



DERİN ÖĞRENME TEKNİKLERİ KULLANARAK RETİNA HASTALIĞI SINIFLANDIRMASI

RETINAL DISEASE CLASSIFICATION USING DEEP LEARNING TECHNIQS

DOI: 10.20854/bujse.1612453

Fatma Ayan¹ Soydan Serttaş^{2,*}

Öz

Bu çalışma, derin öğrenme tekniklerini kullanarak retina hastalıklarının otomatik sınıflandırılmasını ve teşhis süreçlerini iyileştirmeyi amaçlamaktadır. Retina hastalıkları, özellikle diyabetik retinopati, yaşa bağlı maküler dejenerasyon (AMD), glaukom ve retina damar tıkanıklığı, dünya genelinde görme kaybının başlıca nedenleri arasındadır. Bu hastalıkların erken teşhisi ve doğru sınıflandırılması, görme kaybını önlemek adına kritik bir öneme sahiptir. Derin öğrenme tabanlı yaklaşımlar, insan faktörüne bağlı teşhis hatalarını azaltarak yüksek doğruluk oranları sunmakta ve retina görüntüleme yöntemlerinin etkinliğini artırmaktadır.

Çalışmada, evrimsel sinir ağları (CNN) ve transfer öğrenme teknikleri kullanılarak retina hastalıklarının sınıflandırılmasının karşılaştırılması gerçekleştirilmiştir. Fundus ve optik koherens tomografi (OCT) görüntüleri üzerinde yapılan analizler, yüksek doğruluk oranlarıyla bu yöntemlerin etkili olduğunu ortaya koymaktadır. Ağlar, doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metriklerle değerlendirilmiş ve klinik uygulamalardaki potansiyelleri irdelenmiştir.

Araştırma sonuçları, derin öğrenme yöntemlerinin, retina hastalıklarının erken teşhisinde hız ve doğruluk gibi konularda katkı sunduğunu göstermektedir. Özellikle CNN tabanlı ağların performansı, uzman teşhis süreçlerini destekleyerek görme kaybını önlemeye yönelik önemli bir katkı sağlamaktadır. Bu çalışma, tıbbi görüntüleme teknolojilerinde derin öğrenmenin kullanımına dair yeni bir perspektif sunmakta ve sağlık profesyonellerinin iş yükünü azaltacak çözüm önerileri ortaya koymaktadır.

Abstract

This study aims to complete the automatic classification and diagnosis processes of retinal diseases using deep learning technics. Retinal diseases, especially diabetic retinopathy, age-related macular degeneration (AMD), glaucoma and retinal vascular fragments, are among the main causes of vision loss worldwide. Early diagnosis and correct classification of this disease is critical for preventing vision loss. Deep learning-based technics offer high accuracy rates in calculating diagnostic errors due to human factors and the procedures of retinal imaging methods are increased.

In the study, the classification of retinal diseases was carried out using convolutional neural networks (CNN) and transfer learning technics. Analyses performed on fundus and optical coherence tomography (OCT) images reveal that these methods are effective with high accuracy rates. The obtained networks were evaluated with metrics such as accuracy, precision, recall and F1 score and their potential in clinical applications was examined.

The research results show that deep learning technics can provide contributions such as accuracy, precision and reproducibility in the early diagnosis of retinal diseases. In particular, the performance of CNN-based technics provides a significant contribution to increasing their losses by supporting expert diagnosis processes. This study presents a new perspective on the use of deep learning in treatment imaging technologies and suggests solutions that will reduce the workflow of healthcare professionals.

Anahtar Kelime: derin öğrenme, retina hastalıkları, fundus görüntüleme, transfer öğrenme, CNN

Keywords: deep learning, retinal diseases, fundus imaging, transfer learning, CNN

¹Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Bölümü, fatmaoztanir@gmail.com, orcid.org/0009-0009-9750-5788

^{2,*} Sorumlu Yazar: Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Bölümü, soydan.serttas@dpu.edu.tr, orcid.org/0000-0001-8887-8675

1. GİRİŞ

Retina, gözün iç yüzeyinde yer alan ve ışığı algılayarak görme işlemini başlatan hayati bir dokudur. Retina hastalıkları, bu dokunun yapısında ve işlevinde bozulmalara yol açarak ciddi görme kaybına neden olabilir. Özellikle diyabetik retinopati, yaşa bağlı maküler dejenerasyon, glokom ve retina damar tıkanıklığı gibi hastalıklar, retina kaynaklı görme kaybının başlıca nedenleri arasında yer almaktadır. Dünya Sağlık Örgütü'nün (WHO) raporlarına göre, dünya genelinde yaklaşık 2,2 milyar insan bir tür görme bozukluğundan etkilenmekte olup, bu vakaların büyük bir kısmı önlenemez veya tedavi edilebilir durumdadır (World Health Organization, 2019).

Retina hastalıkları genellikle erken dönemde belirgin semptomlar göstermediğinden, teşhis ve tedavide gecikmeler yaşanabilmektedir. Bu nedenle, erken tanı ve doğru sınıflandırma, görme kaybını önlemek adına büyük önem taşımaktadır. Gelişen teknolojilerle birlikte optik koherens tomografi, fundus fotoğrafçılığı ve fluorescein anjiyografi gibi görüntüleme yöntemleri retina hastalıklarının teşhisinde yaygın olarak kullanılmaktadır. Retina görüntülerinin analizinde insan uzmanlara bağlı teşhislerde çeşitli faktörler nedeniyle hata oranları gözlemlenmektedir. Örneğin, araştırmalar retina hastalıklarının teşhisinde göz uzmanları arasında önemli ölçüde değişkenlik bulunduğunu ve bazı durumlarda teşhis doğruluğunun %65 ile %85 arasında değişebileceğini göstermiştir (Ting et al., 2017). Bu hataların başlıca nedenleri arasında görüntüleme kalitesindeki farklılıklar, uzmanların deneyim düzeyindeki çeşitlilik ve yorgunluk gibi insan faktörleri yer almaktadır. Bu bağlamda, yapay zeka ve derin öğrenme tabanlı yaklaşımlar, hata oranlarını azaltarak tutarlı ve hızlı teşhisler sağlamada önemli bir çözüm sunmaktadır. Bununla birlikte, retina görüntülerindeki karmaşıklık ve insan faktörüne bağlı teşhis hatalarını azaltmak amacıyla yapay zeka ve derin öğrenme tabanlı yaklaşımlar giderek daha fazla önem kazanmaktadır (Gulshan et al., 2016). Bu kapsamda, retina hastalıklarının sınıflandırılmasında kullanılan derin öğrenme yöntemleri, görme kaybını önlemeye yönelik tedavi süreçlerini hızlandırmakta ve göz sağlığı alanında yeni ufuklar açmaktadır.

Son yıllarda, yapay zeka ve özellikle derin öğrenme tabanlı yaklaşımlar, retina hastalıklarının sınıflandırılmasında büyük bir ilerleme sağlamıştır. Özellikle evrimsel sinir ağları gibi derin öğrenme yöntemleri, retina görüntülerini otomatik olarak analiz ederek hastalıkları yüksek doğruluk oranlarıyla sınıflandırabilmektedir (Leibig et al., 2017). Bu yöntemler, retina hastalıklarının farklı evrelerini tanımlamak için hem tanı koyma sürecini hızlandırmakta hem de uzmanlar arasındaki teşhis farklılıklarını minimize etmektedir.

Derin öğrenme, makine öğrenmesi alanının bir alt dalı olarak çok katmanlı sinir ağları kullanarak büyük miktarda veriden öğrenme yeteneğine sahiptir. CNN'ler, görüntü işleme alanında yüksek doğruluk oranlarına ulaşmasıyla bilinir ve tıbbi görüntüleme gibi karmaşık uygulamalarda sıklıkla kullanılmaktadır (Litjens et al., 2017). Bu bağlamda, derin öğrenme tabanlı sistemler, doktorların karar verme süreçlerini destekleyen güçlü bir araç haline gelmektedir (Gulshan et al., 2016; Esteva et al., 2017).

2. RETİNA HASTALIKLARININ SINIFLANDIRILMASINDA DERİN ÖĞRENME YAKLAŞIMLARI

Son yıllarda, yapay zeka ve derin öğrenme tabanlı yöntemler, retina görüntülerinin analizinde kullanılmaya başlanmıştır. Bu yöntemler, fundus fotoğrafları ve OCT görüntüleri gibi verilerden otomatik olarak öğrenerek hastalıkların sınıflandırılmasında yüksek doğruluk oranları sunmaktadır (Gulshan et al., 2016). Derin öğrenme tabanlı algoritmalar, diyabetik retinopatinin farklı evrelerini ayırt etmek ve AMD'nin kuru ve ıslak formlarını sınıflandırmak için de kullanılmaktadır (Ting et al., 2017).

Retina hastalıkları sınıflandırma süreçleri, retina görüntülerinden elde edilen verilerin analizine dayanır. Fundus fotoğrafları, mikroanevrizmalar, kanamalar, eksüdalar ve diğer patolojik özelliklerin tanımlanmasında kullanılırken OCT retina tabakasının mikroskobik düzeyde incelenmesini sağlar (Gulshan et al., 2016; Ting et al., 2017). Özellikle OCT görüntüleri, yaşa bağlı maküler dejenerasyon gibi hastalıklarda kuru ve ıslak tiplerin ayırımında kritik bilgiler sunar (Spaide et al., 2008). Ayrıca, fundus autofloresans görüntüleme, retinal pigment epiteli (RPE) hastalıklarını tespit etmek için önemlidir. Görme alanı testleri,

hastaların görsel fonksiyonunu değerlendirmek için kullanılır. (Brown et al., 2019; Wong et al., 2020). Diabetik retinopati, yaşa bağlı AMD ve retinal damar tıkanıklığı gibi hastalıkların erken tespiti için CNN'ler, görüntü işleme ve sınıflandırma alanında büyük başarılar elde etmiştir (Gulshan et al., 2016). Bunun yanı sıra, derin öğrenme ağları arasında U-Net gibi mimariler, segmentasyon görevlerinde kullanılarak, retina yapılarını doğru bir şekilde ayırmada etkili olmuştur (Ronneberger et al., 2015).

Gulshan ve arkadaşları (2016), diabetik retinopatiyi sınıflandırmak için bir CNN yapısı kullanmış ve ağın doğruluğunu %90'ın üzerinde bir başarıyla doğrulamıştır. Çalışmada kullanılan yöntem, fundus görüntülerini analiz ederek, hastalığın erken aşamalarını tespit etmeye yönelik oldukça etkili bir araç sunmuştur.

Ronneberger ve arkadaşları (2015), U-Net'in biyomedikal görüntü segmentasyonundaki başarısını göstererek, retina hastalıklarının daha hassas bir şekilde tanımlanmasını sağlamıştır. U-Net tabanlı yöntem, özellikle damar yapılarının ve lezyonların doğru bir şekilde segment edilmesi konusunda yüksek başarı oranları elde etmiştir. Ayrıca, Lee ve arkadaşları (2017), yaşa bağlı AMD gibi retinal hastalıkları sınıflandırırken, CNN'lerin yanı sıra transfer öğrenmeyi de kullanmış ve %94 doğruluk oranı sağlamıştır.

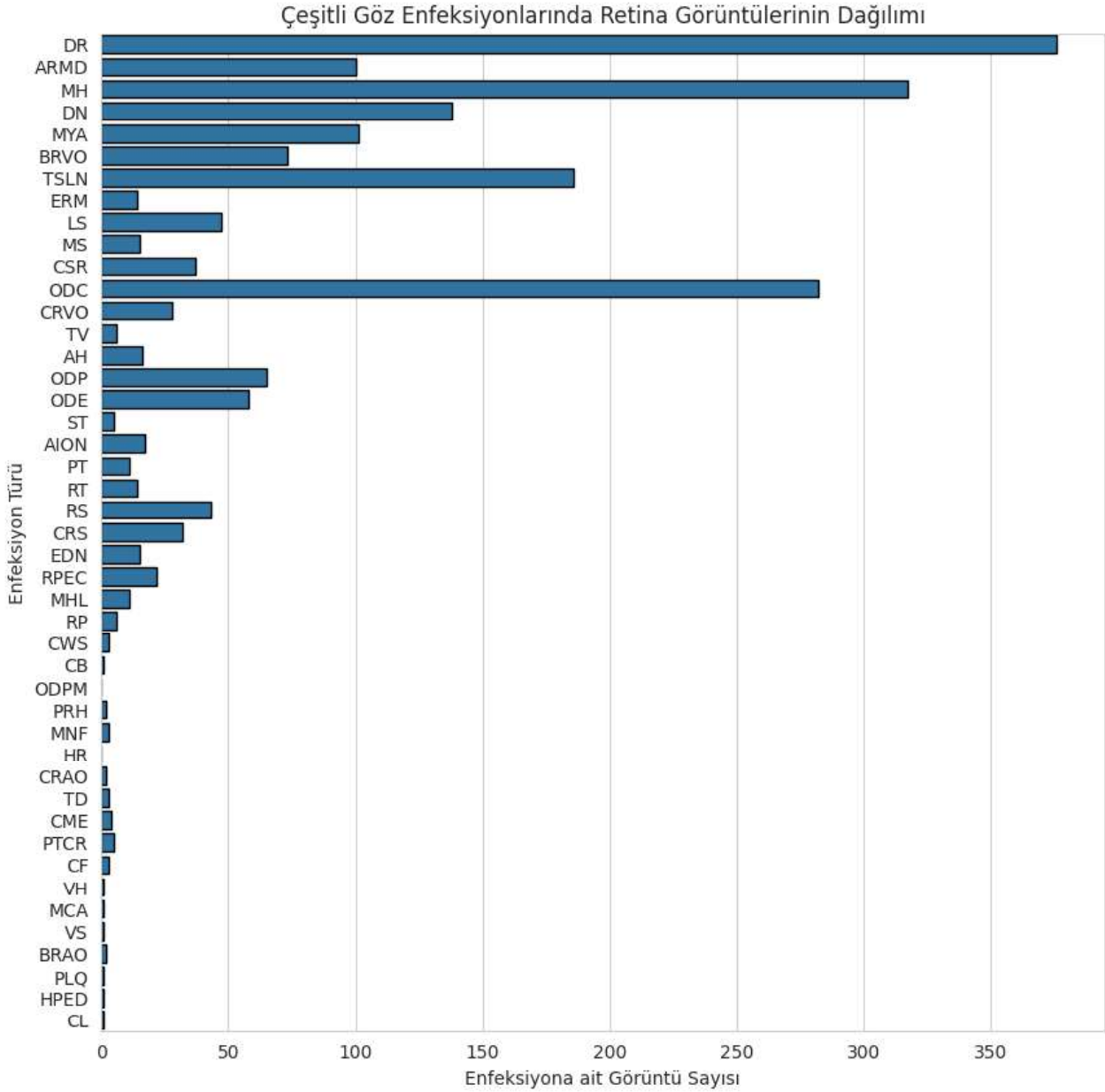
Deye ve arkadaşları (2020), çeşitli görüntüleme tekniklerinden (fundus fotoğrafları, OCT ve fluorescein anjiyografi) elde edilen verileri birleştirerek, derin öğrenme ağlarının doğruluğunu %92 oranına çıkarmıştır.

3. MATERYAL VE YÖNTEM

3.1. Veri Seti

RFMID (Retinal Fundus and Multimodal Imaging Database), retina hastalıklarının sınıflandırılması ve tespiti için kullanılan önemli bir tıbbi görüntü veri setidir. Bu veri seti, fundus görüntüleri, OCT, fluorescein anjiyografi ve diğer multimodal görüntüleme verilerini içeren bir koleksiyondur. Bu veri setinin kullanılması, retina hastalıklarının otomatik tespiti ve sınıflandırılmasında önemli bir adımdır ve klinik ortamda uygulama için büyük potansiyel taşır (Zhang et al., 2021). RFMID, retinal fundus görüntülerinden oluşan ve 45 farklı hastalık veya patolojiyi içeren kapsamlı bir veri setidir. Veri seti, toplamda 3.200 renkli fundus görüntüsünden oluşmakta olup, eğitim, doğrulama ve test setlerine ayrılmıştır. Her görüntü, normal veya anormal (hastalık) durumlarını belirten etiketlere sahiptir. Veri seti içerisinde sıra numarası, hastalık olup olmadığı, hastalığın cinsini belirten 47 sütundan oluşmaktadır. RFMID veri setine "RIADD (ISBI-2021) Grand Challenge" platformu üzerinden erişilebilmektedir. Bu veri seti, retinal fundus görüntüleri üzerinde çoklu hastalık tespiti çalışmaları gerçekleştirmek isteyen araştırmacılar için önemli bir kaynak sunmaktadır (RIADD, 2021).

Şekil 1 RFMID veri seti içindeki hastalık dağılımını göstermektedir. Veri setinde belirgin bir sınıf dengesizliği olması sebebiyle örnek sayısı 300 ve üzerindeki hastalıklar seçilip bunun altındaki sınıflar "Diğer" kategorisinde gruplanmıştır. Çok etiketli sınıflandırma için yalnızca en yaygın görülen iki hastalık Diabetic Retinopathy ve Macular Hole (DR-MH) dikkate alınmıştır.



Şekil 1: RFMID Veri Seti İçindeki Hastalık Dağılımı.

Bu veri seti kullanılarak, öncelikle hastalık varlığı ve yokluğu arasında bir sınıflandırma yapılmıştır. Sınıflandırma işlemi, veri setindeki "Disease_Risk" sütunundaki değerlere dayanarak, hastalık varlığı ve yokluğu arasındaki dağılım inceleyerek gerçekleştirilmiştir. Bu makalede toplamda üç farklı ağ oluşturulmuştur. Oluşturulan iki ağ yalnızca hasta/hasta değil sınıflandırması yaparken üçüncü ağ veri setinde en çok örnek içeren iki hastalığı (DR ve MH) esas alarak bu ikisini tespit etmiştir (Multilabel Classification).

Veri seti eğitim için %60, doğrulama için %20, test için %20 olarak ayrılmıştır. Eğitim için ayrılan kısımda 1920 veri, doğrulama için ayrılan kısımda 640 veri, test için ayrılan kısımda ise 640 veri bulunmaktadır.

3.2. Performans Parametreleri

- Kesinlik (Precision): Gerçek Pozitiflerin, toplam pozitif tahminlere oranıdır. Bu metrik, ağın yanlış pozitif (false positive) tahminlerini ne kadar iyi kontrol edebildiğini göstermektedir (Denklem 1).
- Duyarlılık (Recall): Gerçek Pozitiflerin, toplam pozitif gerçek değerlere oranıdır. Duyarlılık, yanlış negatif (false negative) tahminlerin etkisini değerlendirmek için önemli bir metriktir (Denklem 2).
- F1-skoru (F1-Score), duyarlılık ve kesinlik değerlerinin harmonik ortalamasıdır (Denklem 3).

$$F1\ Skor = \frac{(2 \times Kesinlik \times Duyarlilik)}{(Kesinlik + Duyarlilik)}$$

Denklem (1)

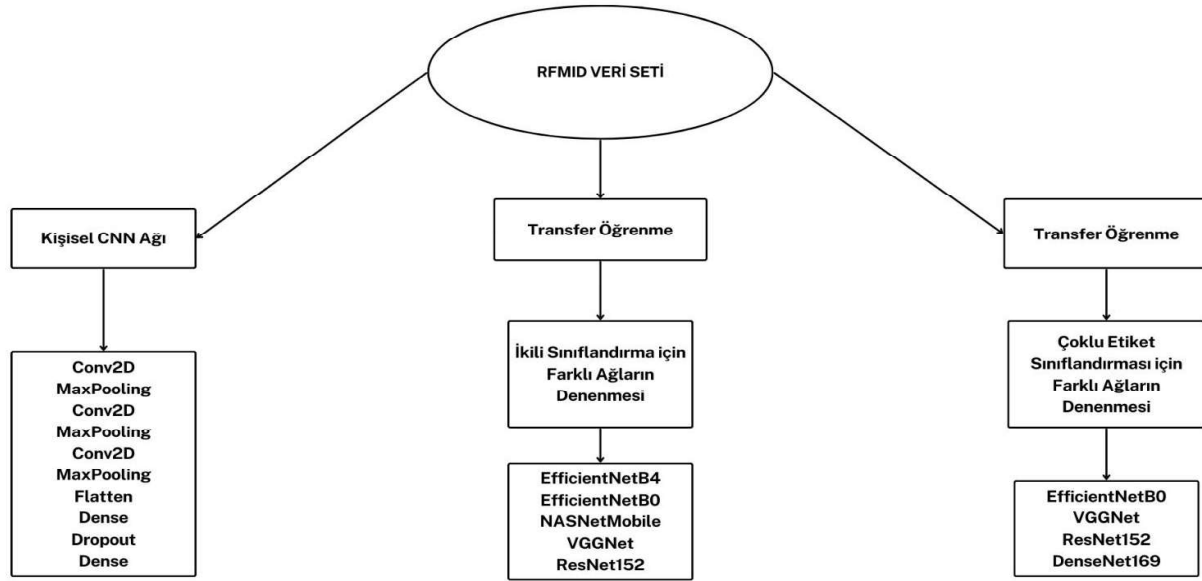
$$Kesinlik = \frac{Gerçek\ Pozitifler}{(Gerçek\ Pozitifler + Yanlış\ Pozitifler)}$$

Denklem (2)

$$Duyarlilik = \frac{Gerçek\ Pozitifler}{(Gerçek\ Pozitifler + Yanlış\ Negatifler)}$$

Denklem (3)

3.3. Kullanılacak Ağ Tasarımı



Şekil 2: Veri Seti Üzerinde Çalışılan Ağ ve Tekniklerin Akış Diyagramı.

Şekil 2 RFMID veri kümesi kullanılarak gerçekleştirilen derin öğrenme çalışmasının genel yapısını göstermektedir. Çalışma, iki ana yaklaşımı içermektedir: kişisel bir CNN ağı geliştirme ve transfer öğrenimi kullanımı. Kişisel CNN Ağı başlığında, temel bileşenler sırasıyla Conv2D, MaxPooling, Flatten, Dense ve Dropout katmanları ile modellenmiştir. Bu yaklaşım, özel bir derin öğrenme ağı tasarımı ve eğitimi üzerine odaklanmaktadır. Transfer Öğrenimi yaklaşımı ise iki alt gruba ayrılmaktadır: ikili sınıflandırma ve çoklu etiket sınıflandırma. Her iki sınıflandırma türü için EfficientNet (B4 ve B0), NASNetMobile, VGGNet, ResNet152 ve DenseNet169 gibi çeşitli önceden eğitilmiş ağ yapıları kullanılarak deneyler gerçekleştirilmiştir. Bu süreçte, farklı ağ yapısı performanslarının değerlendirilmesi amaçlanmıştır.

3.4. Kullanılan Transfer Öğrenme Teknikleri

Transfer öğrenimi (transfer learning), bir ağı bir görevde öğrendiği bilgileri başka bir göreve uyarlamasını sağlayan makine öğrenimi tekniğidir. Bu teknik, özellikle büyük veri setlerine ve hesaplama kaynaklarına erişimin sınırlı olduğu durumlarda, daha küçük veri setleriyle etkili ağlar oluşturmak için kullanılır. Transfer öğreniminde genellikle, önceden büyük ve genel bir veri seti üzerinde eğitilmiş bir ağ yapısı alınır ve bu ağ, hedef görev için yeniden eğitilir (fine-tuning) veya doğrudan kullanılır. Bu yaklaşım, derin öğrenme ağlarının eğitim süresini kısaltırken performansı artırır ve bilgisayarla görüden doğal dil işleme kadar geniş bir alanda başarıyla uygulanmaktadır (Pan ve Yang, 2010). Bu makalede de bu yöntemden faydalanılarak EfficientNetB4, EfficientNetB0, NASNetMobile, VGGNet, ResNet152, DenseNet169 ağları kullanılmıştır.

EfficientNet (B0 ve B4) ağları, Tan ve Le (2019) tarafından geliştirilen ve derin öğrenme tekniği

ölçeklendirmesi için "Compound Scaling" yöntemini kullanan bir ağ mimari ailesidir. Bu teknik, ağın genişliğini, derinliğini ve çözünürlüğünü birlikte optimize ederek daha verimli bir ağ yapısı sunar. EfficientNet-B0, ağ ailesinin temel versiyonu olup daha küçük hesaplama kaynaklarıyla çalışabilirken, EfficientNet-B4 daha karmaşık bir yapıya sahiptir ve doğruluk açısından daha yüksek potansiyel sunmaktadır. Bu ağ yapıları, sınıflandırma problemlerinde yüksek performansı ve düşük hesaplama maliyetleri ile öne çıkmaktadır (Tan & Le, 2019).

NASNetMobile, sınırlı kaynaklara sahip cihazlarda verimli çalışacak şekilde optimize edilmiş bir ağ yapısıdır. Google tarafından geliştirilen bu yapı, "Neural Architecture Search (NAS)" mimarisi kullanılarak tasarlanmıştır. NASNetMobile, mobil cihazlar gibi düşük hesaplama gücüne sahip ortamlar için ideal bir seçimdir ve parametre verimliliği ile dikkat çekmektedir (Zoph et al., 2018).

VGGNet, Simonyan ve Zisserman (2014) tarafından derin yapıli bir sinir ağı olarak geliştirilmiştir. 16 veya 19 katmanlı versiyonlarıyla (örneğin VGG16 ve VGG19) bilinen bu ağ yapısı, ardışık küçük (3×3) filtreler kullanarak basit ancak etkili bir mimariye sahiptir. VGGNet, sınıflandırma problemlerinde sağlam performansı ve genelleme yeteneğiyle tanınır (Simonyan & Zisserman, 2014).

ResNet (152 katmanlı), He ve arkadaşları (2015) tarafından geliştirilen ve "residual connections" kullanarak derin ağlardaki kaybolan gradyan problemini çözen bir yapıdır. ResNet-152, 152 katmanlı yapısıyla derin öğrenme teknikleri arasında öne çıkar ve yüksek doğruluk gerektiren görevlerde etkili sonuçlar sağlar. ResNet'in başarısı, gradyanların daha verimli şekilde yayılmasını sağlaması ve derin ağları eğitilebilir hale getirmesinden kaynaklanmaktadır (He et al., 2016).

DenseNet (DenseNet169), Huang ve arkadaşları (2017) tarafından önerilen ve her katmanın tüm önceki katmanlarla bağlantılı olduğu bir CNN yapısıdır. Bu yoğun bağlantılar, parametre verimliliğini artırırken bilgi ve gradyanların daha iyi iletilmesine olanak tanır. DenseNet169, özellikle veri yetersizliği olan durumlarda etkili sonuçlar üretmek için uygundur ve yüksek doğruluk gerektiren görevlerde yaygın olarak kullanılmaktadır (Huang et al., 2017).

4. BULGULAR

4.1. Oluşturulan CNN Ağ Yapısının Performansının Değerlendirilmesi

Bu çalışmada sınıflandırma, iki sınıflı bir veri kümesinde test edilmiştir. Ağ başarımını değerlendirmek için kesinlik, duyarlılık, F1 skor ve doğruluk gibi temel performans metrikleri kullanılmıştır. Aşağıda, bu metrikler ve sonuçların yorumları detaylı bir şekilde ele alınmaktadır:

Tablo 1: Oluşturulan CNN ağı performans değerlendirilmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
0 (y_test)	0,85	0,38	0,53	134
1 (y_pred)	0,86	0,98	0,92	506
Doğruluk	-	-	0,86	640
Makro ortalama	0,85	0,68	0,72	640
Ağırlıklı ortalama	0,86	0,86	0,83	640

Tablo 1'e bakıldığında duyarlılık değerleri 0,38 ve 0,98'dir. Bu, sınıf 1'e ait örneklerin %98'inin doğru bir şekilde sınıflandırıldığını göstermektedir. Ağ, sınıf 1 için oldukça başarılı bir performans sergilemektedir. F1 skor değerleri 0,53 ve 0,92 olarak bulunmuştur. Bu değerler, ağın sınıf 1 üzerindeki tahminlerinde oldukça başarılı olduğunu göstermektedir.

Makro ortalama (Macro Average), her bir sınıfın metriklerinin eşit ağırlıkta ortalamasını alır (Kesinlik: 0,85, Duyarlılık: 0,68, F1 skor: 0,72). Bu değerler, sınıflar arasında performans dengesizliklerinin olduğunu

göstermektedir.

Ağırlıklı ortalama (Weighted Average), sınıfların veri sayısı değerleriyle ağırlıklandırılmış metrik ortalamalarını ifade eder (Kesinlik: 0,86, Duyarlılık: 0,86, F1 skor: 0,83). Ağırlıklı ortalama, veri kümesindeki sınıf dengesizliklerini dikkate alarak daha genel bir başarı değerlendirmesi sunar. Bu metrikler, ağın genel olarak iyi bir performans sergilediğini göstermektedir.

Bu sonuçlar, ağın sınıf 1'i tespit etmede oldukça başarılı olduğunu ancak sınıf 0 üzerinde geliştirilmesi gereken yönleri bulunduğunu göstermektedir. Özellikle sınıf 0 için duyarlılık değerinin düşük olması, ağ yapısının bu sınıfta yanlış negatif sonuçlar üretmeye daha yatkın olduğunu ortaya koymaktadır. Veri kümesindeki sınıf dengesizliği, bu performans farkının ana nedenlerinden biridir.

4.2. İkili Sınıflandırmada Transfer Öğrenme Performans Değerlendirmesi

4.2.1. EfficientNetB4 ağı

Tablo 2: EfficientNetB4 ağı performans değerlendirme.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,85	0,38	0,53	134
Hasta	0,86	0,98	0,92	506
Doğruluk	-	-	0,86	640
Makro ortalama	0,85	0,68	0,72	640
Ağırlıklı ortalama	0,86	0,86	0,83	640

Tablo 2'ye bakıldığında sınıflar arasındaki veri dağılımına göre ağırlıklandırılmış ortalamalar kesinlik, duyarlılık ve F1 skor için sırasıyla 0,87, 0,85 ve 0,85'tir. Ağırlıklı ortalamalar, ağın genel performansını daha iyi yansıttığı için makro ortalamalardan daha anlamlıdır. Ağ genel olarak iyi bir performans sergilemiş olsa da, "Sağlıklı" sınıfındaki performansın artırılması daha dengeli bir sınıflandırma için gereklidir.

4.2.2. EfficientNetB0 ağı

Tablo 3: EfficientNetB0 ağı performans değerlendirme.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,68	0,43	0,52	134
Hasta	0,86	0,95	0,90	506
Doğruluk	-	-	0,84	640
Makro ortalama	0,77	0,69	0,71	640
Ağırlıklı ortalama	0,82	0,84	0,82	640

Tablo 3'e bakıldığında ağı duyarlılık değerine bakıldığında "Hasta" olanların %95'ini doğru bir şekilde tespit etmiştir. Bu, enfekte olanları tespit etme oranının oldukça yüksek olduğunu gösterir. Makro ortalamaya bakıldığında duyarlılık (0,69) düşük olduğundan, "Sağlıklı" sınıfındaki zayıf performansı yansıtmaktadır. Sınıflar arasındaki veri dağılımına göre ağırlıklandırılmış ortalamalar kesinlik, duyarlılık ve F1 skor için sırasıyla 0,82, 0,84 ve 0,82'dir.

4.2.3. NASNetMobile ağı

Tablo 4: NASNetMobile ağı performans değerlendirme.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,00	0,00	0,00	134
Hasta	0,79	1,00	0,88	506
Doğruluk	-	-	0,79	640
Makro ortalama	0,40	0,50	0,44	640
Ağırlıklı ortalama	0,63	0,79	0,70	640

Tablo 4'e bakıldığında ağın genel doğruluk oranı %79'tur. Bu doğruluk tek başına çok fazla bilgi vermez çünkü "Sağlıklı" sınıfının örnek sayısı "Hasta" sınıfına göre çok daha az olmuştur. Bu durum, ağın çoğunlukla "Hasta" sınıfını tahmin ettiğini ve bu sınıf üzerinde yüksek performans gösterdiğini düşündürmektedir. "Sağlıklı" sınıfı üzerinde sıfır performans gösterildiği için, doğruluk oranı yanıltıcı olabilir. Makro ortalama, her iki sınıfın performansını eşit bir şekilde değerlendirdiğinden, ağın genel başarısının düşük olduğu görülmektedir. Özellikle "Sağlıklı" sınıfının kötü performansı, bu ortalamaları olumsuz etkilemiştir.

4.2.4.VGGNet ağı

Tablo 5: VGGNet ağı performans değerlendirmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,66	0,54	0,60	134
Hasta	0,88	0,93	0,91	506
Doğruluk	-	-	0,85	640
Makro ortalama	0,77	0,74	0,75	640
Ağırlıklı ortalama	0,84	0,85	0,84	640

Tablo 5'e bakıldığında ağın genel doğruluk oranı %79'tur, bu doğruluk tek başına çok fazla bilgi vermez çünkü "Sağlıklı" sınıfının örnek sayısı "Hasta" sınıfına göre çok daha az olmuştur. Bu durum, ağın çoğunlukla "Hasta" sınıfını tahmin ettiğini ve bu sınıf üzerinde yüksek performans gösterdiğini düşündürmektedir.

4.2.5.Değerlendirme

En başarılı ağlar EfficientNetB4 ve VGGNet olmuştur, ancak VGGNet daha dengeli bir performans sergilemiştir. VGGNet her iki sınıfı da daha dengeli bir şekilde tanımlamış ve F1 skoru açısından tatmin edici bir değer sunmuştur.

EfficientNetB4'de "Hasta" sınıfındaki yüksek kesinlik ve duyarlılık değerleri dikkat çekici olmuştur. Ancak "Sağlıklı" sınıfındaki düşük kesinlik (0,60) bu ağın biraz dengesiz olduğunu göstermiştir. VGGNet'te her iki sınıf için de daha dengeli bir başarı oluşmuştur. "Sağlıklı" sınıfı için duyarlılık daha düşük olsa da, kesinlik ve F1 skoru daha yüksek çıkmıştır. "Hasta" sınıfı için de iyi bir performans sergilemiştir.

4.3. Diabetic Retinopathy ve Macular Hole Hastalıkları İçin Multilabel Transfer Öğrenme Değerlendirmesi

4.3.1.EfficientNetB0 ağı

Tablo 6: EfficientNetB0 ağı performans değerlendirmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,93	0,94	0,94	508
Hasta	0,76	0,73	0,75	132
Doğruluk	-	-	0,90	640
Makro ortalama	0,85	0,84	0,84	640
Ağırlıklı ortalama	0,90	0,90	0,90	640

Tablo 6'ya bakıldığında ağ, "Sağlıklı" sınıfında iyi bir performans sergilemiştir. Kesinlik, duyarlılık ve F1 skoru 0,94 civarında çıkmış olup bu sınıfın doğru sınıflandırılması için önemli bir artı olmuştur. "Hasta" sınıfında kesinlik ve duyarlılık değerleri biraz daha düşük olmasına rağmen 0,75 F1 skoru, ağın bu sınıfı başarılı bir şekilde sınıflandırdığını göstermiştir.

4.3.2.VGG16 ağı

Tablo 7: VGG16 ağı performans değerlendirmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
---------------	----------	------------	---------	-------------

Sağlıklı	0,92	0,93	0,92	508
Hasta	0,71	0,68	0,70	132
Doğruluk	-	-	0,88	640
Makro ortalama	0,82	0,81	0,81	640
Ağırlıklı ortalama	0,88	0,88	0,88	640

Tablo 7'ye bakıldığında ağ "Sağlıklı" sınıfında sınıfta iyi bir performans sergilemiştir. Kesinlik, duyarlılık ve F1 skoru 0,92 civarında çıkmış olup bu sınıfın doğru sınıflandırılması için iyi bir değerdir. "Hasta" sınıfı için kesinlik (0,71), duyarlılık (0,68) ve F1 skoru (0,70) ile daha düşük değerlere sahip olmuştur. Bu, ağın bu sınıfı doğru tanımada bazı eksiklikler olduğunu göstermektedir. Özellikle duyarlılık değerinin düşük olması ağın bazı "Hasta" örneklerini yanlış sınıflandırdığını veya gözden kaçırdığını göstermektedir.

4.3.3.DenseNet169 ağı

Tablo 8: DenseNet169 ağı performans değerlendirmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,94	0,93	0,94	508
Hasta	0,75	0,78	0,76	132
Doğruluk	-	-	0,90	640
Makro ortalama	0,84	0,86	0,85	640
Ağırlıklı ortalama	0,90	0,90	0,90	640

Tablo 8'e bakıldığında ağ "Sağlıklı" sınıfında sınıfta iyi bir performans sergilemiştir. Kesinlik, duyarlılık ve F1 skoru 0,94 civarında çıkmış olup bu sınıfın doğru sınıflandırılması için iyi bir değer olmuştur. Ağ, "Hasta" sınıfı için kesinlik (0,75), duyarlılık (0,78) ve F1 skoru (0,76) ile daha düşük değerlere sahip olmuştur. Bu, ağın "Hasta" sınıfını doğru tanıma konusunda bazı eksiklikler yaşadığını göstermekle birlikte yine de iyi bir performans olarak kabul edilebilir.

4.3.4.ResNet152 ağı

Tablo 9: ResNet152 ağı performans değerlendirmesi.

Sınıflandırma	Kesinlik	Duyarlılık	F1 skor	Veri Sayısı
Sağlıklı	0,93	0,91	0,92	508
Hasta	0,69	0,73	0,71	132
Doğruluk	-	-	0,88	640
Makro ortalama	0,81	0,82	0,81	640
Ağırlıklı ortalama	0,88	0,88	0,88	640

Tablo 9'a bakıldığında ağ "Sağlıklı" sınıfında sınıfta iyi bir performans sergilemiştir. Kesinlik, duyarlılık ve F1 skoru 0,92 civarında çıkmış olup bu sınıfın doğru sınıflandırılması için iyi bir değer olmuştur. Ağ "Hasta" sınıfı için kesinlik (0,69), duyarlılık (0,73) ve F1 skoru (0,71) ile daha düşük değerlere sahip olmuştur. Bu, ağın "Hasta" sınıfını doğru tanıma konusunda bazı eksiklikler yaşadığını göstermekle birlikte yine de bu değer yeterli bir performans olarak kabul edilebilir.

5. SONUÇ

Bu çalışmada tasarlanan CNN yapısında, iki sınıflı bir veri seti üzerinde sınıflandırma ağının performansı değerlendirilmiştir. Ağ, genel olarak %86 doğruluk oranı ile başarılı bir sonuç elde etmiş, ancak sınıflar arasında performans dengesizlikleri gözlemlenmiştir. Ağın, sınıf 1 üzerinde oldukça yüksek bir başarı sergilediği görülmektedir. Sınıf 1 için elde edilen kesinlik (%86), duyarlılık (%98) ve F1 skor (%92) değerleri, ağın bu sınıfta doğru pozitif tahminler yapma ve gerçek örnekleri yakalama yeteneğinin oldukça güçlü olduğunu ortaya koymaktadır. Bununla birlikte, sınıf 0 için duyarlılık (%38) ve F1 skor (%53) değerleri

düşüktür, bu da ağın sınıf 0'a ait örneklerin bir kısmını doğru şekilde tespit edemediğini göstermektedir.

Makro ortalama değerlendirmesi, sınıflar arasında performans farklarını daha net bir şekilde yansıtmaktadır. Kesinlik (%85) makul bir seviyede olsa da duyarlılık (%68) ve F1 skor (%72) değerleri, ağın sınıf 0 üzerindeki düşük performansından etkilenmiştir. Ağırlıklı ortalama metrikleri ise sınıfların destek oranlarını göz önüne aldığından, ağın genel performansını daha tutarlı bir şekilde yansıtmaktadır (Kesinlik: %86, duyarlılık: %86, F1 skor: %83).

Veri setindeki sınıf dengesizliği, performans farklılıklarının temel nedenlerinden biri olarak değerlendirilmektedir. Sınıf 1'in toplam test örneklerinin büyük bir kısmını (%79) oluşturması, ağın bu sınıfta daha yüksek başarı göstermesine yol açmıştır. Bu durum, sınıflar arasında daha dengeli bir veri dağılımı sağlanmadığında, ağın azınlık sınıfları (sınıf 0) yeterince iyi öğrenemediğini göstermektedir.

Transfer öğrenme kullanarak yapılan ikili sınıflandırmada, EfficientNetB4 ve VGGNet ağları, genel başarı açısından en yüksek performansı sergileyen ağlar olarak öne çıkmıştır. Bununla birlikte, VGGNet biraz daha dengeli bir performans göstermiştir. EfficientNetB4 ağı, "Hasta" sınıfında iyi bir kesinlik ve duyarlılık değerlerine sahip olmuştur. Ancak, "Sağlıklı" sınıfındaki düşük kesinlik (0,60), ağın dengesiz olduğunu ve bu sınıfta doğru tahminlerde bazen eksiklikler yaşandığını göstermektedir. VGGNet, her iki sınıf için de daha dengeli bir performans sergilemektedir. "Sağlıklı" sınıfı için duyarlılık değeri düşük olsa da, kesinlik ve F1 skoru daha yüksek çıkarak ağın bu sınıfı daha başarılı bir şekilde tanıdığını göstermektedir. "Hasta" sınıfı için de başarılı değerler çıkmıştır.

NASNetMobile, diğer ağlara nazaran düşük performans değerleri göstermiştir. EfficientNetB0 ağı ise genel olarak iyi bir performans göstermektedir. Ancak, "Sağlıklı" sınıfındaki düşük duyarlılık değeri, ağın bu sınıfı tanımda bazı eksiklikler yaşadığını ortaya koymaktadır. EfficientNetB0 ağı, "Sağlıklı" sınıfında yüksek bir kesinlik (0,93) ve duyarlılık (0,91) değerine sahip olup, bu da ağın bu sınıfı doğru bir şekilde sınıflandırma konusunda güçlü bir performans gösterdiğini işaret etmektedir. F1 skoru da oldukça yüksek (0,92) olup, bu da kesinlik ve duyarlılık arasında iyi bir denge sağlandığını göstermektedir. Ancak, "Hasta" sınıfında kesinlik ve duyarlılık oranları sırasıyla 0,69 ve 0,73 ile daha düşük kalmaktadır. Bu, ağın "Hasta" sınıfındaki örnekleri tanımda bazı eksiklikler yaşadığını ve daha fazla iyileştirmeye ihtiyaç duyduğunu göstermektedir. Ağın genel doğruluk oranı %88, makro ortalama ve ağırlıklı ortalama değerleri ise sırasıyla 0,81 ve 0,88'dir ve bu da genel olarak dengeli bir performans sergilediğini ifade etmektedir.

VGG16 ağı, "Sağlıklı" sınıfında çok yüksek bir kesinlik (0,92) ve duyarlılık (0,93) ile başarılı bir performans göstermektedir, ancak "Hasta" sınıfında kesinlik 0,71, duyarlılık ise 0,68 gibi daha düşük değerlere sahiptir. "Hasta" sınıfındaki düşük F1 skoru (0,70), ağın bu sınıfta bazı hatalar yaptığını ve "Hasta" örneklerini daha iyi sınıflandırmak için iyileştirme gerektiğini göstermiştir. Genel doğruluk oranı %88, makro ortalama ve ağırlıklı ortalama değerleri sırasıyla 0,82 ve 0,88 ile güçlü bir performansı işaret etmektedir. Ancak "Hasta" sınıfındaki daha düşük performans bu oranları kısmen etkileyebilir. DenseNet169 ağına, "Sağlıklı" sınıfı için kesinlik 0,60, duyarlılık ise 0,80 olarak ölçülmüş, bu da ağın bu sınıftaki örnekleri doğru tanımda zayıf olduğunu göstermiştir. Özellikle, kesinlik değerinin düşük olması ağın yanlış pozitif tahminler yapma olasılığının yüksek olduğunu ve "Sağlıklı" sınıfındaki örnekleri doğru şekilde tanımda eksiklikler yaşadığını göstermiştir. Diğer yandan, "Hasta" sınıfında oldukça yüksek bir kesinlik (0,94) ve duyarlılık (0,86) değeri elde edilmiştir ki bu da ağın "Hasta" sınıfındaki örnekleri doğru şekilde tanıdığını işaret etmektedir. Bu ağın genel doğruluk oranı %85, makro ortalama ve ağırlıklı ortalama değerleri ise sırasıyla 0,77 ve 0,87'dir. "Sağlıklı" sınıfındaki düşük kesinlik değeri, ağın dengesiz olduğunu ve iyileştirilmeye ihtiyaç duyduğunu göstermektedir. ResNet152 ağı "Sağlıklı" sınıfında oldukça yüksek bir kesinlik (0,94) ve duyarlılık (0,93) elde ederken, "Hasta" sınıfında daha düşük kesinlik (0,75) ve duyarlılık (0,78) değerlerine sahiptir. "Sağlıklı" sınıfındaki yüksek F1 skoru (0,94) ağın bu sınıfı iyi tanıdığını göstermiştir. "Hasta" sınıfındaki düşük F1 skoru (0,76) ise ağın bu sınıftaki örnekleri doğru tanımda bazı eksiklikler yaşadığını işaret etmektedir. Genel doğruluk oranı %90, makro ortalama ve ağırlıklı ortalama değerleri sırasıyla 0,84 ve 0,90 ile güçlü bir performans sergilemiştir. Bu ağ, özellikle "Sağlıklı" sınıfında başarılı iken, "Hasta" sınıfında iyileştirilmeye ihtiyaç duyan bir performans sergilemektedir.

- En Başarılı Ağ Yapısı: EfficientNetB0, VGG16 ve ResNet152'nin genelde yüksek doğruluk oranları ve dengeli performansları, bu ağların en başarılı ağ yapıları olduğunu göstermektedir. Ancak, ResNet152, "Sağlıklı" sınıfında iyi bir performans sergilerken, "Hasta" sınıfında bazı eksiklikler yaşamakta ve iyileştirilmeye ihtiyaç duymaktadır.
- Dengeli Performans: VGG16 ve EfficientNetB0, her iki sınıfı da daha dengeli bir şekilde tanımakta ve F1 skoru açısından tatmin edici sonuçlar elde etmektedir. Bu ağlar, her iki sınıfın da doğru şekilde sınıflandırılmasında başarılıdır.
- Başarılı Performans: Ağların genel başarıları değerlerine bakıldığında, EfficientNetB0, VGG16 ve ResNet152 güçlü performans sergileyen ağ yapıları olarak öne çıkmıştır. DenseNet169 ise "Sağlıklı" sınıfındaki başarısızlık nedeniyle düşük performans göstermiştir.
- Genel Sonuç: Yine ağların genel başarıları göz önünde bulundurulduğunda derin öğrenme tabanlı yaklaşımlar uzman doktorların verimliliğini artıracak karar destek mekanizmaları olarak konumlandırılabilir.

6. KAYNAKÇA

- Brown, G. C., et al. (2019). Retina. Elsevier.
- Deye, J., et al. (2020). Multimodal deep learning for retinal disease diagnosis using fundus images and OCT. *IEEE Transactions on Medical Imaging*, 39(6), 2060–2072. <https://doi.org/10.1109/TMI.2019.2952127>
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- 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 (CVPR)* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>
- 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 (CVPR)*. <https://doi.org/10.1109/CVPR.2017.243>
- Lee, C. H., et al. (2017). Deep learning in retinal disease diagnosis: A review of current applications. *Ophthalmology*, 124(4), 593–604. <https://doi.org/10.1016/j.ophtha.2016.10.048>
- Leibig, C., Allken, V., Ayhan, M. S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scientific Reports*, 7(1), 1–14. <https://doi.org/10.1038/s41598-017-17876-z>
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- RIADD (2021). Retinal Fundus Multi-disease Image Dataset. Erişim adresi: <https://riadd.grand-challenge.org/download-all-classes/>
- Ronneberger, O., et al. (2015). U-Net: Convolutional networks for biomedical image segmentation. In

- Medical Image Computing and Computer-Assisted Intervention (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint. <https://arxiv.org/abs/1409.1556>
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML). <https://proceedings.mlr.press/v97/tan19a.html>
- Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., ... & Wong, T. Y. (2017). Artificial intelligence and deep learning in ophthalmology. The British Journal of Ophthalmology, 103(2), 167-175. <https://doi.org/10.1136/bjophthalmol-2018-313173>
- Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., ... & Schmetterer, L. (2017). Artificial intelligence and deep learning in ophthalmology. The British Journal of Ophthalmology, 101(2), 184-190. <https://doi.org/10.1136/bjophthalmol-2017-311307>
- Wong, T. Y., et al. (2020). Retinal Imaging. Springer.
- World Health Organization. (2019). World report on vision. WHO Press. Retrieved from <https://www.who.int/publications-detail/world-report-on-vision>
- Zhang, J., Liu, Y., Mitsuhashi, T., & Matsuo, T. (2021). Accuracy of Deep Learning Algorithms for the Diagnosis of Retinopathy of Prematurity by Fundus Images: A Systematic Review and Meta-Analysis. Journal of ophthalmology, 2021, 8883946. <https://doi.org/10.1155/2021/8883946>
- Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/CVPR.2018.00916>