluk, %92.8 Kesinlik, %92.6 Hassasiyet, %92.4 F1
Arora ve ark. (2024)	EyePacs	EfficientNet-B0 tabanlı ensemble model	-	%86.53 Doğruluk
Yang ve ark. (2024)	EyePacs	Maskeli Oto kodlayıcı ile ön-eğitilmiş ViT	%93.42 Doğruluk	-
Minarno ve Bagaskara (2025)	Aptos-19	Inception-V3, Veri artırma	-	%82.7 Doğruluk
Akhtar ve ark. (2025)	Messidor-1	RSG-Net + Histogram eşitleme	%99.37 Doğruluk	-

Tablo 1’de literatürdeki benzer çalışmalar özetlenerek verilmiştir. Mevcut literatürde, yöntemler genellikle tek bir model kapsamında değerlendirilmiş olup, DR tespiti veya sınıflandırmasına odaklanmıştır. Bu çalışmada ise, farklı derin öğrenme mimarilerinin DR tespiti ve

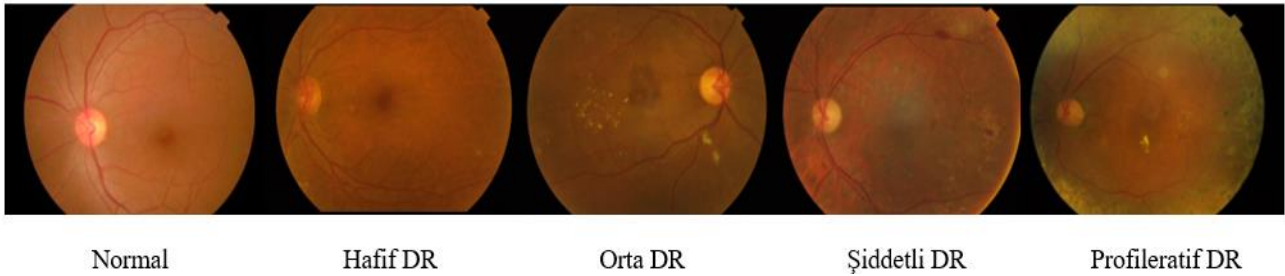
sınıflandırma performansları karşılaştırmalı olarak incelenmiş; böylece modellerin etkinlikleri ve sınıflandırma doğrulukları akademik açıdan kapsamlı biçimde değerlendirilmiştir.

2. Materyal ve Metot

Bu bölümde çalışmada kullanılan veri seti, görüntü işleme adımları, veri artırma yöntemleri, kullanılan model mimarileri, eğitim stratejisi ve performans değerlendirme metrikleri anlatılmaktadır.

2.1. Veri Seti

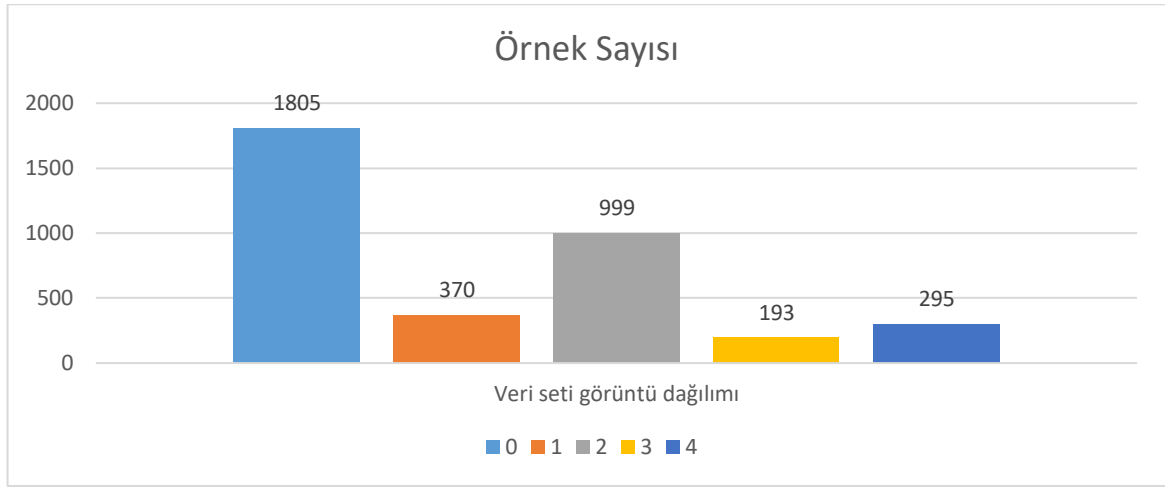
Bu çalışmada DR'nin tespiti ve derecelendirilmesi için APTOS 2019 veri seti kullanılmaktadır. Bu veri setinin bir kısmı açık erişimli olarak yayınlanmıştır. Veri seti fundus görüntülerini içermektedir ve bu görüntüler hastaların retinasındaki damarlar, lezyonlar ve diğer hasarları tespit etmek için kullanılabilir. APTOS 2019 veri setinde her bir görüntü için bir etiket bulunur ve bu etiket, retinopatinin çeşitli derecelerine işaret eder (0'dan 4'e kadar, 0: sağlıklı, 1: hafif, 2: orta, 3: ileri, 4: çok ileri). Bu etiketler modelin DR'yi farklı seviyelerde sınıflandırmasını sağlar.



Şekil 1. Aptos-19 veri setinden örnek görüntüler.

APTOS 2019 veri seti, 3662 fundus görüntüsü içerir ve bu görüntüler modelin eğitim ve test süreçlerinde kullanılmak üzere derin öğrenme modellerinin performansını değerlendirmeye olanak sağlar. Veri setinin etiketlenmiş yapısı DR'nin sınıflandırılması için derin öğrenme tabanlı modellerin geliştirilmesinde yaygın bir şekilde kullanılmaktadır (Kaggle, 2019). Şekil 1'de veri setindeki farklı derecelere ait görüntü örnekleri verilmiştir.

Şekil 2'de APTOS 2019 veri setindeki örneklerin dağılımı grafik ile gösterilmiştir.



Şekil 2. APTOS-2019 veri seti hastalık ciddiyet düzeylerinin dağılımı.

2.2. Görüntü Ön İşleme ve Veri Artırma

Bu çalışmada kullanılan fundus görüntüleri modele verilmeden önce bir dizi ön işleme ve veri artırma adımına tabi tutulmuştur. Bu adımlar, modelin hem görüntü kalitesinden bağımsız olarak tutarlı sonuçlar üretmesini sağlamakta hem de azınlık sınıflar üzerindeki başarımını artırmaktadır. Görüntü işleme süreci sırasıyla kırpma, kontrast sınırlamalı adaptif histogram eşitleme (CLAHE - Contrast Limited Adaptive Histogram Equalization) ve renk düzeltmesini veri artırma ise yansıtma, döndürme, RGB kaydırma parlaklık ayarlama ve Gamma düzeltmesi gibi işlemleri kapsamaktadır. Tüm bu adımlar Şekil 3'te çalışmada kullanılan veri setindeki örnek bir görüntü üzerinden gösterilmiştir.



Şekil 3. Görüntü işleme adımları.

2.3. Retina Bölgesinin Otomatik Kırılması

Fundus görüntülerinde retina dışı bölgeler (siyah arka plan gibi) genellikle sınıflandırma için anlamlı bilgi taşımaz. Bu nedenle ilk adımda, görüntünün yalnızca retina bölgesini içerecek şekilde otomatik kırılması sağlanmıştır. Görüntü gri seviyeye dönüştürülmüş ardından eşikleme yöntemiyle en büyük kontur belirlenmiş ve bu kontur çevresindeki dikdörtgen alan görüntüden çıkarılmıştır. Böylece modelin dikkatinin yalnızca anlamlı bilgi içeren bölgelere yönelmesi sağlanmıştır (Şekil 3, 1-2).

2.4. CLAHE ile Kontrast İyileştirme

Kırılmış görüntü, RGB renk uzayından LAB renk uzayına dönüştürülerek L (Lightness - Aydınlik) kanalı ayrıştırılmıştır. Ardından, sadece L kanalına Kontrast CLAHE yöntemi uygulanmıştır. CLAHE yöntemi, lokal histogram eşitlemesi prensibiyle çalışarak düşük kontrastlı bölgelerdeki ayrıntıların daha belirgin hale gelmesini sağlar ve medikal görüntüleme alanında etkinliği nedeniyle yaygın olarak kullanılmaktadır (Zuiderveld, 1994; Gonzalez & Woods, 2018). Bu işlem sonucunda retina görüntülerindeki damar yapıları, lezyonlar ve optik disk gibi klinik açıdan önemli yapılar daha görünür hale getirilmiştir (Şekil 3, 3).

2.5. Renk Düzeltme

CLAHE işleminin ardından elde edilen görüntüye (gamma) $\gamma=1.2$ parametresi kullanılarak renk düzeltilmesi uygulanmıştır. Renk düzeltmesinde $\gamma > 1$ seçildiğinde görüntünün parlaklık seviyelerini doğrusal olmayan bir dönüşümle değiştirerek koyu alanlardaki detayların daha görünür olmasını sağlar. Böylelikle düşük ışıklı bölgelerdeki ayrıntıların (örneğin retina damarları, mikro anevrizmalar) görünürlüğü artırılarak daha etkili görsel analiz ve tanı imkânı elde edilir. (Şekil 3, 4).

2.6. Veri Artırma

Modelin genelleme kabiliyetini artırmak, aşırı uyumu (overfitting) önlemek ve veri dengesizliği sorununu hafifletmek amacıyla eğitim aşamasında her bir görüntüye çeşitli veri artırma teknikleri uygulanmıştır. Bu dönüşümler aşağıda açıklanmıştır.

2.7. Yansıtma

Görüntü yatay düzlemde ters çevrilerek modelin anatomik simetrilere duyarlılığı geliştirmiştir. Kullanılan parametreler `HorizontalFlip(p=0.5)`; düşey yansıma kullanılmadı (anatomik uyumsuzluk) (Şekil 3, 5).

2.8. Rastgele Döndürme

Görüntüler ± 30 dereceye kadar rastgele döndürülerek farklı açılardan görüntülenen retina yapılarına karşı duyarlılığı artmıştır. Kullanılan parametreler; `Rotate(limit=30, p=0.5, border_mode=cv2.BORDER_REFLECT)` (Şekil 3, 6).

2.9. RGB Renk Kaydırma

Her bir renk kanalına (R, G, B) ayrı ayrı küçük değerler eklenerek renk varyasyonları oluşturulmuştur. Bu, farklı kameralar ve ışık kaynakları nedeniyle oluşabilecek renk değişimlerine karşı modelin dayanıklılığını artırmaktadır. Kullanılan parametreler; `RGBShift(r_shift= $\pm 20$ , g_shift= $\pm 20$ , b_shift= $\pm 20$ , p=0.5)` (Şekil 3, 8).

2.10. Parlaklık ve Kontrast Ayarlama

Görüntülerin parlaklık ve kontrast seviyeleri rasgele değiştirilerek farklı aydınlatma koşullarının simülasyonu sağlanmıştır. Kullanılan parametreler; `RandomBrightnessContrast(brightness_limit=0.2, contrast_limit=0.2, p=0.5)` (Şekil 3, 7).

Görüntü işleme adımları, Python programlama dili kullanılarak gerçekleştirilmiştir. Bu süreçte özellikle OpenCV, Albumentations ve NumPy gibi yaygın olarak kullanılan açık kaynak kütüphanelerden yararlanılmıştır. Tüm görüntüler öncelikle 224×224 piksel boyutuna yeniden ölçeklendirilmiş, ardından normalize edilerek ön işleme aşamasından geçirilmiştir. Normalizasyon sonrası, veri artırma sürecinde, her bir görüntüye uygulanabilecek çeşitli dönüşümler (yansıtma, döndürme, RGB kanal kaydırması, parlaklık-kontrast ayarlamaları vb.) için belirlenen olasılıklar (örneğin $p=0.5$) dahilinde, Bernoulli dağılımına dayalı rastgele bir seçim mekanizması işletilmiştir. Böylece, her dönüşümün görüntüye uygulanıp uygulanmayacağı ayrı ayrı ve rastgele olarak belirlenmiştir. Bu yaklaşım sayesinde her epoch'ta aynı görüntü örneği modele farklı varyasyonları ile sunulmakta ve eğitim veri setine dinamik bir çeşitlilik kazandırılmaktadır. Elde edilen bu çeşitlilik,

modelin aşırı öğrenmeden kaçınmasına ve farklı oryantasyon, renk varyasyonu ve aydınlatma koşullarında daha genel bir performans sergilemesine olanak tanımaktadır.

2.11. Kullanılan Modeller

Bu bölümde çalışmada kullanılan modeller hakkında bilgi verilmektedir. Çalışmada 4 ESA (ResNet18, EfficientNetB4, DenseNet121, VGG16) ve 1 Transformer (Swin) tabanlı olmak üzere 5 model kullanılmıştır.

2.12. ResNet18

ResNet (Residual Network), 2015 yılında He ve arkadaşları tarafından geliştirilmiş olup derin sinir ağlarının eğitiminde karşılaşılan gradyan kaybolması (vanishing gradient) problemini çözmek amacıyla önerilmiştir (He vd., 2016). ResNet mimarisi artıklı bağlantılar (residual connections) kullanarak daha derin modellerin eğitilmesini mümkün kılmıştır. ResNet mimarisi farklı derinliklerde varyantlara sahiptir. ResNet18 bu mimarinin en hafif ve hesaplama açısından en verimli versiyonlarından biridir. ResNet18 modeli toplamda 18 katman içermektedir. Bu katmanlar arasında 17 adet evrişimli katman ve 1 adet tam bağlantılı (fully connected) katman bulunmaktadır. Model, girişte $3 \times 224 \times 224$ boyutunda RGB bir görüntü kabul eder. İlk katmanda 7×7 boyutunda ve 64 filtrelili bir evrişim işlemi uygulanır, ardından toplu normalizasyon (batch normalization-BN), ReLU aktivasyonu ve maksimum havuzlama (max pooling) işlemleri gerçekleştirilir. Bu aşamadan sonra model dört adet artıklı blok (residual block) içerir. Her artıklı blok, sırasıyla iki adet 3×3 evrişim, toplu normalizasyon ve ReLU aktivasyon fonksiyonundan oluşur. Giriş ile çıkış boyutları eşleştiğinde giriş doğrudan çıkışa eklenerek artıklı bağlantı yapılır. Bu yapı sayesinde model derinleştikçe öğrenme kapasitesini kaybetmez ve gradyan geçişi bozulmadan ilerler. Bu mekanizma ağı daha iyi genelleştirmesini ve daha hızlı eğitilmesini sağlar (He vd., 2016). Modelin sonunda bir global ortalama havuzlama katmanı uygulanır. Elde edilen 512 boyutlu vektör tam bağlantılı katmana iletilerek son sınıflandırma yapılır. Genellikle bu katman sınıf sayısına (örneğin ImageNet'te 1000, tez çalışmasında derecelendirme için 5 ve tespit için 2) göre çıktılar üretir.

2.13. VGG16

VGG16, 13 evrişim katmanı ve 3 tam bağlantılı katman olmak üzere toplam 16 öğrenilebilir katmandan oluşan derin bir evrişimsel sinir ağıdır. Bu mimari, Oxford Üniversitesi'ne bağlı Visual Geometry Group tarafından geliştirilmiştir. Model giriş olarak 224×224 piksel boyutlarında üç kanallı

RGB görüntü alır. Evrişim katmanları boyunca 3×3 boyutlu küçük filtreler kullanılması mimarinin temel yeniliklerinden biridir. Bu filtreler çekirdek boyutunu büyütmeden derinliği artırmakta ve daha karmaşık özelliklerin öğrenilmesini sağlamaktadır (Simonyan ve Zisserman, 2014). Model beş adet evrişimsel blok yapısında düzenlenmiştir. Her blokta 2 veya 3 adet ardışık 3×3 evrişim katmanı ve bunu takip eden maksimum havuzlama katmanı yer alır. Bu yapı sayesinde ağ derinleştikçe hem uzamsal çözünürlük düşürülmekte hem de özellik soyutlama düzeyi artmaktadır. Evrişim bloklarının ardından iki adet 4096 nöronlu tam bağlantılı gizli katman ve bir adet 1000 sınıflı çıkış katmanı (genellikle ImageNet veri kümesi için) bulunur.

2.14. DenseNet121

DenseNet121, toplamda 121 katmandan oluşan ve yoğun bağlantı (dense connectivity) ilkesi üzerine inşa edilmiş derin evrişimsel sinir ağı mimarisidir. Modelin temel yapı taşları, ardışık şekilde organize edilmiş dört adet dense blok yapısıdır. Bu bloklardaki katman sayıları sırasıyla 6, 12, 24 ve 16'dır. Her bir dense blok içerisinde yer alan katmanlar birbirine tam bağlantılıdır yani her katman, kendisinden önceki tüm katmanların çıktısını doğrudan girdi olarak alır ve kendi çıktısını da sonraki tüm katmanlara iletir. Bu yapı hem özellik tekrar kullanımını artırmakta hem de gradyan akışını kolaylaştırarak derin ağların daha stabil ve verimli eğitilmesini sağlamaktadır (Huang vd., 2017).

2.15. EfficientNetB4

EfficientNet mimari ailesi, Tan ve Le (2019) tarafından geliştirilen ve görüntü sınıflandırma görevlerinde hem yüksek doğruluk hem de hesaplama verimliliği sunan evrişimli sinir ağı mimarileridir. Bu mimarinin temel yeniliği, genişlik (width), derinlik (depth) ve çözünürlük (resolution) ölçeklendirmesinin dengeli bir şekilde yapılmasını sağlayan bileşik ölçeklendirme (compound scaling) yöntemidir. Bu sayede model hem parametre sayısını verimli kullanmakta hem de performansını artırmaktadır. EfficientNetB4, EfficientNet ailesinin orta büyüklükteki modellerinden biridir ve ImageNet veri seti üzerinde yapılan deneylerde yüksek doğruluk oranları elde etmiştir. EfficientNetB4, EfficientNetB0 modelinin ölçeklendirilmiş hali olarak giriş görüntüsü çözünürlüğünü 380×380 piksele çıkarır ve ağı derinliği ile genişliğini artırarak performansını yükseltir. Literatürde EfficientNetB4 hem genel görüntü sınıflandırma hem de medikal görüntü analizinde yaygın şekilde kullanılmakta olup, düşük hesaplama maliyetiyle yüksek başarı sağlaması nedeniyle tercih edilmektedir (Tan ve Le, 2019; Yang ve ark., 2021). Özellikle medikal görüntü görevlerinde EfficientNet-B4, dengeli ölçeklendirmesi sayesinde etkili sonuçlar vermektedir (Yang ve ark., 2021).

2.16. Swin Transformer

Görsel tabanlı görevlerde derin öğrenme modellerinin başarısı büyük ölçüde modelin hem yerel hem de küresel bağlamdan anlamlı özellikler çıkarabilme yetisine bağlıdır. Özellikle ViT mimarileri görüntüyü sabit boyutlu yama dizilerine ayırarak dizisel bilgi işleme (sequential processing) yöntemini görsel alana uyarlamıştır (Dosovitskiy vd., 2021). Ancak ViT'in doğrudan uyguladığı küresel çok başlı dikkat mekanizması yüksek çözünürlüklü görüntülerde ciddi hesaplama maliyetlerine neden olmaktadır. Bu sorunu çözmek amacıyla (Liu ve ark., 2021), Swin Transformer adlı mimariyi önermiştir. Swin Transformer, dikkat işlemlerini sabit boyutlu yerel pencereler (ör. 7×7) içinde gerçekleştiren pencere tabanlı dikkat (Window-based Multi-head Self-Attention- W-MSA) mekanizmasıyla çalışır. Ayrıca katmanlar arasında pencere kaydırma (Shifted Window-based Multi-head Self-Attention - SW-MSA) stratejisi kullanılarak pencereler arası bilgi akışı sağlanır. Böylece hem yerel hem de daha geniş bağlamlar işlenebilirken hesaplama verimliliği de korunmuş olur. Swin Transformer mimarisi, her bir Swin bloğunda sırasıyla çok başlı dikkat mekanizma, artık bağlantı, katman normu ve çok katmanlı algılayıcı bileşenlerini içerir. Bu yapı Transformer mimarisinin genel yapısına benzetmekle birlikte dikkat mekanizması klasik küresel öz dikkat yerine yerel pencerelerle sınırlanmıştır. Swin Transformer 'da ardışık iki blok çift olarak tasarlanmıştır. Bunlardan ilki W-MSA ve ikincisi SW-MSA mekanizmalarını uygular. Bu yapı sayesinde model hem yerel ilişkileri hem de pencere kaydırarak daha geniş bağlamları öğrenebilir.

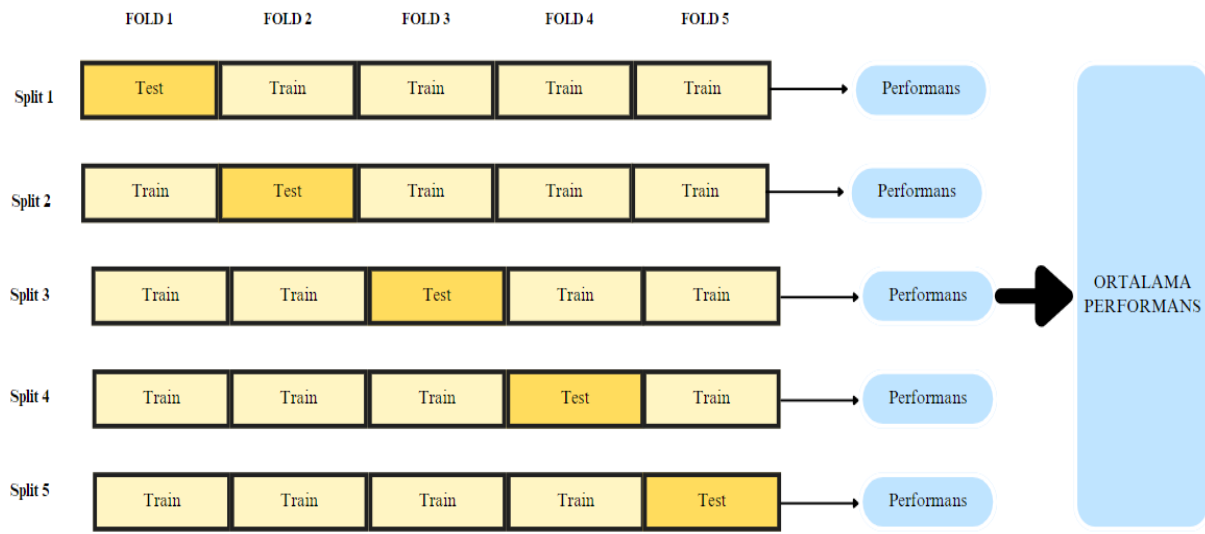
2.17. Eğitim Stratejisi

Model eğitim süreçleri her bir mimari için aynı hiper parametre ayarlarıyla gerçekleştirilmiştir. Bu yöntem farklı modellerin adil bir şekilde kıyaslanabilmesi açısından önemlidir. Modelin başarıyla eğitilebilmesi için kullanılan optimizasyon algoritması, öğrenme oranı planlayıcısı, kayıp fonksiyonu, eğitim süresi ve batch boyutu gibi hiper parametrelerin dikkatle belirlenmesi gerekmektedir. Bu bölümde, çalışmada kullanılan eğitim stratejisi detaylı olarak açıklanmakta ve her bir kararın literatürdeki karşılığı belirtilmektedir.

Tüm modeller AdamW optimizasyon algoritması ile eğitilmiştir. Öğrenme oranı $lr = 1e-4$ ile başlatılmış ve WarmRestarts öğrenme oranı planlayıcısı (scheduler) kullanılarak azaltılmış ardından dönemsel olarak sıfırlanmıştır. Eğitim süreci boyunca maksimum 20 epoch (dönem) belirlenmiş ancak doğrulama kaybı belirli bir sürede iyileşmediğinde erken durdurma stratejisi uygulanmıştır. Erken durdurma için sabır (patience) değeri = 5 olarak belirlenmiştir. Her bir eğitim adımında işlenen örnek sayısı batch size = 8 olarak belirlenmiş ve bu değer GPU belleği sınırları ile model stabilitesi gözetilerek seçilmiştir (Masters ve Lusch, 2018).

Veri setindeki belirgin sınıf dengesizliği nedeniyle kayıp fonksiyonu olarak Focal Loss tercih edilmiştir. Focal Loss için odaklanma parametresi $\gamma=2.0$ olarak belirlenmiş ve her sınıfa özgü ağırlıklar [1.0, 2.0, 1.0, 4.5, 5.0] şeklinde uygulanmıştır. Bu sayede özellikle düşük frekanslı sınıflar (sınıf 3 ve 4) üzerinde hata payı artırılarak modelin söz konusu sınıflara daha fazla dikkat etmesi amaçlanmıştır (Lin vd., 2017).

Model değerlendirmelerinde sınıf dengesini korumak amacıyla K-katlı çapraz doğrulama (Stratified K-Fold Cross Validation) yöntemi kullanılmıştır. Bu sayede her katmanda sınıf dağılımı korunarak adil bir değerlendirme sağlanmıştır. Her alt küme (fold) için model sıfırdan eğitilmiş ve en iyi doğrulama kaybına ulaşıldığında model ağırlıkları kayıt altına alınmıştır. K-Katlı çapraz doğrulama yapısı Şekil 4'te k=5 için gösterilmiştir.



Şekil 4. 5-Katlı çapraz doğrulama yapısı.

2.18. Performans Değerlendirme Metrikleri

Model performansı doğruluk, F1-Skor ve Kappa gibi farklı metriklerle değerlendirilmiştir. Özellikle sınıflar arası dengesiz veri dağılımı nedeniyle bu çalışmada Kappa ve F-1 skor, doğruluk metriğinden daha anlamlı metrikler olarak değerlendirilmiştir.

2.19. Doğruluk

Doğruluk, Denklem (1)'de görüldüğü gibi sınıflandırma modelinin toplam örnekler içerisindeki doğru tahmin oranını ifade eder. En temel ve yaygın olarak kullanılan değerlendirme metriklerinden biridir.

$$Accuracy = Doğruluk = \frac{\text{Doğru Tahmin Sayısı}}{\text{Toplam Örnek Sayısı}} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Doğruluk metriği modelin genel başarımını özetleyen temel bir ölçüt olsa da özellikle sınıf dağılımının dengesiz olduğu durumlarda yanıltıcı sonuçlar verebilir. Çünkü model çoğunluk sınıfına aşırı odaklanarak yüksek doğruluk elde edebilirken azınlık sınıflarda ciddi hata yapabilir. Bu durum modelin genel başarı düzeyini olduğundan yüksek gösterir (Saito ve Rehmsmeier, 2015). Bu nedenle bu çalışmada doğruluk metriği tek başına kullanılmamış bunun yerine sınıf dengesizliğine daha duyarlı metrikler olan F1-skor ve QWK gibi alternatif ölçütlerle birlikte değerlendirme yapılmıştır.

2.20. F-1 Skor

F1-Skoru, Denklem (2)'de görüldüğü gibi bir sınıflandırma modelinin kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. Bu metrik, özellikle sınıf dağılımının dengesiz olduğu veri kümelerinde modelin yalnızca baskın sınıfları değil, azınlık sınıfları da ne ölçüde doğru tahmin edebildiğini göstermek açısından tercih edilir (Saito ve Rehmsmeier, 2015).

$$F - 1 \text{ Skor} = 2 \times \frac{\text{Kesinlik.Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}} \quad (2)$$

Çok sınıflı sınıflandırma problemlerinde her sınıf için ayrı ayrı F1 skoru hesaplanabilir. Ancak sınıflar arasında belirgin örnek sayısı farkları bulunduğu anda modelin azınlık sınıflardaki başarımını daha adil ölçmek için ağırlıklı ortalama (weighted-average) F1 skoru tercih edilir. Bu yöntem, Denklem (3)'te her sınıfın $F1_i$ değerini o sınıfa ait örnek sayısı ile n_i ağırlıklandırarak genel F1 skorunu hesaplar (Sokolova ve Lapalme, 2009):

$$F1_{\text{weighted}} = \sum_{i=1}^C \frac{n_i}{N} \cdot F1_i \quad (3)$$

Burada,

C: Sınıf sayısı,

n_i : i. sınıfa ait örnek sayısı,

$N = \sum_{i=1}^C n_i$: Toplam örnek sayısı,

$F1_i$: i. sınıfa ait F1 skoru.

Bu yaklaşım, her sınıfın katkısını örnek sayısına göre dengeleyerek çoğunluk ve azınlık sınıflara eşit önem atfeder ve dengesiz veri kümelerinde başarıyı daha gerçekçi yansıtır.

2.21. QWK

Bu çalışmada sınıflandırma problemi sıralı (ordinal) sınıflar içerdiği için modelin performansı yalnızca doğru/yanlış tahmin sayısına değil, tahminin ne kadar uzak sınıflara yapıldığına da duyarlı bir metrik olan QWK ile değerlendirilmiştir. QWK, Cohen'in Kappa istatistiğinin bir genellemesi olup Denklem (4)'te her sınıf hatasına karelenmiş uzaklığa dayalı bir ceza uygular (Ben-David, 2008; Cohen, 1968).

$$K = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}} \quad (4)$$

$O_{i,j}$: Gözlenen frekans matrisi (gerçek sınıf i, tahmin edilen sınıf j)

$E_{i,j}$: Beklenen frekans (rastgele tahmin varsayımıyla)

$W_{i,j}$: Ağırlık fonksiyonu Denklem (5)'te, sınıflar arası farkın karesine dayalıdır:

$$W_{i,j} = \frac{(i-j)^2}{(C-1)^2} \quad (5)$$

QWK değeri -1 ile 1 arasında değişir:

K =1 : Mükemmel uyum

K=0: Şansa bağlı tahmin

K<0: Rastgeleden daha kötü performans

Bu yaklaşım sayesinde örneğin gerçek sınıfı 4 olan bir sınıfın 3 olarak tahmin edilmesi, 0 olarak tahmin edilmesinden daha az cezalandırılır. Bu da sıralı etiketli görevlerde daha gerçekçi bir başarımlı ölçümü sağlar.

3. Bulgular ve Tartışma

Bu bölümde DR tespiti için farklı derin öğrenme modellerinin performansları, APTOS 2019 veri seti üzerinde çoklu ve ikili sınıflandırma görevleri kapsamında değerlendirilmiştir. Deneyler, öncelikle herhangi bir görüntü işleme veya veri artırma uygulanmadan gerçekleştirilmiş, ardından bu tekniklerin eklenmesiyle model performanslarındaki değişim incelenmiştir.

Tablo 2'de ön işleme ve veri artırma adımları olmadan modellerin çoklu sınıflandırma performansları sunulmaktadır. Bu sonuçlara göre, Swin Transformer modeli %83.67 doğruluk, %89.23 QWK ve %83.10 ağırlıklı F1 skoru ile diğer modellere göre üstün performans sergilemiştir.

EfficientNetB4 ve DenseNet121 modelleri de yüksek doğruluk ve QWK skorları ile Swin Transformer'ı takip etmiştir.

Tablo 2. Modellerin ön işleme ve artırma adımları olmadan çoklu sınıflandırma performansları.

Model	Doğruluk (%)	QWK (%)	Ağırlıklı F1 (%)
ResNet18	80.21	84.08	79.30
EfficientNetB4	83.45	87.65	83.02
VGG16	82.01	87.52	81.27
DenseNet121	82.08	86.92	81.02
Swin Transformer	83.67	89.23	83.10

Ön işleme ve veri artırma teknikleri uygulandıktan sonra elde edilen sonuçlar Tablo 3'te verilmiştir. Tüm modellerde hem doğruluk hem de QWK ve F1 skorlarında belirgin iyileşmeler gözlenmiştir. Özellikle Swin Transformer modelinde doğruluk %85.00'e, QWK %91.37'ye yükselmiş ve ağırlıklı F1 skoru %83.35 olarak ölçülmüştür. Bu artışlar, kullanılan görüntü işleme ve veri artırma yöntemlerinin model performansına olumlu katkısını göstermektedir.

Tablo 3. Modellerin ön işleme ve artırma adımları ile çoklu sınıflandırma performansları.

Model	Doğruluk (%)	QWK (%)	Ağırlıklı F1 (%)
ResNet18	81.65	88.03	81.50
EfficientNetB4	84.04	90.21	83.12
VGG16	83.42	89.39	82.23
DenseNet121	83.05	89.13	82.13
Swin Transformer	85.00	91.37	83.35

Tablo 4'te ise modellerin ikili sınıflandırma görevindeki performansları verilmiştir. Bu kategoride de veri artırma ve ön işleme sonrası sonuçlar daha yüksek başarı oranları ortaya koymaktadır. En yüksek doğruluk %98.41 ile VGG16 modelinde görülmüş, QWK skorlarında ise %96'nın üzerinde performans değerleri elde edilmiştir. Genel olarak tüm modeller ikili sınıflandırmada çoklu sınıflandırmaya kıyasla daha yüksek doğruluk ve QWK skorları göstermiştir.

Tablo 4. Modellerin ikili sınıflandırma performansları.

Model	Doğruluk (%)	Doğruluk (%) *	QWK (%)	QWK (%) *
ResNet18	97.21	97.62	94.41	95.24
EfficientNetB4	98.04	98.35	96.42	96.70
VGG16	98.36	98.41	96.71	96.82
DenseNet121	97.95	98.33	95.89	96.66
Swin Transformer	97.89	98.33	95.78	96.66

Not: * Veri artırma ve ön işleme uygulandıktan sonra elde edilen sonuçları ifade etmektedir.

Tablo 5'te APTOS 2019 veri seti kullanılarak yapılan farklı çalışmalarda kullanılan derin öğrenme modellerinin performans metriklerini karşılaştırmalı olarak gösterilmektedir. Majumder ve

Kehtarnavaz (2021) tarafından kullanılan SE-DenseNet201 modeli, %85 doğruluk ve %90 QWK skorları ile iyi bir temel performans sunmuştur. Bodapati ve arkadaşları (2021) Xception ve VGG16 modellerinin hibrit kombinasyonunu kullanarak %82 doğruluk ve %79 QWK elde etmiş, ancak F1 skoruna yer vermemiştir. Taufiqurrahman ve ark. (2021) ise MobileNetV2 ve SVM birleşimiyle %85 doğruluk ve %92 QWK değerlerine ulaşarak yüksek sınıflandırma tutarlılığı göstermiştir. Boulubi ve arkadaşları (2023) Swin Transformer mimarisini kullanarak %89 doğruluk ve %83 F1 skoru ile güçlü sonuçlar elde etmiş, bu da Transformer tabanlı modellerin etkinliğini ortaya koymuştur. Minarno ve Bagaskara (2025) InceptionV3 modeli ile veri artırma tekniklerini birleştirerek %83 doğruluk ve %82 F1 skoru seviyelerinde performans sağlamışlardır.

Tablo 5. Benzer çalışmaların APTOS 2019 veri seti üzerindeki performansı

Çalışma	Model	Doğruluk (%)	QWK (%)	F1 Skor (%)
Majumder ve Kehtarnavaz (2021)	SE-DenseNet201	85	90	72
Bodapati ve ark. (2021)	Xception & VGG16	82	79	-
Taufiqurrahman ve ark. (2021)	MobileNetV2 + SVM	85	92	-
Boulubi ve ark. (2023)	Swin Transformer	89	-	83
Minarno ve Bagaskara (2025)	InceptionV3, Veri artırma	83	-	82
Bu çalışma (2025)	Swin Transformer	85	91	84

Bu çalışmada ise Swin Transformer tabanlı model kullanılarak %85 doğruluk, yaklaşık %91.37 QWK ve %84 F1 skoru elde edilmiştir. Özellikle QWK ve F1 skorlarındaki iyileşme, modelin farklı DR evrelerini dengeli ve doğru şekilde sınıflandırma kapasitesini yansıtmaktadır. Ayrıca, Transformer tabanlı modellerin medikal görüntü sınıflandırmasında artan popülaritesi ve etkinliği bu sonuçlarla desteklenmektedir. Çalışmanın performansının literatürdeki benzer modellerle karşılaştırıldığında rekabetçi olduğu ve özellikle sınıf uyumu açısından üstünlük sağladığı söylenebilir.

4. Sonuçlar ve Öneriler

Bu çalışmada, DR tespiti amacıyla farklı derin öğrenme modelleri, APTOS 2019 veri seti üzerinde hem çoklu hem de ikili sınıflandırma görevleri kapsamında karşılaştırılmıştır. Elde edilen sonuçlar, özellikle Swin Transformer modelinin diğer modellerden daha yüksek doğruluk, QWK ve F1 skoru elde ettiğini göstermiştir. Ayrıca, görüntü işleme ve veri artırma tekniklerinin uygulanması model başarısını anlamlı ölçüde artırmıştır.

İkili sınıflandırma görevinde yüksek performans değerleri, derin öğrenme tabanlı yaklaşımların klinik uygulamalarda hastalığın varlığının hızlı ve doğru tespiti için uygun olduğunu desteklemektedir. Çoklu sınıflandırmada ise sınıf dengesizliği ve benzerlik gibi zorluklar göz önünde bulundurularak gelecekte bu alanlarda iyileştirmeler yapılması gerekliliği ortaya çıkmaktadır.

Çalışmanın bulguları ışığında öneriler şunlardır:

- **Model Birleştirme ve Topluluk Yöntemleri:** Bu çalışmada tekli modeller değerlendirilmiştir. Gelecekte en iyi performans gösteren modellerin bir topluluk (ensemble) oluşturacak şekilde birleştirilmesi düşünülebilir. Farklı mimarilerin hataları farklı olabileceğinden bir oylama veya ağırlıklı ortalama mekanizması ile birleştirme yapmak genel doğruluğu daha da artırabilir. Özellikle sağlık uygulamalarında kritik hataları en aza indirmek için ensemble modelleri faydalı olabilir.
- **Daha Geniş ve Çeşitli Veri Setleri:** Modelin genellenebilirliğini artırmak için eğitim veri setinin genişletilmesi önemlidir. Mevcut APTOS 2019 veri setine ek olarak EyePACS gibi binlerce görüntü içeren açık veri setleri ile eğitimin tekrarlanması modelin farklı popülasyonlarda da başarılı olmasını sağlayacaktır. Gelecekte mümkün olan en geniş kapsamlı retina görüntü koleksiyonuyla model yeniden eğitilebilir.
- **Birleşik Modeller ve Çoklu Çıkış:** DR tanısı koymanın yanı sıra belirli lezyon türlerinin tespiti de (mikro anevrizma sayısı, makula ödemi varlığı gibi) klinik açıdan önemlidir. Gelecekte çok-görevli (multi-task) bir öğrenme yaklaşımıyla aynı modelin hem DR seviyesini sınıflandırması hem de varsa ilgili lezyonları tespit etmesi sağlanabilir.
- **Model Açıklanabilirliği:** Model kararlarının görsel veya sayısal açıklamaları önem taşır. Gelecekte eğitilmiş modellere Grad-CAM gibi teknikler uygulanarak hangi bölgeleri önemsedikleri ortaya koyulabilir. Eğer model yanlış bir sınıflandırma yaptıysa Grad-CAM ısı haritaları hatanın nedenini anlamaya yardımcı olabilir (örneğin optik diski bir lezyon sanmak gibi).

Sonuç olarak, derin öğrenme tabanlı DR sınıflandırma modelleri, erken teşhis ve tedavi sürecine katkı sağlayarak görme kayıplarının önlenmesinde önemli bir araç olarak kullanılabilir.

Teşekkür

Bu çalışma, Atatürk Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü (BAP) tarafından, FYL-2024-14050 numaralı proje kapsamında desteklenmiştir.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Akhtar, S., Aftab, S., Ali, O., Ahmad, M., Khan, M. A., Abbas, S., & Ghazal, T. M. (2025). A deep learning based model for diabetic retinopathy grading. *Scientific Reports*, 15(1), 3763.
- Antonetti, D. A., Klein, R., & Gardner, T. W. (2012). Diabetic retinopathy. *New England Journal of Medicine*, 366(13), 1227–1239. <https://doi.org/10.1056/NEJMra1005073>
- Arora, L., Singh, S. K., Kumar, S., Gupta, H., Alhalabi, W., Arya, V., ... & Gupta, B. B. (2024). Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy. *Scientific Reports*, 14(1), 30554.
- Batool, S., Gilani, S. O., Waris, A., Iqbal, K. F., Khan, N. B., Khan, M. I., ... & Awwad, F. A. (2023). Deploying efficient net batch normalizations (BNs) for grading diabetic retinopathy severity levels from fundus images. *Scientific Reports*, 13(1), 14462.
- Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications*, 34(2), 825-832.
- Bodapati, J. D., Shaik, N. S., & Naralasetti, V. (2021). Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. *Journal of Ambient Intelligence and Humanized Computing*, 12(10), 9825-9839.
- Boulaabi, M., Gader, T. B. A., Echi, A. K., & Bouraoui, Z. (2025). Enhancing DR Classification with Swin Transformer and Shifted Window Attention. *arXiv preprint arXiv:2504.15317*.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4), 213.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houshy, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Fong, D. S., Aiello, L., Gardner, T. W., King, G. L., Blankenship, G., Cavallerano, J. D., Ferris, F. L. ve Klein, R., 2004. Retinopathy in diabetes. *Diabetes Care*, 27(Suppl 1), S84-S87.
- Elzennary, A., Soliman, M., & Ibrahim, M. (2020, November). Early deep detection for diabetic retinopathy. In *2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (pp. 1-5). IEEE.
- Gonzalez, R. C. (2009). *Digital image processing*. Pearson education india.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., & others. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402-2410.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). Brussels, Belgium: International Diabetes Federation. <https://www.diabetesatlas.org>
- İncir, R., & Bozkurt, F. (2024). A study on effective data preprocessing and augmentation method in diabetic retinopathy classification using pre-trained deep learning approaches. *Multimedia Tools and Applications*, 83(4), 12185-12208.
- Kaggle. (2019). *APTOS 2019 Blindness Detection* [Veri seti]. <https://www.kaggle.com/competitions/aptos2019-blindness-detection>
- Lavanya, R. V., Sumesh, E. P., Jayakumari, C., & Isaac, R. (2020, November). Detection and classification of diabetic retinopathy using raspberry PI. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1688-1691). IEEE.
- LeCun, Y., Bengio, Y. ve Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436-444.

- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).
- Majumder, S., & Kehtarnavaz, N. (2021). Multitasking deep learning model for detection of five stages of diabetic retinopathy. *IEEE Access*, 9, 123220-123230.
- Masters, D., & Luschi, C. (2018). Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*.
- Minarno, A. E., Bagaskara, A. D., Bimantoro, F., & Suharso, W. (2025). Classification of Diabetic Retinopathy Based on Fundus Image Using InceptionV3. *JOIV: International Journal on Informatics Visualization*, 9(1), 23-28.
- Nguyen, Q. H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., Nguyen, B. P., & Chua, M. (2020, January). Diabetic retinopathy detection using deep learning. In *Proceedings of the 4th international conference on machine learning and soft computing* (pp. 103-107).
- Pamadi, A. M., Ravishankar, A., Nithya, P. A., Jahnavi, G., & Kathavate, S. (2022, March). Diabetic retinopathy detection using mobilenetv2 architecture. In *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)* (pp. 1-5). IEEE.
- Patel, R., & Chaware, A. (2020, June). Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy. In *2020 international conference for emerging technology (INCET)* (pp. 1-4). IEEE.
- Rao, M., Zhu, M., & Wang, T. (2020). Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy. *arXiv preprint arXiv:2010.11692*.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3), e0118432.
- Saranya, P., Devi, S. K., & Bharanidharan, B. (2022, March). Detection of diabetic retinopathy in retinal fundus images using densenet based deep learning model. In *2022 international mobile and embedded technology conference (MECON)* (pp. 268-272). IEEE.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., ... & Magliano, D. J. (2022). IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research and clinical practice*, 183, 109119.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- Taufiqurrahman, S., Handayani, A., Hermanto, B. R., & Mengko, T. L. E. R. (2020, November). Diabetic retinopathy classification using a hybrid and efficient MobileNetV2-SVM model. In *2020 IEEE Region 10 Conference (Tencon)* (pp. 235-240). IEEE.
- Thota, N. B., & Reddy, D. U. (2020, August). Improving the accuracy of diabetic retinopathy severity classification with transfer learning. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1003-1006). IEEE.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- Yang, J., Shi, R., & Ni, B. (2021, April). Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 191-195). IEEE.
- Yang, Y., Cai, Z., Qiu, S., & Xu, P. (2024). Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image. *Plos one*, 19(3), e0299265.
- Yau, J. W., Rogers, S. L., Kawasaki, R., Lamoureux, E. L., Kowalski, J. W., Bek, T., ... & Meta-Analysis for Eye Disease (META-EYE) Study Group. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes care*, 35(3), 556-564.
- Yi, S. L., Yang, X. L., Wang, T. W., She, F. R., Xiong, X., & He, J. F. (2021). Diabetic retinopathy diagnosis based on RA-EfficientNet. *Applied Sciences*, 11(22), 11035.

- Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Graphics Gems IV (pp. 474–485). Academic Press. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- Zhang, P., Zhang, X., Brown, J., Vistisen, D., Sicree, R., Shaw, J., & Nichols, G. (2010). Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes research and clinical practice*, 87(3), 293-301.
- World Health Organization. (2020). *WHO/Europe is focusing on eye screening for people with diabetes*. <https://www.who.int/europe/news/item/13-11-2020-who-europe-is-focusing-on-eye-screening-for-people-with-diabetes> (Eriřim Tarihi: 22.11.2024).
- World Health Organization. (2023). *Blindness and vision impairment*. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (Eriřim Tarihi: 22.11.2024).