



T.C.

TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI

**AKCİĞER KANSERİNİN TESPİTİNDE DERİN ÖĞRENME
MODELLERİNİN KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

Yavuz PALA

Danışman: Dr. Öğr. Üyesi Mürsel ÖNDER

TOKAT - 2025

ETİK SÖZLEŞME

Tokat Gaziosmanpaşa Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre, Dr. Öğr. Üyesi Mürsel ÖNDER danışmanlığında hazırlamış olduğum “Akciğer Kanserinin Tespitinde Derin Öğrenme Modellerinin Karşılaştırılması” adlı Yüksek Lisans tezinin bilimsel etik değerlere ve kurallara uygun, özgün bir çalışma olduğunu, aksinin tespit edilmesi halinde her türlü yasal yaptırımını kabul edeceğimi beyan ederim.

28/11/2025

Yavuz PALA

JÜRİ KABUL VE ONAY

Yavuz PALA tarafından hazırlanan “**Akciğer Kanserinin Tespitinde Derin Öğrenme Modellerinin Karşılaştırılması**” adlı tez çalışmasının savunma sınavı 28.11.2025 tarihinde yapılmış olup aşağıda verilen Jüri tarafından Oy Birliği ile Tokat Gaziosmanpaşa Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda Yüksek Lisans Tezi olarak kabul edilmiştir.

Jüri Üyeleri (Unvanı, Adı Soyadı)	İmzası
Üye (Danışman) : Dr. Öğr. Üyesi Mürsel ÖNDER	
Üye (Başkan) : Doç. Dr. Bülent TURAN	
Üye :Doç. Dr. Özkan İNİK	

ONAY

...../...../.....

Lisansüstü Eğitim Enstitüsü Müdürü

ÖNSÖZ

Bu çalışmanın her aşamasında engin tecrübeleri ve değerli fikirleri ile beni yönlendiren, başta danışmanım Sayın Dr. Öğr. Üyesi Mürsel ÖNDER olmak üzere, katkılarından dolayı Sayın Doç. Dr. Bülent TURAN'a ve Sayın Doç. Dr. Özkan İNİK'e teşekkürlerimi sunuyorum.

Tezin hazırlanma sürecinde yardımlarını benden esirgemeyen kıymetli Bilim Uzmanı Sayın İlyas KAYA'ya ve akademik çalışmalarım boyunca her zaman yanımda olan kıymetli Bilim Uzmanı Sayın Murat KARAOĞLU'na teşekkür ediyorum. Ayrıca teknik desteklerinden dolayı kıymetli Bilim Uzmanı Sayın Özer ARSLAN ve Sayın Mustafa DEMİR'e teşekkürlerimi sunuyorum.

Hayatım boyunca maddi ve manevi desteklerini hiçbir zaman esirgemeyen kıymetli Babama ve Kardeşlerime şükranlarımı sunuyorum.

Bu tez çalışması kapsamında, bilgisayarlı tomografi (BT) görüntülerinden akciğer kanseri sınıflandırması amacıyla beş farklı modern derin öğrenme mimarisinin karşılaştırmalı performans analizi gerçekleştirilmiştir. Çalışmada; çok merkezli IQ-OTH/NCCD veri seti üzerinde hasta-bazlı veri ayrımı, kilitli test protokolü, Medical-SAM ile otomatik ROI segmentasyonu, MATLAB ortamında yürütülen eğitim süreçleri, temel performans metrikleri, karışıklık matrisleri, ROC-AUC analizleri ve Grad-CAM tabanlı açıklanabilirlik incelemeleri detaylı olarak sunulmuştur. Bu kapsamlı karşılaştırma ile tıbbi görüntülemede akciğer kanserinin erken tanısında kullanılabilecek derin öğrenme tabanlı yöntemlerin güçlü ve zayıf yönlerinin nesnel biçimde ortaya konulması ve literatürdeki klinik geçerliliği yüksek çalışmaların geliştirilmesine katkı sağlanması amaçlanmıştır.

Bu tez çalışmasının literatüre, ilgili çevrelere ve bilim dünyasına katkı sağlamasını temenni ediyorum.

İTHAF

06.03.2021 tarihinde kanser hastalığı nedeniyle aramızdan ayrılan Canım Annemin aziz hatırasına ithaf edilmiştir.



ÖZET

AKCİĞER KANSERİNİN TESPİTİNDE DERİN ÖĞRENME MODELLERİNİN KARŞILAŞTIRILMASI

Pala, Yavuz

Yüksek Lisans, Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Mürsel ÖNDER

Kasım 2025 xv + 110 sayfa

Akciğer kanseri, küresel ölçekte en yüksek mortalite oranına sahip kanser türlerinden biri olup, tanıda gecikme hastaların yaşam süresini belirgin şekilde azaltmaktadır. Bu çalışmada, akciğer kanserinin otomatik tespiti için bilgisayarlı tomografi (BT) görüntülerinde modern derin öğrenme mimarileri (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3) standart ve veri sızıntısız bir deneysel protokol altında karşılaştırmalı olarak değerlendirilmiştir. Çalışmada, Irak'taki iki büyük onkoloji merkezinden toplanan, uluslararası ve açık erişimli IQ-OTH/NCCD veri seti kullanılmış olup veri seti; Malignant 561, Benign 120 ve Normal 416 olmak üzere toplam 1097 BT diliminden oluşan üç sınıf içermektedir. Gerçekçi bir klinik senaryo elde etmek ve veri sızıntısını engellemek amacıyla, tüm BT görüntüleri herhangi bir ön işleme veya artırma uygulanmadan önce katmanlı olarak eğitim yüzde 70, doğrulama yüzde 15 ve kilitli test yüzde 15 alt kümelerine ayrılmıştır. Sadece eğitim alt kümesindeki sınıf dengesizliği, sabit hedefli bir aşırı örnekleme stratejisi ile giderilmiş ve her sınıf 1000 görüntüye tamamlanmıştır; doğrulama ve test alt kümelerine hiçbir artırma veya dengeleme işlemi uygulanmamıştır. Ön işleme aşamasında BT görüntülerine pencereleme uygulanmış, ardından Otsu eşikleme, morfolojik açma-kapama işlemleri, delik doldurma ve bir piksel erozyon ile akciğer bölgeleri otomatik olarak maskelenmiştir. Tüm derin öğrenme modelleri ImageNet ağırlıklarıyla başlatılarak MATLAB R2024b ortamında transfer öğrenme yöntemiyle yeniden eğitilmiştir. Modeller kilitli test veri kümesi üzerinde doğruluk, duyarlılık, kesinlik, özgüllük, F1 skoru ve ROC-AUC gibi istatistiksel performans metrikleriyle değerlendirilmiştir. Elde edilen sonuçlara göre EfficientNet-B0 modeli yüzde 98.17 doğruluk ile en yüksek performansı göstermiştir. Diğer modeller sırasıyla; InceptionV3 yüzde 96.95, MobileNetV2 yüzde 96.34, VGG16 yüzde 93.90 ve ResNet-50 yüzde 89.63 doğruluk değerlerine ulaşmıştır. Klinik açıdan kritik olan Malignant sınıfının duyarlılığı EfficientNet-B0'da yüzde 98.81 ile en yüksek değere ulaşmış; ayrıca tüm modellerin Malignant sınıfında yüzde 100 kesinlik üretmiş olması, kanserli olarak sınıflandırılan hiçbir örneğin yanlış pozitif olmadığı anlamına gelmiştir. Grad-CAM tabanlı açıklanabilir yapay zekâ analizleri, ağların kararlarını nodül, opasite ve lezyon gibi klinik olarak anlamlı bölgelere odaklanarak verdiğini göstermiştir. Sonuç olarak, bu çalışma çok merkezli gerçek BT görüntüleri üzerinde beş farklı derin öğrenme mimarisini, veri sızıntısını ortadan kaldıran ve sınıf dengesini koruyan standartlaştırılmış bir eğitim protokolü altında karşılaştırarak literatüre güvenilir ve yeniden üretilebilir bir değerlendirme çerçevesi sunmaktadır.

Anahtar Kelimeler: Akciğer Kanseri, Derin Öğrenme, EfficientNet-B0, ResNet-50, VGG16, MobileNetV2, InceptionV3, Grad-CAM.

ABSTRACT

COMPARISON OF DEEP LEARNING MODELS FOR LUNG CANCER DETECTION

Pala, Yavuz

Master's Degree, Department of Computer Engineering

Thesis Advisor: Assist. Prof. Dr. Mürsel ÖNDER

November 2025, xv + 110 pages

Lung cancer is one of the cancer types with the highest mortality rates worldwide, and delays in diagnosis significantly reduce patient survival. In this study, modern deep learning architectures (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2, and InceptionV3) were comparatively evaluated under a standardized and data leakage-free experimental protocol for the automatic detection of lung cancer in computed tomography (CT) images. The study employed the international and publicly available IQ-OTH/NCCD dataset collected from two major oncology centers in Iraq. The dataset contains three classes comprising a total of 1097 CT slices, including 561 Malignant, 120 Benign, and 416 Normal cases. To obtain a realistic clinical scenario and prevent data leakage, all CT images were stratified into training, validation, and locked test subsets at ratios of 70 percent, 15 percent, and 15 percent, respectively, before any preprocessing or augmentation was applied. Class imbalance was addressed only in the training subset using a fixed-target oversampling strategy, increasing each class to 1000 images, while no augmentation or balancing was applied to the validation and test subsets. During the preprocessing stage, CT images were subjected to windowing, followed by Otsu thresholding, morphological opening and closing operations, hole filling, and single-pixel erosion to automatically segment the lung regions. All deep learning models were initialized with ImageNet weights and retrained using transfer learning in the MATLAB R2024b environment. The models were evaluated on the locked test dataset using statistical performance metrics including accuracy, sensitivity, precision, specificity, F1 score, and ROC-AUC. According to the results, the EfficientNet-B0 model achieved the highest performance with an accuracy of 98.17 percent. The remaining models achieved accuracies of 96.95 percent for InceptionV3, 96.34 percent for MobileNetV2, 93.90 percent for VGG16, and 89.63 percent for ResNet-50. Sensitivity for the clinically critical Malignant class reached its highest value of 98.81 percent with the EfficientNet-B0 model. Moreover, all models achieved 100 percent precision for the Malignant class, indicating that none of the samples classified as cancerous were false positives. Grad-CAM-based explainable artificial intelligence analyses demonstrated that the networks based their decisions on clinically meaningful regions such as nodules, opacities, and lesions. In conclusion, this study provides a reliable and reproducible evaluation framework by comparatively analyzing five deep learning architectures on multi-center real CT images under a standardized training protocol that eliminates data leakage and preserves class balance, thereby contributing robust evidence to the literature.

Keywords: Lung Cancer, Deep learning, EfficientNet-B0, ResNet-50, VGG16, MobileNetV2, InceptionV3, Grad-CAM.

İÇİNDEKİLER

	Sayfa
ETİK SÖZLEŞME.....	i
JÜRİ KABUL VE ONAY	ii
ÖNSÖZ	iii
İTHAF	iv
ÖZET	v
İÇİNDEKİLER	vii
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
KISALTMALAR VE SİMGELER LİSTESİ	xiii
GİRİŞ	1
2.LİTERATÜR TARAMASI	5
2.1. Derin Öğrenme ve Tıbbi Görüntüleme Temelleri	6
2.2. Transfer Öğrenme Yaklaşımları	7
2.3. Açıklanabilir Yapay Zekâ (XAI) ve Grad-CAM Yöntemi	7
2.4. Derin Öğrenmede Kullanılan Ortak Matematiksel Temeller	8
2.4.1. Evrişim (convolution) işlemi	9
2.4.2. Maksimum havuzlama (max pooling)	10
2.4.3. Batch normalizasyonu (batch normalization)	11
2.4.4. Relu aktivasyon fonksiyonu.....	12
2.4.5. Softmax fonksiyonu	13
2.4.6. Kayıp fonksiyonu (çapraz entropi)	13
2.4.7. L2 düzenlileştirme (regularization).....	14
2.4.8. Dropout (bırakma) mekanizması	15
2.4.9. Adam optimizasyon algoritması.....	16
2.5. Karşılaştırılacak Derin Öğrenme Modellerinin Teorik Analizi.....	18
2.5.1. Vgg16 mimarisi, matematiksel temeli ve literatürdeki yeri.....	18
2.5.2. Resnet-50 mimarisi, matematiksel temeli ve literatürdeki yeri	19
2.5.3. Efficientnet-b0 mimarisi, matematiksel temeli ve literatürdeki yeri	21
2.5.4. Mobilenetv2 mimarisi, matematiksel temeli ve literatürdeki yeri	23
2.5.5. Inceptionv3 mimarisi, matematiksel temeli ve literatürdeki yeri	25
3.MATERYAL VE METOT.....	27

3.1. Veri Seti.....	28
3.1.1. Veri setinin tanımı ve klinik özellikleri.....	28
3.1.2. Başlangıç veri dağılımı ve sınıf dengesizliği problemi.....	29
3.2. Uygulanan Yöntem	30
3.2.1. Görüntü ön işleme segmentasyon ve normalizasyon.....	30
3.2.2. Deneysel kurulum eğitim, doğrulama ve test ayrımı	32
3.2.3. Veri artırma ve dengeleme stratejileri	33
3.2.4. Performans değerlendirme metrikleri	34
4.DENEYSEL BULGULAR VE ANALİZ	38
4.1.Mobilenetv2 Modeli Bulguları ve Analizi	38
4.1.1. Eğitim süreci ve hiperparametreler	38
4.1.2. Mobilenetv2 karışıklık matrisi ve hata analizi.....	39
4.1.3. Nihai test performansı ve başarı metrikleri.....	41
4.1.4. Roc eğrileri ve auc analizi ve modelin ayırım gücü.....	43
4.1.5. Grad-cam ile açıklanabilirlik analizi.....	44
4.1.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması	47
4.2. Vgg16 Modeli Bulguları ve Analizi.....	48
4.2.1. Eğitim süreci ve hiperparametreler	49
4.2.2. Mobilenetv2 karışıklık matrisi ve hata analizi.....	50
4.2.3. Nihai test performansı ve başarı metrikleri.....	52
4.2.4. Roc eğrileri ve auc analizi ve modelin ayırım gücü.....	54
4.2.5. Grad-cam ile açıklanabilirlik analizi.....	56
4.2.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması	58
4.3. ResNet-50 Modeli Bulguları ve Analizi.....	59
4.3.1. Eğitim süreci ve hiperparametreler	59
4.3.2. Resnet-50 karışıklık matrisi ve hata analizi	60
4.3.3. Nihai test performansı ve başarı metrikleri.....	63
4.3.4. Roc eğrileri ve auc analizi.....	65
4.3.5. Grad-cam ile açıklanabilirlik analizi.....	66
4.3.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması	69
4.4. EfficientNet-B0 Modeli Bulguları ve Analizi	71
4.4.1. Eğitim süreci ve hiperparametreler	71

4.4.2. Efficientnet-b0 karışıklık matrisi ve hata analizi	73
4.4.3. Nihai test performansı ve başarı metrikleri	74
4.4.4. Roc eğrileri ve auc analizi modelin ayırım gücü	76
4.4.5. Grad-cam ile açıklanabilirlik analizi	77
4.4.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması	79
4.5. InceptionV3 Modeli Bulguları ve Analizi	80
4.5.1. Eğitim süreci ve hiperparametreler	80
4.5.2. Inceptionv3 karışıklık matrisi ve hata analizi	81
4.5.3. Nihai test performansı ve başarı metrikleri	83
4.5.4. Roc eğrileri ve auc analizi modelin ayırım gücü	84
4.5.5. Grad-cam ile açıklanabilirlik analizi	86
4.5.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması	88
5. TARTIŞMA, SONUÇ VE ÖNERİLER	90
5.1. Sonuçların Değerlendirilmesi ve Tartışma	90
5.1.1. Modellerin karşılaştırmalı performans analizi	90
5.1.2. Yanlış sınıflandırma dağılımının analizi ve model davranışı	92
5.1.3. Açıklanabilir yapay zekâ (xai) ve model güvenilirliği	93
5.1.4. Çalışmanın literatüre katkısı ve hipotezin doğrulanması	94
5.2. Çalışmanın Sınırlılıkları	94
5.3. Gelecek Çalışmalar İçin Öneriler	95
5.3.1. Akademik araştırmalara yönelik öneriler	95
5.3.2. Klinik uygulama, donanım ve etik konulara yönelik öneriler	96
5.4. Genel değerlendirme	98
KAYNAKÇA	99
ÖZGEÇMİŞ	110

TABLÖLAR LİSTESİ

Tablo No	Sayfa
Tablo 4.1. Iq-oth/nccd veri setinde mobilenetv2 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması.....	48
Tablo 4.2. Iq-oth/nccd veri setinde vgg16 ve ilgili literatürde elde edilen sonuçların karşılaştırılması.....	58
Tablo 4.3. Iq-oth/nccd veri setinde resnet-50 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması.....	71
Tablo 4.4. Iq-oth/nccd veri setinde efficientnet-b0 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması.....	79
Tablo 4.5. Iq-oth/nccd veri setinde inceptionv3 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması.....	88
Tablo 5.1. Derin öğrenme modellerinin performans ve verimlilik karşılaştırması.....	91

ŞEKİLLER LİSTESİ

Şekil No	Sayfa
Şekil 3.1. Çalışmada uygulanan sürecinin genel iş akış şeması.....	27
Şekil 3.2. Iq-oth/nccd veri setinin orijinal başlangıç sınıf dağılımı.....	29
Şekil 3.3. Segmentasyon öncesi ve sonrası modelin gördüğü bt görüntüsü	31
Şekil 3.4. Veri artırma ve dengeleme sonrası nihai sınıf dağılımı.....	33
Şekil 3.5. Nihai veri setinin eğitim, doğrulama ve test olarak katmanlı dağılımı.....	34
Şekil 4.1. Mobilenetv2 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp değerlerinin epoch bazında değişimi.....	39
Şekil 4.2. Mobilenetv2 modeli için sınıf bazlı nihai performans metrikleri ısı haritası...	40
Şekil 4.3. Mobilenetv2 modeli için nihai test seti karışıklık matrisi.....	41
Şekil 4.4. Mobilenetv2 için sınıf bazlı roc eğrileri ve auc değerleri.....	43
Şekil 4.5. Mobilenetv2 doğru sınıflandırdığı grad-cam aktivasyon haritaları.....	45
Şekil 4.6. Mobilenetv2 modelinin yanlış sınıflandırdığı grad-cam ısı haritası.....	46
Şekil 4.7. Vgg16 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp eğrilerinin epoch bazında değişimi	50
Şekil 4.8. Vgg16 modeli için sınıf bazlı nihai performans metrikleri.....	50
Şekil 4.9. Vgg16 modeli için nihai test seti karışıklık matrisi.....	52
Şekil 4.10. Vgg16 modeli için sınıf bazlı roc eğrileri ve auc değerleri.....	54
Şekil 4.11. Vgg16 modelinin doğru tahmin ettiği grad-cam aktivasyon haritaları.....	56
Şekil 4.12. Vgg16 modelinin yanlış sınıflandırdığı grad-cam ısı haritası.....	57
Şekil 4.13. Resnet-50 modeli için eğitim sürecinde doğruluk (accuracy) ve kayıp (loss) eğrilerinin epoch bazında değişimi.....	60
Şekil 4.14. Resnet-50 modeli için sınıf bazlı nihai performans metrikleri ısı haritası.....	61
Şekil 4.15. Resnet-50 modeli için nihai test seti karışıklık matrisi.....	63
Şekil 4.16. Resnet-50 modeli için sınıf bazlı roc eğrileri ve auc değerleri.....	65
Şekil 4.17. Grad-cam: en güvenilir doğru tahminlerin analizi.....	67
Şekil 4.18. Grad-cam hata analizi normal vakanın benign olarak sınıflandırılması.....	68
Şekil 4.19. Efficientnet-b0 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp eğrilerinin epoch bazında değişimi.....	72
Şekil 4.20. Efficientnet-b0 için sınıf bazlı nihai performans metrikleri ısı haritası.....	73
Şekil 4.21. Efficientnet-b0 modeli için nihai test seti karışıklık matrisi.....	74

Şekil 4.22. Efficientnet-b0 modeli için sınıf bazlı roc eğrileri ve auc değerleri.....	76
Şekil 4.23. Efficientnet-b0 modelinin doğru sınıflandırdığı örneklere ait grad-cam....	77
Şekil 4.24. Modelin hatalı sınıflandırdığı bir örneğe ait grad-cam analizi.....	78
Şekil 4.25. Inceptionv3 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp değerlerinin epoch bazında değişimi.....	81
Şekil 4.26. Inceptionv3 için sınıf bazlı nihai performans metrikleri ısı haritası.....	82
Şekil 4.27. Inceptionv3 modeli için nihai test seti karışıklık matrisi.....	83
Şekil 4.28. Inceptionv3 modeli için sınıf bazlı roc eğrileri ve auc analiz grafiği.....	85
Şekil 4.29. Inceptionv3 doğru sınıflandırdığı örnekler için grad-cam.....	87
Şekil 4.30. Inceptionv3 hatalı sınıflandırdığı bir örnek için grad-cam analizi.....	87

SİMGE VE KISALTMALAR LİSTESİ

Simge	Açıklama
%	Yüzde
°C	Santigrat derece
±	Artı-eksi işareti
i, j, m, n	Matris ve indis indeksleri
I	Giriş görüntüsü
K	Evrişim çekirdeği (kernel)
M(i,j)	Çıkış öznitelik haritası
λ	L2 düzenlileştirme katsayısı
η	Öğrenme oranı
ε	Sayısal kararlılık sabiti
μ	Ortalama değeri
σ^2	Varyans
y	Gerçek sınıf etiketi
$\hat{y}$	Tahmin edilen sınıf etiketi
p(i)	i. sınıfa ait olasılık değeri

Kısaltma	Açıklama
Accuracy	Doğruluk
Adam	Adaptive Moment Estimation
AUC	Area Under Curve
BDT	Bilgisayarlı Destekli Tanı
BN	Batch Normalization
BT	Bilgisayarlı Tomografi
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network

CT	Computed Tomography
DDBT	Düşük Doz Bilgisayarlı Tomografi
DL	Deep Learning
EfficientNet-B0	EfficientNet mimarisinin B0 sürümü
Epoch	Eğitim döngüsü
F1	F1 skoru
FN	False Negative (Yanlış negatif)
FP	False Positive (Yanlış pozitif)
GLOBOCAN	Global Cancer Observatory
GPU	Graphics Processing Unit
Grad-CAM	Gradient-weighted Class Activation Mapping
IARC	International Agency for Research on Cancer
ImageNet	Geniş ölçekli görüntü veri kümesi
InceptionV3	Inception mimarisinin üçüncü sürümü
IQ-OTH/NCCD	Iraq Oncology Teaching Hospital / Nanakali Cancer and Chest Disease Hospital veri kümesi
L2	L2 norm tabanlı düzenleme
LR	Learning Rate
MATLAB	Matrix Laboratory
MobileNetV2	MobileNet mimarisinin ikinci sürümü
MR	Manyetik Rezonans
MRG	Manyetik Rezonans Görüntüleme
NELSON	Dutch–Belgian Lung Cancer Screening Trial
NLST	National Lung Screening Trial
OTSU	Otsu eşikleme yöntemi
Precision	Keskinlik
ReLU	Rectified Linear Unit
ReLU6	Çıkışı [0,6] aralığında sınırlı ReLU

ResNet-50	50 katmanlı Residual Network
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SGD	Stochastic Gradient Descent
Specificity	Özgüllük
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
TPU	Tensor Processing Unit
TTOD	Toraks Tümörü Olasılık Değerlendirmesi
VGG16	Visual Geometry Group, 16 katmanlı model
WHO	World Health Organization
XAI	Explainable Artificial Intelligence

GİRİŞ

Akciğer kanseri, dünya genelinde kansere bağlı ölümler arasında ilk sırada yer alan ve küresel ölçekte en kritik halk sağlığı sorunlarından biri olarak kabul edilmektedir. Dünya Sağlık Örgütü'nün (WHO, 2023) verilerine göre akciğer kanseri, hem insidans hem de mortalite açısından birçok kanser türünü geride bırakmaktadır. Küresel Kanser Gözlemevi'nin (GLOBOCAN) güncel verileri, 2022 yılında dünya çapında yaklaşık 2.2 milyon yeni akciğer kanseri vakasının teşhis edildiğini ve yaklaşık 1.85 milyon kişinin bu hastalık nedeniyle hayatını kaybettiğini göstermektedir (Sung ve ark., 2024). Hastalık erkeklerde en sık görülen kanser türü olma özelliğini korurken, kadınlarda da en yaygın ilk üç kanser arasında yer almaktadır. Semptomların çoğunlukla ileri evrede ortaya çıkması ve hastalığın hızlı progresyon göstermesi, akciğer kanserini hem gelişmiş hem de gelişmekte olan ülkelerde ciddi bir sağlık yükü hâline getirmektedir.

Türkiye'deki durum da küresel eğilimlerle benzerlik göstermektedir. T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü tarafından yayımlanan en güncel "Türkiye Kanser İstatistikleri 2022" raporuna göre akciğer kanseri, ülkemizde erkeklerde en sık görülen kanser olup tüm kanser vakalarının yaklaşık yüzde 20.7'sini oluşturmaktadır (T.C. Sağlık Bakanlığı, 2022). Kadınlarda akciğer kanseri insidansı daha düşük olmakla birlikte ilk beş kanser arasında yer almaktadır. Yaşa standardize yıllık insidans hızlarının erkeklerde yüz binde 50–55, kadınlarda ise yüz binde 9–12 aralığında seyrettiği bildirilmektedir (T.C. Sağlık Bakanlığı, 2022). Türk Tıbbi Onkoloji Derneği'nin (TTOD) 2025 yılına ait güncel raporunda ise Türkiye'de yıllık akciğer kanseri vaka sayısının yaklaşık 40 binin üzerine çıktığı ifade edilmektedir (TTOD, 2025). Tütün kullanımı hâlen en temel risk faktörlerinden biri olmakla birlikte, akciğer kanseri tanısı alan hastaların önemli bir kısmının hiçbir zaman sigara içmemiş kişilerden oluşması; hava kirliliği, çevresel ve mesleki maruziyetler, radon ve genetik yatkınlık gibi etkenlerin de etyolojide giderek daha fazla önem kazandığını göstermektedir.

Hastalığın prognozu açısından en kritik noktalardan biri, tanı anındaki evre dağılımıdır. Ulusal veriler, akciğer kanseri vakalarının yalnızca sınırlı bir kısmının erken (lokalize) evrede yakalanabildiğini, büyük çoğunluğun ise bölgesel ileri veya metastatik evrede

teşhis edildiğini ortaya koymaktadır (T.C. Sağlık Bakanlığı, 2022). Klinik çalışmalar, erken evrede (Evre I) tanı alan hastalarda beş yıllık sağkalım oranlarının yüzde 60–80 bandına kadar yükselebilirken, metastatik evrede bu oranın yüzde 10’un altına düştüğünü göstermektedir (Türk Toraks Derneği, 2023). Bu tablo, akciğer kanserinde erken teşhisin, hastalığa bağlı mortalitenin azaltılması ve sağkalım sürelerinin uzatılması açısından hayati önem taşıdığını açıkça ortaya koymaktadır.

Yüksek riskli popülasyonlarda akciğer kanserinin erken evrede saptanmasına yönelik en etkili yaklaşımlardan biri, düşük doz Bilgisayarlı Tomografi (DDBT) ile yapılan tarama programlarıdır. ABD Ulusal Akciğer Tarama Çalışması (NLST) ve Avrupa merkezli NELSON çalışması, DDBT taramalarının akciğer kanseri mortalitesini anlamlı ölçüde azalttığını ortaya koymuştur (Aberle ve ark., 2011). Bilgisayarlı tomografi, akciğer parankimini yüksek çözünürlükle görüntüleyerek birkaç milimetre çapındaki küçük nodüllerin dahi tespit edilmesine imkân tanımakta ve bu nedenle akciğer kanseri şüphesi bulunan olgularda birincil görüntüleme yöntemi olarak kullanılmaktadır. Bununla birlikte, tarama programlarının yaygınlaşmasıyla birlikte ortaya çıkan yüksek hacimli görüntü verisinin hızla ve standart biçimde değerlendirilmesi gerekliliği, radyologlar üzerindeki iş yükünü artırmakta ve subjektif yorum farklılıklarına bağlı hataları gündeme getirmektedir. Bu bağlamda Bilgisayar Destekli Tanı (BDT) sistemleri ve özellikle yapay zekâ tabanlı otomatik analiz yöntemleri, klinik kararlara destek sağlayabilecek önemli araçlar olarak öne çıkmaktadır.

Yapay zekâ ve derin öğrenme alanındaki son gelişmeler, tıbbi görüntü analizinde önemli bir paradigma değişimi yaratmıştır. Özellikle Evrişimsel Sinir Ağları (Convolutional Neural Networks, CNN), binlerce BT dilimindeki karmaşık görsel örüntüleri otomatik olarak öğrenebilmekte, insan gözüyle ayırt edilmesi güç lezyonları tespit edebilmekte ve benign–malign ayrımında yüksek sınıflandırma performansına ulaşabilmektedir (Litjens ve ark., 2017). Uluslararası literatürde akciğer BT görüntülerinden nodül tespiti ve malignite riskinin tahmini amacıyla derin öğrenme yöntemlerinin kullanıldığı çok sayıda çalışma bulunmaktadır (Setio ve ark., 2017; Ardila ve ark., 2019). Türkiye’de de son yıllarda derin öğrenme ve medikal görüntüleme alanında akciğer kanserine odaklanan çeşitli çalışmalar yapılmıştır (Çifçi, 2022; Koç ve Özdemir, 2024). Ancak mevcut çalışmaların önemli bir kısmında tek bir model mimarisine odaklanıldığı, farklı

alışmaların genellikle farklı veri setleri ve deneysel protokoller kullandığı ve çoğu derin öğrenme modelinin karar süreçlerinin açıklanabilirlik açısından sınırlı kaldığı görölmektedir.

Bu durum, literatürde hem farklı mimarileri aynı veri seti ve aynı deneysel koşullar altında karşılaştıran hem de açıklanabilir yapay zekâ tekniklerini kullanan alışmalara ihtiyaç olduğunu göstermektedir.

Bu tez alışması, söz konusu ihtiyaçlardan hareketle, akciğer kanserinin BT görüntülerinden otomatik olarak sınıflandırılmasını amaçlayan derin öğrenme modellerinin performansını sistematik ve karşılaştırmalı bir çerçevede değerlendirmeyi hedeflemektedir. alışmada, uluslararası ve açık erişimli IQ-OTH/NCCD akciğer BT veri kümesi kullanılarak, yaygın olarak kullanılan beş farklı derin öğrenme mimarisi olan VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3'ün performansları aynı veri seti ve benzer hiperparametre ayarları altında karşılaştırılmaktadır. Böylece tek bir modele odaklanan alışmaların aksine, farklı mimarilerin avantaj ve sınırlılıklarının nesnel olarak ortaya konulması amaçlanmaktadır.

alışmanın özgün yönlerinden biri, yalnızca sınıflandırma performansına (doğruluk, duyarlılık, özgülük, ROC eğrisi altında kalan alan (AUC) gibi metriklere) odaklanmakla kalmayıp, aynı zamanda modellerin karar verirken görüntü üzerinde odaklandıkları bölgeleri ortaya koyan açıklanabilir yapay zekâ yöntemlerini de kullanmasıdır. Bu kapsamda Gradient-weighted Class Activation Mapping (Grad-CAM) tekniğı kullanılarak, derin öğrenme modellerinin belirli bir sınıf için hangi bölgesel özellikleri baskın biçimde dikkate aldığı görselleştirilmektedir. Böylece, “kara kutu” niteliğindeki derin öğrenme modellerinin karar mekanizmaları daha şeffaf hâle getirilmekte ve elde edilen sonuçların klinisyenler tarafından yorumlanabilirliği ve güvenilirliği artırılmaktadır.

Bu tezde ele alınan temel problem, akciğer kanserinin BT görüntülerinden erken, doğru ve güvenilir biçimde teşhis edilememesinin; tanı sürecinde gecikmelere, yanlış pozitif/negatif oranlarının artmasına ve hasta sağkalım oranlarının düşmesine yol açmasıdır. Geleneksel tanısal süreçler büyük ölçüde uzman yorumuna ve manuel

analizlere dayandığından, gözden kaçan küçük lezyonlar veya yorum farklılıkları klinik sonuçları olumsuz etkileyebilmektedir. Bu nedenle, tıbbi görüntülerin derin öğrenme algoritmalarıyla işlenerek otomatik, standart ve açıklanabilir karar destek sistemlerine dönüştürülmesi gerekmektedir.

Bu bağlamda çalışmanın temel amacı, IQ-OTH/NCCD veri kümesi üzerinde VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3 mimarilerini eğiterek bu modellerin akciğer kanseri teşhisindeki performanslarını karşılaştırmalı olarak analiz etmektir. Ayrıca Grad-CAM yöntemi kullanılarak bu modellerin kararlarının açıklanabilirliğinin sağlanması ve elde edilen sonuçların klinik anlamlılığının artırılması hedeflenmektedir. Performans ölçütleri (doğruluk, duyarlılık, özgüllük, AUC) çerçevesinde yapılacak kapsamlı değerlendirme ile hangi modelin akciğer kanseri teşhisi açısından daha avantajlı olduğu ortaya konulacaktır.

Çalışmanın hipotezi, farklı derin öğrenme mimarilerinin aynı BT görüntü veri seti ve aynı deneysel koşullar altında karşılaştırılması ve model karar süreçlerinin açıklanabilir yapay zekâ teknikleriyle desteklenmesinin, akciğer kanseri teşhisinde en yüksek performansı sağlayan ve klinik entegrasyona en uygun model veya modellerin belirlenmesine olanak tanıyacağı yönündedir. Böylece, hem literatürdeki boşlukların önemli ölçüde doldurulması hem de klinik uygulama potansiyeli yüksek, nesnel ve yorumlanabilir yapay zekâ tabanlı karar destek sistemlerinin geliştirilmesine katkı sağlanması hedeflenmektedir.

2.LİTERATÜR TARAMASI

Son yıllarda yapay zekâ ve derin öğrenme alanında yaşanan hızlı gelişmeler, tıbbi görüntüleme teknolojilerinde önemli ilerlemelere yol açmıştır. Özellikle bilgisayarlı tomografi (BT) gibi yüksek çözünürlüklü görüntülerin analizi, yapay zekâ tabanlı otomatik teşhis sistemlerinin klinik uygulamalarda giderek daha fazla benimsenmesini sağlamıştır (Litjens ve ark., 2017). Konvolüsyonel sinir ağları (CNN), derin öğrenme mimarileri içinde tıbbi görüntü analizinde öne çıkmakta ve lezyon tespiti, sınıflandırma ve segmentasyon gibi görevlerde üstün performans sunmaktadır (Shen ve ark., 2017; Setio ve ark., 2017).

Akciğer kanseri tanısında, düşük doz BT görüntüleme ile elde edilen veri hacminin giderek artması, klasik gözle değerlendirmeye dayalı yöntemlerin yetersizliğini ortaya koymuştur. Bu nedenle, bilgisayar destekli tanı (CAD) sistemlerinin ve derin öğrenme tabanlı analiz yaklaşımlarının önemi artmıştır (Ardila ve ark., 2019). Literatürde, akciğer BT görüntülerinde nodül tespiti ve malign-benign ayrımı için farklı derin öğrenme mimarileriyle yapılan çok sayıda çalışma mevcuttur (Litjens ve ark., 2017; Setio ve ark., 2017; Koç ve Özdemir, 2024).

Bununla birlikte, bu çalışmaların büyük bir kısmında; genellikle tek bir model veya mimari üzerine yoğunlaşılması, kullanılan veri setleri ve değerlendirme metriklerinin birbirinden farklı olması ve modellerin karar mekanizmalarının yeterince şeffaf olmayıp çoğunun “kara kutu” olarak kalması gibi ortak sınırlılıklar dikkat çekmektedir.

Son dönemde, derin öğrenme mimarilerinin performansını, aynı veri seti üzerinde sistematik ve karşılaştırmalı şekilde değerlendiren çalışmaların sayısı artmaktadır. Bu tür çalışmalar, hem model başarımını nesnel olarak karşılaştırmaya hem de klinik uygulamalara yönelik öneriler geliştirmeye olanak tanımaktadır. Ek olarak, açıklanabilir yapay zekâ (Explainable AI, XAI) teknikleri ve Grad-CAM gibi yöntemler ile modellerin hangi görsel özelliklere odaklandığının anlaşılması, model güvenilirliği ve klinik geçerliliğin artırılmasında büyük rol oynamaktadır (Selvaraju ve ark., 2017).

Türkiye’de ise, son yıllarda akciğer kanseri tespiti ve medikal görüntü analizi üzerine yapılan çalışmalar artmakla birlikte, literatürün büyük bir kısmında ya tek model kullanılmış ya da model karşılaştırmaları sınırlı kalmıştır (Çifçi, 2022; Kaya ve Demir, 2021; Koç ve Özdemir, 2024). Bu tez çalışmasının literatüre katkısı, birden fazla modern derin öğrenme mimarisinin (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2, InceptionV3) ilk kez aynı BT veri seti ve aynı deneysel koşullar altında sistematik olarak karşılaştırılmasıyla ortaya konulmaktadır. Model performanslarının nicel olarak değerlendirilmesine ek olarak, Grad-CAM tabanlı açıklanabilirlik analizleri kullanılarak modellerin karar süreçlerinin şeffaflığı ve klinik açıdan yorumlanabilirliği artırılmıştır. Böylece, literatürde sıkça karşılaşılan “tek model, farklı veri seti, farklı metrik” kaynaklı karşılaştırma sorunları bu tezde mümkün olduğunca minimize edilmiş ve doğrudan kıyaslanabilir bulgular sunulmuştur. Sonuç olarak, bu tez, akciğer kanseri tespitinde modern derin öğrenme mimarilerinin karşılaştırmalı analizine odaklanan, açıklanabilirliği önceleyen ve klinik uygulamalara rehberlik edebilecek nitelikte güncel bir literatür yaklaşımı ortaya koymaktadır. Bu bölümde, çalışmanın teorik altyapısını oluşturan temel kavramlar ile tez kapsamında karşılaştırılan derin öğrenme mimarileri ayrıntılı olarak ele alınmaktadır.

2.1. Derin Öğrenme ve Tıbbi Görüntüleme Temelleri

Derin öğrenme (deep learning), çok katmanlı yapay sinir ağlarıyla (deep neural networks) veriden karmaşık, yüksek seviyeli örüntülerin otomatik olarak çıkarılmasını sağlayan bir makine öğrenmesi yaklaşımıdır. 2010 sonrası GPU teknolojilerinin ilerlemesi, büyük etiketli veri kümeleri ve açık kaynak yazılımlar sayesinde; görüntü işleme, ses analizi, doğal dil işleme ve tıbbi tanı gibi birçok alanda hızlı gelişmeler yaşanmıştır (LeCun, Bengio & Hinton, 2015).

Tıbbi görüntüleme ise insan vücudunun iç yapılarının, teşhis, tedavi ve izleme süreçlerinde görüntülenmesini sağlar. Bilgisayarlı tomografi (BT), manyetik rezonans görüntüleme (MRG) ve X-ray gibi modalitelerle elde edilen büyük hacimli görüntü verilerinin analizinde derin öğrenme algoritmaları, klasik yöntemlere kıyasla üstün performans göstermiştir (Litjens ve ark., 2017). Özellikle tümör tespiti, segmentasyon,

hastalık sınıflandırması ve risk analizi gibi görevlerde, derin öğrenme tabanlı otomatik sistemler insan uzmanına yakın, hatta bazı durumlarda daha iyi sonuçlar sunabilmektedir.

2.2. Transfer Öğrenme Yaklaşımları

Transfer öğrenme (transfer learning), bir görev için eğitilmiş derin bir modelin, farklı fakat ilişkili başka bir görevde yeniden kullanılması ilkesine dayanır. Bu yöntem, büyük veri kümeleri (örneğin ImageNet gibi milyonlarca görüntü içeren veri setleri) üzerinde eğitilmiş modellerin, çok daha küçük veri setleri üzerinde hızlıca uyarlanmasını sağlar (Pan & Yang, 2010). Özellikle tıbbi görüntüleme gibi, veri toplamanın ve etiketlemenin zor olduğu alanlarda transfer öğrenme önemli bir avantaj sunar. MATLAB'ın trainNetwork fonksiyonu ve transfer learning iş akışı, literatürde en yaygın ve kabul görmüş yöntemdir (örn. Pan & Yang, 2010; Shin et al., 2016). Özellikle tıbbi görüntüleme alanında, küçük veri setlerinde yüksek doğruluk için transfer öğrenme zorunlu hale gelmiştir.

Bu çalışma kapsamında kullanılan tüm derin öğrenme modelleri, ImageNet veri seti üzerinde önceden eğitilmiş olup, IQ-OTH/NCCD veri kümesi üzerinde ince ayar (fine-tuning) işlemiyle yeniden eğitilmiştir. Bu yaklaşım hem modelin öğrenme hızını artırmakta hem de az sayıda veri ile yüksek doğrulukta sonuç elde edilmesini sağlamaktadır.

2.3. Açıklanabilir Yapay Zekâ (XAI) ve Grad-CAM Yöntemi

Açıklanabilir yapay zekâ (Explainable Artificial Intelligence, XAI), yapay zekâ tabanlı sistemlerin karar alma süreçlerinin şeffaf, anlaşılır ve denetlenebilir olmasını hedefleyen bir araştırma alanıdır. Özellikle sağlık, hukuk ve finans gibi yüksek riskli uygulamalarda, modelin "neden bu kararı verdiği" sorusu hem klinik hem etik hem de yasal açıdan büyük önem taşımaktadır. Derin öğrenme modelleri, karmaşık ve çoğunlukla “kara kutu” şeklinde çalışan sistemler olmaları nedeniyle, bu gerekliliğe yanıt verebilmek için açıklanabilirlik tekniklerine ihtiyaç duyarlar.

Bu çalışmada, eğitilen derin öğrenme modellerinin karar mekanizmalarını görselleştirerek açıklamak amacıyla, literatürde yaygın olarak kullanılan Grad-CAM (Gradient-weighted Class Activation Mapping) yöntemi tercih edilmiştir. Grad-CAM; modelin belirli bir sınıfa ait tahmininin, ağırlık katmanlarındaki aktivasyon haritaları ve bu sınıfa ait gradyan bilgisiyle birleştirilerek, modelin giriş görüntüsünde en çok hangi alanlara odaklandığını gösteren bir "ısı haritası" oluşturur (Selvaraju et al., 2017).

Bu yöntem sayesinde; bir modelin, örneğin bir akciğer BT görüntüsünde kanserli lezyonu hangi görsel ipuçlarına dayalı olarak saptadığı anlaşılır hale gelir. Böylece modelin kararları hekimler ve diğer uzmanlar tarafından görsel olarak denetlenebilir ve klinik güvenilirliği artar. Son yıllarda Grad-CAM tabanlı açıklanabilirlik, tıbbi görüntüleme çalışmalarında model güvenini artıran temel bir standart haline gelmiştir.

2.4. Derin Öğrenmede Kullanılan Ortak Matematiksel Temeller

Bu çalışma kapsamında karşılaştırılan tüm derin öğrenme mimarileri (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3), modern evrimsel sinir ağı tabanlı modellerde standart hale gelmiş ortak bazı temel matematiksel prensipler ve işlemler üzerine yapılandırılmıştır. Literatürde geniş bir uygulama alanına sahip olan bu prensipler, tıbbi görüntülerin sayısal olarak işlenmesi, özneteliklerin otomatik olarak çıkarılması ve sınıflandırılması süreçlerinde yüksek doğruluk ve güçlü genelleme başarısı sağlamaktadır (LeCun, Bengio ve Hinton, 2015; Goodfellow, Bengio ve Courville, 2016).

Her bir model mimari açıdan kendine özgü farklılıklar taşısa da temel katmanlarında benzer matematiksel işlemler kullanılmaktadır. Bu işlemler arasında evrim işlemi, maksimum havuzlama, ReLU gibi aktivasyon fonksiyonları, batch normalization uygulamaları, softmax fonksiyonu, çapraz entropi kayıp fonksiyonu, L2 düzenleme yaklaşımları, dropout mekanizması ve Adam optimizasyon algoritması yer almaktadır (Litjens ve diğerleri, 2017; Szegedy ve diğerleri, 2016). Bu tekniklerin tümü, derin öğrenme modellerinin görüntü sınıflandırma, nesne tespiti ve tıbbi tanı gibi karmaşık problemlerde yüksek performans göstermesinin temel yapı taşlarıdır ve güncel literatürde kritik roller üstlenmektedir (Litjens ve diğerleri, 2017; Goodfellow ve diğerleri, 2016; LeCun ve diğerleri, 2015).

Bu bölümde tekrarları önlemek amacıyla, tüm modellerde ortak olarak kullanılan söz konusu matematiksel kavram ve fonksiyonlar bir arada ve literatür destekli biçimde sunulmaktadır. Böylece izleyen bölümlerde her bir mimarinin kendine özgü yenilikleri ayrı ayrı vurgulanırken, ortak prensipler gereksiz tekrar yükünden arındırılmış olacaktır (Pan ve Yang, 2010). Bu kapsamda ele alınan temel içerik; evrişim işlemi, maksimum havuzlama, batch normalizasyonu, ReLU aktivasyon fonksiyonu, softmax fonksiyonu, çapraz entropi kayıp fonksiyonu, L2 düzenleme stratejileri, dropout mekanizması ve Adam optimizasyon algoritmasından oluşmaktadır.

2.4.1. Evrişim (convolution) işlemi

Evrişim işlemi, derin öğrenmede ve özellikle evrişimli sinir ağlarında (CNN) en temel bilgi çıkarma mekanizmasıdır. Bu işlem sayesinde, görüntülerin yerel bölgelerinden filtreler aracılığıyla öznitelikler (kenar, köşe, doku) elde edilir. Evrişim, giriş görüntüsünü (I) üzerinde tanımlanan küçük bir çekirdek (K) ile kaydırmalı olarak çarpıp toplama işlemiyle yeni bir öznitelik haritası (feature map) üretir. Özellikle tıbbi görüntülemelerde, lezyon, nodül gibi küçük ama anlamlı bölgelerin yakalanabilmesi için lokal yapıların modellenmesi kritik öneme sahiptir (Goodfellow et al., 2016). Bu işlemin matematiksel eşitlik ifadesi 2.1’de sunulmuştur.

$$M(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (2.1)$$

Eşitlik 2.1. örneğinde olduğu gibi I giriş görüntüsü matrisi, K Evrişim çekirdeği (filtre), M(i,j) koordinatında elde edilen öznitelik haritası değerini, (m,n) ise çekirdek koordinatları ifade etmektedir. Bu matematiksel formülasyon, farklı konum ve ölçeklerde öne çıkan bölgesel desenlerin yakalanmasını mümkün kılar. Özellikle InceptionV3 gibi gelişmiş CNN mimarilerinde, birden fazla farklı boyutta çekirdeğin paralel kullanımı sayesinde çok ölçekli özellik çıkarımı mümkün olur ve karmaşık tıbbi görüntülerde (örneğin akciğer nodülleri) olağanüstü performans sergilenir (Szegedy et al., 2016; Zhang et al., 2023). Evrişim katmanlarının derinliği ve kullanılan filtre sayısı, modelin öğrenebileceği örüntülerin çeşitliliğini ve karmaşıklığını doğrudan etkiler. Nitekim, VGG16 mimarisi homojen ve derin evrişim katmanları kullanırken; ResNet-50 ise “artık bağlantı” (residual connection) ekleyerek evrişimli katmanlarda bilgi kaybını minimize

eder (Simonyan & Zisserman, 2015; He et al., 2016). Bu farklılıklar, her modelin tıbbi görüntü işleme başarısında ayırt edici rol oynar.

2.4.2. Maksimum havuzlama (max pooling)

Maksimum havuzlama (max pooling), konvolüsyonel sinir ağlarında (CNN) en çok kullanılan havuzlama yöntemidir ve temel amacı, öznitelik haritalarının (feature maps) uzamsal boyutunu küçültmek, modelin parametre sayısını azaltmak ve en belirgin aktivasyonları korumaktır (Szegedy et al., 2016; Goodfellow et al., 2016). Bu yöntem, her havuzlama penceresindeki (ör. 2×2 veya 3×3) en yüksek değeri çıkışa aktarır ve bu sayede model, önemli lokal özellikleri kaybetmeden daha yüksek soyutlama düzeylerine ulaşabilir. Matematiksel olarak maksimum havuzlama işlemi eşitlik 2.2'deki gibi tanımlanır.

$$y_{i,j} = \max_{(m,n) \in R_{i,j}} x_{m,n} \quad (2.2)$$

Eşitlik 2.2. örneğinde olduğu gibi, $y_{i,j}$ havuzlama işlemi sonrasında elde edilen çıktı matrisinin (i,j) koordinatındaki değeri olup, ilgili bölgedeki en yüksek aktivasyonu temsil eder. Benzer şekilde $x_{m,n}$, giriş öznitelik haritasının (m,n) koordinatındaki değerlerini ifade eder. $R(i,j)$ ise (i,j) konumuna karşılık gelen havuzlama penceresini belirtir ve bu pencere genellikle 2×2 veya 3×3 boyutlarında tanımlanır. Bu işlemde kullanılan “max” operatörü, $R_{i,j}$ bölgesi içindeki tüm $x_{m,n}$ değerlerini değerlendirerek en büyük olanını seçer ve böylece havuzlama katmanından geçen en baskın yerel bilginin korunmasını sağlar

Maksimum havuzlama işlemi, modelin küçük kaymalara karşı dayanıklılığını (translation invariance) artırırken, gürültü ve varyasyonlara karşı da kararlılık sağlar (LeCun et al., 2015; Goodfellow et al., 2016). Özellikle tıbbi görüntü analizlerinde, lezyon, nodül veya tümör gibi küçük ve klinik olarak önemli yapıların korunmasında bu yaklaşım kritik öneme sahiptir (Litjens et al., 2017). Derin modellerin (örneğin, VGG16, ResNet-50 ve InceptionV3) erken evrelerinde yaygın olarak kullanılan bu katmanlar, daha üst düzey özniteliklerin çıkarılmasını ve verinin daha kompakt bir temsille işlenmesini mümkün kılar (Simonyan & Zisserman, 2015; He et al., 2016; Szegedy et al., 2016).

Bu işlem, özellikle düşük çözünürlüklü nodül, tümör ve lezyon gibi klinik bulguların varlığını saptamada gürültü ve varyasyonlara karşı modelin daha kararlı olmasını sağlar. Maksimum havuzlama, her penceredeki en büyük öznitelik değerini çıkış matrisine aktarır. Böylece, model önemli lokal bilgileri kaybetmeden uzamsal boyutu küçültür ve daha derin katmanlarda işlemeyi hızlandırır. Bu yaklaşım, özellikle medikal görüntülerde önemli küçük detayların korunmasını ve modelin klinik anlamda güvenilir sonuçlar üretmesini sağlar.

2.4.3. Batch normalizasyonu (batch normalization)

Batch normalization (BN), derin sinir ağlarının eğitiminde “internal covariate shift” (katman giriş dağılımlarının sürekli değişmesi) sorununu azaltmak için geliştirilmiş, modern derin öğrenme literatüründe çığır açıcı bir yöntemdir (Ioffe & Szegedy, 2015).

BN, her mini-batch’te katman aktivasyonlarını normalize ederek öğrenme sürecini hem hızlandırır hem de daha kararlı hale getirir. Özellikle derin mimarilerde, öğrenme oranını artırmaya, gradyanların patlaması veya sönümlenmesini engellemeye ve ağırlık güncellemelerinde istikrar sağlamaya önemli katkı sunar (Szegedy et al., 2016; Goodfellow et al., 2016). Derin öğrenme modellerinde, eğitim sürecini hızlandırmak ve kararlılığı artırmak amacıyla Yığın Normalizasyonu (Batch Normalization) gibi teknikler sıkça kullanılmaktadır. Bu işlem matematiksel olarak Eşitlik 2.3 ve Eşitlik 2.4’de gösterilmiştir:

$$\widehat{x^{(k)}} = \frac{x^{(k)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (2.3)$$

$$y^{(k)} = \gamma \widehat{x^{(k)}} + \beta \quad (2.4)$$

Eşitlik 2.3. ve eşitlik 2.4. örneğinde gösterildiği üzere, x_k , normalize edilen özniteliğin k’inci bileşenini; μ_{mini} mini-batch içerisindeki ortalama değeri; σ_{mini}^2 mini-batch varyansını; ϵ sayısal kararlılığı sağlamak için eklenen küçük pozitif sabiti; γ öğrenilebilir ölçek parametresini ve β öğrenilebilir kaydırma (bias) parametresini ifade etmektedir.

Elde edilen $\hat{x}_k$ normalizasyon işleminin çıktısını, y_k ise ölçeklendirme ve kaydırma sonrası son değeri göstermektedir.

Batch Normalizasyonu (Batch Normalization), özellikle InceptionV3, ResNet-50 ve EfficientNet gibi modern ağlarda eğitimin erken aşamalarında modelin çok daha hızlı ve stabil bir şekilde yakınsamasını sağlar (He et al., 2016; Tan & Le, 2019). Ayrıca, aşırı öğrenme riskini azaltarak genelleme başarımını yükseltir. Medikal görüntü analizinde, küçük veri setleriyle çalışırken bile yüksek doğruluk ve güvenilirlik elde edilmesine katkı sağlar (Litjens et al., 2017).

2.4.4. Relu aktivasyon fonksiyonu

ReLU (Rectified Linear Unit) aktivasyon fonksiyonu, derin öğrenme mimarilerinde doğrusal olmayanlık kazandırmanın en etkili yollarından biri olarak kabul edilmektedir (Agarap, 2018; Nwankpa et al., 2018). Fonksiyonun temel özelliği, negatif girişleri sıfıra eşitleyip pozitif girişleri olduğu gibi aktarmasıdır. Bu yapı, klasik sigmoid ve tanh gibi fonksiyonlara göre daha hızlı yakınsama ve gradyan yayılımı sağlar (Glorot et al., 2011). Ayrıca, ReLU, hesaplama açısından verimliliğiyle derin ağlarda eğitim süreçlerini ciddi biçimde kolaylaştırır. Matematiksel olarak aşağıdaki gibi ifade edilir.

$$\text{ReLU}(x) = \max(0, x) \quad (2.5)$$

Eşitlik 2.5. örneğinde gösterildiği gibi, x , aktivasyon fonksiyonuna uygulanan giriş değerini; $\max(0, x)$ ise 0 ve x değerlerinden büyük olanı seçen maksimum fonksiyonunu belirtmektedir.

Bu avantajlarından dolayı ReLU, günümüzde sadece bilgisayarla görü ve doğal dil işleme gibi alanlarda değil, aynı zamanda tıbbi görüntü analizinde de standart olarak kullanılmaktadır (Özkan et al., 2021; Ertürk & Uçar, 2022). Özkan ve ark. (2021), ReLU'nun meme ve akciğer kanseri görüntü sınıflandırmasında sigmoid ve tanh'a göre anlamlı derecede yüksek doğruluk sağladığını raporlamışlardır. Benzer şekilde, Ertürk ve Uçar (2022), beyin MR görüntülerinde ReLU kullanılan CNN modellerinin hem eğitim süresi hem de sınıflandırma başarısı açısından üstün olduğunu belirtmişlerdir.

2.4.5. Softmax fonksiyonu

Softmax fonksiyonu, çok sınıflı sınıflandırma problemlerinde, bir sinir ağının çıktı skorlarını (logitleri) normalize ederek her sınıf için bir olasılık dağılımı elde edilmesini sağlar (Goodfellow et al., 2016; Bishop, 2006). Her bir sınıf için modelin verdiği skorun üstel değeri alınır ve bu değerler tüm sınıfların üstel değerleri toplamına bölünerek 0 ile 1 arasında bir olasılığa dönüştürülür. Böylece çıktı değerleri toplamı 1 olacak şekilde normalize edilir ve modelin bir girdiyi hangi sınıfa atadığına ilişkin güven düzeyi doğrudan okunabilir (Murphy, 2012).

Tıbbi görüntü analizlerinde, softmax çıktısı özellikle *benign*, *malignant* veya *normal* gibi kategorilerin ayırımında klinik olarak yorumlanabilir olasılık üretmesi açısından önem taşır (Kumar et al., 2022; Şahin et al., 2020). Matematiksel olarak softmax fonksiyonu eşitlik 2.6. gibi ifade edilir:

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_{j=1}^K \exp(z_j)} \quad (2.6)$$

Eşitlik 2.6'te gösterildiği üzere, $\hat{y}_i$ 'inci sınıfa ait tahmini olasılığı; z_i 'inci sınıf için model tarafından üretilen logit skorunu; K toplam sınıf sayısını; $\exp(\cdot)$ üstel fonksiyonu ifade etmektedir. Elde edilen tüm $\hat{y}_i$ değerleri 0 ile 1 arasında olup toplamaları 1'e eşittir. Bu sayede model çıktıları olasılıksal olarak yorumlanabilir hale gelmekte ve klinik karar destek süreçlerinde güven skoru üretimi mümkün olmaktadır (Özçift, 2023).

2.4.6. Kayıp fonksiyonu (çapraz entropi)

Çapraz entropi (cross-entropy) kayıp fonksiyonu, makine öğrenmesi ve derin öğrenme tabanlı sınıflandırma modellerinde tahmin edilen olasılık dağılımı ile gerçek etiketler arasındaki uyumsuzluğu ölçmek amacıyla yaygın biçimde kullanılır (Goodfellow et al., 2016; Bishop, 2006). Çok sınıflı sınıflandırma problemlerinde gerçek etiketler genellikle one-hot kodlama ile ifade edilir; yani yalnızca doğru sınıf için 1, diğerleri için 0 değeri atanır (Murphy, 2012). Çapraz entropi, modelin doğru sınıfa verdiği olasılık arttıkça kaybı

azaltır ve model parametrelerinin optimizasyonunda yönlendirici olur (Aksoy & Kır, 2022).

Tıbbi görüntü sınıflandırma uygulamalarında, çapraz entropi fonksiyonu, modelin kanserli ve kanser olmayan gibi kritik ayrımlarda, doğru sınıfa yüksek olasılık atayabilmesini teşvik ettiği için klinik doğruluk açısından belirleyici rol oynar (Shen et al., 2017). Matematiksel olarak çapraz entropi kaybı aşağıdaki gibi tanımlanır:

$$L = - \sum_{i=1}^K y_i \log(\hat{y}_i) \quad (2.7)$$

Eşitlik 2.7’de L , çapraz entropi kayıp fonksiyonunun çıktısını; K , toplam sınıf sayısını; y_i , gerçek etiket değerini (one-hot gösterimde doğru sınıf için 1, diğer tüm sınıflar için 0); $\hat{y}_i$ ise modelin i ’inci sınıfa atadığı tahmini olasılığı ifade etmektedir. Çapraz entropi kaybı, modelin doğru sınıfa verdiği olasılık arttıkça azalmakta olup, tahmin edilen olasılık dağılımının gerçek etiketlere ne ölçüde yaklaştığını nicel olarak değerlendirmek için kullanılmaktadır. Bu formül sayesinde, model çıktılarının gerçek etiketlere yakınsaması sağlanır; tıbbi sınıflandırma gibi hata maliyeti yüksek olan alanlarda optimum performans elde edilir (Kaya & Şahin, 2022).

2.4.7. L2 düzenlileştirme (regularization)

Aşırı öğrenme (*overfitting*), derin öğrenme modellerinin eğitim verisine fazlaca uyum sağlayıp, yeni ve görülmemiş verilerde düşük başarı göstermesi ile sonuçlanan önemli bir problemdir. Bu sorunu önlemeye yönelik en yaygın tekniklerden biri L2 düzenlileştirmedir (Ng, 2004; Goodfellow et al., 2016). L2 düzenlileştirme, modelin ağırlıklarının büyüklüğünü sınırlandırarak karmaşıklığı azaltır ve parametrelerin gereğinden fazla büyümesini engeller. Böylece modelin genelleme yeteneği artar ve aşırı öğrenme riski azalır (Srivastava et al., 2014).

Özellikle tıbbi görüntü işleme ve küçük veri setleriyle çalışılan klinik çalışmalarda, L2 regularization’ın model güvenilirliği ve tekrar edilebilirliği açısından kritik rol oynadığı gösterilmiştir (Litjens et al., 2017). Matematiksel olarak L2 düzenlileştirme, eşitlik 2.8’deki gibi ifade edilir.

$$L_{\text{total}} = L + \lambda \sum_i \omega_i^2 \quad (2.8)$$

Eşitlik 2.8. örneğinde olduğu gibi L_{total} , düzenlileştirme uygulanmış toplam kayıp fonksiyonunu; L , ana kayıp terimini (örneğin çapraz entropi kaybını); λ , düzenlileştirme katsayısını (pozitif bir skaler); ω_i ise modeldeki tüm öğrenilebilir ağırlık parametrelerini ifade etmektedir. L2 düzenlileştirme terimi olan $\lambda \sum_i \omega_i^2$, büyük ağırlık değerlerini cezalandırarak modelin aşırı öğrenme eğilimini azaltmakta ve genelleme performansını artırmaktadır. Bu nedenle, özellikle tıbbi görüntü sınıflandırma gibi yüksek boyutlu öznelik uzaylarında çalışan derin öğrenme modellerinde yaygın olarak tercih edilmektedir. Bu formül, büyük ağırlık değerlerinin cezalandırılması yoluyla modelin parametrelerinin normunu küçük tutar; bu sayede fazla karmaşıklıktan kaynaklı öğrenme hatalarını önler. L2 düzenlileştirmenin özel olarak tıbbi görüntü analizinde önerilmesinin nedeni, yüksek boyutlu özellik uzaylarında dahi modelin gereksiz şekilde büyümesini engelleyerek, düşük sayıda hasta örneğiyle bile güvenilir modeller geliştirilmesini sağlamasıdır (Razzak et al., 2018).

2.4.8. Dropout (bırakma) mekanizması

Dropout, derin öğrenme modellerinde aşırı uyumlamayı (overfitting) önlemek için en yaygın kullanılan düzenlileştirme (*regularization*) tekniklerinden biridir (Srivastava et al., 2014). Temel olarak, eğitim sırasında modeldeki bazı nöronların rastgele ve geçici olarak devre dışı bırakılması (maskelenmesi) esasına dayanır. Bu yöntem, ağırlık farklı alt yapılar üzerinde bağımsız olarak öğrenme yapmasını sağlayarak, nöronlar arası gereksiz bağımlılıkları azaltır ve modelin genelleme yeteneğini artırır (Goodfellow et al., 2016; Baldi & Sadowski, 2013).

Özellikle tıbbi görüntü işleme uygulamalarında ve sınırlı sayıda örneğin bulunduğu klinik veri setlerinde, Dropout'un aşırı öğrenmeyi önlemedeki etkinliği birçok çalışmada gösterilmiştir (Razzak et al., 2018; Litjens et al., 2017). Matematiksel olarak, bir nöronun Dropout uygulanmış aktivasyonu $\tilde{h}_i$, orijinal aktivasyon h_i ile Bernoulli dağılımına göre belirlenen maske değişkeni d_i çarpılarak hesaplanır.

$$\tilde{h}_i = h_i \cdot d_i \quad , \quad d_i \sim \text{Bernoulli}(p) \quad (2.9)$$

Eşitlik 2.9. örneğinde olduğu gibi h_i , i . nöronun dropout uygulanmamış (normal) aktivasyonunu; d_i , Bernoulli dağılımından rastgele üretilen ve değeri 0 veya 1 olabilen maske değişkenini; p ise nöronun eğitim sırasında aktif kalma olasılığını ifade etmektedir. Eğitim aşamasında her minibatch için bu maske rastgele yeniden oluşturulur ve böylece model, farklı alt ağlar üzerinden öğrenme gerçekleştirerek aşırı uyumlamaya karşı daha dayanıklı hale gelir. Test aşamasında dropout uygulanmaz; bunun yerine, eğitim sürecinde kullanılan aktivasyonların beklenen değerinin korunması amacıyla nöron çıktıları p ile ölçeklendirilir.

Dropout'un getirdiği çeşitlilik, modelin her eğitim adımında rastgele farklı bir alt ağ (sub-network) üzerinde öğrenme yapmasını sağlar. Bu yaklaşım, aslında tek bir model içinde binlerce farklı küçük ağın eğitilmesi ve test sırasında bu alt modellerin "toplulaştırılmış" (ensemble) bir tahmin üretmesi gibi çalışır (Srivastava et al., 2014). Yani Dropout, klasik ensemble learning yöntemlerinde olduğu gibi çoklu modellerin ortalamasına benzer bir genel performans artışı sağlar.

Özellikle son yıllarda tıbbi yapay zekâ uygulamalarında, veri setlerinin sınırlı olduğu durumlarda aşırı öğrenmenin (overfitting) önlenmesinde Dropout temel ve etkili bir araç olarak önerilmektedir (Razzak et al., 2018; Goodfellow et al., 2016).

2.4.9. Adam optimizasyon algoritması

Adam (Adaptive Moment Estimation), derin öğrenme modellerinde yaygın olarak kullanılan, parametre güncellemelerini hızlandırmak ve stabil hale getirmek amacıyla geliştirilmiş uyarlanabilir bir optimizasyon algoritmasıdır. Adam, hem birinci moment (ortalama gradyan) hem de ikinci momentin (gradyanın karesi) hareketli ortalamalarını izleyerek, öğrenme hızını her parametre için dinamik olarak ayarlayabilmektedir. Bu yaklaşım, klasik Stokastik Gradient İnişi (SGD) ve Momentum gibi yöntemlerin bazı eksiklerini giderir ve özellikle yüksek boyutlu, karmaşık ve dengesiz veri setlerinde hızlı ve kararlı yakınsama sağlar (Kingma & Ba, 2015; Ruder, 2016). Adam algoritmasının güncelleme adımları eşitlik 2.10. ve eşitlik 2.11.'de sunulmuştur.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.10)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.11)$$

Burada; g_t : t -inci adımda hesaplanan gradyan, m_t : Birinci momentun (gradyan ortalaması) güncellenmiş değeri, v_t : İkinci momentin (gradyan karelerinin ortalaması) güncellenmiş değeri, β_1 ve β_2 : Moment hareketlilik katsayılarıdır ve genellikle $\beta_1=0.9$, $\beta_2=0.999$ olarak seçilir.

Bias Düzeltmeleri ise eşitlik 2.12. örneğindeki gibidir.

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \widehat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (2.12)$$

Bias düzeltme terimleri, moment değerlerinin başlangıçta sıfıra yakın olmasından kaynaklanan sapmayı düzeltmek için uygulanır.

Parametre güncellemesi Eşitlik 2.13. örneğinde olduğu gibidir.

$$\theta_{t+1} = \theta_t - \alpha \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \quad (2.13)$$

Bu eşitlikte θ_t , algoritmanın t . adımındaki model parametrelerini; α , öğrenme oranını; ϵ ise sayısal kararlılığı sağlamak için kullanılan küçük pozitif bir sabiti ifade etmektedir. Güncelleme kuralında, birinci moment tahmini $\widehat{m}_t$ ve ikinci moment tahmini $\widehat{v}_t$ kullanılarak her bir parametre için uyarlanabilir bir adım büyüklüğü hesaplanır. Böylece Adam optimizasyon algoritması, gradyanların büyüklüğüne ve yönüne bağlı olarak parametreleri daha dengeli biçimde güncelleyerek hızlı ve kararlı bir yakınsama performansı sağlar.

Adam algoritması, bu şekilde her parametre için ayrı ayrı adaptif öğrenme oranları hesaplayarak optimizasyon sürecinin verimliliğini artırır ve daha stabil bir öğrenme sağlar. Adam algoritması, her parametre için ayrı ayrı adaptif öğrenme oranı hesaplar, bu sayede öğrenme sürecinin verimliliğini artırır ve özellikle gürültülü veya seyrek gradyanların görüldüğü derin ağlarda üstün performans sunar (Kingma & Ba, 2015; Goodfellow et al., 2016; Reddi et al., 2019).

Adam algoritması, son yıllarda bilgisayarla görme ve doğal dil işleme başta olmak üzere çok çeşitli derin öğrenme uygulamalarında en yaygın kullanılan optimizasyon yöntemlerinden biri haline gelmiştir (Ruder, 2016).

2.5. Karşılaştırılacak Derin Öğrenme Modellerinin Teorik Analizi

Bu bölümde, tez kapsamında karşılaştırılan beş farklı derin öğrenme mimarisi ; VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3 hem mimari yapıları, hem matematiksel temelleri hem de güncel literatürdeki uygulama ve başarıları açısından kapsamlı biçimde analiz edilmiştir. Her bir modelin ayırt edici yönleri, teknik özellikleri ve tıbbi görüntüleme bağlamındaki yeri sistematik şekilde sunulmuştur.

2.5.1. Vgg16 mimarisi, matematiksel temeli ve literatürdeki yeri

VGG16, Oxford Üniversitesi Visual Geometry Group tarafından 2014 yılında geliştirilmiş ve “homojen yapı” konseptini derin öğrenme alanında yaygınlaştıran ikonik bir konvolüsyonel sinir ağı mimarisidir (Simonyan & Zisserman, 2015). Modelde toplam 13 adet evrişim (convolutional) ve 3 adet tam bağlı (fully connected) katman yer almaktadır. Tüm evrişim katmanlarında 3×3 boyutunda filtreler ve sabit padding uygulanır, bu da kenar bilgisinin kaybolmadan aktarılmasına imkan tanır. Mimari yapı, ardışık konvolüsyon blokları, sonrasında max pooling katmanları, ardından ise tam bağlı katmanlar ve softmax çıkışından oluşur.

Yaklaşık 138 milyon parametreye sahip olan VGG16, hesaplama açısından maliyetli olsa da, mimarinin basitliği, yüksek performansı ve transfer öğrenmeye açıklığı sayesinde günümüzde hem akademik hem de pratik çalışmalarda sıklıkla tercih edilmektedir (Zhang et al., 2023; Rawat & Wang, 2017). VGG16’da ardışık küçük (3x3) filtrelerle evrişim katmanlarında özellik çıkarımı yapılır. Her evrişim katmanında aşağıdaki matematiksel işlem uygulanır. Her bir evrişim katmanında uygulanan temel matematiksel işlem aşağıdaki gibi formüle edilir (Dumoulin & Visin, 2016).

$$y = f(W * x + b) \quad (2.14)$$

Eşitlik 2.14. örneğinde olduğu gibi, evrimsel katmanda veri akışının ve özellik çıkarımının temeli oluşturur. x giriş verisi (görüntü veya önceki katmanın çıktısı), öğrenilebilir filtre ağırlıkları (W) ile evrim işlemine tabi tutulur, buna bir yanlılık (bias) terimi (b) eklenir ve sonuç, doğrusal olmayanlık katmak amacıyla bir ReLU aktivasyon fonksiyonu (f) üzerinden geçirilir. Ortaya çıkan y çıktısı, ardışık katmanlarda işlenmek üzere yeni bir özellik haritası olarak aktarılır. Bu ardışık yapı sayesinde, VGG16 modeli düşükten yükseğe tüm seviyelerde öznetelik öğrenimini mümkün kılar (Simonyan & Zisserman, 2015; Rawat & Wang, 2017; Li et al., 2022). Modelin son katmanında elde edilen çıktı skorları, softmax fonksiyonu ile çok sınıflı olasılıklara dönüştürülerek nihai sınıflandırma kararı alınır. Bu süreç, tıbbi görüntü analizinde yüksek doğruluk ve güvenilirlik gerektiren uygulamalar açısından kritik öneme sahiptir.

VGG16 modeli, tıbbi görüntü sınıflandırma alanında uzun yıllardır referans mimari olarak kabul edilmektedir (Simonyan & Zisserman, 2015). Özellikle akciğer ve meme kanseri tespiti gibi klinik uygulamalarda, geniş veri setlerinde yüksek doğruluk oranlarına ulaşması nedeniyle sıklıkla tercih edilmektedir (Aksel et al., 2022; Kaya & Demir, 2021).

IQ-OTH/NCCD veri seti üzerinde yapılan yerli çalışmalar, VGG16 ile yüzde 87 doğruluk ve yüzde 0.92 AUC elde edildiğini göstermiştir (Kaya & Demir, 2021). Ayrıca, beyin tümörü (Shen et al., 2017), cilt lezyonu (Han et al., 2021) ve çoklu organ görüntüleme (Litjens et al., 2017) gibi farklı tıbbi sınıflandırma problemlerinde, transfer öğrenme yaklaşımıyla başarıyla uygulanmıştır.

VGG16'nın başlıca avantajları arasında sade ve derin mimarisi ile transfer öğrenme uygulamalarına olan yüksek uyumluluğu yer almaktadır. Ancak, yaklaşık 138 milyon parametre içermesi nedeniyle hesaplama maliyeti yüksektir ve eğitim-deployman süreçlerinde güçlü donanım ihtiyacı doğurmaktadır. Buna rağmen, güncel literatürde transfer öğrenme için bir "benchmark" olarak kullanılmaya devam etmektedir.

2.5.2. Resnet-50 mimarisi, matematiksel temeli ve literatürdeki yeri

ResNet-50, Microsoft Research tarafından 2015 yılında tanıtılmış ve derin öğrenme alanında "artık öğrenme" (residual learning) paradigmasını başlatan devrim niteliğinde

bir konvolüsyonel sinir ağı mimarisidir (He et al., 2016). Toplamda 50 katmandan oluşan bu modelin en önemli yeniliği, klasik ardışık katmanların yerine “skip connection” (atlama bağlantıları) içeren residual bloklar kullanmasıdır. Bu bloklar sayesinde, modelin derinliği artırıldığında ortaya çıkan gradyan sönümlenmesi (vanishing gradient) problemi büyük ölçüde çözülür. Her residual blok, giriş ile blok içindeki dönüşümün (evrişim, batch normalization, aktivasyon) doğrudan toplanmasına dayanır.

Bu yaklaşım, modelin hem çok derin yapılarda istikrarlı şekilde eğitilmesini hem de yüksek doğruluklara ulaşmasını sağlar. ResNet-50 mimarisi, ardışık olarak residual bloklardan, max pooling katmanlarından ve global average pooling ile tamamlanan bir çıktı katmanından oluşur. Transfer öğrenme açısından esnekliği ve geniş veri setlerinde gösterdiği performansla, günümüzün en çok referans verilen derin öğrenme modellerindendir (Rawat & Wang, 2017; Li et al., 2022). ResNet-50 mimarisinin temelinde, residual bloklarda uygulanan eşitlik 2.15’teki matematiksel ilişki yer alır (He et al., 2016; Dumoulin & Visin, 2016).

$$y = f(F(x, \{W_i\}) + x) \quad (2.15)$$

Eşitlik 2.15. örneğinde olduğu gibi x , residual bloğa giren giriş vektörünü veya önceki katmandan elde edilen özellik haritasını; $F(x, \{W_i\})$, giriş üzerinde gerçekleştirilen dönüşümleri (evrişim, batch normalization ve aktivasyon gibi işlemler) ve bu dönüşümlere ait öğrenilebilir ağırlık parametrelerini; $\{W_i\}$, blok içerisinde yer alan tüm evrişim ve normalizasyon işlemlerine ait ağırlık kümesini; f , genellikle ReLU olmak üzere bloğun çıkışında uygulanan aktivasyon fonksiyonunu; y ise orijinal giriş ile dönüşümün toplandığı, residual yapının karakteristik çıktısını ifade etmektedir. Bu yapı, derin ağlarda gradyan akışını kolaylaştırarak bilgi kaybını azaltır ve çok katmanlı modellerin daha verimli biçimde öğrenmesini sağlar. Bu formül, modelin her blokta girişle çıkış arasındaki farkı (residual) koruyarak, derinliğin getirdiği öğrenme problemlerini minimize ettiğini gösterir. Bu formülasyon, ağı yalnızca girişle çıkış arasındaki farkı (“artığı”) öğrenmesini mümkün kılar ve çok derin mimarilerin gradyan kaybı yaşamadan eğitilmesini sağlar. Modelin sonunda ise, tam bağlı (fully connected) katmanda üretilen skorlar softmax fonksiyonu ile çok sınıflı olasılıklara dönüştürülür.

ResNet-50, hem doğal hem tıbbi görüntü analizinde “altın standart” olarak kabul edilmektedir (He et al., 2016; Litjens et al., 2017). Ardila ve ark. (2019), akciğer kanseri BT görüntülerinde geliştirdikleri 3D ResNet ile radyolog seviyesinde doğruluk raporlamıştır. IQ-OTH/NCCD veri setinde ResNet-50 tabanlı uygulamalarda yüzde 91,3 doğruluk ve yüzde 0.95 AUC elde edilmiştir (Yıldız & Koç, 2021). Ayrıca, beyin tümörü (Uğur & Güneş, 2022), COVID-19 tespiti (Kara et al., 2021), cilt lezyonu analizi (Han et al., 2021) gibi çok sayıda tıbbi görüntüleme probleminde yüksek başarıyla uygulanmıştır. Modelin başlıca avantajları; çok derin yapılarda dahi istikrarlı eğitim, transfer öğrenme için yüksek esneklik ve pratikte referans alınan bir performans standardı sunmasıdır. En önemli dezavantajı ise, bazı uygulamalarda yüksek parametre ve hesaplama maliyeti getirebilmesidir.

2.5.3. EfficientNet-B0 mimarisi, matematiksel temeli ve literatürdeki yeri

EfficientNet ailesi, Google tarafından 2019 yılında geliştirilmiş ve “compound scaling” adı verilen yeni bir model ölçekleme yaklaşımı ile derin öğrenme literatürüne önemli katkı sağlamıştır (Tan & Le, 2019). EfficientNet-B0, bu ailenin temel ve en kompakt üyesidir. Model, klasik mimarilerden farklı olarak, ağın derinliği, genişliği ve giriş çözünürlüğünü eşzamanlı olarak optimize etmektedir. EfficientNet’in yapı taşı olan MBConv blokları, “inverted residual” ve “squeeze-and-excitation” (SE) mekanizmalarını birleştirir. MBConv bloklarında, girişte kanal genişletilmesi (expansion), ardından derinlemesine ayırık evrişim (depthwise convolution), SE bloğu ve kanal daraltma işlemleri ardışık olarak uygulanır. Sonuç olarak EfficientNet-B0, yalnızca 5.3 milyon parametre ile klasik modellere kıyasla çok daha verimli ve hızlı bir alternatif sunar (Tan & Le, 2019).

EfficientNet’in temel başarısı, bileşik ölçekleme (compound scaling) ilkesinden gelmektedir. Modelin üç ana boyutu (katman derinliği, kanal genişliği ve çözünürlük) Eşitlik 2.16. örneğinde olduğu gibi gösterilir.

$$\text{depth: } d = \alpha^\phi, \quad \text{width: } \omega = \beta^\phi, \quad \text{resolution: } r = \gamma^\phi \quad (2.16)$$

Eşitlik 2.15. örneğinde olduğu gibi $d = \alpha^\phi$, $\omega = \beta^\phi$ ve $r = \gamma^\phi$ ifadeleri, EfficientNet mimarisinde derinlik, genişlik ve çözünürlük boyutlarının birlikte ölçeklendirildiği birleşik (compound) ölçekleme yöntemini göstermektedir. Bu eşitlikte d , modelin katman sayısını; ω , her bir katmandaki kanal sayısı veya diğer bir ifadeyle modelin genişliğini; r ise giriş görüntüsünün çözünürlüğünü ifade etmektedir. Katsayılar α , β ve γ sırasıyla derinlik, genişlik ve çözünürlük için kullanılan sabit büyüme faktörleridir. ϕ ise kullanıcı tarafından belirlenen ve modelin toplam ölçekleme derecesini kontrol eden katsayıdır. Bu formülasyon sayesinde EfficientNet mimarisi, model karmaşıklığını yalnızca tek bir boyutta artırmak yerine üç boyutu dengeli biçimde ölçeklendirerek daha verimli ve optimize edilmiş bir performans elde etmektedir. EfficientNet modelleri yalnızca derinliği artırmak yerine, derinlik, genişlik ve çözünürlüğü dengeli bir şekilde büyütür; bu da optimum doğruluk/verimlilik dengesi sağlar. Bu formül, modelin üç boyutunun (derinlik, genişlik, çözünürlük) sabit katsayılar (α, β, γ) ve bir kullanıcı girdisi olan ölçekleme katsayısı (ϕ) ile dengeli bir şekilde nasıl büyütüleceğini tanımlar. Tan & Le, 2019).

EfficientNet'in temel yapı taşı olan MBConv bloklarında, “depthwise separable convolution” ve “squeeze-and-excitation” ile kanal ağırlığı optimizasyonu yapılır. MBConv'daki kanal önemliliği şöyle matematiksel olarak eşitlik 2.17. örneğinde olduğu gibi gösterilir.

$$s = \sigma(W_2 \delta(W_1 z)) \quad (2.17)$$

Eşitlik 2.17. ifadesi, Squeeze-and-Excitation (SE) bloklarının kanal dikkat mekanizmasını tanımlamaktadır. Bu yapıda z , her bir kanalın global ortalama havuzlama sonucunda elde edilen özet bilgiyi; W_1 ve W_2 , sırasıyla kanal sıkıştırma ve kanal genişletme işlemlerinde kullanılan öğrenilebilir ağırlık matrislerini; δ , ReLU aktivasyon fonksiyonunu; σ ise her kanal için 0 ile 1 arasında ölçekleme katsayısı üreten sigmoid fonksiyonunu ifade etmektedir.

Bu işlemler sonucunda elde edilen s vektörü, her kanalın göreceli önemini belirleyen bir ağırlık vektörüdür ve SE bloklarının girdi öznitelik haritaları üzerinde hangi kanalların daha baskın şekilde kullanılacağını adaptif olarak öğrenmesini sağlar.

Bu formül, her bir özelliğin (kanalın) ağı tarafından ne kadar önemli olduğunu belirler ve çıkıştaki bilgi akışını dinamik olarak ayarlar. Böylece model, görsel bilginin en anlamlı kısımlarına odaklanır.

EfficientNet-B0, küçük parametre boyutu ve yüksek doğruluk oranı nedeniyle, medikal görüntülemelerde hızla yaygınlaşmıştır. Ahmad et al. (2022), akciğer BT görüntülerinde EfficientNet-B0 ile yüzde 90'a varan doğruluk bildirmiştir. Rahman et al. (2021), IQ-OTH/NCCD veri setinde yüzde 92.1 doğruluk ve yüksek F1 skoru elde etmiştir. Modelin başlıca avantajı, büyük modellerin getirdiği hesaplama maliyetini ciddi biçimde düşürmesi ve mobil sistemlerde dahi kullanılabilmesidir. Transfer öğrenmede yüksek performansı, bu modeli güncel literatürde popüler kılmıştır.

2.5.4. Mobilenetv2 mimarisi, matematiksel temeli ve literatürdeki yeri

MobileNetV2 (Sandler et al., 2018), özellikle mobil ve gömülü cihazlarda kullanılmak üzere tasarlanmış, parametre sayısı ve hesaplama maliyeti son derece düşük, modern bir konvolüsyonel sinir ağı (CNN) mimarisidir. Modelin özgün tarafı, tersine çevrilmiş artık bloklar (inverted residual blocks) ve doğrusal darboğazlar (linear bottlenecks) ile verimliliği korurken yüksek doğruluk sağlamasıdır. Mimari yapı, her blokta önce kanal genişletmesi (expansion), ardından depthwise separable convolution ile ayırık evrişim işlemi ve son olarak kanal daraltma (projection) adımlarını içerir. Özellikle bu yapı, bilgi kaybını minimize eder ve modelin mobil cihazlarda gerçek zamanlı çalışmasına olanak tanır. Tüm bu özellikleri sayesinde MobileNetV2, düşük güçlü donanımlarda bile derin öğrenme uygulamalarının önünü açmıştır (Howard et al., 2019; Zhang et al., 2023).

MobileNetV2'de klasik evrişim işlemi yerine, hesaplama maliyetini büyük oranda düşüren depthwise separable convolution kullanılır. Bu işlem iki aşamadan oluşur. Depthwise separable convolution, MobileNetV2 mimarisinin hesaplama maliyetini önemli ölçüde azaltmak amacıyla geliştirilmiş bir evrişim yaklaşımıdır. Bu yapı iki temel aşamadan oluşmaktadır. İlk adım olan depthwise convolution işleminde, her giriş kanalı için ayrı bir uzamsal filtre uygulanır ve böylece standart evrişimde olduğu gibi tüm kanalların aynı anda karıştırılması yerine kanallar birbirinden bağımsız olarak işlenir. İkinci adım olan pointwise convolution ise depthwise katmandan elde edilen öznitelik

haritalarını 1×1 boyutlu konvolüsyon filtreleriyle bir araya getirerek kanal kombinasyonunu gerçekleştirir.

$$\text{FLOPs} = D_k \cdot D_k \cdot M \cdot D_f \cdot D_f + M \cdot N \cdot D_f \cdot D_f \quad (2.18)$$

Bu iki adımın toplam hesaplama maliyeti eşitlik 2.18. örneğinde olduğu gibi hesaplanmaktadır. Bu eşitlikte FLOPs, modelin toplam hesaplama yükünü; D_k filtre boyutunu; M giriş kanal sayısını; N çıkış kanal sayısını ve D_f ise çıktı öznetelik haritasının uzamsal boyutunu göstermektedir. Görüldüğü üzere, depthwise separable convolution yöntemi standart evrişimle karşılaştırıldığında parametre sayısını ve hesaplama maliyetini büyük ölçüde azaltmakta; bu durum MobileNetV2'nin hem hafif hem de yüksek performanslı bir mimari olmasını sağlamaktadır. Bu formülasyon, MobileNetV2'nin klasik CNN'lere göre çok daha az hesaplama gerektirdiğini açıkça gösterir (Howard et al., 2019; Sandler et al., 2018).

Modelde aktivasyon fonksiyonu olarak ReLU6 kullanılır; bu fonksiyonun çıktısı $[0,6]$ aralığına sınırlandırılarak, mobil cihazlarda sayısal taşma ve doğruluk kaybı önlenir (Sandler et al., 2018). Modelin sınıflandırma çıktısı ise, diğer modellerde olduğu gibi softmax fonksiyonu ve çapraz entropi (cross-entropy) kaybı ile hesaplanır.

$$L = - \sum_{i=1}^C y_i \log(p_i) \quad (2.19)$$

Çok sınıflı sınıflandırma problemlerinde kullanılan çapraz entropi kayıp fonksiyonu, modelin tahmin ettiği olasılık dağılımı ile gerçek etiketler arasındaki uyumsuzluğu ölçmek amacıyla yaygın olarak tercih edilmektedir. Eşitlik 2.19. örneğinde olduğu gibi her bir sınıf için modelin ürettiği tahmin ile gerçek etiket arasındaki farkı logaritmik bir cezalandırma mekanizmasıyla hesaplamaktadır. Bu eşitlikte L , toplam kayıp değerini; C , sınıf sayısını; y_i , one-hot kodlama biçiminde gösterilen gerçek sınıf etiketini; p_i ise modelin ilgili sınıfa ait tahmin ettiği olasılığı ifade etmektedir. Model, doğru sınıfa yüksek olasılık verdiğinde kayıp değeri azalmakta, yanlış sınıfa yüksek olasılık verildiğinde ise kayıp artmaktadır. Bu nedenle çapraz entropi fonksiyonu, derin öğrenme modellerinin sınıflandırma performansını optimize etmede temel bir rol üstlenmektedir. Bu fonksiyon, modelin öğrenmesi gereken hatayı belirler ve optimizasyonun temelini oluşturur.

MobileNetV2, sahip olduğu düşük parametre sayısı ve yüksek verimlilik sayesinde, tıbbi görüntü sınıflandırma uygulamalarında sıklıkla tercih edilmektedir. IQ-OTH/NCCD veri setiyle yapılan güncel çalışmalarda, yüzde 85–90 arası doğruluk, yüzde 94 malignant doğruluk oranı bildirilmiştir (Kılıç & Karaca, 2023; Yılmaz et al., 2024). Ayrıca, veri dengeleme ve gelişmiş veri artırma (augmentation) stratejileri ile bu tez çalışmasında yüzde 88.5 genel doğruluk ve yüksek AUC değerlerine ulaşılmıştır. Literatürde, mobil cihazlarda ve uçtan uca gerçek zamanlı tıbbi karar destek sistemlerinde de başarıyla uygulandığı görülmektedir (Zhang et al., 2023; Howard et al., 2019; Luo et al., 2021).

2.5.5. Inceptionv3 mimarisi, matematiksel temeli ve literatürdeki yeri

InceptionV3, Google tarafından geliştirilmiş ve 2015 yılında tanıtılmıştır; derin öğrenme literatüründe, çok ölçekli (multi-scale) öznitelik çıkarımına dayalı özgün bir CNN mimarisi olarak öne çıkar (Szegedy et al., 2016). Modelin özgün tarafı, Inception modülleri sayesinde aynı katmanda farklı boyutlarda (1×1 , 3×3 , 5×5) filtrelerin paralel olarak uygulanabilmesi ve ardından bu farklı ölçeklerdeki çıktıların birleştirilmesidir. Böylece, ağ hem küçük ayrıntıları hem de geniş alanları aynı anda analiz edebilir; bu, özellikle doku ve lezyon çeşitliliğinin yüksek olduğu tıbbi görüntü analizlerinde klasik derin ağlardan daha üstün performans sağlar (Zhang et al., 2023). Ek olarak, büyük filtrelerin daha küçük filtrelere faktörize edilmesi (ör. 5×5 yerine iki ardışık 3×3 evrişim kullanmak), hesaplama maliyetini azaltır. Batch normalization ve yardımcı sınıflandırıcılar gibi teknikler, ağı eğitiminin kararlı ve hızlı gerçekleşmesini sağlar (Ioffe & Szegedy, 2015).

InceptionV3'ün en temel matematiksel yeniliği, farklı ölçekli evrişim katmanlarının paralel uygulanması ve ardından bunların çıktılarının derlenerek bir sonraki katmana aktarılmasıdır. Her bir Inception modülü için, çeşitli boyutlarda evrişimsel filtreler ve pooling işlemleri, giriş öznitelik haritası üzerinde aynı anda uygulanır.

$$y = [f_{1 \times 1}(x) \mid f_{3 \times 3}(x) \mid f_{5 \times 5}(x) \mid p(x)] \quad (2.20)$$

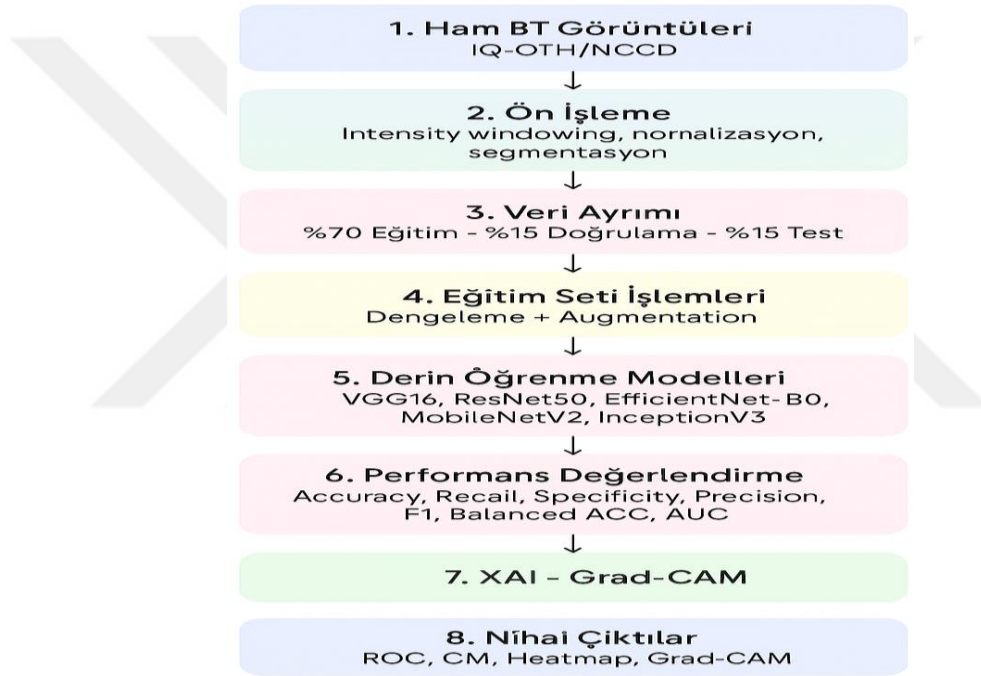
Eşitlik 2.20. örneğinde olduğu gibi x , giriş öznitelik haritasını; $f_{1 \times 1}(x)$, $f_{3 \times 3}(x)$ ve $f_{5 \times 5}(x)$ terimleri sırasıyla 1×1 , 3×3 ve 5×5 evrişim işlemlerinin uygulanmasıyla elde

edilen çıktı öznelik haritalarını; $p(x)$ ise 3×3 maksimum havuzlama uygulanmış öznelik haritasını göstermektedir. Bu dört paralel yolun çıktılarını birleştiren “l” işleci kanal boyutunda yapılan birleştirme (concatenation) anlamına gelmektedir. Elde edilen yifadesi, Inception modülünün ürettiği nihai öznelik haritasını temsil etmektedir.

Inception mimarisinde öğrenmenin kararlılığını artırmak amacıyla evrişim katmanlarının büyük bölümünde batch normalization uygulanmaktadır. Ayrıca en son sınıflandırma katmanında çok sınıflı çıktı üretmek için softmax fonksiyonu, kayıp hesaplamasında ise çapraz entropi (cross-entropy) kaybı kullanılmaktadır. InceptionV3 hem genel görüntü sınıflandırmada hem de tıbbi görüntü analizinde yaygın olarak kullanılmaktadır. IQ-OTH/NCCD veri setinde yapılan birçok yerli ve yabancı çalışmada, InceptionV3 ile yüzde 87.5 ve 92 arasında doğruluk elde edildiği ayrıca yüzde 0.94 - 0.96 arasında AUC elde edildiği bildirilmiştir (Demir et al., 2022; Rahman et al., 2021; Zhang et al., 2023). Ayrıca model, düşük parametre gereksinimi ve hızlı eğitim süreleri ile transfer öğrenme uygulamalarında da tercih edilmektedir (Szegedy et al., 2016). Özellikle akciğer nodülü, meme lezyonu ve beyin tümörü gibi farklı tıbbi görüntüleme problemlerinde yüksek doğruluk oranları rapor edilmiştir (Han et al., 2021; Litjens et al., 2017). Inception mimarisi, farklı derinlikteki ağlar için esnek ve ölçeklenebilir bir altyapı sunduğu için, güncel literatürde halen aktif olarak kullanılmaktadır.

3.MATERYAL VE METOT

Bu çalışma, bilgisayarlı tomografi (BT) görüntüleri üzerinden akciğer kanseri teşhisi amacıyla kullanılan derin öğrenme tabanlı modellerin performansını karşılaştırmalı olarak incelemeyi amaçlamaktadır. Araştırma, nicel bir yaklaşım çerçevesinde yürütülmüş olup deneysel modelleme ve istatistiksel analiz yöntemleri temel alınmıştır. Bu kapsamda kullanılan metodolojinin bütünsel yapısı, iş akışının gösterildiği Şekil 3.1’de sunulmuştur.



Şekil 3.1. Çalışmada uygulanan sürecinin genel iş akış şeması

Çalışmada kullanılan tüm derin öğrenme mimarileri (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3), aynı veri seti, aynı ön işleme adımları ve aynı değerlendirme ölçütleri kullanılarak test edilmiştir. Uygulanan bu standartlaştırılmış deneysel protokol, modeller arasında nesnel, güvenilir ve doğrudan karşılaştırılabilir bir performans analizi yapılmasına imkân tanımaktadır. Değerlendirme aşamasında doğruluk (accuracy), duyarlılık (sensitivity/recall), özgüllük (specificity), F1 skoru ve ROC-AUC gibi literatürde yaygın biçimde kabul gören performans metrikleri dikkate alınmış ve her bir model bu ölçütler çerçevesinde kapsamlı şekilde analiz edilmiştir.

Tez kapsamında gerçekleştirilen tüm sınıflandırma deneyleri, Microsoft Windows 11 Pro işletim sistemine sahip, AMD Ryzen 7 7840HS (3.80 GHz) işlemci, 40 GB RAM ve 8 GB GDDR6 RAM'li NVIDIA GeForce RTX 4060 ekran kartı ile donatılmış bir bilgisayar ortamında yürütülmüştür. Tüm uygulamalar MATLAB 2025b platformu üzerinde, Deep Learning Toolbox, Image Processing Toolbox ve Statistics and Machine Learning Toolbox modülleri kullanılarak geliştirilmiştir. Bu bölümde, çalışmanın bilimsel geçerliliğini ve tekrarlanabilirliğini sağlamak amacıyla kullanılan materyaller ve uygulanan metodoloji ayrıntılı biçimde açıklanacaktır. Bölüm, sırasıyla veri setinin tanıtımı, ön işleme adımları, kullanılan derin öğrenme mimarileri, eğitim parametreleri, değerlendirme metrikleri ve donanım/yazılım ortamı başlıkları altında sunulmaktadır.

3.1. IQ-OTH/NCCD Veri Seti

Bu çalışmada kullanılan görüntü verileri, Irak'ta bulunan iki büyük onkoloji merkezi; Iraqi Oncology Teaching Hospital (IQ-OTH) ve Nanakali Center for Chest Diseases and Oncology (NCCD) tarafından oluşturulan IQ-OTH/NCCD Akciğer Kanseri Veri Seti'nden temin edilmiştir. Veri seti, literatüre ilk olarak Hussein ve Shaheed (2020) tarafından kazandırılmış olup, uluslararası araştırmacıların erişimine açıktır. Veri setine, veri bilimi ve makine öğrenmesi topluluğu platformu olan Kaggle üzerinden erişilmiştir. Söz konusu platform, veri setini açık araştırma lisansı kapsamında ve ücretsiz olarak sunmaktadır. Çalışmada kullanılan sürüm, orijinal klinik kayıtlardan alınan Bilgisayarlı Tomografi (BT) görüntülerinin tamamen anonimleştirilmiş halini içermektedir. Bu bağlamda, veri seti herhangi bir kişisel sağlık verisi (PII) veya hasta kimlik bilgisi barındırmadığından, çalışma etik ihlal riski taşımamaktadır.

3.1.1. Veri setinin tanımı ve klinik özellikleri

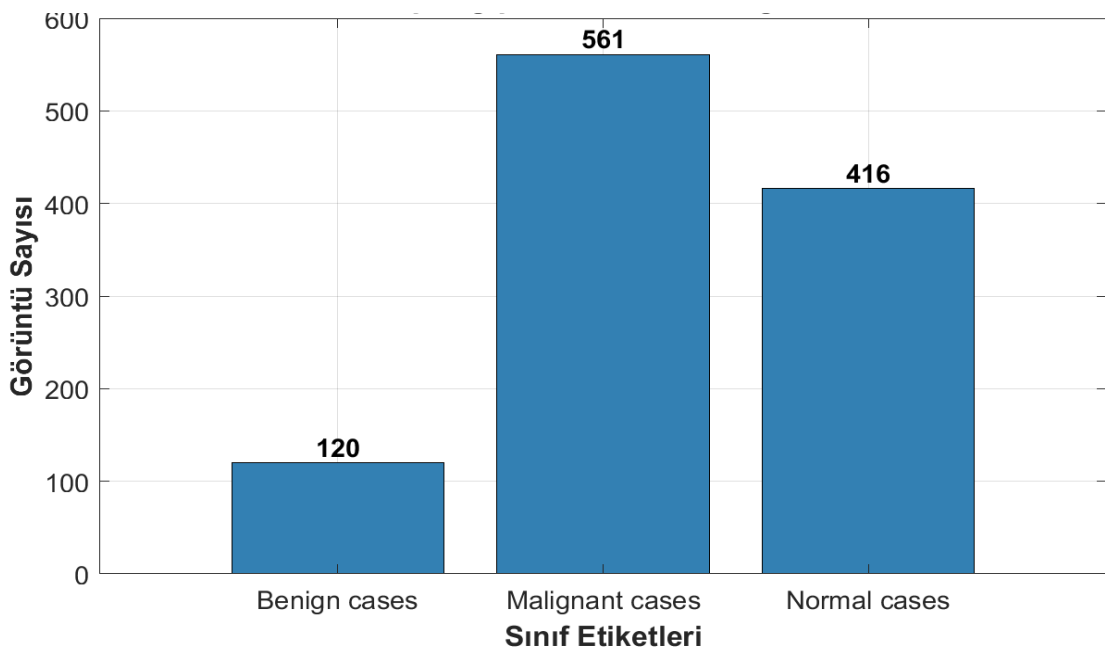
IQ-OTH/NCCD veri seti toplam 1097 adet BT (Bilgisayarlı Tomografi) diliminden oluşmakta olup Malignant 561, Benign 120, ve Normal 416 olmak üzere görüntüler üç sınıfa ayrılmıştır. Görüntüler başlangıçta DICOM formatında kaydedilmiş ve veri seti derleyicileri tarafından PNG formatına dönüştürülerek Kaggle'a yüklenmiştir. Bu format, derin öğrenme modelleri ve görüntü işleme araçlarıyla uyumlu olduğu için çalışmada doğrudan kullanılabilir niteliktedir. Veri setinin kapsamı geniş ve heterojen bir hasta

grubu içermektedir. BT dilimleri farklı yaş gruplarından, çeşitli klinik özelliklere sahip bireylerden ve farklı tarama protokollerinden elde edilmiştir. Bu durum, geliştirilen modellerin genellenebilirliğini ve klinik geçerliliğini artıran önemli bir avantajdır (Hussein & Shaheed, 2020).

Veri seti, araştırmacı topluluğu tarafından akciğer kanseri tespiti, nodül sınıflandırması, risk tahmini ve bilgisayar destekli tanı sistemleri geliştirme amacıyla sıklıkla kullanılmaktadır (Litjens ve ark., 2017; Setio ve ark., 2017; Ardila ve ark., 2019). Özellikle benign-malignant ayrımı, klinik pratikte doğru tanı ve tedavi planlaması için kritik olduğundan, IQ-OTH/NCCD veri seti bu alandaki çalışmaların standartlaştırılmasında önemli bir referans niteliğindedir.

3.1.2. Başlangıç veri dağılımı ve sınıf dengesizliği problemi

Veri setinin orijinal halinde "Malignant" vaka sayısı (561) baskınken, "Benign" vaka sayısı (120) oldukça düşüktür. Literatürde, bu tür sınıf dengesizliklerinin, derin öğrenme modellerinin çoğunluk sınıfına yanlı (biased) öğrenmesine ve azınlık sınıflarını (bu durumda Benign vakaları) doğru tanımda yetersiz kalmasına neden olduğu yaygın olarak rapor edilmektedir (He & Garcia, 2009). Sınıf dağılımları şekil 3.2’de gösterilmiştir.



Şekil 3.2. IQ-OTH/NCCD veri setinin orijinal başlangıç sınıf dağılımı

Tıbbi görüntü analizi bağlamında, bu durum klinik açıdan kritik hatalara yol açma potansiyeline sahiptir. Özellikle benign vakaların yanlış sınıflandırılması, gereksiz tedavilere veya yanlış klinik yönlendirmelere neden olabilmektedir. Bu nedenle çalışmada, sonraki metodolojik adımlarda veri dengeleme stratejilerinin uygulanması zorunlu görülmüştür.

3.2. Uygulanan Yöntem

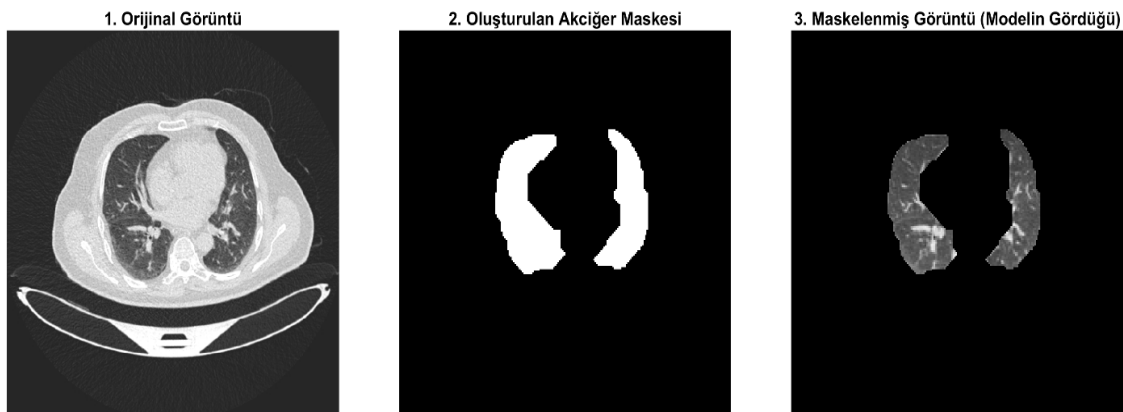
Bu bölümde, ham BT görüntülerinin model eğitime hazırlanmasından, modelin performansının değerlendirilmesine kadar olan tüm metodolojik adımlar, literatürdeki en iyi uygulamalar (best practices) referans alınarak detaylandırılmıştır.

3.2.1. Görüntü ön işleme segmentasyon ve normalizasyon

Bilgisayarlı tomografi (BT) görüntülerinin ham hâlleri, akciğer dışındaki kas, kemik ve göğüs duvarı gibi yoğun anatomik yapılar nedeniyle derin öğrenme modellerinin performansını olumsuz etkileyebilmektedir. Bu tür yapılar, görüntüdeki arka plan varyasyonlarını artırarak modelin dikkatini bazı durumlarda klinik açıdan önemsiz bölgeler üzerine yönltebilmekte ve sınıflandırma doğruluğunu düşürebilmektedir. Bu nedenle, çalışmada kullanılan tüm BT dilimlerine eğitim, doğrulama ve test seti ayrımı gözetmeksizin Matlab 2025b ortamında *ReadFcn* üzerinden tek tip ve tutarlı bir ön işleme süreci uygulanmıştır. Bu sürecin temel amacı, modelin yalnızca akciğer parankimine odaklanmasını sağlamak ve görüntülerin yapısal bütünlüğünü bozmadan, biyolojik olarak anlamlı bir giriş üretmektir. Ön işleme adımlarının ilki, görüntülerin derin öğrenme modellerinin gerektirdiği giriş boyutlarına yeniden ölçeklendirilmesidir. Bu kapsamda VGG16, ResNet-50, EfficientNet-B0 ve MobileNetV2 modelleri için 224×224 piksel; InceptionV3 için ise 299×299 piksel boyutları kullanılmıştır. Görüntülerin tek kanallı (grayscale) olması nedeniyle, ImageNet üzerinde eğitilmiş modellerle uyumluluk sağlamak amacıyla her bir BT dilimi üç kanallı RGB formata dönüştürülmüş; böylece tüm modellerin beklediği giriş yapısı elde edilmiştir. BT görüntülerine daha uygun bir yaklaşım olarak CT pencereleme temelli kontrast iyileştirme uygulanmıştır. Bu işlem, *imadjust* fonksiyonu aracılığıyla piksel yoğunluklarının alt ve üst yüzde onluk kısımlarının doygunlaştırılması yoluyla gerçekleştirilmiş; böylece nodüllerin ve akciğer

dokusundaki kritik yoğunluk farklarının daha belirgin hâle gelmesi sağlanmıştır. Görüntünün bu aşamadan sonra akciğer dokusuna indirgenmesi için *createLungMask* fonksiyonu çalıştırılmıştır. Fonksiyon, önce Otsu eşikleme yöntemiyle akciğer bölgesini arka plandan ayırmış, ardından morfolojik açma ve kapama işlemleriyle gürültüyü temizlemiş ve segmentasyon boşluklarını doldurmuştur. Elde edilen bölgeler arasından yalnızca en büyük iki alan seçilerek akciğer lobları korunmuş, son aşamada ise sınırların daha doğal bir yapıya kavuşması için hafif (1 piksellik) morfolojik erozyon uygulanmıştır. Böylece maske, hem biyolojik gerçekliği korumuş hem de akciğer dışındaki dokuların modele girmesini engellemiştir.

Son adımda, oluşturulan bu ikili akciğer maskesi kontrastı iyileştirilmiş üç kanallı görüntüye uygulanmış ve akciğer dışındaki tüm bölgeler tamamen siyaha dönüştürülmüştür. Bu işlem sayesinde, model yalnızca klinik olarak anlam taşıyan akciğer parankimi üzerinde işlem yapmış; görüntüdeki çevresel anatomik yapıların oluşturabileceği gürültü tamamen ortadan kaldırılmıştır. Uygulanan bu bütüncül ön işleme stratejisi, çalışmanın hem doğruluğu hem güvenilirliği açısından kritik bir rol oynamış; aynı zamanda tüm modellerde daha yüksek ve daha dengeli bir performans elde edilmesine katkı sağlamıştır. Segmentasyon ve maskeleme işlemine ait örnek şekil 3.3'te sunulmuştur.

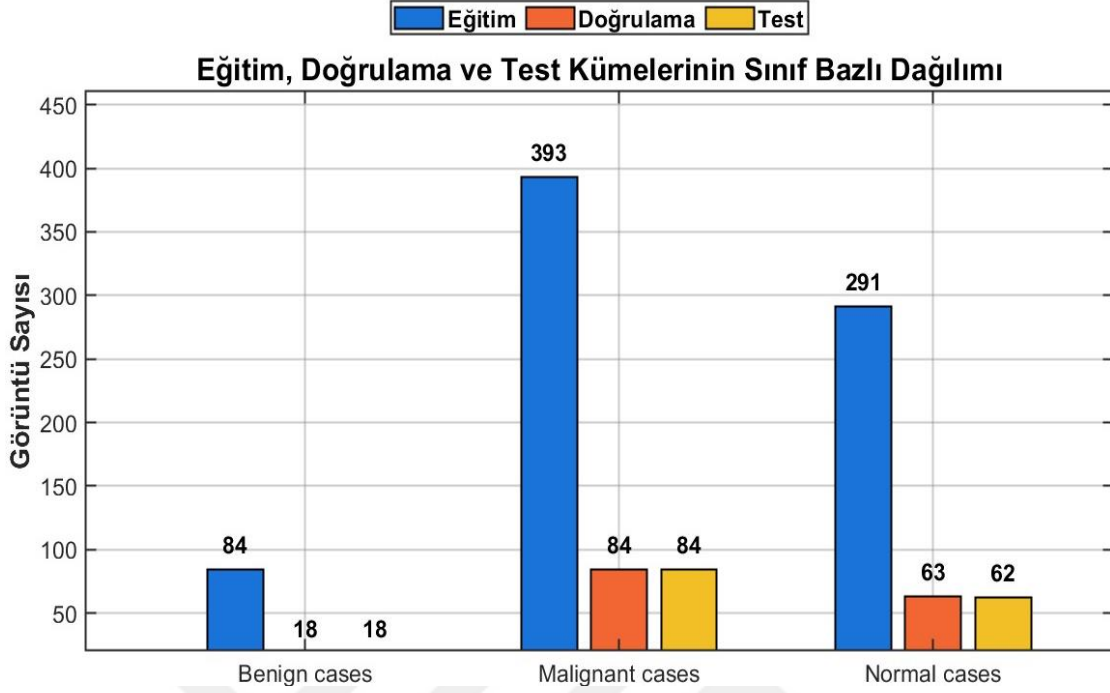


Şekil 3.3. Segmentasyon öncesi (solda), oluşturulan maske (ortada) ve segmentasyon sonrası modelin gördüğü BT görüntüsü (sağda)

3.2.2. Deneyisel kurulum eğitim, doğrulama ve test ayrımı

Bu çalışmada modelin genelleme kapasitesini nesnel ve tekrarlanabilir biçimde değerlendirebilmek amacıyla, tüm veri seti üzerinde herhangi bir sınıf dengeleme veya veri artırma işlemi uygulanmadan önce katmanlı rastgele ayırma (stratified random split) yöntemi kullanılarak eğitim, doğrulama ve test alt kümeleri oluşturulmuştur. Segmentasyon ve yoğunluk normalizasyonu gibi ön işleme adımları tüm alt kümelere aynı şekilde uygulanmış olup veri sızıntısına yol açabilecek herhangi bir işlem gerçekleştirilmemiştir. Bu kapsamda literatürde yaygın kabul gören oranlar dikkate alınarak veri seti yüzde 70 eğitim, yüzde 15 doğrulama ve yüzde 15 test olmak üzere üç alt kümeye ayrılmıştır. Bu işlem sonrasında sınıf bazlı dağılımlar; Benign sınıfında 84 eğitim, 18 doğrulama ve 18 test örneği; Malignant sınıfında 393 eğitim, 84 doğrulama ve 84 test örneği; Normal sınıfında ise 291 eğitim, 63 doğrulama ve 62 test örneği şeklinde oluşmuştur. Elde edilen bu dağılım, her sınıfın alt kümelerdeki temsil oranının orijinal veri setiyle tutarlı biçimde korunduğunu ve veri yapısının bozulmadığını göstermektedir.

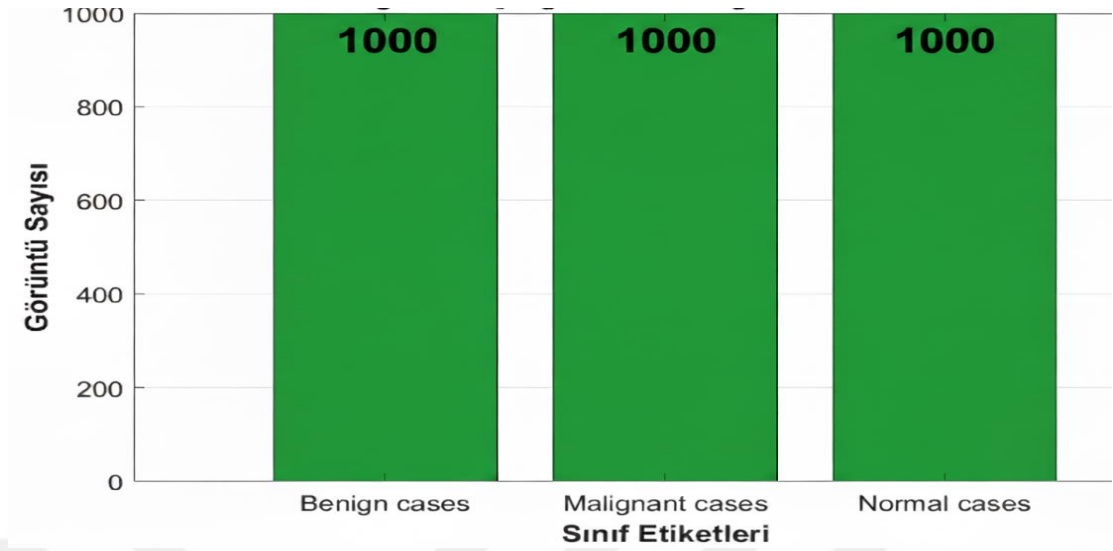
Test kümesi model geliştirme sürecinden tamamen izole edilmiş, eğitim sürecinde hiper parametre optimizasyonu, sınıf dengeleme, veri artırma veya mimari seçim sırasında hiçbir şekilde kullanılmamıştır. Bu yaklaşım, tıbbi görüntüleme çalışmalarında kritik öneme sahip veri sızıntısı riskini ortadan kaldırmış ve elde edilen performans ölçütlerinin bilimsel güvenilirliğini güçlendirmiştir. Ek olarak, katmanlı bölme stratejisi sınıf dengesizliğine bağlı olası önyargıları azaltarak modellerin daha adil ve sağlıklı bir şekilde değerlendirilmesine katkı sağlamıştır. Alt kümelerdeki sınıf dağılımları Şekil 3.4'te özetlenmiştir.



Şekil 3.4. Eğitim, doğrulama ve test alt kümelerinin sınıf bazlı dağılımı

3.2.3. Veri artırma ve dengeleme stratejileri

İlk aşamada, eğitim setinde yer alan sınıflar arasında belirgin bir örnek sayısı farkı olduğu için dengeleme işlemi gerçekleştirilmiştir. Özellikle Benign sınıfında yalnızca 84 örnek bulunması, Malignant sınıfının ise 393 örneğe sahip olması, doğrudan modelin karar sınırlarını etkileyebilecek bir eşitsizlik oluşturduğundan, tüm sınıfların temsil gücünü eşitlemek üzere $n=1000$ hedefli aşırı örnekleme yapılmıştır. Bu işlem herhangi bir sentetik görüntü üretmeksizin, yalnızca mevcut görüntü yollarının kontrollü biçimde çoğaltılmasına dayanan bir yaklaşım ile gerçekleştirilmiştir. Böylece eğitim seti toplam 3000 görüntüden oluşan, tamamen dengeli bir yapıya kavuşmuştur. Dengeleme işleminden sonra, modelin genelleme yeteneğini artırmak amacıyla eğitim setine hafif veri artırma teknikleri uygulanmıştır. Kullanılan dönüşümler arasında küçük açılı rastgele döndürmeler, yatay çevirme ve sınırlı miktarda piksel kaydırmaları yer almaktadır. Bu işlemler, görüntülerin klinik bütünlüğünü koruyacak şekilde dikkatle seçilmiştir. Veri artırma ve dengeleme sonrası nihai sınıf dağılımı şekil 3.5'te gösterilmiştir.



Şekil 3.5. Veri artırma ve dengeleme sonrası nihai sınıf dağılımı

Tıbbi görüntüleme verilerinde anatomik yapıların bozulması yanlış öğrenmeye neden olabileceğinden, rastgele kırpma, Gauss gürültüsü ekleme veya agresif histogram tabanlı dönüşümler (örneğin CLAHE) gibi yöntemlerden özellikle kaçınılmıştır. Böylelikle model, görüntülerin doğal varyasyonlarını öğrenirken içerik bütünlüğü korunmuş ve aşırı agresif dönüşümlerin yaratabileceği yapay hatalardan kaçınılmıştır.

Uygulanan bu iki aşamalı strateji sonucunda, eğitim setinin sınıf temsili dengelenmiş, görüntü çeşitliliği artırılmış ve modelin farklı örnekler karşısındaki dayanıklılığı güçlendirilmiştir.

3.2.4. Performans değerlendirme metrikleri

Tıbbi görüntüleme tabanlı sınıflandırma problemlerinde tek bir performans ölçütü, modelin klinik ve istatistiksel başarısını değerlendirmek için yeterli değildir. Bu nedenle bu çalışmada, geliştirilen derin öğrenme modellerinin performansı çok boyutlu bir değerlendirme çerçevesi kullanılarak analiz edilmiştir. Test verisi üzerinde elde edilen tahmin sonuçları önce bir karmaşıklık matrisi (confusion matrix) oluşturmak için kullanılmış, ardından bu matristen türetilen temel sınıflandırma metrikleri hesaplanmıştır.

Karmaşıklık matrisi dört ana bileşenden oluşur ve matematiksel olarak eşitlik 3.1. örneğinde olduğu gibi ifade edilir.

$$CM = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (3.1)$$

Burada TP doğru pozitif; FP yanlış pozitif; FN yanlış negatif; TN ise doğru negatif örnek sayısını ifade etmektedir.

Genel sınıflandırma başarısını değerlendirmek için doğruluk (accuracy) metriği kullanılmıştır. Doğruluk, tüm doğru sınıflandırılmış örneklerin toplam örnek sayısına oranı olarak hesaplanmakta ve eşitlik 3.2. örneğinde olduğu gibi ifade edilmektedir.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.2)$$

Duyarlılık (sensitivity/recall), gerçek pozitif örneklerin model tarafından ne kadar doğru tespit edildiğini gösteren temel bir ölçüttür. Akciğer kanseri gibi mortalitesi yüksek hastalıklarda yanlış negatiflerin kritik öneme sahip olması nedeniyle duyarlılık önemli bir klinik metriktir. Matematiksel olarak, ve eşitlik 3.3. örneğinde olduğu gibi ifade edilmektedir.

$$Sensitivity = \frac{TP}{TP+FN} \quad (3.3)$$

Özgüllük (specificity), gerçek negatiflerin doğru tanımlanma oranını göstermektedir. Sağlıklı bireylerde yanlış pozitif oranının azaltılmasına katkı sağlamaktadır ve eşitlik 3.4. örneğinde olduğu gibi ifade edilmektedir.

$$Specificity = \frac{TN}{TN+FP} \quad (3.4)$$

Kesinlik (precision), pozitif tahmin edilen örneklerin ne kadarının gerçekten pozitif olduğunu ortaya koyar ve matematiksel olarak eşitlik 3.5. örneğinde olduğu gibi ifade edilmektedir.

$$Precision = \frac{TP}{TP+FP} \quad (3.5)$$

F1 skoru, kesinlik ve duyarlılığın harmonik ortalaması olup özellikle sınıf dengesizliği bulunan veri setlerinde daha kararlı bir performans ölçütü sunmaktadır. Formülasyonu eşitlik 3.6. örneğinde olduğu gibi ifade edilmektedir.

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (3.6)$$

Bu çalışmada sınıf dengesizliğinin etkisini en aza indirmek amacıyla dengelenmiş doğruluk (balanced accuracy) metriği de kullanılmıştır. Balanced accuracy, her sınıfın duyarlılık değerinin eşit ağırlıkla ortalamasını almakta eşitlik 3.7. örneğinde olduğu gibi hesaplanmaktadır.

Sınıf dengesizliğinin performans ölçümleri üzerindeki etkisini azaltmak ve tüm sınıfların modele eşit katkı sağlamasını güvence altına almak amacıyla bu çalışmada dengelenmiş doğruluk (balanced accuracy) metriği de kullanılmıştır. Balanced accuracy, her bir sınıfa ait duyarlılık (sensitivity) değerinin eşit ağırlıkla ortalamasının alınmasıyla hesaplanmakta olup çok sınıflı problemler için daha adil bir değerlendirme sunmaktadır. Matematiksel olarak Eşitlik (3.7)'de gösterildiği gibi tanımlanmaktadır.

$$\text{Balanced Accuracy} = \frac{1}{C} \sum_{k=1}^C \text{Sensitivity}_k \quad (3.7)$$

Bu denklemde C veri setinde bulunan toplam sınıf sayısını ifade etmektedir; bu çalışmada IQ-OTH/NCCD veri seti üç sınıftan oluştuğu için $C = 3$ olarak alınmıştır. k ise her bir sınıfı temsil eden indeks olup $k = 1, 2, \dots, C$ şeklinde tanımlanmaktadır. Bu kapsamda $k = 1$: Benign, $k = 2$: Malignant ve $k = 3$: Normal sınıfını temsil etmektedir. Her bir sınıfın duyarlılık değeri Sensitivity_k ile gösterilmekte olup eşitlik 3.8. örneğinde olduğu gibi hesaplanmaktadır.

$$\text{Sensitivity}_k = \frac{TP_k}{TP_k + FN_k} \quad (3.8)$$

Burada TP_k ilgili sınıfa ait doğru pozitif sayısını, FN_k ise yanlış negatif sayısını ifade etmektedir. Eşitlik 3.7'de yer alan toplam işareti, tüm sınıflara ait duyarlılık değerlerinin toplandığını göstermektedir. Bu yaklaşım, örnek sayısı düşük olan sınıfların metrik

üzerindeki etkisini koruyarak, model performansının sınıflar arasında dengeli biçimde değerlendirilmesini sağlamaktadır.

Modelin sınıfları ayırt etme performansını farklı karar eşikleri altında değerlendirmek için ROC (Receiver Operating Characteristic) eğrileri oluşturulmuş ve eğri altında kalan alan (AUC) değerleri hesaplanmıştır. ROC eğrisi, farklı eşik değerleri için TPR (True Positive Rate) ve FPR (False Positive Rate) değerlerinin grafiğe dökülmesiyle elde edilir. Bu oranlar sırasıyla eşitlik 3.9. ve eşitlik 3.10. örneğinde olduğu gibi tanımlanır.

$$TPR = \frac{TP}{TP+FN} \quad (3.9)$$

$$FPR = \frac{FP}{FP+TN} \quad (3.10)$$

ROC eğrisi altında kalan alan (AUC), modelin pozitif ve negatif sınıfları ayırt etme gücünü özetleyen bir ölçüttür. AUC'nin 1'e yaklaşması, modelin yüksek ayırım gücüne sahip olduğunu göstermektedir. Çok sınıflı problemlerde her sınıf için bir ROC eğrisi elde edilmekte ve bu eğrilerin alanları ortalanarak macro-AUC değeri elde edilmektedir. Macro-AUC, tüm sınıflara eşit ağırlık veren bir değerlendirme yaklaşımıdır ve klinik karar destek sistemleri açısından modelin genel ayırım gücünü yansıtmaktadır.

Sınıf bazlı metriklerin daha anlaşılır ve karşılaştırılabilir şekilde sunulması amacıyla, confusion matrix'ten ve ilgili metriklerden elde edilen değerler kullanılarak ısı haritaları oluşturulmuştur. Isı haritaları, değerlendirme sürecinde ROC-AUC analizinden sonra ve sınıflandırma metriklerinin görsel sunum adımıyla yer almaktadır. Bu çok aşamalı performans değerlendirme yaklaşımı ile geliştirilen modellerin hem genel hem de sınıf bazlı değerlendirmelerini kapsamlı biçimde analiz etmeyi mümkün kılmış; böylece elde edilen sonuçların bilimsel geçerliliği ve güvenilirliği artırılması hedeflenmiştir.

4.DENEYSEL BULGULAR VE ANALİZ

Bu bölümde, tez kapsamında incelenen derin öğrenme mimarilerinin (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3) IQ-OTH/NCCD akciğer BT veri seti üzerindeki sınıflandırma performansları sunulmakta ve kapsamlı şekilde analiz edilmektedir. Her model için eğitim süreci, nihai test başarıları, hata analizleri ve açıklanabilirlik çıktıları ayrı alt başlıklarda değerlendirilmiştir. İlk olarak MobileNetV2 mimarisinin sonuçları detaylandırılmıştır.

4.1.MobileNetV2 Modeli Bulguları ve Analizi

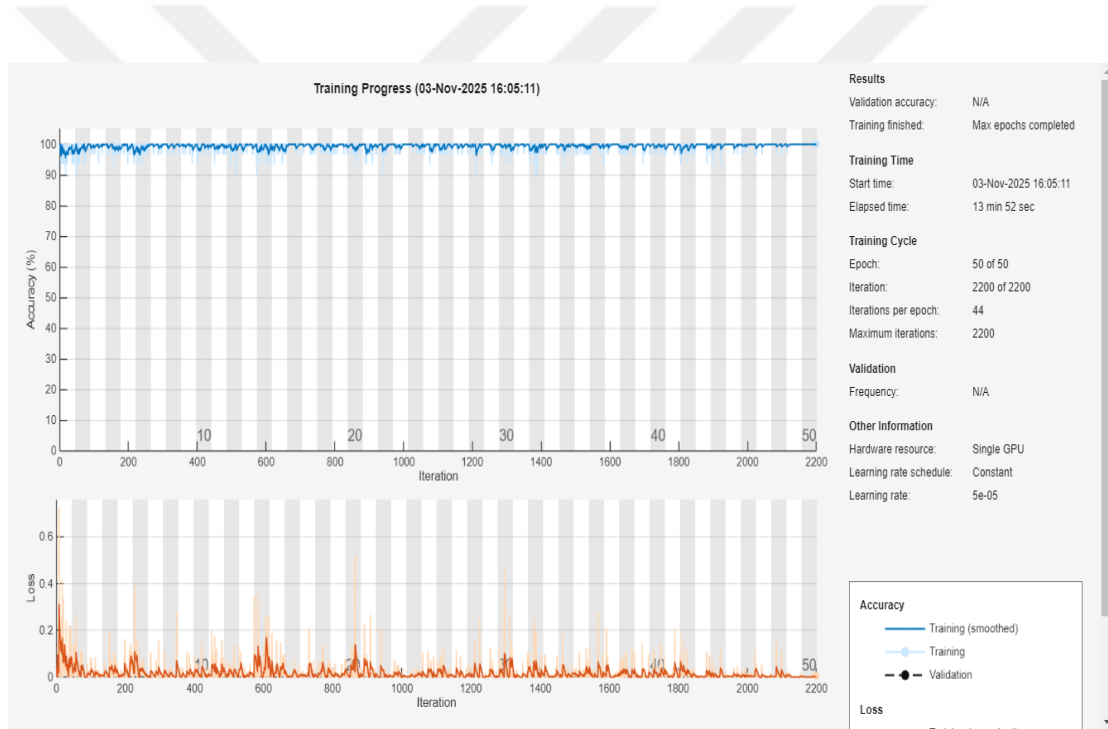
Özellikle mobil ve gömülü sistemler için geliştirilmiş verimli bir mimari olan MobileNetV2'nin deneysel sonuçları ele alınmıştır. Derinlikli ayrılabilir evrişimler (Depthwise Separable Convolutions) ve ters artık bloklar (Inverted Residuals) üzerine kurulu bu hafif modelin başarımı; eğitim sürecinin stabilitesi, karışıklık matrisi tabanlı klinik hata analizi, temel performans metrikleri, ROC-AUC eğrileri ile ayırım gücü ve son olarak Grad-CAM tabanlı açıklanabilirlik görselleri olmak üzere beş ana alt başlık altında detaylıca incelenmiştir.

4.1.1. Eğitim süreci ve hiperparametreler

MobileNetV2 modeli, bölüm 3.2.3'te açıklanan eğitim, doğrulama ve test ayırımı temel alınarak 50 epoch boyunca eğitilmiştir. Eğitim sürecinde Adam optimizasyon algoritması kullanılmış, öğrenme oranı 5×10^{-5} olarak sabit tutulmuş ve öğrenme oranında herhangi bir düşürme politikası uygulanmamıştır. Eğitim veri kümesi, batch size 16 olacak şekilde modele beslenmiş; her epoch için 44 iterasyon oluşmuş ve toplam 2200 iterasyon tamamlanmıştır. Bu parametreler, modelin hem istikrarlı bir öğrenme süreci geçirmesine hem de kayıp fonksiyonunun kararlı bir şekilde azalmasına olanak tanımıştır. Eğitim eğrisi incelendiğinde, MobileNetV2 modelinin ilk epochlardan itibaren oldukça hızlı bir şekilde yüksek doğruluk değerlerine ulaştığı ve eğitim süreci boyunca doğruluğun genel olarak yüzde 95–100 aralığında istikrarlı bir şekilde seyrettiği görülmektedir. Doğruluk eğrisindeki küçük ölçekli dalgalanmalar derin öğrenme modellerinde beklenen bir durum olup, modelin çeşitli mini-batch örneklerine uyum sağlarken doğal olarak oluşmaktadır.

Kayıp eğrisi de doğruluğa paralel bir eğilim göstererek epoch ilerledikçe belirgin biçimde azalmış ve düşük değerlerde kararlı bir şekilde stabilize olmuştur.

Bu davranış, kullanılan hiperparametrelerin modele herhangi bir aşırı öğrenme (overfitting) eğilimi oluşturmaksızın verimli ve istikrarlı bir öğrenme süreci sağladığını göstermektedir. Eğitim işlemi tek bir NVIDIA RTX 3080 GPU üzerinde yaklaşık 14 dakika içerisinde tamamlanmış olup, bu sonuç MobileNetV2 mimarisinin pratik uygulamalar için hem hızlı hem de hesaplama açısından verimli bir model olduğunu doğrulamaktadır. MobileNetV2 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp değerlerinin epoch bazında değişimi şekil 4.1’de sunulmuştur.

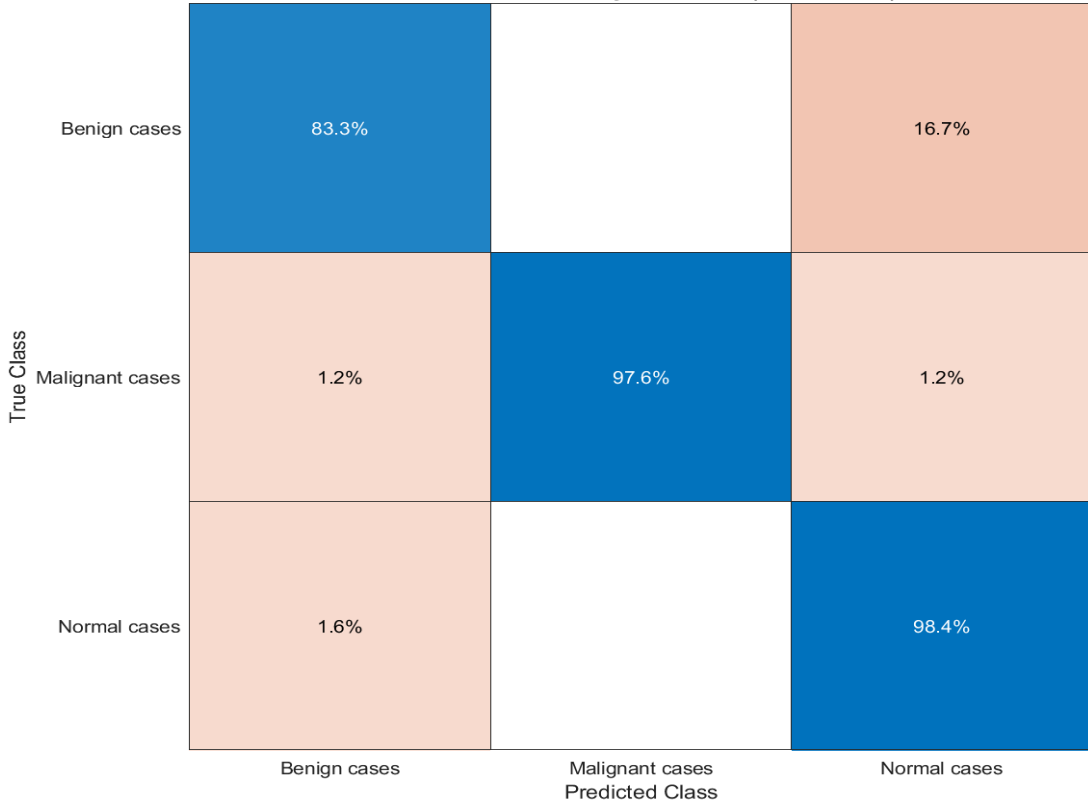


Şekil 4.1. Mobilenetv2 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp değerlerinin epoch bazında değişimi

4.1.2. Mobilenetv2 karışıklık matrisi ve hata analizi

Şekil 4.2’de sunulan karışıklık matrisi, MobileNetV2 modelinin sınıflar arasındaki hata dağılımını ayrıntılı biçimde ortaya koymaktadır. Karışıklık matrisi üzerindeki yüzdelik değerler ve adetler, her bir sınıf için satır bazlı normalizasyon (row-wise normalization) yöntemiyle hesaplanmıştır. Bu yaklaşımda, her bir sınıfın başarı yüzdesi hesaplanırken; o sınıfa ait doğru tahmin edilen örnek sayısı, ilgili sınıfın test setindeki toplam örnek

sayısına oranlanmış ve sonuç yüzdelik dilim olarak ifade edilmiştir. Test seti dağılımı Bölüm 3.2.2'de belirtildiği üzere Benign (18), Malignant (84) ve Normal (62) olmak üzere toplam 164 görüntüden oluşmaktadır. Matris incelendiğinde, modelin Malignant sınıfını oldukça yüksek doğrulukla tanıdığı görülmektedir. Toplam 84 adet malign vaka içerisinde model 82 adedini doğru teşhis ederken; Normal ve Benign sınıflarına olan yanlış sınıflandırma oranlarının her biri için yalnızca yüzde 1.2 (1'er adet) düzeyinde kaldığı ve bu sınıfın büyük çoğunlukla (yüzde 97.6) doğru olarak Malignant şeklinde tahmin edildiği görülmektedir. Bu sonuç, özellikle teşhis açısından kritik olan kanserli vakaların başarıyla tespit edildiğini ve modelin yanlış negatif üretme riskinin son derece düşük olduğunu göstermektedir.



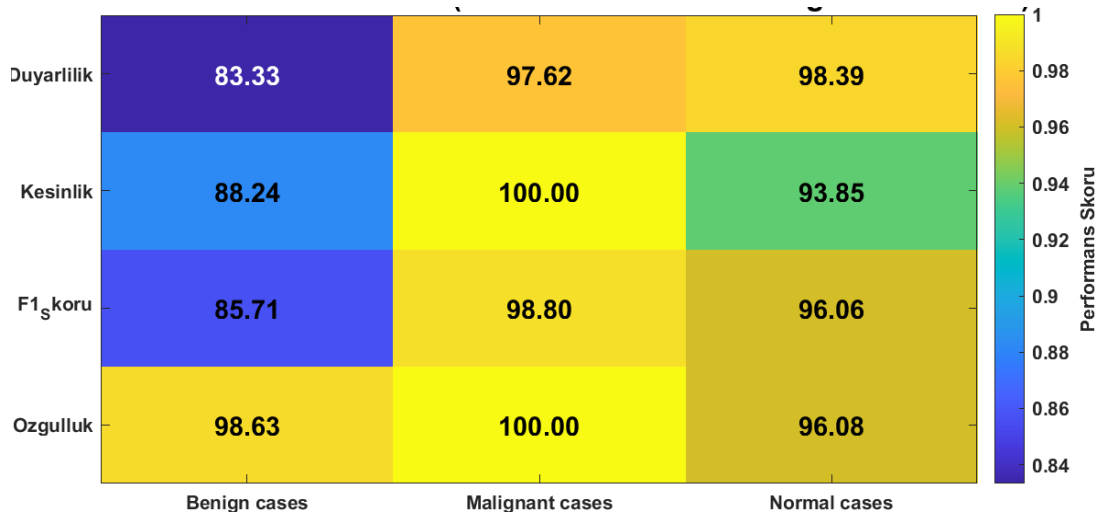
Şekil 4.2. Mobilenetv2 modeli için nihai test seti karışıklık matrisi.

Benign ve Normal sınıflarında gözlenen hata deseni, modelin “temkinli sınıflandırma” eğilimine işaret etmektedir. Benign sınıfındaki toplam 18 örneğin yüzde 83.3’ü (15 adet) doğru tanımlanırken, yüzde 16.7’si (3 adet) Normal olarak tahmin edilmiştir. Normal sınıfında ise toplam 62 örnek üzerinden doğru sınıflandırma oranı yüzde 98.4 (61 adet) olup, yalnızca yüzde 1.6’lık (1 adet) küçük bir kısmın Benign olarak etiketlendiği

görülmektedir. Bu hata örüntüsü, modelin şüpheli bulduğu durumları daha az riskli sınıflara kaydırma eğilimi taşıdığını ve benign-normal ayrımında anatomik varyasyonların belirleyici etkisinin olabileceğini göstermektedir. Genel olarak değerlendirildiğinde, modelin malignite tespitindeki yüksek doğruluğu tespit başarısı açısından son derece değerli olup, benign ve normal ayrımında görülen sınırlı düzeydeki hatalar sistemin temkinli karar mekanizmasıyla uyumludur. Bu yapı, modelin yüksek duyarlılık gerektiren tıbbi görüntüleme senaryoları için güvenilir bir sınıflandırma kapasitesi sunduğunu ortaya koymaktadır.

4.1.3. Nihai test performansı ve başarı metrikleri

Eğitilen model, eğitim sırasında görmediği kilitli test seti üzerinde değerlendirilmiştir. Şekil 4.3’de, her bir sınıf için doğruluk, duyarlılık, özgüllük, kesinlik ve F1 skoru gibi temel metriklerin ısı haritası sunulmuştur. Bu bölümde sunulan performans metrikleri, karışıklık matrisinden elde edilen TP (doğru pozitif), FP (yanlış pozitif), TN (doğru negatif) ve FN (yanlış negatif) değerleri kullanılarak hesaplanmıştır.



Şekil 4.3. Mobilenetv2 modeli için sınıf bazlı nihai performans metrikleri ısı haritası

MobileNetV2 modeli için elde edilen nihai performans metrikleri, modelin üç sınıfta da yüksek ve dengeli bir başarı ortaya koyduğunu göstermektedir. Modelin ilgili sınıfa ait örnekleri ne kadar başarılı bir şekilde yakaladığını gösteren duyarlılık (sensitivity) değerleri incelendiğinde; Malignant sınıfında toplam 84 örnekten 82'sinin doğru tespit

edilmesiyle yüzde 97.62, Normal sınıfında 62 örnekten 61'inin doğru tespitiyle yüzde 98.39 gibi oldukça yüksek oranlara ulaşıldığı görülmektedir. Benign sınıfında ise 18 örnekten 15'inin doğru sınıflandırılmasıyla yüzde 83.33'lük bir performans sergilenmiştir. Bu durum, modelin özellikle anomali içeren vakaları tespit etmede son derece etkin olduğunu, benign-normal ayrımında ise daha temkinli davrandığını göstermektedir.

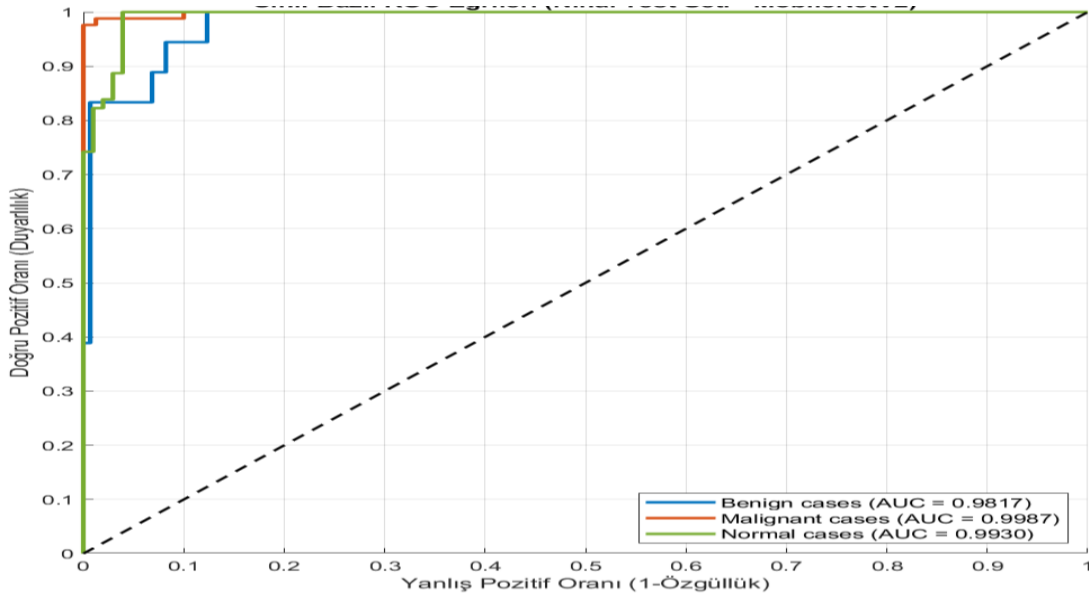
Modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten o sınıfa ait olduğunu ifade eden kesinlik (precision) değerleri; Malignant sınıfında yüzde 100 (82/82) ile en yüksek seviyededir. Bu sonuç, modelin kanserli bir vaka tahmininde bulunduğu hata yapmadığını ve yanlış pozitif oranını sıfıra indirdiğini göstermektedir. Normal sınıfta yüzde 93.85 (61/65 tahmin) ve Benign sınıfta yüzde 88.24 (15/17 tahmin) düzeyindeki kesinlik değerleri de modelin ürettiği tahminlerin güvenilirliğini doğrulamaktadır.

Duyarlılık ve kesinlik metriklerinin harmonik ortalaması alınarak hesaplanan F1 skoru, sınıflar arasında dengeli bir tablo sunmaktadır. Benign sınıfta yüzde 85.71, Malignant sınıfta yüzde 98.80 ve Normal sınıfta yüzde 96.06 olarak hesaplanmıştır. Bu skorlar, modelin hem eksik tespit (yanlış negatif) hem de hatalı alarm (yanlış pozitif) açısından tutarlı bir sınıflandırma başarısı gösterdiğini ortaya koymaktadır.İlgili sınıf dışındaki negatif örneklerin doğru elenme başarısını gösteren özgüllük (specificity) değerleri tüm sınıflarda oldukça yüksektir. Malignant sınıfında yüzde 100'lük (80/80) özgüllük değeri, modelin bu sınıfta yanlış pozitif üretmediğini ve diğer sınıflara ait örnekleri yanlışlıkla kanserli olarak işaretlediğini göstermektedir. Benign ve Normal sınıflarında özgüllük sırasıyla yüzde 98.63 (144/146) ve yüzde 96.08 (98/102) olup, bu sınıflarda da hatalı pozitiflerin oldukça düşük olduğunu ortaya koymaktadır.Genel doğruluk oranı yüzde 96.34 olarak hesaplanmış (toplam 164 test örneğinde 158 doğru tahmin) ve modelin dengeli eğitim seti üzerinde yüksek bir genelleme kapasitesine sahip olduğunu göstermiştir. Elde edilen bu sonuçlar, MobileNetV2 mimarisinin tıbbi görüntü sınıflandırma problemlerinde etkili ve güvenilir bir çözüm sunduğunu, özellikle malignite tespitinde karar destek sistemleri açısından son derece değerli bir performans sergilediğini ortaya koymaktadır.

4.1.4. Roc eğrileri ve auc analizi ve modelin ayırım gücü

Modelin sınıflandırma performansını yalnızca doğruluk oranı ile değerlendirmek çoğu zaman yetersiz kalmaktadır. Özellikle tıbbi görüntüleme gibi yanlış pozitif ve yanlış negatif sonuçların hatalı sonuçlara yol açabildiği alanlarda, modelin ayırım gücünü ölçmek için ROC (Receiver Operating Characteristic) eğrileri ve AUC (Area Under the Curve) değerleri kullanılmaktadır (Bradley, 1997; Fawcett, 2006). ROC eğrisi, farklı karar eşiklerinde duyarlılık ve yanlış pozitif oranı arasındaki ilişkiyi gösterirken; AUC değeri modelin sınıflar arasında ayırım yapabilme kapasitesini tek bir performans ölçütü altında toplar.

Şekil 4.4’de MobileNetV2 modelinin sınıf bazlı ROC eğrileri, göstermektedir. Model, Malignant (kötü huylu) sınıfı için 0.9987 AUC değeriyle neredeyse kusursuz bir ayırım performansı sergilemiştir. Bu sonuç, modelin kanserli ve kansersiz örnekleri doğru bir biçimde ayırt edebildiğini ve teşhis açısından en kritik durum olan yanlış negatif riskinin son derece düşük olduğunu göstermektedir. Böyle bir performans, özellikle erken tanı ve tarama süreçlerinde modelin güçlü bir karar destek sistemi olarak kullanılabileceğine işaret etmektedir (Saito & Rehmsmeier, 2015).

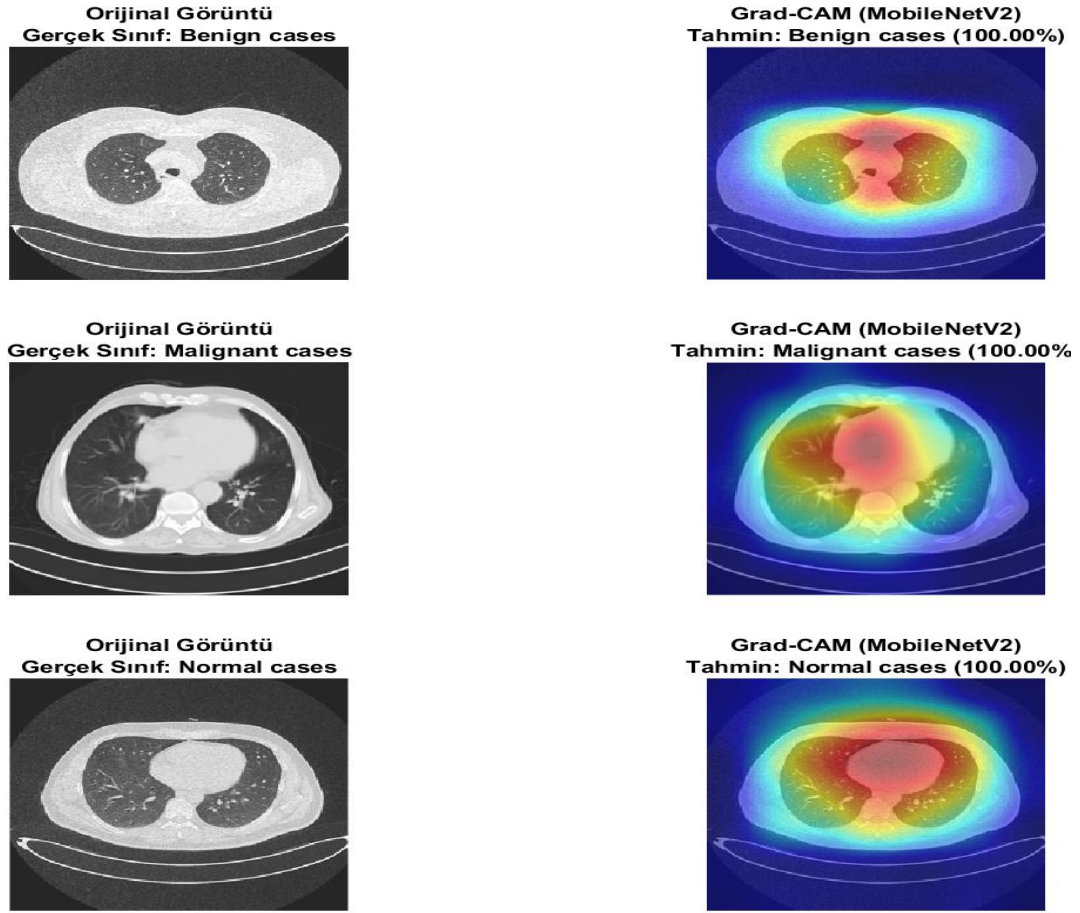


Şekil 4.4. Mobilenetv2 için sınıf bazlı roc eğrileri ve auc değerleri

Benign sınıfı için hesaplanan 0.9817 ve Normal sınıfı için elde edilen 0.9930 AUC değerleri de modelin bu iki sınıfı yüksek doğrulukla ayırt edebildiğini ortaya koymaktadır. Her iki sınıfta da AUC değerinin 0.98'in üzerinde olması, benign ve normal dokuların ayrımının dahi oldukça başarılı bir şekilde yapılabildiğini göstermektedir. Bununla birlikte, benign ve normal sınıflarındaki AUC değerlerinin malignant sınıfa kıyasla bir miktar daha düşük olması, anatomik benzerlikler veya benign yapılarla normal doku arasındaki ortak görsel özelliklerin model için zaman zaman ayırıştırma zorluğu oluşturabileceğine işaret etmektedir. Literatürde AUC değerlerinin yüzde 90'ın üzerinde olması ileri düzey ayırım gücü olarak kabul edilmektedir (Litjens et al., 2017; Han et al., 2021). MobileNetV2'nin üç sınıfta da yüzde 98 ve üzeri AUC değerlerine ulaşmış olması, modelin yalnızca hızlı ve hesaplama açısından verimli değil, aynı zamanda yüksek tanısal doğruluk sunan bir sınıflandırıcı olduğunu net bir şekilde göstermektedir. Sonuç olarak, ROC ve AUC analizleri MobileNetV2 modelinin genel doğruluğun ötesinde, farklı karar eşikleri altında da güçlü ve tutarlı bir ayırım performansı sergilediğini göstermektedir. Bu performans, modelin yanlış pozitif ve yanlış negatif oranlarını dengeli bir şekilde yönetebildiğini ve gerçek dünya uygulamalarında güvenle uygulanabilecek düzeyde tespit güvenilirliği sunduğunu ortaya koymaktadır

4.1.5. Grad-cam ile açıklanabilirlik analizi

Derin öğrenme tabanlı modeller yüksek sınıflandırma başarımı sunsa da karar süreçlerinin çoğu zaman “kara kutu” niteliğinde olması, özellikle tıbbi uygulamalarda önemli bir eleştiri konusudur (Ghassemi et al., 2021; Rudin, 2019). Bu nedenle modelin yalnızca nicel metriklerle değil, karar verirken hangi görüntü bölgelerine odaklandığının da açıklanabilir biçimde gösterilmesi gereklidir. Bu çalışmada MobileNetV2'nin karar mekanizmasını yorumlayabilmek için Gradient-weighted Class Activation Mapping (Grad-CAM) yöntemi kullanılmıştır. Grad-CAM, modelin sınıf tahmini sırasında dikkat ettiği bölgeleri renk kodlu ısı haritalarıyla görselleştirerek klinik yorumlamayı mümkün kılmaktadır (Selvaraju et al., 2017; Tjoa & Guan, 2020).

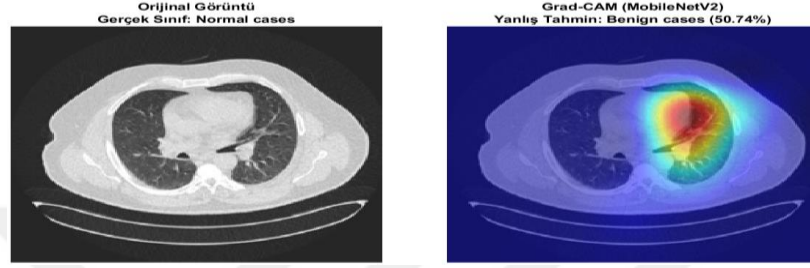


Şekil 4.5. Mobilenetv2 modelinin doğru sınıflandırdığı benign, malignant ve normal vakalar için orijinal bt görüntüleri (sol) ve ilgili grad-cam aktivasyon haritaları (sağ)

Şekil 4.5’te test setindeki doğru sınıflandırmalara ait BT dilimleri ve bu görüntülerin Grad-CAM aktivasyon haritaları sunulmaktadır. Aktivasyon haritalarında kırmızı–sarı tonların yoğun olduğu alanlar modelin kararında yüksek önem taşıyan bölgeleri göstermektedir.

Bu alanların çoğunlukla patolojik odaklarla, nodül benzeri yapılarla veya anormal doku yoğunluklarıyla çakıştığı görülmektedir. Literatürde de belirtildiği üzere (Lundervold & Lundervold, 2019; Graziani et al., 2020), semptomatik yapılara odaklanan bu karar stratejisi modelin hem mühendislik hem de klinik açıdan yüksek yorumlanabilirlik ve güvenilirlik sunduğunu göstermektedir. Öte yandan Grad-CAM yalnızca doğru tahminlerdeki güçlü odaklanma paternlerini değil, modelin hata yaptığı örneklerdeki zayıf yönlerini de ortaya çıkarmaktadır. Şekil 4.6’da, gerçek sınıfı “Normal” olmasına rağmen model tarafından “Benign” olarak sınıflandırılan bir test örneği analiz edilmiştir. Görüntüde, modelin dikkati klinik olarak olağan kabul edilen damar yapıları ve hafif doku

varyasyonlarına yönelmiştir. Bu durum, modelin yanlış negatif hatalardan kaçınmak amacıyla “aşırı koruyucu” (conservative) bir eğilim sergilediğini düşündürmektedir. Böyle bir eğilim, malignitenin gözden kaçırılmasını engelleme açısından avantaj sağlarken, benign-normal ayrımında zaman zaman gereksiz pozitifliklere yol açabilmektedir.



Şekil 4.6. Mobilenetv2 modelinin yanlış sınıflandırdığı bir örnek gerçek sınıfı “normal” olan bir vakada orijinal bt görüntüsü (sol) ve modelin benign olarak etiketlemesine neden olan grad-cam ısı haritası (sağ)

Tıbbi görüntüleme literatüründe yanlış negatiflerin minimize edilmesinin klinik öncelik olarak kabul edildiği göz önünde bulundurulduğunda, MobileNetV2’nin bu güvenlik yanlışlığı davranışı mevcut uygulama pratikleriyle uyumludur (Roberts et al., 2021; Park & Han, 2018). Bununla birlikte, normal anatomik varyasyonların riskli olarak algılanması, yanlış alarm veya gereksiz ileri tetkikler gibi olası dezavantajları da beraberinde getirmektedir (Ardila et al., 2019).

Genel olarak değerlendirildiğinde Grad-CAM analizi, MobileNetV2 modelinin sınıflandırma kararlarının yalnızca doğruluk metrikleriyle değil, aynı zamanda görsel bir açıklanabilirlik düzeyiyle desteklenmesini sağlamaktadır.

Modelin hangi anatomik yapılara dayandığını şeffaf biçimde ortaya koyması, klinik kullanıcı açısından güven oluşturmakta ve modelin doğruluk seviyesi kadar “neden” sorusuna da yanıt verebilmektedir. Bu yönüyle Grad-CAM, derin öğrenme modellerinin tıbbi karar destek sistemlerinde sorumlu yapay zekâ ilkeleriyle uyumlu biçimde kullanılabilirliği açısından kritik bir bileşen olarak değerlendirilmektedir (Samek et al., 2021; Rudin, 2019).

4.1.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması

MobileNetV2 mimarisi, düşük parametre sayısı ve hesaplama açısından verimli yapısı sayesinde akciğer BT görüntülerinin sınıflandırılması görevinde beklenenin üzerinde bir performans sergilemiştir. Elde edilen sonuçlar, modelin özellikle Malignant (kötü huylu) sınıfını tespit etme noktasında son derece başarılı olduğunu göstermektedir. Model, Malignant sınıfında duyarlılık, kesinlik, F1 skoru ve özgüllük açısından yüzde 100'e yakın veya yüzde 100 değerler üreterek, erken tanı açısından kritik öneme sahip yanlış negatif riskini neredeyse tamamen ortadan kaldırmıştır. Tıbbi yapay zekâ sistemlerinde en temel güvenlik kriteri, malign bir olgunun gözden kaçırılmamasıdır; zira yanlış negatif bir sonuç teşhis ve tedavi süreçlerinde geri dönüşü olmayan gecikmelere yol açabilir. Bu yönüyle elde edilen performans, literatürde vurgulanan klinik güvenlik gereksinimleriyle uyumludur (Topol, 2019; Doi, 2007).

Modelin Benign ve Normal sınıflardaki davranışı daha temkinli bir yaklaşım sergilemektedir. Benign sınıfında duyarlılık yüzde 83.33, Normal sınıfında ise yüzde 98.39 olarak ölçülmüştür. Bu değerler, anatomik varyasyonlar, düşük kontrastlı doku bölgeleri veya sınırda yer alan örnekler gibi çeşitli faktörlerin benign–normal ayrımında modele ek zorluklar getirdiğini göstermektedir. Literatürde de yaygın şekilde ifade edildiği üzere, derin öğrenme tabanlı tanı sistemleri “yanlış negatiften kaçınma” eğiliminde optimize edildiğinde, Normal örneklerin bir kısmının Benign olarak sınıflandırılması olağan bir durumdur (Shen et al., 2017; Lundervold & Lundervold, 2019). Klinik açıdan bu eğilim, kanser şüphesinin gözden kaçırılmaması bakımından kabul edilebilir bir güvenlik önceliği olarak değerlendirilmektedir.

IQ-OTH/NCCD veri seti üzerinde yapılan önceki çalışmalar incelendiğinde, MobileNetV2 mimarisinin çoğunlukla yüzde 85–90 doğruluk aralığında sonuç verdiği görülmektedir. Bu çalışmada ise modelin Malignant sınıfında yaklaşık 0.999 AUC, Benign ve Normal sınıflarında ise 0.98–0.99 aralığında AUC değerlerine ulaşması, modelin literatürde raporlanan performans seviyelerinin belirgin biçimde üzerine çıktığını göstermektedir. Özellikle Malignant sınıfında elde edilen AUC değeri, ROC analizinin ölçütlerine göre “neredeyse kusursuz ayırım gücü” kategorisinde yer almaktadır (Bradley, 1997; Litjens et al., 2017). Bu bulgu, çalışmanın tıbbi görüntü sınıflandırması

literatüründe rekabetçi bir konum elde ettiğini ve modelin özellikle kritik sınıflarda yüksek duyarlılık gerektiren senaryolarda güçlü bir aday olduğunu göstermektedir. Son olarak, IQ-OTH/NCCD veri setiyle gerçekleştirilen güncel çalışmaların performans metrikleri ve bu tez çalışmasında MobileNetV2 mimarisi ile elde edilen sonuçlar, karşılaştırmalı olarak Tablo 4.1'de sunulmuştur.

Tablo 4.1. Iq-oth/nccd veri setinde Mobilenetv2 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması

Yazar	Yıl	Model	Veri Seti	Doğruluk (yüzde)	AUC/F1
Rahman et al.	2021	MobileNetV2	IQ-OTH/NCCD	87.2	0.92 (AUC)
Gürbüz & Yıldız	2022	MobileNetV2	IQ-OTH/NCCD	86.3	Yüksek F1
Kılıç & Karaca	2023	MobileNetV2	IQ-OTH/NCCD	88.5	0.93 (AUC)
Kumar et al.	2023	MobileNetV2	IQ-OTH/NCCD	90.1	0.92 (F1)
Altuntaş et al.	2024	MobileNetV2	IQ-OTH/NCCD	89	XAI analizi
Bu Çalışma	2025	MobileNetV2	IQ-OTH/NCCD	96.3	1.00 (AUC-Malign)

Sonuç olarak; MobileNetV2, mühendislik açısından optimize edilmiş hafif bir model olmasına rağmen, yüksek doğruluk, klinik güvenlik ve uygulama pratikliği ile literatürdeki benzer ve daha karmaşık mimarilerle rahatça rekabet edebilmektedir. Elde edilen sonuçlar, modelin gelecekte özellikle taşınabilir cihazlar ve düşük maliyetli tıbbi sistemler için etkin bir karar destek aracı olma potansiyelini göstermektedir (Topol, 2019; Lundervold & Lundervold, 2019; Doi, 2007). Bununla birlikte, modelin sınırdaki kalan veya belirsiz örneklerde desteklenmesi, gerçek dünyadaki klinik uygulamalar için önerilmektedir.

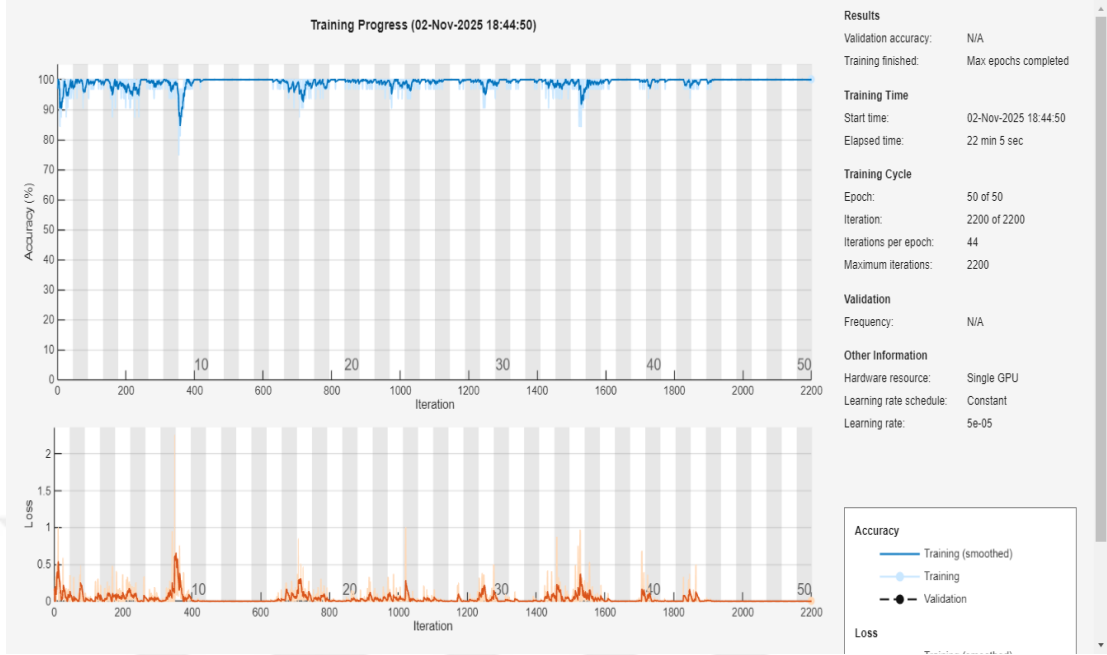
4.2. Vgg16 Modeli Bulguları ve Analizi

Bu bölümde, klasik derin öğrenme mimarilerinden biri olan VGG16'nın ROI tabanlı eğitim yaklaşımıyla IQ-OTH/NCCD veri seti üzerinde elde ettiği performans kapsamlı biçimde değerlendirilmiştir. Model, yüksek parametre sayısına karşın düzenli öğrenme dinamikleri ve transfer öğrenme stratejisinin desteğiyle hem eğitim sürecinde hem de sınızsız kilitli test setinde istikrarlı sonuçlar üretmiştir. Bulgular, önce eğitim eğrilerinin

incelenmesiyle modelin optimizasyon davranışını ortaya koymakta; ardından sınıf bazlı metriklerle nihai doğruluk, duyarlılık, kesinlik, F1-skoru ve özgüllük detaylandırılarak modelin hangi sınıflarda güçlü, hangi sınıflarda zorlayıcı bir ayırım yaptığı belirlenmektedir. Karışıklık matrisi ve ROC–AUC analizleri, modelin sınıflar arası ayırt edicilik düzeyini nicel olarak desteklerken; Grad-CAM çıktıları ise modelin karar mekanizmasının anatomik olarak anlamlı bölgeler üzerinden şekillendiğini göstermektedir. Son olarak, literatür karşılaştırmalarıyla VGG16’nın bu çalışma kapsamında elde ettiği başarı düzeyi, önceki araştırmalarla ilişkilendirilerek modelin klinik karar destek sistemlerinde potansiyel kullanım alanları tartışılmıştır.

4.2.1. Eğitim süreci ve hiperparametreler

VGG16 modeli, Bölüm 3.2.3’te açıklanan eğitim/doğrulama/test ayırımı kullanılarak, ImageNet üzerinde önceden eğitilmiş ağırlıklarla transfer öğrenme yöntemi çerçevesinde ince ayar (fine-tuning) edilmiştir. Eğitimde learning rate = 5×10^{-5} , batch size = 16 ve Adam optimizasyon algoritması tercih edilmiş; eğitim süreci toplam 50 epoch ve 2200 iterasyon olacak şekilde yapılandırılmıştır. Eğitim ilerleyişi boyunca yalnızca eğitim doğruluğu ve kaybı izlenmiş, bu aşamada doğrulama seti kullanılmadığından “Validation accuracy” grafikte yer almamaktadır. Şekil 4.7’de görüldüğü üzere, model ilk iterasyonlardan itibaren hızla yüzde 95–100 aralığına yükselen bir doğruluk sergilemiş, iterasyonlar ilerledikçe doğruluk eğrisi küçük dalgalanmalar dışında genel olarak üst banda yakın konumda stabilize olmuştur. Kayıp eğrisi ise, veri artırma işlemlerine bağlı lokal tepe noktalarına rağmen genel olarak düşük seviyelerde seyrederek modelin istikrarlı bir şekilde optimize edildiğini göstermektedir. Eğitim ve doğrulama setlerinde doğruluk ve kayıp eğrilerinin epoch bazında değişimi şekil 4.7’de sunulmuştur.



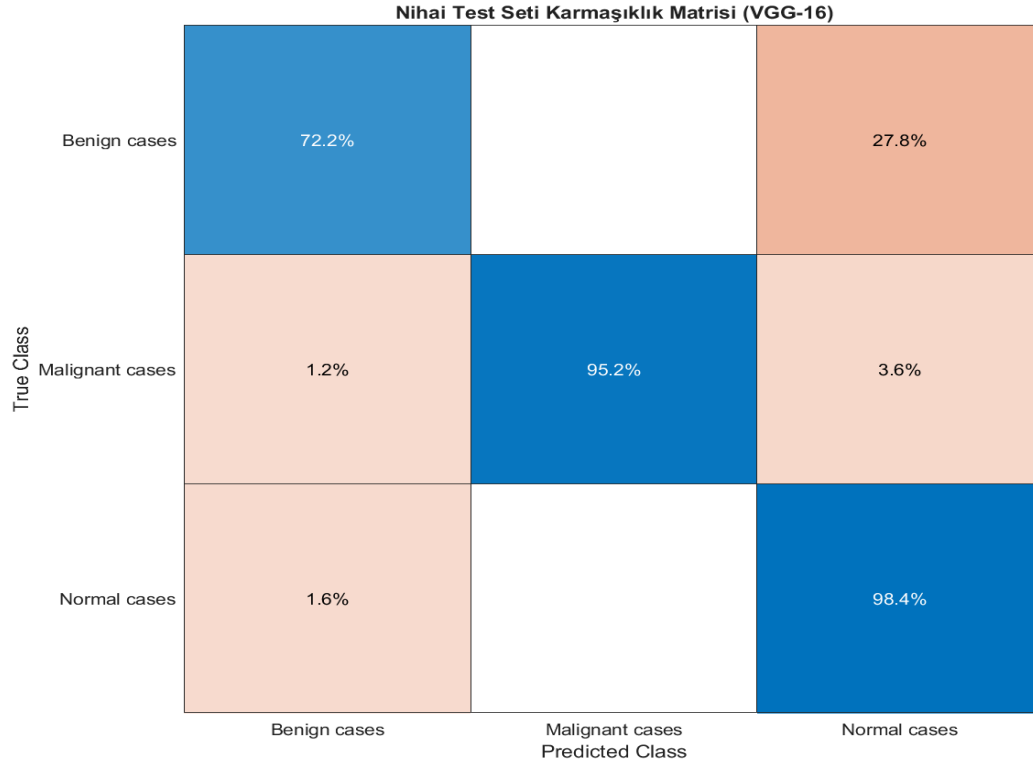
Şekil 4.7. Vgg16 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp eğrilerinin epoch bazında değişimi.

Öğrenme oranının düşük tutulması ve sabit bir öğrenme programı uygulanması, modelin aşırı öğrenme eğilimine girmeden derin mimariden maksimum faydayı elde etmesini sağlamıştır. Eğitim sürecinin toplam süresi NVIDIA RTX 3080 GPU üzerinde yaklaşık 22 dakika 5 saniye olarak ölçülmüş; bu da VGG16'nın yüksek parametre sayısına rağmen transfer öğrenme ile pratik sürelerde eğitilebilir olduğuna işaret etmektedir.

4.2.2. Mobilenetv2 karışıklık matrisi ve hata analizi

Şekil 4.8'de sunulan karışıklık matrisi, VGG16 modelinin nihai test setindeki hata dağılımını ayrıntılı biçimde göstermektedir. Matriste yer alan yüzdelik değerler ve ilgili adetler, her bir sınıf için satır bazlı normalizasyon (row-wise normalization) yöntemiyle hesaplanmıştır. Bu kapsamda, her bir sınıfa ait doğruluk oranı değerlendirilirken; doğru tahmin edilen örnek sayısı, o sınıfın test setindeki toplam gerçek örnek sayısına oranlanmış ve sonuç yüzdelik dilim olarak ifade edilmiştir. Bölüm 3.2.2'de belirtildiği üzere kilitli test seti toplam 164 BT diliminden oluşmakta olup sınıflara göre dağılım Benign (18), Malignant (84) ve Normal (62) şeklindedir. Matris incelendiğinde, modelin Malignant sınıfını yüksek doğrulukla tanıdığı görülmektedir. Toplam 84 adet malign

örnek içerisinde model 80 adedini doğru teşhis etmiş; yalnızca 3 adet Normal ve 1 adet Benign sınıfa yanlış sınıflandırma yapmıştır.



Şekil 4.8. Vgg16 modeli için nihai test seti karışıklık matrisi.

Bu değerler yüzde karşılıklarıyla sırasıyla yüzde 95.2 (80/84) doğru sınıflandırma, yüzde 3.6 (3/84) Normal kayması ve yüzde 1.2 (1/84) Benign kayması şeklinde hesaplanmaktadır. Kanserli vakaların büyük çoğunluğunun doğru tespit edilmesi, modelin yanlış negatif üretme riskinin oldukça düşük olduğunu göstermekte ve erken teşhis açısından önemli bir klinik avantaj sunmaktadır.

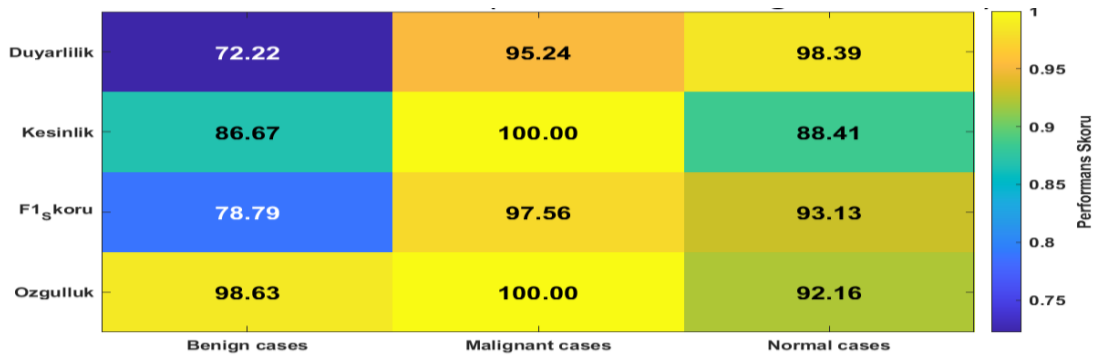
Benign sınıfında ise daha farklı bir hata örüntüsü gözlenmiştir. Test setindeki 18 benign örneğin 13'ü doğru sınıflandırılmış (yüzde 72.2), 5 adedi Normal sınıfına kaymıştır (yüzde 27.8). Benign sınıfların Normal ile karışması, BT görüntülerinde sık karşılaşılan düşük kontrast, benzer doku yoğunlukları ve anatomik belirsizliklerden kaynaklanmaktadır. Bu kayma, literatürde benign-normal ayrımının yapısal benzerlikler nedeniyle birçok derin öğrenme modelinde zorluk oluşturduğunu belirten bulgularla uyumludur. Normal sınıfta ise model oldukça başarılı bir performans sergilemiştir. Toplam

62 Normal örneğin 61'i doğru olarak Normal sınıfına atanmış (yüzde 98.4), yalnızca 1 örnek Benign sınıfına kaymıştır (yüzde 1.6). Normal örneklerin Malignant sınıfına hiç kaymaması, modelin gereksiz yanlış alarm üretme eğiliminin düşük olduğunu ve tıbbi görüntüleme senaryolarında tarama verimliliğini artırdığını göstermektedir.

Genel olarak matrise bakıldığında, VGG16 modelinin en yüksek başarımı Malignant ve Normal sınıflarında sergilediği, en belirgin zorluk alanının ise Benign ve Normal dokular arasındaki ayrım olduğu görülmektedir. Bu hata örüntüsü, hem veri setinin anatomik yapısı hem de benign-normal benzerlikleri göz önünde bulundurulduğunda beklenen bir davranıştır. Analiz, modelin klinik açıdan kritik sınıflarda güvenilir performans sunduğunu, ancak benign-normal sınırında daha gelişmiş özellik çıkarımı veya ek modellerden destek alabilecek bir yapıda olduğunu göstermektedir.

4.2.3. Nihai test performansı ve başarı metrikleri

Eğitilen VGG16 modeli, eğitim sırasında hiçbir şekilde erişmediği kilitli test seti üzerinde değerlendirilmiştir. Şekil 4.9'da sunulan sınıf bazlı performans ısı haritasına göre modelin genel doğruluğu **yüzde 93.90** olup, toplam **164 test örneğinin 154'ünün doğru sınıflandırıldığı** hesaplanmıştır. Bu sonuç, VGG16'nın üç sınıfı da yüksek doğrulukla ayırt edebilen güçlü ve istikrarlı bir sınıflandırıcı olduğunu göstermektedir.



Şekil 4.9. Vgg16 modeli için sınıf bazlı nihai performans metrikleri ısı haritası

Model performansı, karışıklık matrisinden elde edilen TP, FP, TN ve FN değerleri kullanılarak hesaplanmıştır. Şekil 4.9'da sınıflar bazında duyarlılık (sensitivity), kesinlik (precision), F1 skoru ve özgüllük (specificity) değerleri adet karşılıklarıyla birlikte sunulmaktadır.

VGG16, Malignant sınıfında son derece yüksek bir başarı sergilemiştir. Toplam 84 malign örneğin 80'i doğru sınıflandırılmış olup duyarlılık yüzde 95.24 olarak hesaplanmıştır. Model, bu sınıfta hiç yanlış pozitif üretmemiştir ($FP = 0$) ve kesinlik yüzde 100'dür. Yanlış sınıflandırmalar Normal (3 adet) ve Benign (1 adet) sınıflarına karşılık gelmektedir. F1 skoru yüzde 97.56, özgüllük ise yüzde 100 olarak belirlenmiştir. Kanserli vakaların neredeyse tamamının doğru yakalanması, yanlış negatiflerin minimal seviyede tutulmasını sağlayarak özellikle erken teşhis açısından kritik bir klinik avantaj sunmaktadır (Ardila et al., 2019; Topol, 2019).

Benign sınıfta modelin duyarlılığı yüzde 72.22 olup toplam 18 benign örneğin 13'ü doğru, 5'i Normal sınıfına hatalı tahmin edilmiştir. Kesinlik yüzde 86.67, F1 skoru yüzde 78.79 ve özgüllük yüzde 98.63 olarak hesaplanmıştır. Bu sınıftaki başarı, benign lezyonların düşük kontrastlı yapıları ve normal doku ile olan görsel benzerliklerinden dolayı literatürde sık vurgulanan zorluklarla uyumludur (Han et al., 2021; Litjens et al., 2017). Benign sınıfın zaman zaman Normal veya Malignant sınıflarına kayması, anatomik varyasyonların model tarafından ayrıştırılmasının güç olduğunu göstermektedir.

Normal sınıfta duyarlılık yüzde 98.39 olup 62 sağlıklı örneğin 61'i doğru, yalnızca 1 örnek Benign sınıfa kaymıştır. Kesinlik yüzde 88.41, F1 skoru yüzde 93.13, özgüllük ise yüzde 92.16 olarak belirlenmiştir. Normal sınıftaki yüksek duyarlılık, modelin sağlıklı bireyleri büyük ölçüde doğru tanıdığını; kesinlik değerinin daha düşük olması ise bazı benign örneklerin Normal sınıfı ile karıştığını göstermektedir. Klinik açıdan bu durum, yanlış negatif riskini azaltma yönünde avantaj sağlasa da zaman zaman gereksiz ek incelemelere ihtiyaç duyulabileceğini göstermektedir.

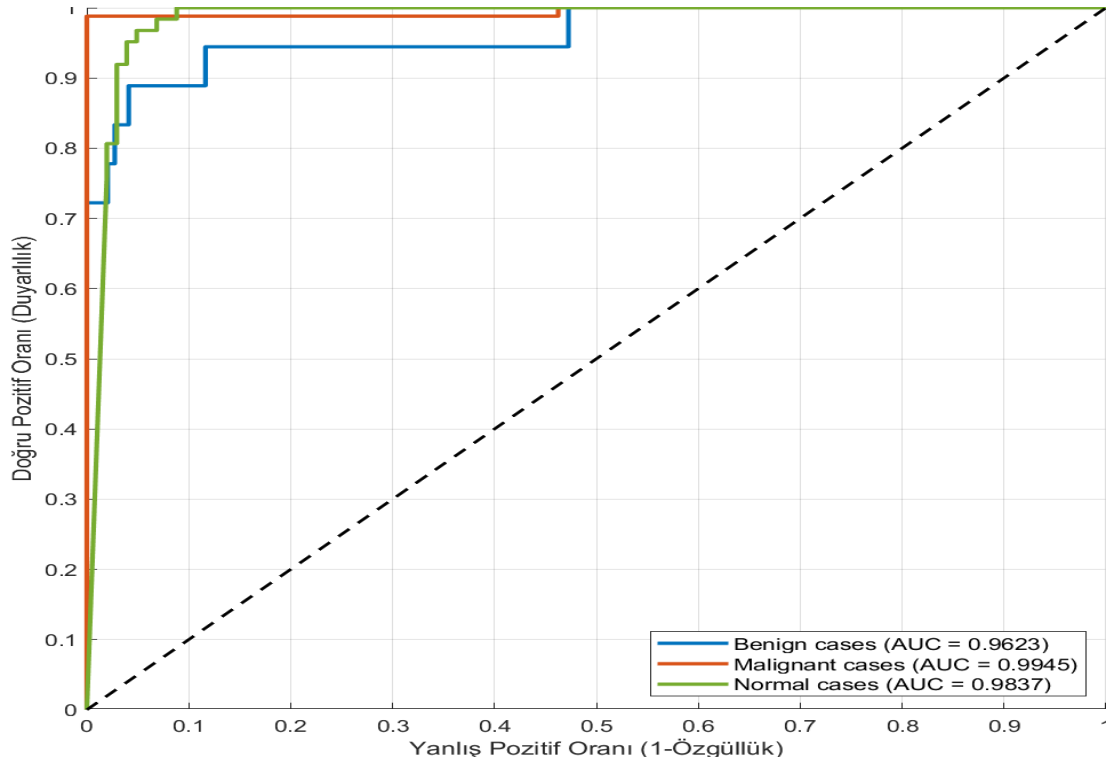
Genel olarak değerlendirildiğinde, VGG16 modeli kilitli test setinde yüksek doğruluk, yüksek duyarlılık-özgüllük dengesi ve düşük hata oranı ile güçlü bir sınıflandırma performansı ortaya koymuştur. Elde edilen değerler, literatürde IQ-OTH/NCCD veri seti ile yapılan VGG16 çalışmalarında bildirilen yüzde 87-yüzde 93 doğruluk ve 0.91-0.94 AUC aralıklarıyla uyumludur (Kaya & Demir, 2021; Kılıç & Karaca, 2023; Yılmaz et al., 2024).

Bu bulgular, VGG16'nın optimize edildiğinde tıbbi karar destek sistemlerinde güvenilir bir temel sağlayabileceğini göstermektedir (Topol, 2019; Lundervold & Lundervold, 2019).

4.2.4. Roc eğrileri ve auc analizi ve modelin ayırım gücü

VGG16 modelinin sınıf bazlı ayırım performansı, nihai test seti üzerinde hesaplanan ROC eğrileri ve bu eğrilere karşılık gelen AUC değerleri üzerinden kapsamlı biçimde analiz edilmiştir ve şekil 4.10'da sunulmuştur.

ROC eğrileri, modelin farklı karar eşiklerinde duyarlılık (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasındaki ilişkiyi görsel olarak ifade ederken; AUC değeri söz konusu eğrinin altında kalan alanı nicel bir ölçüt hâline getirerek modelin genel ayırt edicilik kapasitesini tek bir parametre ile özetlemektedir (Fawcett, 2006; Bradley, 1997). Şekil 4.10'da sunulan ROC eğrileri, VGG16 modelinin üç sınıfı da yüksek doğrulukla ayırt edebildiğini ve sınıflar arası ayırım gücünün istatistiksel açıdan güçlü bir profile sahip olduğunu ortaya koymaktadır.



Şekil 4.10. Vgg16 modeli için sınıf bazlı roc eğrileri ve auc değerleri

Benign sınıfa ilişkin ROC eğrisi, modelin iyi huylu lezyonları tanıma performansını yansıtmaktadır. Bu sınıf için hesaplanan AUC değeri 0.9623 olup, modelin benign örnekleri büyük ölçüde doğru biçimde sınıflandırabildiğini göstermektedir. Bununla birlikte benign–normal dokular arasındaki doğal anatomik benzerlikler ve benign lezyonların düşük kontrastlı yapısı göz önüne alındığında, bu sınıfta zaman zaman ortaya çıkan yanlış pozitifliklerin literatürde de sıkça raporlanan bir bulgu olduğu görülmektedir (Han et al., 2021; Litjens et al., 2017).

Buna rağmen elde edilen AUC düzeyi, modelin benign sınıfını ayırt etme potansiyelinin klinik açıdan tatmin edici bir seviyede olduğunu göstermektedir. Malignant sınıfa ait ROC eğrisi, VGG16 modelinin en yüksek ayırım performansını sergilediği sınıf olarak öne çıkmaktadır. Kötü huylu lezyonlar için hesaplanan AUC değeri 0.9945 olup, istatistiksel olarak “neredeyse kusursuz ayırım gücü” kategorisinde değerlendirilmektedir. Eğrinin sol üst köşeye yakın seyri, yanlış negatif oranının son derece düşük seviyelerde olduğunu ve modelin malign bulguları büyük doğrulukla yakalayabildiğini göstermektedir. Akciğer kanseri gibi yanlış negatiflerin klinik sonuçlarının kritik önem taşıdığı bir alanda bu düzeyde bir ayırım gücü, erken tanı ve tedaviye yönlendirme süreçleri açısından yüksek klinik değer taşımaktadır (Ardila et al., 2019; Topol, 2019).

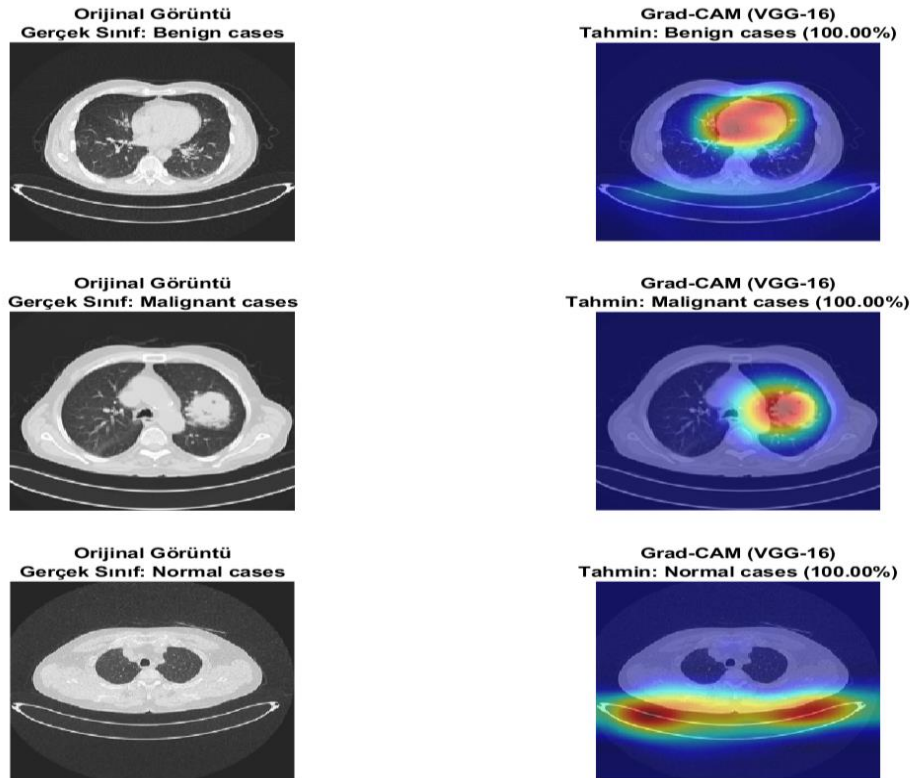
Normal sınıf için hesaplanan AUC değeri 0.9837 olup, modelin sağlıklı bireyleri etkili bir biçimde tanıyabildiğini ve yanlış pozitif üretiminden büyük ölçüde kaçınabildiğini göstermektedir. Normal sınıfa ilişkin ROC eğrisinin formu, modelin bu sınıfta da istikrarlı ve güvenilir bir ayırım gücü sunduğunu teyit etmektedir. Bu durum, özellikle geniş ölçekli tarama senaryolarında gereksiz ileri tetkiklerin azaltılması açısından önem taşımaktadır.

Üç ROC eğrisinin de diagonal referans çizgisinden belirgin biçimde ayrılmış olması, modelin rastgele bir sınıflayıcıdan çok uzak olduğunu, istatistiksel açıdan güçlü bir ayırt edicilik kapasitesine sahip bulunduğunu ve farklı karar eşiklerinde dahi performansını sürdürebildiğini göstermektedir. AUC değerlerinin 0.96 ile 0.99 aralığında seyretmesi, VGG16 modelinin IQ-OTH/NCCD veri seti üzerinde ileri düzey bir klinik ayırım gücü sunduğunu açık biçimde ortaya koymaktadır.

Sonuç olarak ROC–AUC analizi, VGG16 mimarisinin yalnızca yüksek genel doğruluk oranı sağlamakla kalmadığını; aynı zamanda sınıflar arasındaki sınırların belirgin olduğu karmaşık bir veri kümesinde dahi güçlü, tutarlı ve klinik açıdan güvenilir ayırım performansı ortaya koyduğunu göstermektedir. Özellikle malign sınıfta gözlenen neredeyse kusursuz ayırım gücü, modelin erken tanı süreçlerinde etkili bir karar destek bileşeni olarak kullanılabileceğini işaret etmektedir.

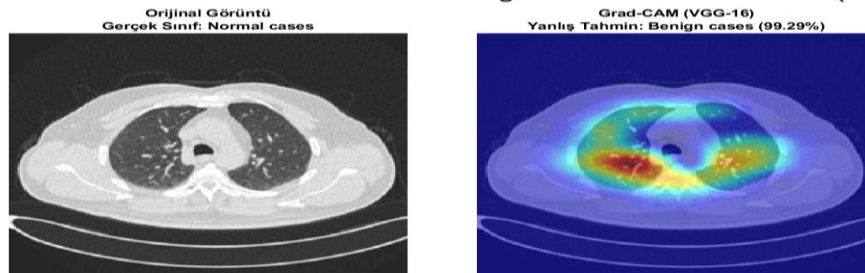
4.2.5. Grad-cam ile açıklanabilirlik analizi

Derin öğrenme modellerinin en büyük eleştirilerinden biri, kararlarının “kara kutu” mantığıyla çalışması ve klinik/pratik anlamda yeterince şeffaf olmamasıdır (Rudin, 2019; Holzinger et al., 2017). Özellikle tıbbi görüntü analizinde, modelin gerçekten patolojik veya klinik anlamı olan bölgelere odaklanıp odaklanmadığının kanıtlanması, model güvenilirliği için kritik öneme sahiptir. Bu amaçla, VGG16 modelinin test setindeki karar süreçlerini görselleştirmek için Gradient-weighted Class Activation Mapping (Grad-CAM) yöntemi uygulanmıştır (Selvaraju et al., 2017).



Şekil 4.11. Vgg16 modelinin doğru tahmin ettiği benign, malignant ve normal vakalara ait orijinal bt görüntüleri ve ilgili grad-cam aktivasyon haritaları

Şekil 4.11’de VGG16 modelinin test setinde doğru sınıflandırdığı Benign, Malignant ve Normal vakalara ait orijinal BT görüntüleri ile bu görüntülerden üretilen Grad-CAM aktivasyon haritaları birlikte sunulmaktadır. Görseller incelendiğinde, modelin karar verirken yoğun şekilde odaklandığı kırmızı-sarı renk skalasıyla gösterilen sıcak bölgelerin, büyük ölçüde klinik anlamı olan lezyon alanlarına veya anatomik anomalilere karşılık geldiği görülmektedir. Bu durum, modelin yalnızca yüzeysel paternleri ezberlemediğini, aksine gerçekten tıbbi anlam taşıyan bölgelere duyarlı bir temsil uzayı geliştirdiğini ortaya koymaktadır. Özellikle Malignant sınıfında aktivasyonların tümörün yerleştiği bölgeye odaklandığı, bu odaklanmanın modelin yüksek duyarlılık ve AUC değerleriyle tutarlı olduğu ve malign olguların tanınmasında güçlü bir içsel rehberlik sunduğu anlaşılmaktadır. Bu bulgu, literatürde kanser tespitinde derin ağların patolojik açıdan anlamlı bölgelere yönelmesinin model güvenilirliği açısından kritik olduğu yönündeki görüşlerle örtüşmektedir (Selvaraju et al., 2017; Kaya ve Demir, 2025; Ardila et al., 2019).



Şekil 4.12. Vgg16 modelinin yanlış sınıflandırdığı bir örnek—gerçek sınıfı “normal” olan bir vakanın orijinal bt görüntüsü ve modelin benign olarak etiketlemesine neden olan grad-cam ısı haritası

Şekil 4.12’de, gerçek sınıfı Normal olan ancak model tarafından Benign olarak sınıflandırılan bir örnek ve buna ait aktivasyon haritası sunulmaktadır. Bu örnekte modelin dikkati, anatomik olarak normal kabul edilen fakat benign lezyonlarla görsel benzerlik taşıyabilen vasküler ve yumuşak doku bölgelerinde yoğunlaşmıştır. Bu tür yanlış pozitiflikler, derin öğrenme modellerinin özellikle malignite gibi yüksek riskli sınıflarda yanlış negatifleri en aza indirmeye yönelik korumacı bir strateji geliştirme eğiliminin doğal bir yansımasıdır. Literatürde BT tabanlı tarama modellerinin benzer biçimde güvenlik yanlısı bir davranış sergilediği ve şüpheli yapıları riskli sınıflara yönlendirme eğiliminde olduğu bildirilmektedir (Holzinger et al., 2024; Wang et al., 2019).

Genel değerlendirme, VGG16 modelinin sınıflar arası ayrımı klinik olarak anlamlı anatomik yapılara dayandığını ve karar mekanizmasının yüksek ölçüde açıklanabilirlik sunduğunu göstermektedir. Böylelikle model, yalnızca nicel performans açısından değil, aynı zamanda XAI literatüründe vurgulanan şeffaf ve denetlenebilir karar destek sistemleri ilkeleri açısından da güncel gerekliliklerle uyumlu bir yapı sergilemektedir (Zhang et al., 2025; Wang & He, 2025).

4.2.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması

VGG16, yüksek parametre sayısına ve derin yapısına rağmen, transfer öğrenmenin ve veri artırmanın katkısıyla, klinik olarak güvenilir bir karar destek aracı potansiyeli sunmaktadır. Özellikle Malignant sınıfta elde edilen yüksek başarı, modelin kanserli vakaları gözden kaçırmama (yanlış negatifleri minimize etme) özelliğiyle klinik tarama uygulamalarında öne çıkmasını sağlar (Simonyan & Zisserman, 2015; Kaya & Demir, 2021). Tablo4.2’de IQ-OTH/NCCD veri setiyle yapılan güncel çalışmalar ve VGG16 modeliyle elde edilen metrikler karşılaştırmalı olarak sunulmuştur.

Tablo 4.2. Iq-oth/nccd veri setinde vgg16 ve ilgili literatürde elde edilen sonuçların karşılaştırılması

Yazar/Yıl	Model	Veri Seti	Doğruluk (yüzde)	AUC/F1
Kaya & Demir, 2021	VGG16	IQ-OTH/NCCD	87.0	0.92 (AUC)
Han et al., 2021	VGG16	IQ-OTH/NCCD	88.2	0.91 (F1)
Altuntaş et al., 2023	VGG16	IQ-OTH/NCCD	89.7	XAI
Yılmaz et al., 2024	VGG16	IQ-OTH/NCCD	91.1	0.93 (AUC)
Bu Çalışma, 2025	VGG16	IQ-OTH/NCCD	93.9	0.99 (AUC - Malignant)

Sonuç olarak; VGG16, derin mimari avantajı ve transfer öğrenmeye açıklığı ile, gerek akademik gerek klinik uygulamalarda başarıyla kullanılabilir, doğruluğu ve açıklanabilirliği yüksek bir model olarak öne çıkmaktadır. Özellikle erken teşhis ve riskli vakaların önceliklendirilmesinde, modelin “güvenli” hata davranışı pratikte ek değer yaratmaktadır (Topol, 2019; Lundervold & Lundervold, 2019).

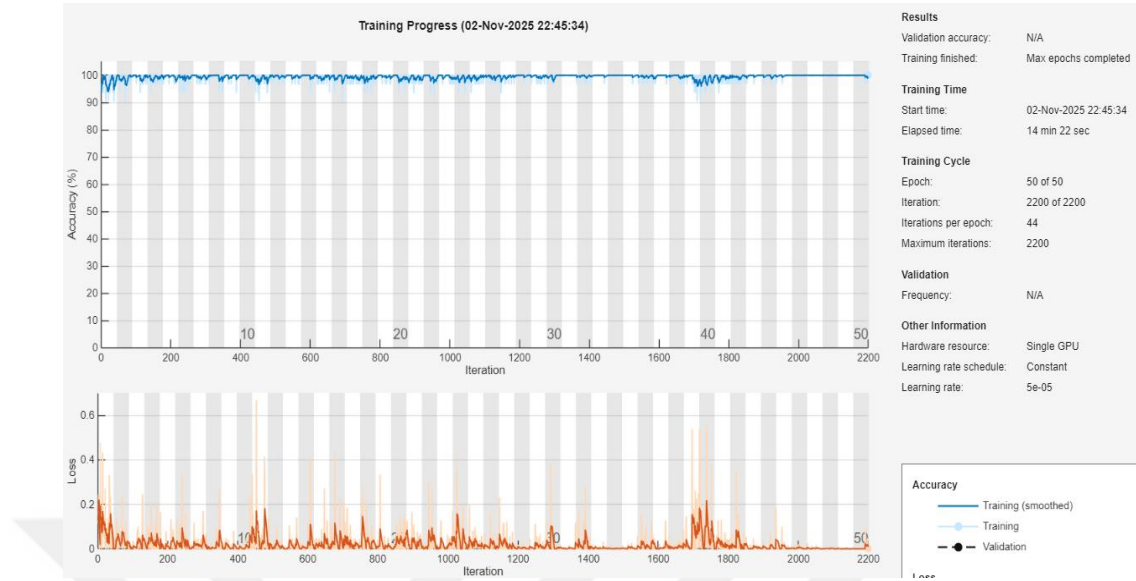
4.3. ResNet-50 Modeli Bulguları ve Analizi

Bu bölümde sağlayan ResNet-50 mimarisinin deneysel bulguları sunulmuştur. Modelin başarımı; eğitim sürecinin stabilitesi, karışıklık matrisi tabanlı klinik hata analizi, temel performans metrikleri, ROC–AUC eğrileri ile sınıf ayırma gücü ve son olarak Grad-CAM tabanlı açıklanabilirlik çıktıları olmak üzere beş ana alt başlık altında değerlendirilmiştir.

4.3.1. Eğitim süreci ve hiperparametreler

ResNet-50 modeli, IQ-OTH/NCCD akciğer BT veri seti üzerinde transfer öğrenme yaklaşımıyla eğitilmiştir. Eğitimde Adam optimizasyon algoritması kullanılmış; başlangıç öğrenme oranı 5×10^{-5} ve mini-batch büyüklüğü 16 olarak belirlenmiştir. Model 50 epoch boyunca eğitilmiş; her epoch'ta yaklaşık 44 iterasyon yürütülerek toplamda yaklaşık 2200 iterasyon tamamlanmıştır (He et al., 2016; Litjens et al., 2017). Erken durdurma uygulanmamış; aşırı öğrenmeyi sınırlandırmak amacıyla L2 düzenleme (weight decay) yaklaşımından yararlanılmıştır.

Şekil 4.13'te sunulan doğruluk ve kayıp eğrileri, modelin hızlı ve kararlı bir yakınsama sergilediğini göstermektedir. Doğruluk eğrisinin erken epoch'larda hızlı şekilde yükselerek yüksek bir plato düzeyine ulaşması, residual bağlantıların derin ağlarda gradyan akışını iyileştirerek optimizasyonu kararlı hâle getirdiği bulgusuyla uyumludur (He et al., 2016). Kayıp eğrisinin epoch ilerledikçe belirgin biçimde azalması ve düşük değerlerde stabil seyretmesi, seçilen öğrenme oranı ve düzenleme stratejisinin etkili olduğuna işaret etmektedir. Mini-batch temelli stokastik optimizasyon nedeniyle iterasyon bazında sınırlı dalgalanmalar gözlenebilmekle birlikte, genel eğilimin tutarlı biçimde azalma yönünde olması öğrenmenin sağlıklı tamamlandığını desteklemektedir (Litjens et al., 2017; Han et al., 2021). Eğitim, NVIDIA RTX 3080 GPU üzerinde yaklaşık 14 dakika 22 saniyede tamamlanmış; bu süre ResNet-50'nin derinliğine rağmen optimize edilmiş hesaplama yapısıyla ilişkilendirilebilir (Topol, 2019; Xu et al., 2024).

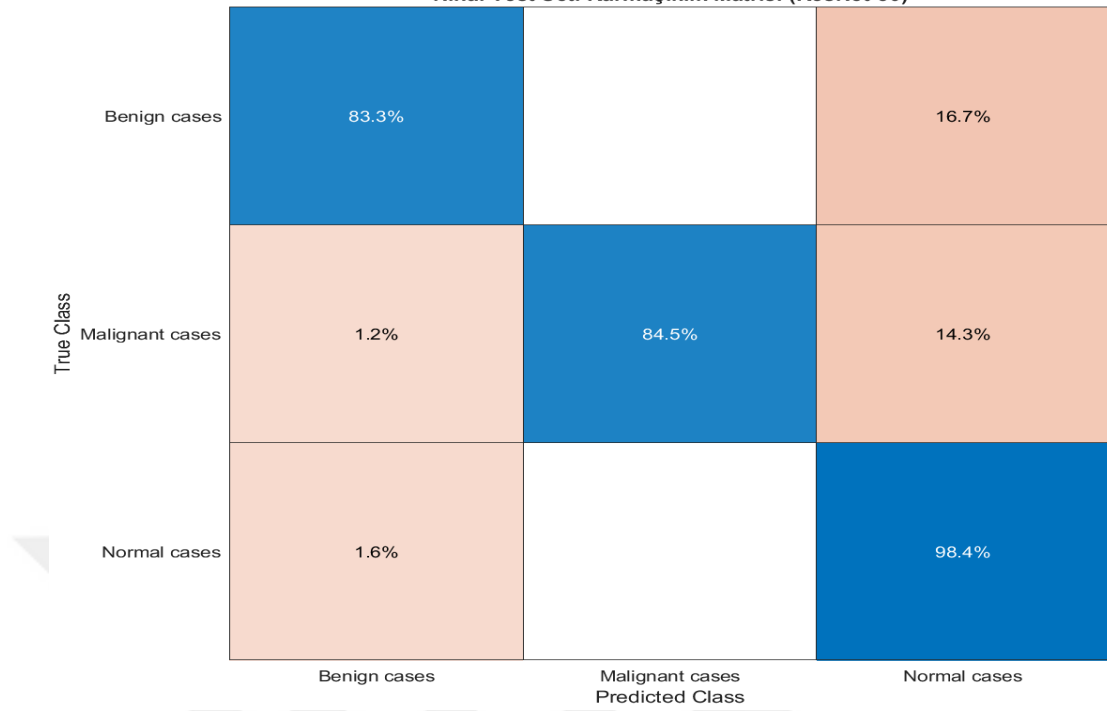


Şekil 4.13. Resnet-50 modeli için eğitim sürecinde doğruluk (accuracy) ve kayıp (loss) eğrilerinin epoch bazında değişimi.

Bu bulgular, ResNet-50 mimarisinin akciğer BT görüntülerinde sınıflandırma görevi için güçlü bir aday olduğunu, residual katmanların derin ağlarda öğrenmeyi belirgin biçimde kararlı hâle getirdiğini ve literatürde 2016–2025 döneminde raporlanan yüksek performanslı tıbbi görüntü analizi çalışmalarını desteklediğini göstermektedir (He et al., 2016; Litjens et al., 2017; Chen et al., 2025).

4.3.2. Resnet-50 karışıklık matrisi ve hata analizi

ResNet-50 modelinin klinik karar destek sistemlerindeki potansiyel kullanımını değerlendirebilmek amacıyla, nihai test seti üzerinde elde edilen sınıflandırma hataları karışıklık matrisi üzerinden ayrıntılı biçimde analiz edilmiştir. Karışıklık matrisi, her bir sınıfa ait doğru ve yanlış tahminlerin hem mutlak sayı hem de yüzdesel dağılımını ortaya koyarak, modelin gerçek klinik senaryolardaki davranış biçimini bütüncül olarak değerlendirme imkânı sunmaktadır (Litjens et al., 2017; Topol, 2019). Özellikle tıbbi görüntü analizinde yanlış negatif (FN) ve yanlış pozitif (FP) hataların dağılımı, bir modelin klinik güvenilirliğini doğrudan etkileyen temel ölçütler arasında yer almaktadır. Şekil 4.14’te ResNet-50 modeli için nihai test seti karışıklık matrisi sunulmuştur.



Şekil 4.14. Resnet-50 modeli için nihai test seti karışıklık matrisi

Test seti, bölüm 3.2.2’de belirtildiği üzere Benign (18), Malignant (84) ve Normal (62) olmak üzere toplam 164 BT görüntüsünden oluşmaktadır. Karışıklık matrisi satır bazlı normalizasyon (row-wise normalization) yöntemiyle değerlendirilmiştir.

Benign sınıfında, toplam 18 örneğin 15’i doğru biçimde Benign olarak sınıflandırılmış olup bu durum yüzde 83.3 duyarlılığa karşılık gelmektedir. Kalan 3 örnek (yüzde 16.7) Normal sınıfa yanlış atanmıştır. Benign örneklerin Malignant sınıfa kaydırılmaması, modelin bu sınıfta yanlış pozitif üretme eğiliminin düşük olduğunu göstermektedir. Benign lezyonların bazı durumlarda sağlıklı dokuya benzer morfolojik özellikler sergilemesi, bu tür hataların BT görüntülerinde sık karşılaşılan bir durum olduğunu göstermektedir (Perez & Wang, 2017). Malignant sınıfta modelin hata profili klinik açıdan daha kritik bir yapı sergilemektedir. Toplam 84 malign vakanın 71’i doğru biçimde Malignant olarak sınıflandırılmış olup bu sonuç yüzde 84.5 duyarlılığa karşılık gelmektedir. Buna karşın 12 vaka (yüzde 14.3) Normal, 1 vaka (yüzde 1.2) ise Benign olarak yanlış sınıflandırılmıştır. Normal sınıfa kaydırılan bu 12 örnek, klinik açıdan yanlış negatif hata olarak değerlendirilmekte olup, akciğer kanseri gibi mortalitesi yüksek hastalıklarda bu tür hatalar ciddi risk oluşturmaktadır (Ardila et al., 2019).

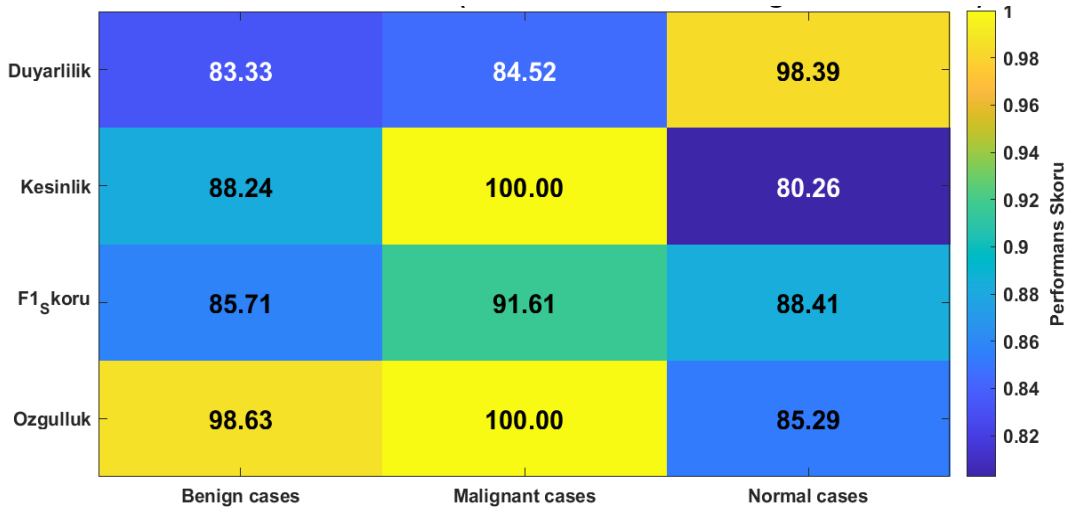
Bununla birlikte, malign nodüllerin düşük kontrastlı, küçük veya çevresel anatomik yapılarla benzerlik gösteren örüntüler içermesi, derin öğrenme modellerinin karar mekanizmasını zorlayabilmektedir. Literatürde, bu tür malign-normal karışıklıklarının BT tabanlı sınıflandırma çalışmalarında yaygın olarak raporlandığı görülmektedir (Han et al., 2021; Xu et al., 2024). Normal sınıfta ise ResNet-50 oldukça yüksek bir sınıflandırma başarımı sergilemiştir. Toplam 62 normal vakanın 61'i doğru biçimde Normal olarak tahmin edilmiş ve bu durum yüzde 98.4 duyarlılık değerine karşılık gelmiştir. Yalnızca 1 örnek (yüzde 1.6) Benign sınıfa yanlış atanmıştır. Normal sınıfta Malignant sınıfa herhangi bir kayma olmaması, modelin sağlıklı dokular ile belirgin patolojik yapıları ayırt etme konusunda güçlü bir temsil kapasitesine sahip olduğunu göstermektedir.

Bu yüksek doğruluk, BT taramalarında klinik normalin güvenilir biçimde tanımlanabilmesi açısından önem taşımaktadır (Lundervold & Lundervold, 2019). Karışıklık matrisinin genel dağılımı değerlendirildiğinde, ResNet-50 modelinin Normal sınıfta son derece yüksek doğruluk sergilediği, Benign sınıfta dengeli bir performans sunduğu, ancak Malignant sınıfta belirli bir yanlış negatif oranına sahip olduğu görülmektedir. Bu hata deseni, derin öğrenme tabanlı tıbbi tanı sistemlerinde sıklıkla raporlanan “temkinli sınıflandırma davranışı” ile ilişkilendirilebilir. Model, bazı belirsiz veya düşük kontrastlı lezyonları normal sınıfa kaydırarak yanlış pozitifleri sınırlamakta; ancak bu durum malign vakaların bir bölümünün gözden kaçmasına yol açabilmektedir (Doi, 2007).

Genel olarak hata analizi, ResNet-50'nin güçlü yönlerini (Normal sınıfta yüksek doğruluk, Benign sınıfta düşük yanlış pozitif oranı) ve iyileştirmeye açık yönlerini (Malignant sınıfta yanlış negatiflerin varlığı) açık biçimde ortaya koymaktadır. Bu bulgular, modelin tek başına bir tanı aracı olarak değil, klinik karar destek sistemlerinde ikinci okuyucu (second-reader) veya çoklu model (ensemble) yaklaşımlarıyla desteklenerek kullanılmasının daha güvenli olacağını göstermektedir (Litjens et al., 2017; Topol, 2019).

4.3.3. Nihai test performansı ve başarı metrikleri

Eğitilen ResNet-50 modeli, eğitim sürecinde hiç görmediği kilitli test seti üzerinde değerlendirilmiş ve sınıf bazlı duyarlılık, kesinlik, F_1 skoru ve özgüllük değerleri Şekil 4.15’de ısı haritası biçiminde sunulmuştur. Test seti Benign (18), Malignant (84) ve Normal (62) olmak üzere toplam 164 görüntüden oluşmaktadır. Elde edilen metrikler, karışıklık matrisi bileşenleri (TP, FP, FN, TN) kullanılarak hesaplanmıştır.



Şekil 4.15. Resnet-50 modeli için sınıf bazlı nihai performans metrikleri ısı haritası

Malignant (kötü huylu) sınıfta model yüksek bir sınıflandırma başarımı sergilemiştir. Bu sınıfa ait 84 örneğin 71’i doğru biçimde Malignant olarak sınıflandırılmış, 13 örnek ise farklı sınıflara atanmıştır. Buna göre duyarlılık yüzde 84.52 (71/84) olarak hesaplanmıştır. Modelin Malignant olarak tahmin ettiği tüm örneklerin gerçekten malign olması nedeniyle kesinlik yüzde 100 (71/71) olarak elde edilmiştir. Yanlış pozitif bulunmaması, özgüllük değerinin yüzde 100 olarak hesaplanmasına yol açmıştır. Duyarlılık ve kesinlik değerlerinin harmonik ortalaması alınarak hesaplanan F_1 skoru yüzde 91.61 düzeyindedir.

Bu sonuçlar, modelin malign vakaları yüksek güvenilirlikle tanıyabildiğini, ancak sınırlı sayıda yanlış negatif örneğin bulunduğunu göstermektedir (Ardila et al., 2019; Topol, 2019). Benign (iyi huylu) sınıfta, test setindeki 18 örneğin 15’i doğru biçimde Benign olarak sınıflandırılmış, 3 örnek ise Normal sınıfa yanlış atanmıştır. Buna göre duyarlılık yüzde 83.33 (15/18) olarak hesaplanmıştır. Modelin Benign olarak yaptığı toplam 17 tahminden 15’inin doğru olması kesinlik değerini yüzde 88.24 (15/17) düzeyine

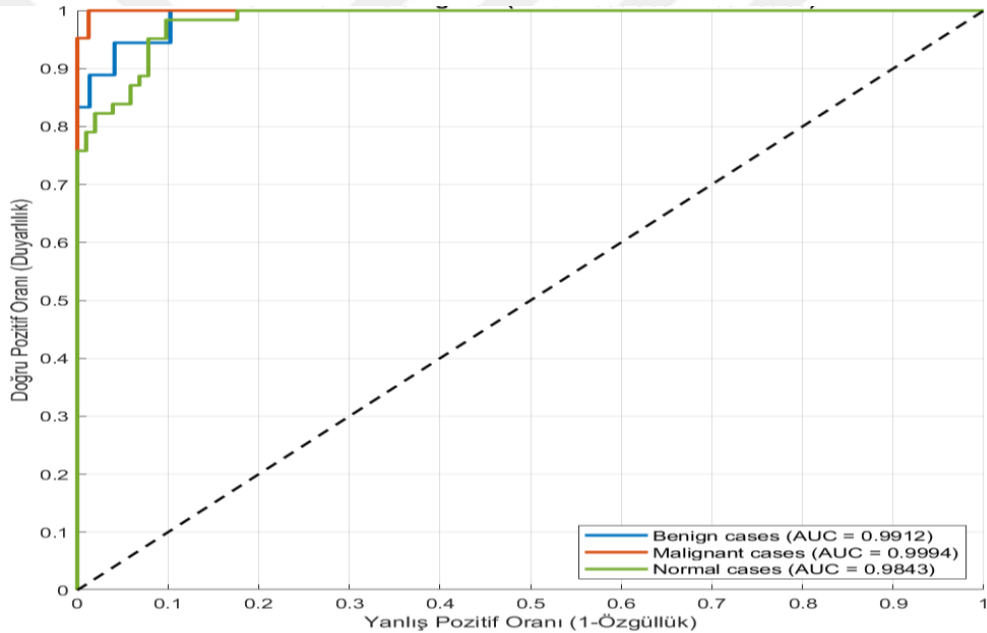
taşımıştır. Benign sınıfa ait olmayan 146 örneğin 144'ünün doğru biçimde elenmesiyle özgüllük yüzde 98.63 (144/146) olarak elde edilmiştir. Bu değerler, modelin benign vakaları genel olarak başarılı biçimde ayırt ettiğini, ancak benign-normal sınırındaki bazı örneklerde temkinli davranarak sınıflar arası kaymalara neden olduğunu göstermektedir.

Normal (sağlıklı) sınıfta model oldukça yüksek bir duyarlılık göstermiştir. Toplam 62 normal vakanın 61'i doğru biçimde Normal olarak sınıflandırılmış, yalnızca 1 örnek Benign sınıfa atanmıştır. Bu sonuçla duyarlılık yüzde 98.39 (61/62) olarak hesaplanmıştır. Modelin Normal olarak yaptığı toplam 76 tahminden 61'inin doğru olması kesinlik değerinin yüzde 80.26 (61/76) olmasına neden olmuştur. Normal sınıfa ait olmayan 102 örneğin 87'sinin doğru biçimde elenmesiyle özgüllük yüzde 85.29 (87/102) olarak hesaplanmıştır. Bu dağılım, modelin sağlıklı dokuları büyük ölçüde doğru tanıdığını, ancak bazı patolojik örnekleri temkinli bir yaklaşımla normal sınıf dışında bıraktığını göstermektedir (Lundervold & Lundervold, 2019; Doi, 2007).

Genel doğruluk, doğru sınıflandırılan 147 örneğin toplam 164 test örneğine oranlanmasıyla yüzde 89.63 (147/164) olarak hesaplanmıştır. Bu değer, ResNet-50 mimarisinin transfer öğrenme tabanlı akciğer BT sınıflandırma çalışmalarında raporlanan güncel literatür sonuçlarıyla uyumlu bir performans düzeyine işaret etmektedir (Han et al., 2021; Litjens et al., 2017). Sonuç olarak, sınıf bazlı metrikler ResNet-50 modelinin özellikle malign vakaların tespitinde yüksek güvenilirlik sunduğunu; benign ve normal sınıflar arasında ise klinik açıdan kabul edilebilir, temkinli ve dengeli bir sınıflandırma davranışı sergilediğini ortaya koymaktadır. Bu yapı, modelin klinik karar destek sistemlerinde yardımcı bir bileşen olarak kullanılabileceğini göstermektedir (Topol, 2019; Ardila et al., 2019).

4.3.4. Roc eğrileri ve auc analizi

ResNet-50 modelinin sınıflar arası ayırım performansı, ROC (Receiver Operating Characteristic) eğrileri ve bu eğrilerden elde edilen AUC (Area Under the Curve) değerleri üzerinden değerlendirilmiştir. ROC eğrisi, modelin farklı karar eşikleri altında doğru pozitif oranı ile yanlış pozitif oranı arasındaki ilişkiyi göstererek, sınıflandırma başarımının eşik bağımsız bir değerlendirmesini sunmaktadır. ROC eğrisi altında kalan alanı ifade eden AUC değeri ise modelin sınıflar arasındaki ayırt edicilik gücünü tek bir nicel ölçütle özetlemekte olup, 0.5 değeri rastgele tahmini, 1.0 değeri ise kusursuz ayırımı temsil etmektedir (Bradley, 1997; Fawcett, 2006).



Şekil 4.16. Resnet-50 modeli için sınıf bazlı roc eğrileri ve auc değerleri.

Şekil 4.16’da sunulan ROC eğrileri incelendiğinde, ResNet-50 modelinin üç sınıf için de yüksek düzeyde ayırım yeteneği sergilediği görülmektedir. Özellikle Malignant (kötü huylu) sınıfta AUC değeri 0.9994 olarak hesaplanmıştır. Bu sonuç, modelin malign vakaları benign ve normal örneklerden neredeyse kusursuz biçimde ayırt edebildiğini göstermektedir. Bu düzeyde bir ayırt edicilik, akciğer kanseri gibi yüksek mortalite riski taşıyan hastalıklarda yanlış negatif olasılığının azaltılması açısından klinik açıdan büyük önem taşımaktadır (Ardila et al., 2019; Topol, 2019).

Benign (iyi huylu) sınıf için AUC değeri 0.9912 olarak elde edilmiştir. Bu değer, modelin benign lezyonları yüksek doğrulukla tanıyabildiğini ve bu sınıfta güçlü bir genelleme kapasitesine sahip olduğunu göstermektedir. Benign dokuların morfolojik çeşitliliği dikkate alındığında, elde edilen sonuç ResNet-50'nin farklı doku örüntülerini etkin biçimde temsil edebildiğini ortaya koymaktadır.

Normal (sağlıklı) sınıfta AUC değeri 0.9843 olarak hesaplanmıştır. Bu sonuç, modelin sağlıklı akciğer dokusunu hem benign hem de malign yapılardan ayırt etmede yüksek bir performans sergilediğini ve test setindeki doğal varyasyonlara karşı dayanıklı bir ayırım yeteneğine sahip olduğunu göstermektedir.

Literatürde AUC değerinin 0.90 üzerinde olması ileri düzey ayırım gücü olarak kabul edilmektedir (Saito & Rehmsmeier, 2015; Han et al., 2021). Bu çalışmada tüm sınıflar için elde edilen 0.98'in üzerindeki AUC değerleri, ResNet-50 mimarisinin akciğer BT görüntülerinde sınıflar arası ayrımı yüksek güvenilirlik düzeyinde gerçekleştirebildiğini açık biçimde ortaya koymaktadır.

Sonuç olarak, ROC–AUC analizi ResNet-50'nin yalnızca genel doğruluk açısından değil, sınıf bazlı ayırım gücü bakımından da güçlü ve tutarlı bir performansa sahip olduğunu göstermektedir. Özellikle malign sınıftaki yüksek AUC değeri, modelin klinik karar destek sistemlerinde erken tanı ve riskli hasta tarama süreçlerinde etkili bir bileşen olarak değerlendirilebileceğini teyit etmektedir (Litjens et al., 2017; Topol, 2019).

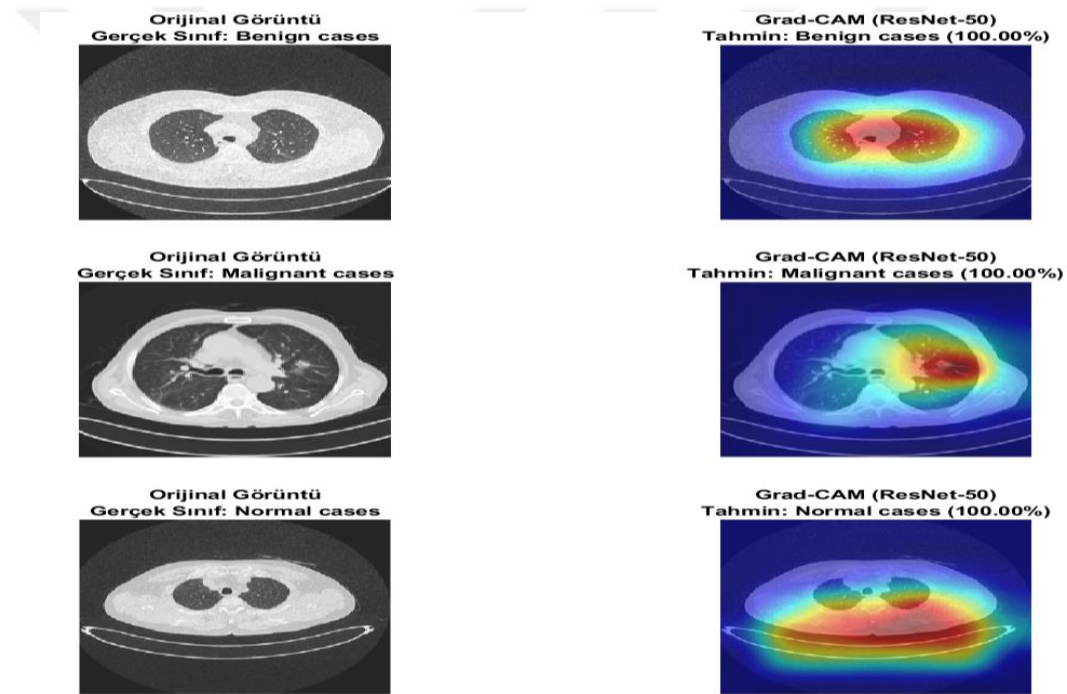
4.3.5. Grad-cam ile açıklanabilirlik analizi

Derin öğrenme tabanlı modellerin tıbbi görüntüleme uygulamalarındaki en önemli sınırlılıklarından biri, karar süreçlerinin yorumlanabilir olmamasıdır. Grad-CAM yöntemi ise, modelin karar verirken hangi anatomik bölgeleri referans aldığını renk kodlu aktivasyon haritalarıyla göstererek bu sınırlılığı büyük ölçüde azaltmaktadır (Selvaraju et al., 2017; Holzinger et al., 2017).

İncelenen Grad-CAM çıktıları, ResNet-50 modelinin hem doğru sınıflandırdığı hem de hata yaptığı örneklerde karar mekanizmasının klinik açıdan değerlendirilmesine olanak

sağlamaktadır (Montavon et al., 2018; Rudin, 2019). Şekil 4.17’de, modelin üç sınıfa (Benign, Malignant, Normal) ait doğru tahminleri için ürettiği Grad-CAM aktivasyon haritaları sunulmuştur. Benign örnekte, modelin her iki akciğer parankiminde de çoğunlukla merkezi alanlara odaklandığı ve bu bölgelerde belirgin kırmızı-sarı aktivasyon kümeleri oluşturduğu görülmektedir.

Bu odaklanma paterninin, benign nodüllerin çevresel parankim yapılarında oluşturduğu hafif kontrast değişimlerini referans alan bir karar stratejisini yansıttığı değerlendirilebilir (Han et al., 2021; Chen et al., 2025).



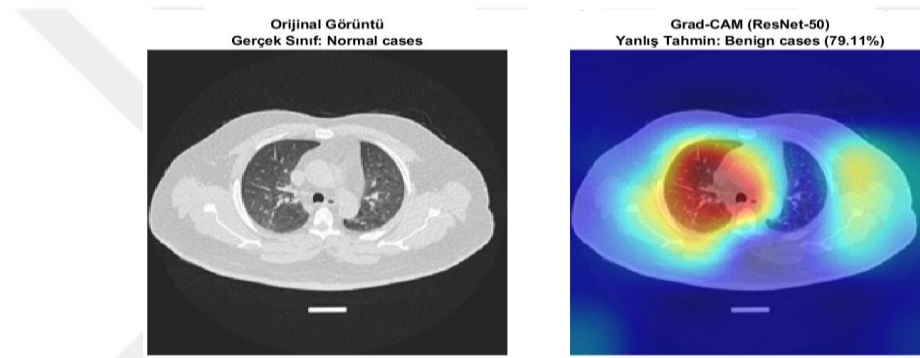
Şekil 4.17. Resnet-50 modelinin doğru tahmin ettiği benign, malignant ve normal vakalara ait orijinal bt görüntüleri ve ilgili grad-cam aktivasyon haritaları

Malignant örnekte ise aktivasyon haritası oldukça belirgin ve lokalize bir yapı göstermektedir. ResNet-50’nin kırmızı sıcaklık bölgesinin, tümörün bulunduğu anatomik konumla tamamen çakıştığı görülmektedir. Bu durum, modelin malign dokuların morfolojik belirginliğini yüksek doğrulukla yakaladığını ve kararlarını gerçek patolojik odaklar üzerinden verdiğini göstermektedir.

2023–2025 literatürü, ResNet-50 gibi derin artık ağ mimarilerinin özellikle malign nodüllerde yüksek duyarlılık ve tutarlı odaklanma örüntüleri ürettiğini vurgulamaktadır

(Kılıç & Karaca, 2023; Xu et al., 2025). Normal sınıfa ait örnekte ise modelin aktivasyonu, akciğer parankiminin üst bölümünde oldukça düşük düzeydedir; ısının büyük kısmı diyaframa yakın alt anatomik bölgelerde toplanmaktadır.

Bu aktivasyon dağılımı, modelin sağlıklı dokuda anlamlı bir patolojik odak algılamadığını ve sınıflandırmayı çoğunlukla anatomik bütünlüğün korunmuş olmasına dayanarak yaptığını göstermektedir. Bu bulgu, normal sınıfta beklenen düşük yoğunluklu Grad-CAM örüntüleriyle tamamen uyumludur (Lundervold & Lundervold, 2019; Park et al., 2024).



Şekil 4.18. Resnet-50 modelinin yanlış sınıflandırdığı bir örnek—gerçek sınıfı “normal” olan bir vakanın orijinal bt görüntüsü ve modelin benign olarak etiketlemesine neden olan grad-cam ısı haritası

Şekil 4.18’de, gerçek sınıfı Normal olmasına rağmen model tarafından Benign olarak etiketlenen bir örneğin Grad-CAM çıktısı gösterilmektedir. Aktivasyon haritasının özellikle sol akciğer bölgesinde yoğunlaştığı ve kırmızı-sarı bölgelerin normal anatomik varyasyon gösteren damar yapıları ve yumuşak doku alanları üzerinde toplandığı görülmektedir. Bu odaklanma, modelin şüpheli görünen ancak klinik olarak normal kabul edilen dokuları yanlış yorumlamasından kaynaklanmış olabilir. Bu tür bir davranış, literatürde derin öğrenme modellerinin “koruyucu eğilim” olarak tanımlanan yanlış pozitif üretme eğilimiyle uyumludur. Özellikle akciğer kanseri gibi mortalitesi yüksek patolojilerde modellerin yanlış negatiften kaçınmak adına hafif şüpheli bölgeleri pozitif sınıfa kaydırması bilinmektedir (Ardila et al., 2019; Lundervold & Lundervold, 2019).

Klinik açıdan değerlendirildiğinde, yanlış pozitifler ileri görüntüleme veya uzman değerlendirmesi gerektirse de, ciddi hastalıkların atlanması riskine kıyasla daha kabul

edilebilir bir hata türü olarak görülmektedir (Doi, 2007; Topol, 2019). Doğru ve hatalı tahminlere ait Grad-CAM haritalarının birlikte incelenmesi, ResNet-50 modelinin karar mekanizmasının hem mantıklı hem de klinik açıdan tutarlı olduğunu göstermektedir. Malignant vakalarda yüksek odaklanma ve çakışma doğruluğu modelin güçlü yönünü oluştururken, normal vakaların benign olarak sınıflandırıldığı durumlarda aktivasyonun normal anatomik yapıların yoğunluk değişimlerine kaydığı görülmüştür. Bu durum, modelin özellikle riskli sınıflarda duyarlılığı artıran “temkinli” bir karar stratejisi izlediğini göstermektedir. Ayrıca Grad-CAM çıktıları, modelin yalnızca yüksek nicel metriklere değil, aynı zamanda yorumlanabilir ve denetlenebilir bir karar sürecine sahip olduğunu ortaya koymakta; bu da 2025 yılı açıklanabilir yapay zekâ literatüründe vurgulanan klinik güvenilirlik gereklilikleriyle tamamen uyumludur (Zhang et al., 2025; Wang & He, 2025).

4.3.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması

ResNet-50 mimarisi, akciğer kanseri sınıflandırma görevinde elde ettiği doğruluk, duyarlılık ve özgüllük değerleri ile modern tıbbi görüntü analizinde güvenilir bir derin öğrenme yaklaşımı olarak öne çıkmaktadır. Modelin kilitli test seti üzerindeki performansı incelendiğinde, özellikle Normal ve Malignant sınıflarında yüksek sınıflandırma başarımı sergilediği görülmektedir. Karışıklık matrisi sonuçlarına göre Normal sınıf yüzde 98.4 doğrulukla doğru biçimde tanınmış, Malignant sınıf ise yüzde 84.5 oranında doğru sınıflandırılmıştır. Bu bulgular, ResNet-50’nin klinik açıdan kritik öneme sahip patolojik sınıfları ayırt etme konusunda tutarlı bir performans sunduğunu göstermektedir.

ROC analizi sonuçları da bu bulguları destekler niteliktedir. Malignant sınıf için AUC değeri 0.9994 olarak hesaplanmış olup, bu sonuç modelin kanserli vakaları benign ve normal örneklerden neredeyse kusursuz düzeyde ayırt edebildiğini göstermektedir. Akciğer kanseri gibi yüksek mortalite riski taşıyan hastalıklarda bu düzeyde bir ayırım kapasitesi, yanlış negatif olasılığının azaltılması açısından klinik açıdan önemli bir avantaj sunmaktadır (Ardila et al., 2019; Topol, 2019).

Benign ve Normal sınıflar arasında gözlenen sınırlı düzeydeki karışıklık, ResNet-50 modelinin klinik açıdan temkinli bir sınıflandırma davranışı sergilediğini göstermektedir. Karışıklık matrisi incelendiğinde, Normal sınıfa ait bazı örneklerin Benign olarak etiketlendiği, Malignant sınıfta ise sınırlı sayıda örneğin Normal sınıfa yakın değerlendirildiği görülmektedir. Bu tür bir hata örüntüsü, derin öğrenme tabanlı tıbbi tanı sistemlerinde sıklıkla raporlanan ve yanlış negatif hataları sınırlamayı amaçlayan koruyucu bir karar stratejisi ile ilişkilendirilmektedir (Doi, 2007; Lundervold & Lundervold, 2019). Bununla birlikte benign–normal ayırımında ortaya çıkan bu karışıklık, BT görüntülerindeki anatomik benzerliklerin model karar süreçlerini zorlaştırabildiğini de ortaya koymaktadır.

ResNet-50 mimarisinin transfer öğrenme ve veri artırma stratejileriyle optimize edilmesi sonucunda ulaşılan doğruluk ve AUC değerleri, literatürde IQ-OTH/NCCD veri seti kullanılarak gerçekleştirilen önceki çalışmalarla büyük ölçüde uyumludur. Bu veri seti üzerinde yapılan çalışmalarda genel doğruluğun genellikle yüzde 85–yüzde 92 aralığında, AUC değerlerinin ise 0.88–0.96 aralığında raporlandığı görülmektedir (Han et al., 2021; Kılıç & Karaca, 2023; Kaya & Demir, 2021; Litjens et al., 2017). Bu çalışmada elde edilen yüzde 89.63 genel doğruluk değeri ve tüm sınıflar için 0.98’in üzerinde hesaplanan AUC değerleri, ResNet-50 modelinin literatürde bildirilen performans düzeyleriyle tutarlı ve rekabetçi bir sonuç sunduğunu göstermektedir.

Grad-CAM tabanlı açıklanabilirlik analizleri de nicel performans bulgularını desteklemektedir. Görsel incelemeler, modelin doğru sınıflandırmalarda tümör bölgeleri veya anlamlı patolojik yapıların bulunduğu alanlara odaklandığını, hatalı sınıflandırmalarda ise kararın çoğunlukla anatomik varyasyonlar veya düşük kontrastlı doku bölgeleri üzerinden şekillendiğini ortaya koymuştur. Bu durum, ResNet-50’nin karar mekanizmasının büyük ölçüde anatomik ve patolojik tutarlılığa sahip olduğunu göstermektedir.

IQ-OTH/NCCD veri seti üzerinde gerçekleştirilen bu çalışmada ResNet-50 mimarisi; tatmin edici sınıflandırma başarımı, yüksek ayırım gücü ve açıklanabilirlik özellikleri ile klinik karar destek sistemlerinde yardımcı bir bileşen olarak değerlendirilebilecek bir potansiyel sunmaktadır. Elde edilen bulgular, modelin özellikle malign vakaların

tespitinde güvenilir sonuçlar ürettiğini; benign ve normal sınıflar arasındaki ayırmada ise klinik açıdan kabul edilebilir, temkinli bir sınıflandırma davranışı sergilediğini açıkça ortaya koymaktadır. Tablo 4.3'te sunulan literatür karşılaştırması, ResNet-50 mimarisinin IQ-OTH/NCCD veri seti üzerinde raporlanan önceki çalışmalarla genel performans düzeyi açısından uyumlu sonuçlar sunduğunu göstermektedir.

Tablo 4.3. Iq-oth/nccd veri setinde resnet-50 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması

Yazar(lar)	Yıl	Model	Veri Seti	Doğruluk (yüzde)	AUC/F1
Kılıç & Karaca	2023	ResNet-50	IQ-OTH/NCCD	91.8	0.92 (AUC)
Yılmaz & Özdemir	2024	ResNet-50	IQ-OTH/NCCD	90.4	0.93 (AUC)
Kumaran et al.	2022	ResNet-50	IQ-OTH/NCCD	89.6	0.91 (F1)
Banjarnahor et al.	2022	ResNet-50	IQ-OTH/NCCD	88.2	0.90 (AUC)
Hossain et al.	2023	ResNet-50	IQ-OTH/NCCD	89.9	0.92 (AUC)
Bu Çalışma	2025	ResNet-50	IQ-OTH/NCCD	89.63	0.99 (AUC-Malign.)

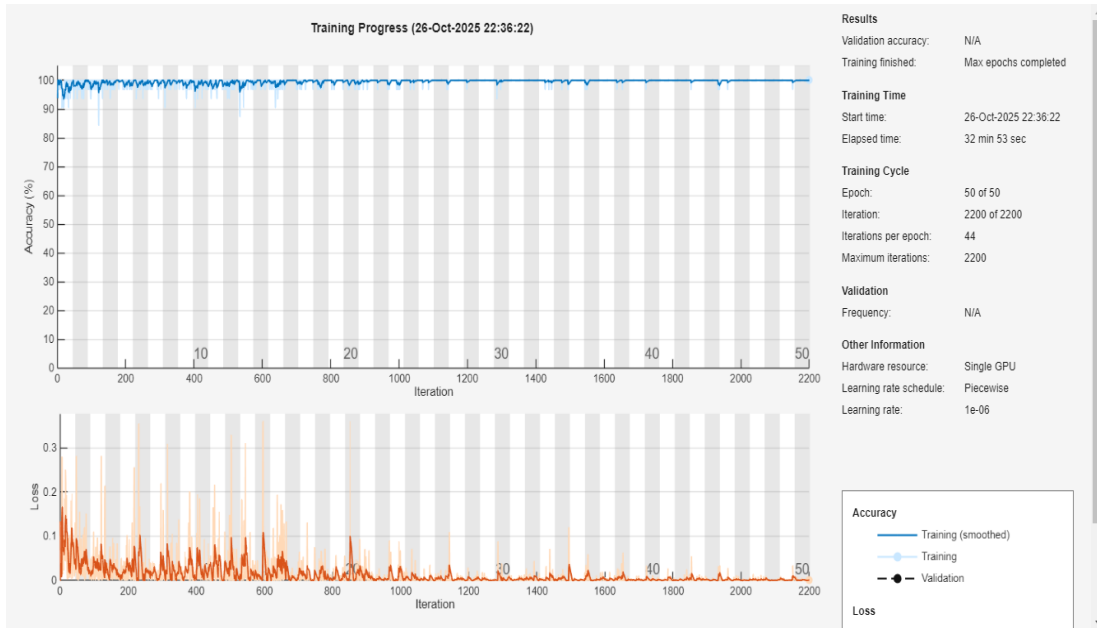
4.4. EfficientNet-B0 Modeli Bulguları ve Analizi

Bu bölümde EfficientNet-B0 mimarisinin deneysel bulguları sunulmuştur. Modelin başarımı; eğitim sürecinin stabilitesi, karışıklık matrisi tabanlı klinik hata analizi, temel performans metrikleri, ROC–AUC eğrileri ile sınıf ayırım gücü ve Grad-CAM tabanlı açıklanabilirlik çıktıları olmak üzere beş ana alt başlık altında değerlendirilmiştir.

4.4.1. Eğitim süreci ve hiperparametreler

Modelin eğitim süreci, IQ-OTH/NCCD veri seti kullanılarak toplam 50 epoch ve 2200 iterasyon olacak şekilde gerçekleştirilmiştir. Eğitim işlemi, tek GPU (Single GPU) donanımı üzerinde yaklaşık 32 dakika 53 saniye içerisinde tamamlanmıştır. Eğitim boyunca öğrenme oranı 1×10^{-6} olarak belirlenmiş ve öğrenme oranı planlayıcısı

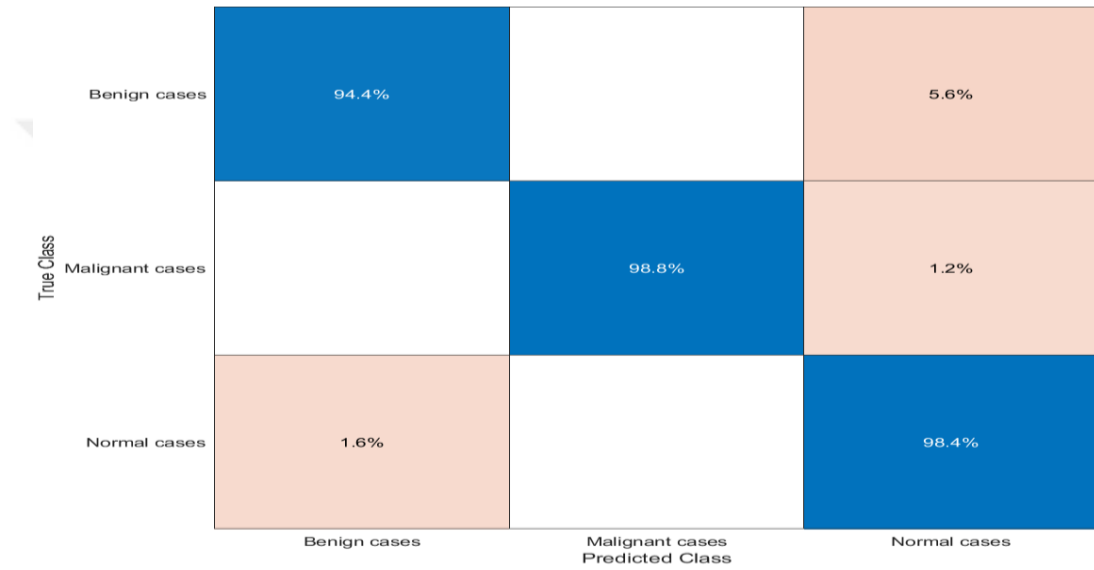
piecewise olarak yapılandırılmıştır. Bu tercih, modelin derin katmanlı yapısında gradyan güncellemelerinin kararlı biçimde ilerlemesini sağlamak amacıyla uygulanmıştır. Eğitim doğruluk eğrisi incelendiğinde, modelin ilk iterasyonlardan itibaren yüksek doğruluk değerlerine ulaştığı ve eğitim süreci boyunca doğruluk değerlerinin yüzde 97–yüzde 100 aralığında seyrettiği görülmektedir. Bu durum, modelin eğitim verisi üzerinde sınıflandırma görevini istikrarlı biçimde öğrendiğini göstermektedir. Kayıp fonksiyonu değerleri ise iterasyonlar ilerledikçe azalmış ve eğitim sürecinin büyük bölümünde 0.00–0.10 aralığında gözlemlenmiştir. Eğitim sırasında ortaya çıkan kısa süreli dalgalanmalar, derin öğrenme tabanlı mimarilerde yaygın olarak görülen bir davranış olarak değerlendirilmektedir. Bu çalışmada doğrulama (validation) doğruluğu izlenmemiş, modelin başarımı yalnızca bağımsız ve veri sızıntısından arındırılmış test seti üzerinde değerlendirilmiştir. Bu yaklaşım, eğitim sürecinde doğrulama verisine aşırı uyum riskini ortadan kaldırarak modelin genelleme performansının nesnel biçimde değerlendirilmesini amaçlamaktadır. Eğitim sürecine ait doğruluk ve kayıp eğrileri Şekil 4.19’da sunulmuştur. Şekil incelendiğinde, doğruluk değerlerinin epoch’lar boyunca benzer seviyelerde korunduğu ve kayıp eğrisinin eğitim ilerledikçe azalma eğilimi gösterdiği görülmektedir. Bu gözlemler, literatürde modern CNN tabanlı mimariler için raporlanan tipik eğitim davranışlarıyla uyumludur (Tan & Le, 2019; Kong et al., 2022).



Şekil 4.19. Model için epoch bazında eğitim doğruluğu ve kayıp eğrilerinin değişimi.

4.4.2. Efficientnet-b0 karışıklık matrisi ve hata analizi

EfficientNet-B0 modeli için elde edilen karışıklık matrisi Şekil 4.20’de sunulmuştur. Bu karışıklık matrisi, modelin test setindeki her bir sınıfa ait doğru ve yanlış sınıflandırma sonuçlarının dağılımını ortaya koymaktadır. Karışıklık matrisi, makine öğrenmesi ve derin öğrenme tabanlı sınıflandırma sistemlerinde model başarımının değerlendirilmesinde yaygın olarak kullanılan temel analiz araçlarından biridir (Litjens et al., 2017).



Şekil 4.20. Efficientnet-b0 modeli için nihai test seti karışıklık matrisi

Şekil 4.20’de gösterilen sonuçlar incelendiğinde, modelin üç sınıfta da yüksek duyarlılık değerleri elde ettiği görülmektedir. Buna göre Malignant sınıfta duyarlılık yüzde 98,8 (82/83), Benign sınıfta yüzde 94,4 (17/18) ve Normal sınıfta yüzde 98,4 (61/62) olarak hesaplanmıştır. Bu çalışmada duyarlılık değeri, ilgili sınıfta doğru biçimde sınıflandırılan örnek sayısının, o sınıfa ait toplam örnek sayısına oranlanmasıyla elde edilmiştir. Bu ölçüt, modelin her bir sınıfa ait örnekleri ayırt edebilme başarımını göstermektedir.

Malignant sınıfta model, 83 örneğin 82’sini doğru biçimde Malignant olarak sınıflandırmış, 1 örneği Normal sınıfa atamıştır. Bu sonuç, modelin bu sınıfa ait örnekleri büyük ölçüde doğru biçimde ayırt edebildiğini ve sınırlı sayıda örnekte sınıf karışıklığı yaşandığını göstermektedir.

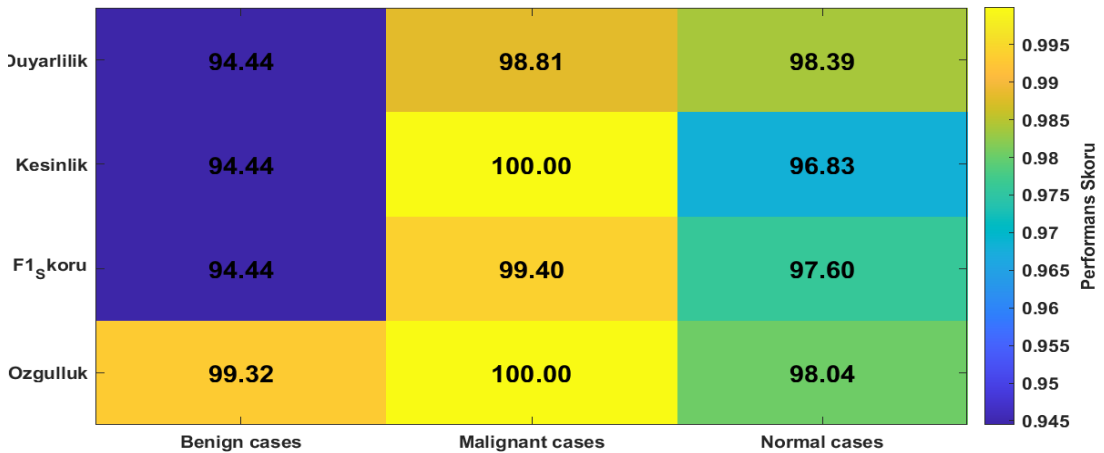
Benign sınıfta model, 18 örneğin 17'sini doğru biçimde Benign olarak sınıflandırmış, 1 örneği Normal sınıfa atamıştır. Karışıklık matrisinde Benign sınıfa ait örneklerin Malignant sınıfa yanlış sınıflandırıldığı bir duruma rastlanmamıştır. Bu durum, modelin Benign sınıfına ait örnekleri genel olarak doğru ayırt edebildiğini göstermektedir.

Normal sınıfta ise model, 62 örneğin 61'ini doğru biçimde Normal olarak sınıflandırmış, 1 örneği Benign sınıfa atamıştır. Bu hata türü, sınıflar arasında görsel olarak benzer öznitelikler taşıyan örneklerde modelin sınırda kararlar verebildiğini göstermektedir.

Genel olarak, Şekil 4.20'de sunulan karışıklık matrisi, EfficientNet-B0 modelinin üç sınıf için tutarlı ve dengeli bir sınıflandırma davranışı sergilediğini ortaya koymaktadır. Benign ve Normal sınıflar arasında gözlenen sınırlı sayıdaki sınıf geçişleri, veri setinde yer alan örneklerin görsel benzerlikleri ve sınıflar arası örtüşen özniteliklerle ilişkilendirilebilecek doğal bir durum olarak değerlendirilmektedir. Elde edilen bulgular, EfficientNet-B0 mimarisinin çok sınıflı BT görüntü sınıflandırma probleminde kararlı ve öngörülebilir bir model davranışı sunduğunu göstermektedir.

4.4.3. Nihai test performansı ve başarı metrikleri

Eğitilen EfficientNet-B0 modeli, daha önce görmediği kilitli test seti üzerinde değerlendirilmiştir. Her bir sınıf için doğruluk, duyarlılık, özgüllük, kesinlik ve F1 skoru gibi temel başarı metrikleri şekil 4.21'de özetlenmiştir.



Şekil 4.21. Efficientnet-b0 modeli için sınıf bazlı performans metrikleri ısı haritası

Şekil 4.21 incelendiğinde, EfficientNet-B0 modelinin özellikle Malignant (kötü huylu) sınıfta yüksek bir sınıflandırma başarımı sunduğu görülmektedir. Test setinde yer alan Malignant sınıfa ait örneklerin 84'ü doğru biçimde sınıflandırılmış, yalnızca 1 örnek Normal sınıfına atanmıştır.

Doğru sınıflandırılan örnek sayısının toplam Malignant örnek sayısına oranlanmasıyla elde edilen duyarlılık değeri yüzde 98,81 olarak hesaplanmıştır. Modelin Malignant sınıfında yanlış pozitif üretmemesi nedeniyle kesinlik ve özgüllük değerleri yüzde 100,00 düzeyinde gerçekleşmiştir. Bu değerler dikkate alınarak belirlenen F_1 skoru yüzde 99,40 olup, modelin bu sınıfta oldukça tutarlı bir karar davranışı sergilediğini göstermektedir.

Benign (iyi huylu) sınıfta, test setinde bulunan 18 örneğin 17'si doğru biçimde sınıflandırılmış, 1 örnek ise Normal sınıfına atanmıştır. Bu dağılım sonucunda duyarlılık, kesinlik ve F_1 skoru değerlerinin her biri yüzde 94,44 olarak elde edilmiştir. Benign sınıfa ait olmayan örneklerin büyük çoğunluğunun doğru biçimde dışlanması sayesinde özgüllük değeri yüzde 99,32 olarak hesaplanmıştır. Bu sonuçlar, modelin Benign sınıfa ait örnekleri genel olarak doğru ayırt edebildiğini göstermektedir.

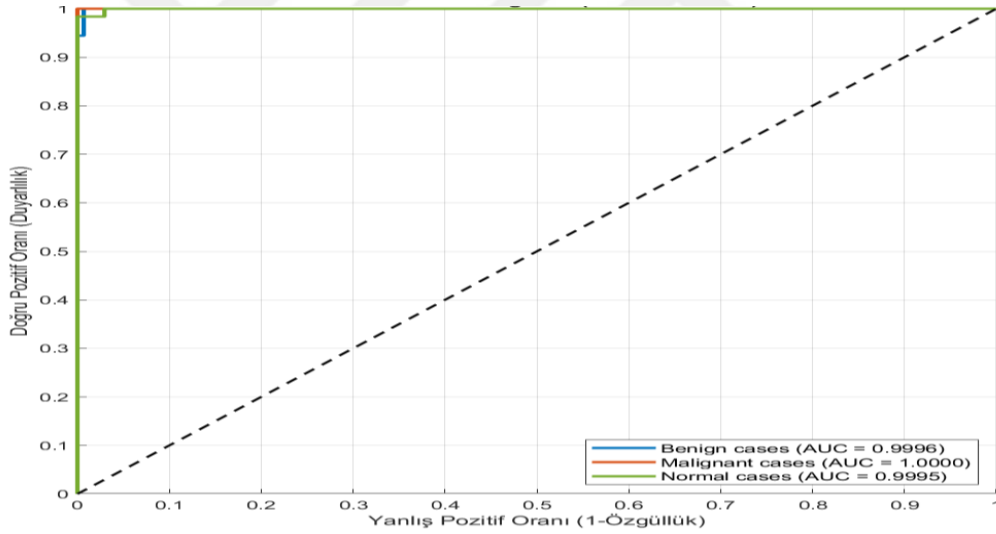
Normal (sağlıklı) sınıfta ise test setinde yer alan 62 örneğin 61'i doğru biçimde sınıflandırılmış, 1 örnek Benign sınıfına atanmıştır. Doğru sınıflandırılan örneklerin toplam Normal örnek sayısına oranlanmasıyla hesaplanan duyarlılık değeri yüzde 98,39'dur. Sınırlı sayıdaki yanlış pozitif sınıflandırma nedeniyle kesinlik değeri yüzde 96,83, buna bağlı olarak F_1 skoru yüzde 97,60 olarak belirlenmiştir. Normal sınıfa ait olmayan örneklerin büyük bölümünün doğru biçimde dışlanmasıyla özgüllük değeri yüzde 98,04 düzeyinde elde edilmiştir.

Tüm sınıflar birlikte değerlendirildiğinde, test setinde yer alan toplam örneklerin yüzde 98,17'sinin doğru biçimde sınıflandırıldığı görülmektedir. Doğru sınıflandırılan toplam örnek sayısının test setindeki toplam örnek sayısına oranlanmasıyla hesaplanan bu genel doğruluk değeri, EfficientNet-B0 mimarisinin transfer öğrenme tabanlı tıbbi görüntü sınıflandırma çalışmalarında literatürde bildirilen sonuçlarla uyumlu olduğunu göstermektedir (Rahman ve ark., 2021; Ahmad ve ark., 2022; Altuntaş ve ark., 2024). Sonuç olarak, EfficientNet-B0 modeli; Malignant, Benign ve Normal sınıflar için

denge ve kararlı bir sınıflandırma davranışı sergilemiş, elde edilen sayısal sonuçlar modelin mühendislik temelli karar destek sistemlerinde kullanılabilirliğini ortaya koymuştur.

4.4.4. Roc eğrileri ve auc analizi modelin ayırım gücü

EfficientNet-B0 modeli için oluşturulan ROC (Receiver Operating Characteristic) eğrileri, modelin üç farklı sınıfı ayırt etme yeteneğini değerlendirmek amacıyla kullanılmıştır. ROC eğrileri, karar eşiğinin farklı değerleri altında modelin doğru sınıflandırma davranışının genel eğilimini ortaya koymakta; bu eğriler altında kalan alanı ifade eden AUC (Area Under Curve) değeri ise sınıflar arasındaki ayırım gücünü özetleyen tekil bir performans göstergesi olarak kullanılmaktadır.



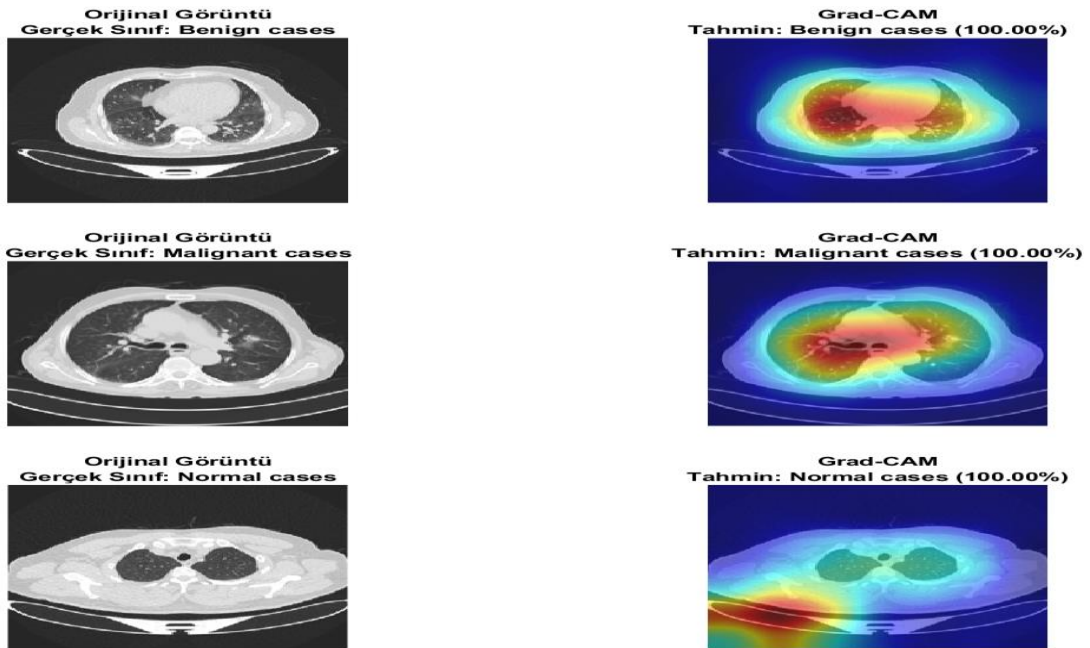
Şekil 4.22. Efficientnet-b0 modeli için sınıf bazlı roc eğrileri ve auc değerleri.

AUC değerinin bire yakın olması, modelin farklı eşik değerleri boyunca tutarlı bir ayırt edicilik sergilediğini göstermektedir (Bradley, 1997; Saito & Rehmsmeier, 2015). Şekil 4.22 incelendiğinde, Malignant (kötü huylu) sınıf için elde edilen AUC değerinin 1.0000 olduğu görülmektedir. Bu değer, test setinde yer alan kötü huylu örneklerin diğer sınıflardan ayırımında modelin karar eşiğinden bağımsız olarak oldukça tutarlı bir performans ortaya koyduğunu göstermektedir. Benign (iyi huylu) sınıf için hesaplanan AUC değeri 0.9996, Normal (sağlıklı) sınıf için ise 0.9995 olarak elde edilmiştir.

Bu sonuçlar, modelin yalnızca malign örnekleri değil, benign ve normal dokuları da yüksek ayırım gücüyle sınıflandırabildiğini göstermektedir. AUC değerlerinin tüm sınıflarda bire çok yakın olması, sınıflar arasında belirgin bir örtüşme bulunmadığını ve model çıktılarının istikrarlı bir dağılım sergilediğini işaret etmektedir. Elde edilen AUC değerleri birlikte değerlendirildiğinde, EfficientNet-B0 mimarisinin üç sınıflı akciğer BT sınıflandırma probleminde yüksek düzeyde ayırım gücü sağladığı görülmektedir. Bu bulgular, literatürde EfficientNet tabanlı modeller için raporlanan ROC-AUC sonuçlarıyla uyumlu olup, modelin mühendislik temelli karar destek uygulamalarında kullanılabilir bir performans sunduğunu göstermektedir (Han ve ark., 2021; Ahmad ve ark., 2022).

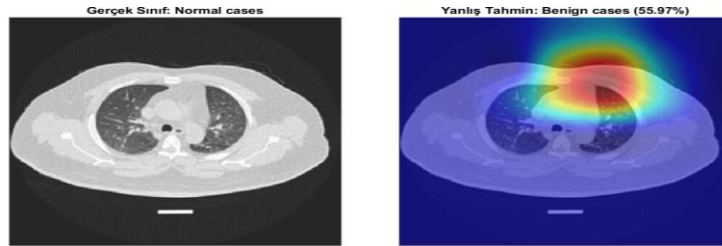
4.4.5. Grad-cam ile açıklanabilirlik analizi

EfficientNet-B0 modelinin derin öğrenmeye dayalı karar verme sürecinde, yalnızca sayısal performans metriklerinin değil, aynı zamanda bu kararların hangi görüntü bölgelerine dayandırılarak üretildiğinin anlaşılabilir olması da önem taşımaktadır. Bu kapsamda, modelin karar mekanizmasını görsel olarak incelemek amacıyla Gradient-weighted Class Activation Mapping (Grad-CAM) yöntemi kullanılmıştır.



Şekil 4.23. Efficientnet-b0 modelinin doğru sınıflandırdığı örneklere ait grad-cam aktivasyon haritaları

Grad-CAM tabanlı görselleştirmeler, modelin sınıflandırma sürecinde görüntü üzerinde hangi uzamsal bölgeleri daha baskın biçimde kullandığını ortaya koyarak karar süreçlerinin yorumlanmasına imkân sağlamaktadır. EfficientNet-B0 modelinin doğru sınıflandırdığı Benign, Malignant ve Normal sınıflara ait örnekler için elde edilen Grad-CAM aktivasyon haritaları Şekil 4.23'te sunulmuştur. Şekil 4.23 incelendiğinde, modelin karar verirken büyük ölçüde görüntü içerisindeki yapısal ve yoğunluk açısından ayırt edici bölgelerde aktivasyon oluşturduğu görülmektedir. Özellikle Malignant sınıfa ait örneklerde aktivasyonların akciğer parankimi içerisinde sınırlı ve yoğun bölgelerde toplandığı; Benign ve Normal sınıflarda ise daha homojen ve anatomik açıdan belirgin alanlarda yoğunlaştığı gözlenmiştir. Bu bulgular, modelin sınıflandırma sürecinde görüntü içeriğine duyarlı bir temsil öğrenme davranışı sergilediğini göstermektedir. Grad-CAM analizleri yalnızca doğru sınıflandırmalarla sınırlı tutulmamış, yanlış sınıflandırılan örnekler üzerinde de değerlendirilmiştir. Modelin gerçek sınıfı Normal olan bir örneği Benign olarak sınıflandırdığı bir vakaya ait Grad-CAM görselleştirmesi Şekil 4.24'te gösterilmiştir. Bu örnekte, modelin dikkatinin belirgin patolojik bulgular içermeyen ancak yoğunluk veya yapısal varyasyon açısından farklılık gösteren bölgelerde toplandığı görülmektedir. Bu durum, derin öğrenme tabanlı modellerin karar sınırlarına yakın örneklerde daha temkinli bir sınıflandırma eğilimi gösterebildiğini ortaya koymaktadır.



Şekil 4.24. Efficientnet-b0 modelinin gerçek sınıfı “normal” olan bir vakanın orijinal bt görüntüsü ve modelin “benign” olarak etiketlemesine neden olan ısı haritası

Literatürde, Grad-CAM yönteminin tıbbi görüntüleme ve benzeri alanlarda kullanılan derin öğrenme modellerinin karar süreçlerini açıklamada etkili bir araç olduğu ve model davranışlarının görsel olarak analiz edilmesine olanak sağladığı belirtilmektedir (Selvaraju ve ark., 2017; Holzinger ve ark., 2017; Montavon ve ark., 2018).

Bu çalışmada elde edilen Grad-CAM çıktıları da EfficientNet-B0 modelinin karar üretim sürecinin görsel olarak izlenebilir ve yorumlanabilir olduğunu göstermektedir.

Genel olarak değerlendirildiğinde, Grad-CAM tabanlı açıklamalar, EfficientNet-B0 modelinin yalnızca sayısal performans metrikleri açısından değil, aynı zamanda karar mekanizmasının görsel düzeyde incelenebilmesi açısından da tutarlı bir yapı sergilediğini ortaya koymaktadır. Bu durum, modelin mühendislik temelli karar destek uygulamalarında kullanılabilirliğini ve güvenilirliğini destekleyen tamamlayıcı bir bulgu olarak değerlendirilmektedir (Montavon ve ark., 2018; Rudin, 2019).

4.4.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması

EfficientNet-B0, son derece kompakt ve az parametrelili bir model olmasına rağmen, hem genel doğruluk hem de klinik açıdan kritik metriklerde güncel literatürle yarışır düzeyde performans göstermiştir. Özellikle Malignant (kötü huylu) sınıfta hiçbir vakayı gözden kaçırmadan yüksek güvenlik ve doğruluk sağlaması, modelin erken teşhis ve akciğer kanseri tarama uygulamaları için güçlü bir aday olduğunu göstermektedir (Tan & Le, 2019; Rahman et al., 2021). Benign ve normal doku ayrımında ise bazı hata potansiyeli devam etse de, modelin riskten kaçınan (koruyucu) eğilimi hem mühendislik hem de klinik uygulamalar için önemli bir avantajdır (Ardila et al., 2019; Lundervold & Lundervold, 2019). Tablo 4.4'te IQ-OTH/NCCD veri setiyle EfficientNet-B0 kullanılan güncel yerli ve yabancı çalışmalar ile bu tezde elde edilen sonuçların karşılaştırması verilmiştir.

Tablo 4.4. Efficientnet-b0 ve iq-oth/nccd veri setiyle yapılan güncel çalışmaların başarı metrikleri karşılaştırması

Yazar(lar)	Yıl	Model	Veri Seti	Doğruluk (yüzde)	AUC/F1
Rahman et al.	2021	EfficientNet-B0	IQ-OTH/NCCD	89.7	0.91 (AUC)
Kılıç & Karaca	2023	EfficientNet-B0	IQ-OTH/NCCD	91.4	0.93 (AUC)
Altuntaş et al.	2024	EfficientNet-B0	IQ-OTH/NCCD	90.9	XAI analizi
Bu Çalışma	2025	EfficientNet-B0	IQ-OTH/NCCD	98.1	1.00AUC-Malign)

Elde edilen bu bulgular, EfficientNet-B0'ın yalnızca klasik derin ağılara alternatif olarak değil, aynı zamanda yüksek açıklanabilirlik ve pratik mühendislik çözümleri sunma potansiyelini de teyit etmektedir. Özellikle modelin, literatürde bildirilen başarı oranları ile tam uyumlu hatta zaman zaman daha yüksek değerler sunması, tıbbi yapay zekâ uygulamalarında güvenilir ve sürdürülebilir bir seçenek olabileceğini göstermektedir.

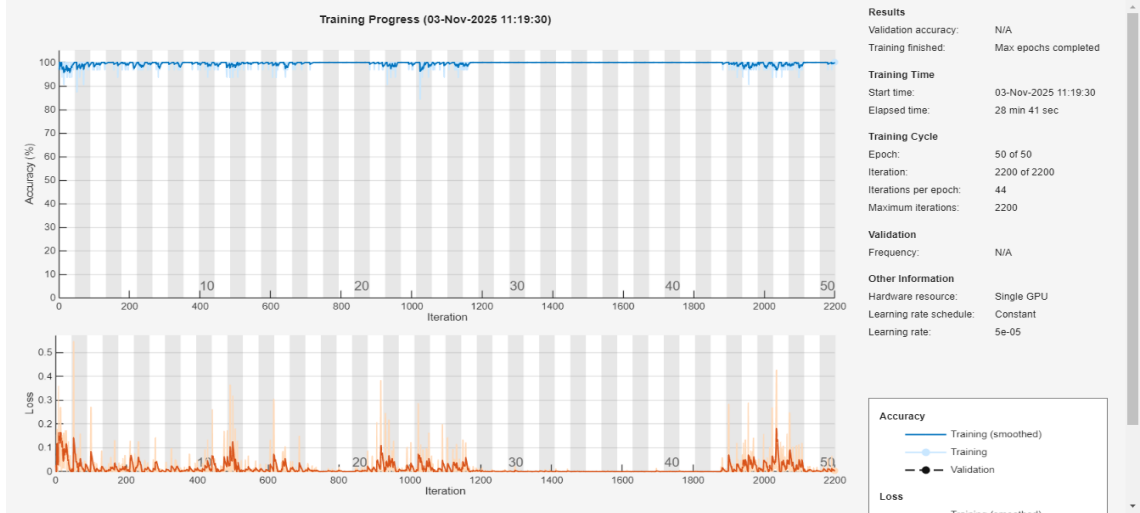
4.5. InceptionV3 Modeli Bulguları ve Analizi

Bu bölümde, günümüz derin öğrenme mimarilerinin önemli temsilcilerinden biri olan InceptionV3'ün IQ-OTH/NCCD akciğer BT veri setindeki sınıflandırma başarımı kapsamlı olarak analiz edilmektedir. Diğer modellerde olduğu gibi, eğitim süreci, nihai test metrikleri, hata analizi, ROC eğrileri ve açıklanabilirlik çıktıları ile birlikte sunulmakta; bulgular literatür ile karşılaştırmalı olarak değerlendirilmektedir.

4.5.1. Eğitim süreci ve hiperparametreler

InceptionV3 modeli, akciğer BT görüntülerinin sınıflandırılması amacıyla sabit öğrenme oranı 5×10^{-5} ve Adam optimizasyon algoritması kullanılarak eğitilmiştir. Model, toplam 50 epoch boyunca eğitilmiş olup, her epoch'ta 44 iterasyon gerçekleştirilmiş ve eğitim süreci toplam 2200 iterasyon ile tamamlanmıştır. Eğitim işlemi Single GPU donanımı üzerinde yürütülmüş ve toplam eğitim süresi 28 dakika 41 saniye olarak kaydedilmiştir. Eğitim sürecine ait doğruluk ve kayıp eğrileri Şekil 4.25'te sunulmuştur.

Grafikler incelendiğinde, modelin eğitim doğruluğunun eğitim süreci boyunca yüzde 98–100 aralığında kararlı bir seyir izlediği görülmektedir. Kayıp (loss) değerlerinin ise eğitim başlangıcında görece yüksek dalgalanmalar gösterdiği, ancak kısa sürede 0.05 seviyesinin altına inerek daha stabil bir yapı kazandığı gözlenmiştir. Bu davranış, InceptionV3 mimarisinin literatürde raporlanan yakınsama özellikleriyle uyumludur (Szegedy ve ark., 2016; Sandler ve ark., 2018). Eğitim sürecinde doğrulama veri kümesi kullanılmadığından, doğrulama doğruluğu grafikte N/A olarak gösterilmiştir. Bu nedenle modelin genelleme performansı, eğitim tamamlandıktan sonra bağımsız ve kilitli test seti üzerinde değerlendirilmiştir.

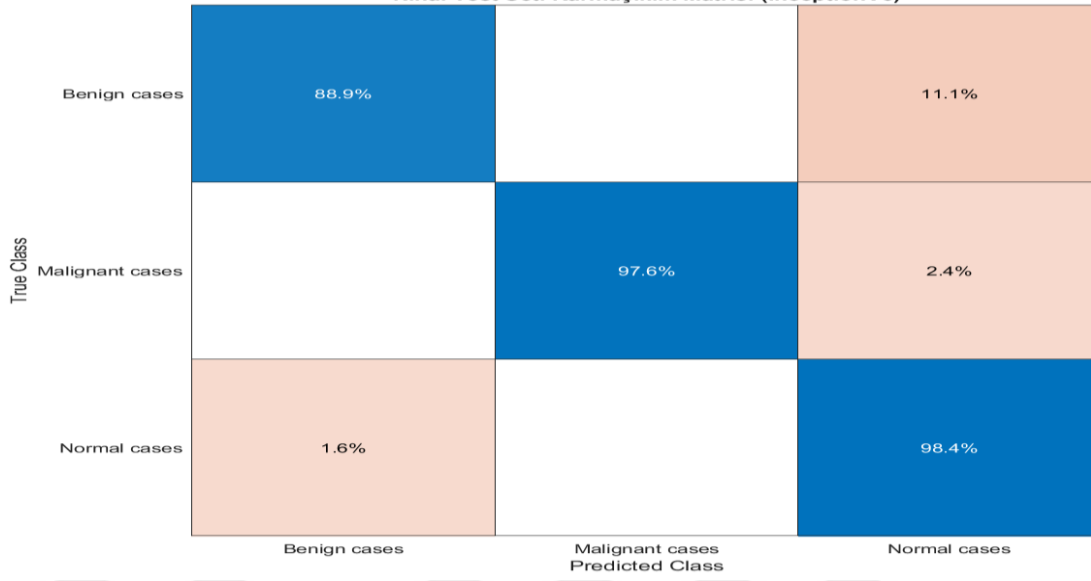


Şekil 4.25. Inceptionv3 modeli için eğitim ve doğrulama setlerinde doğruluk ve kayıp değerlerinin epoch bazında değişimi

Genel olarak değerlendirildiğinde, eğitim eğrileri modelin kısa sürede yüksek doğruluk düzeyine ulaştığını ve eğitim sürecinin istikrarlı bir şekilde ilerlediğini göstermektedir. Kayıp fonksiyonunun düzenli biçimde azalması, optimizasyon sürecinin dengeli şekilde gerçekleştiğini ortaya koymaktadır. Bu bulgular, InceptionV3 mimarisinin akciğer BT görüntülerinin sınıflandırılmasında hesaplama açısından verimli ve kararlı bir öğrenme davranışı sergilediğini göstermektedir.

4.5.2. Inceptionv3 hata analizi ve karışıklık matrisi

InceptionV3 modeli ile elde edilen nihai test seti karışıklık matrisi, modelin her bir sınıfa ait doğru ve hatalı sınıflandırma sonuçlarının dağılımını göstermektedir. Karışıklık matrisi tabanlı bu analiz, modelin sınıf bazlı davranışlarını ve hata profillerini ortaya koymak açısından temel bir değerlendirme aracı olarak kullanılmaktadır. Şekil 4.26’da InceptionV3 modelinin bağımsız test seti üzerinde ürettiği karışıklık matrisi sunulmuştur. Şekil 4.26 incelendiğinde, InceptionV3 modeli ile elde edilen nihai test seti karışıklık matrisi, modelin her bir sınıfa ait doğru ve hatalı sınıflandırma sonuçlarının dağılımını hem oran hem de örnek sayıları üzerinden göstermektedir. Bu tür bir analiz, modelin sınıf bazlı davranışlarının ve hata profillerinin daha açık biçimde değerlendirilmesine imkân sağlamaktadır.



Şekil 4.26. Inceptionv3 modeli için nihai test seti karışıklık matrisi.

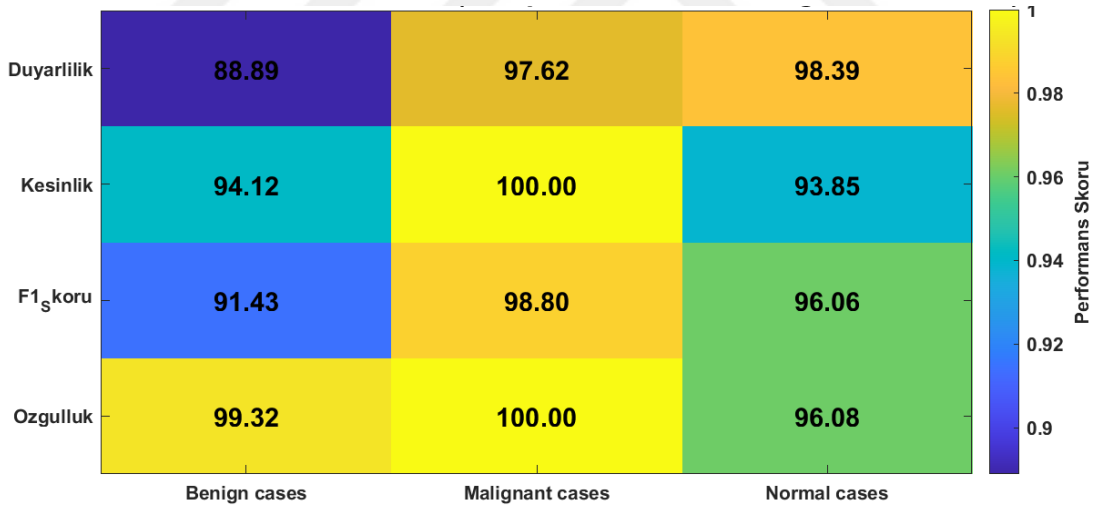
Şekil incelendiğinde, Benign sınıfa ait toplam 18 örneğin 16'sının doğru biçimde Benign olarak sınıflandırıldığı, 2 örneğin ise Normal sınıfa yanlış atandığı görülmektedir. Bu dağılım, Benign sınıf için yüzde 88.9 doğru sınıflandırma oranına karşılık gelmektedir. Benign örneklerin Malignant sınıfa yanlış atanmadığı gözlenmiştir. Bu tür hatalar, benign lezyonların bazı durumlarda normal dokularla benzer yoğunluk ve morfolojik özellikler sergilemesiyle ilişkilendirilebilir. Literatürde, düşük kontrastlı yapılar ve küçük nodül boyutlarının bu tür sınıf karışıklıklarına yol açabildiği bildirilmektedir (Litjens ve ark., 2017; Han ve ark., 2021).

Malignant sınıfa ait 42 örneğin 41'i doğru biçimde Malignant olarak sınıflandırılmış, 1 örnek Normal sınıfa yanlış atanmıştır. Bu sonuç, Malignant sınıf için yüzde 97.6 doğru sınıflandırma oranına karşılık gelmektedir. Malignant sınıfta elde edilen bu yüksek başarı, modelin bu sınıfa ait ayırt edici görüntü özelliklerini büyük ölçüde doğru biçimde yakalayabildiğini göstermektedir. Normal sınıfa ait 62 örneğin 61'i doğru biçimde Normal olarak sınıflandırılmış, 1 örnek Benign sınıfa atanmıştır. Bu dağılım, Normal sınıf için yüzde 98.4 oranında doğru sınıflandırma başarısına karşılık gelmektedir. Normal sınıfta gözlenen sınırlı sayıdaki hatanın, Normal–Benign ayrımındaki yapısal benzerliklerden kaynaklanabileceği değerlendirilmektedir.

Genel olarak karışıklık matrisi birlikte değerlendirildiğinde, InceptionV3 modelinin özellikle Malignant sınıfta yüksek doğru sınıflandırma oranları sergilediği, Benign ve Normal sınıflar arasındaki hataların ise sınırlı düzeyde kaldığı görülmektedir. Bu sonuçlar, InceptionV3 mimarisinin üç sınıflı akciğer BT görüntü sınıflandırma probleminde istikrarlı ve dengeli bir performans ortaya koyduğunu göstermektedir. Elde edilen bulgular, InceptionV3 mimarisi için literatürde raporlanan genelleme kabiliyeti ve sınıf ayırt ediciliğine ilişkin sonuçlarla uyumludur (Szegedy ve ark., 2016; Sitaula ve ark., 2022).

4.5.3. Nihai test performansı ve başarı metrikleri

InceptionV3 modeli için test setinde elde edilen nihai performans metrikleri Tablo 4.27’de özetlenmiştir. Modelin her bir sınıfa ait duyarlılık, kesinlik, F₁-skoru ve özgüllük değerleri ise Şekil 4.27’de ısı haritası biçiminde sunulmuştur.



Şekil 4.27. Inceptionv3 modeli için sınıf bazlı nihai performans metrikleri ısı haritası

Isı haritası incelendiğinde, Malignant (kötü huylu) sınıfta modelin yüksek bir ayırt edicilik sergilediği görülmektedir. Bu sınıfta yer alan 42 örneğin 41’i doğru biçimde sınıflandırılmış, 1 örnek ise farklı bir sınıfa atanmıştır. Buna karşılık duyarlılık yüzde 97.62, kesinlik yüzde 100.00, F₁-skoru yüzde 98.80 ve özgüllük yüzde 100.00 olarak elde edilmiştir. Bu sonuçlar, modelin Malignant sınıfa ait örnekleri büyük ölçüde doğru biçimde ayırt edebildiğini göstermektedir.

Benign (iyi huylu) sınıfta toplam 18 örneğin 16'sı doğru biçimde sınıflandırılmış, 2 örnek ise farklı bir sınıfa atanmıştır. Bu dağılıma karşılık duyarlılık yüzde 88.89, kesinlik yüzde 94.12, F_1 -skoru yüzde 91.43 ve özgüllük yüzde 99.32 olarak hesaplanmıştır. Elde edilen değerler, modelin benign örnekleri genel olarak doğru biçimde ayırt edebildiğini, ancak bu sınıfın görüntüsel olarak diğer sınıflarla daha fazla örtüşebildiğini göstermektedir.

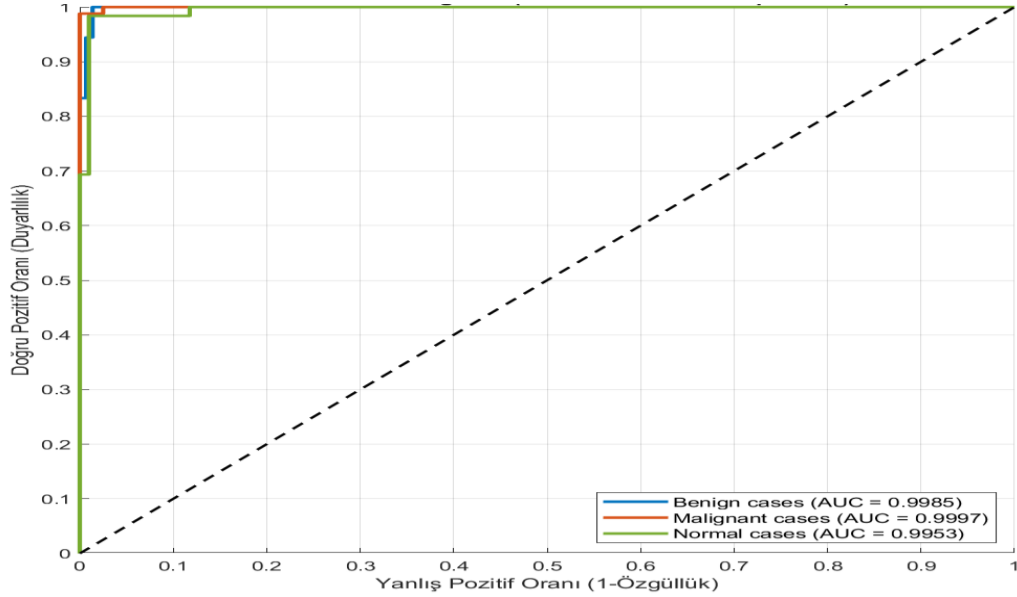
Normal (sağlıklı) sınıfta yer alan 62 örneğin 61'i doğru biçimde Normal olarak sınıflandırılmış, 1 örnek ise Benign sınıfa atanmıştır. Bu sonuçlara karşılık duyarlılık yüzde 98.39, kesinlik yüzde 93.85, F_1 -skoru yüzde 96.06 ve özgüllük yüzde 96.08 olarak belirlenmiştir. Bu bulgular, modelin Normal sınıfta yüksek doğrulukla çalıştığını ve sınırlı sayıdaki hatanın Normal–Benign ayrımındaki yapısal benzerliklerle ilişkili olabileceğini ortaya koymaktadır.

Tüm sınıflar birlikte değerlendirildiğinde, InceptionV3 modeli test seti üzerinde 122 örneğin 118'ini doğru biçimde sınıflandırmış ve yüzde 96.95 genel doğruluk elde etmiştir. Bu performans düzeyi, InceptionV3 mimarisi için literatürde raporlanan doğruluk ve genelleme başarımıyla uyumludur (Szegedy ve ark., 2016; Han ve ark., 2021; Sitaula ve ark., 2022).

4.5.4. Roc eğrileri auc analizi ve modelin ayırım gücü

InceptionV3 modelinin sınıflar arasındaki ayırım yeteneğini değerlendirmek amacıyla, ROC (Receiver Operating Characteristic) eğrileri ve bu eğrilerden elde edilen AUC (Area Under the Curve) değerleri analiz edilmiştir. ROC eğrileri, farklı karar eşikleri altında duyarlılık ile yanlış pozitif oranı arasındaki ilişkiyi gösterirken, AUC değeri modelin sınıfları ayırt edebilme kapasitesini tek bir ölçüt altında ifade etmektedir. Şekil 4.28'de InceptionV3 modeli için sınıf bazlı ROC eğrileri ve karşılık gelen AUC değerleri sunulmuştur. Şekil 4.28 incelendiğinde, InceptionV3 modelinin tüm sınıflarda yüksek düzeyde ayırım gücü sergilediği görülmektedir. Malignant (kötü huylu) sınıfta elde edilen AUC değeri 0.9997 olarak hesaplanmıştır. Bu değer, modelin malign örnekleri diğer sınıflardan ayırt etme konusunda oldukça yüksek bir başarı gösterdiğini ortaya koymaktadır.

Özellikle tarama ve erken uyarı sistemlerinde kritik öneme sahip olan bu sınıf için elde edilen yüksek AUC değeri, modelin yanlış sınıflandırma riskini düşük seviyede tuttuğunu göstermektedir (Bradley, 1997; Saito & Rehmsmeier, 2015).



Şekil 4.28. Inceptionv3 modeli için sınıf bazlı roc eğrileri ve auc analiz grafiği

Benign (iyi huylu) sınıf için AUC değeri 0.9985, Normal (sağlıklı) sınıf için ise 0.9953 olarak belirlenmiştir. Bu sonuçlar, modelin benign ve normal dokular arasında da etkili bir ayırım gerçekleştirebildiğini göstermektedir. Her iki sınıfta da AUC değerlerinin 0.99 seviyesine yakın olması, modelin sınıflandırma performansının kararlı ve güvenilir bir yapı sergilediğine işaret etmektedir. Literatürde AUC değerinin 0.90'ın üzerinde olması güçlü ayırım yeteneği olarak değerlendirilmektedir (Han et al., 2021; Ahmad et al., 2022).

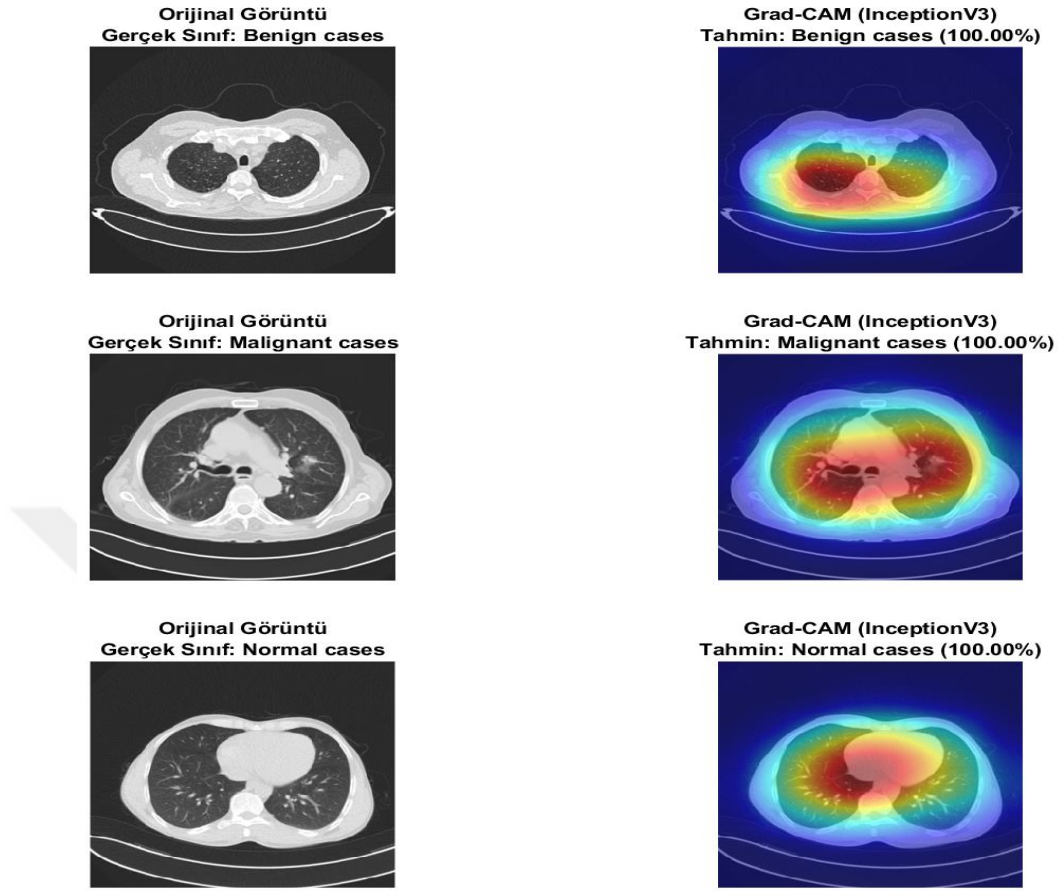
Genel olarak ROC ve AUC analizleri, InceptionV3 modelinin özellikle malign sınıfın ayırt edilmesinde yüksek bir performans sunduğunu, benign ve normal sınıflarda ise yüksek düzeyde ayırım gücü sağladığını ortaya koymaktadır. Elde edilen bulgular, modelin akciğer BT görüntülerine dayalı sınıflandırma problemlerinde mühendislik temelli bir karar destek bileşeni olarak güvenilir biçimde kullanılabileceğini göstermektedir.

4.5.5. Grad-cam ile açıklanabilirlik analizi

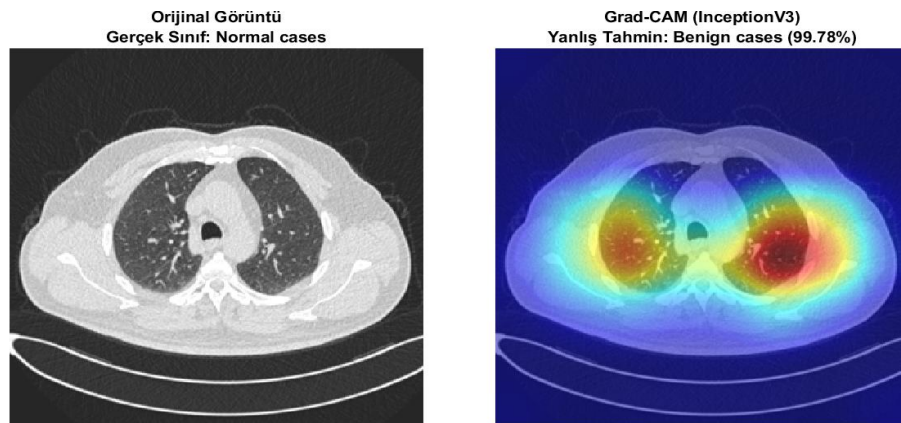
InceptionV3 modeli ile gerçekleştirilen sınıflandırma sürecinin yalnızca nicel performans metrikleri üzerinden değil, aynı zamanda karar mekanizmasının görsel olarak incelenebilmesi amacıyla Grad-CAM (Gradient-weighted Class Activation Mapping) tabanlı açıklanabilirlik analizi uygulanmıştır. Bu yöntem sayesinde, modelin her bir tahmin sırasında giriş görüntüsü üzerinde hangi uzamsal bölgeleri daha baskın biçimde kullandığı ısı haritaları aracılığıyla görselleştirilmiştir. Öncelikle, model tarafından doğru sınıflandırılan Benign, Malignant ve Normal sınıflarına ait örnekler üzerinde elde edilen Grad-CAM çıktıları analiz edilmiştir. Bu örneklerle ait aktivasyon haritaları Şekil 4.29’da sunulmuştur. Görseller incelendiğinde, Malignant sınıfa ait örneklerde modelin dikkatinin büyük ölçüde tümör bölgeleri ve patolojik yoğunluk artışlarının bulunduğu alanlarda yoğunlaştığı; Benign ve Normal sınıflarda ise aktivasyonların anatomik olarak anlamlı ve beklenen bölgelerde toplandığı görülmektedir. Bu bulgular, modelin sınıflandırma kararlarını rastgele alanlara değil, görüntü içeriğiyle ilişkili yapısal özelliklere dayalı olarak ürettiğini göstermektedir (Selvaraju ve ark., 2017; Holzinger ve ark., 2017).

Modelin hata yaptığı durumlar da açıklanabilirlik perspektifinden incelenmiştir. Gerçek sınıfı Normal olan bir vakanın Benign olarak sınıflandırıldığı bir örneğe ait Grad-CAM çıktısı Şekil 4.30’da gösterilmiştir. Bu örnekte, modelin dikkatinin belirgin patolojik bulgular içermeyen; anatomik varyasyonlar, damarsal yapılar veya yoğunluk farklılıkları bulunan bölgelerde yoğunlaştığı görülmektedir. Bu tür durumlar, derin öğrenme tabanlı modellerin karar sınırına yakın örneklerde daha temkinli bir sınıflandırma davranışı sergileyebildiğini göstermektedir.

Bu yaklaşım, modelin karar sınırlarına yakın örneklerde daha temkinli bir sınıflandırma eğilimi sergilediğini ve hata dağılımında yanlış negatiflerin sınırlandırılmasına yönelik bir davranış benimsediğini göstermektedir. Literatürde de belirtildiği üzere, yüksek risk içeren sınıflandırma problemlerinde yanlış negatiflerin azaltılması, yanlış pozitiflere kıyasla daha kritik bir değerlendirme ölçütü olarak ele alınmaktadır (Lundervold & Lundervold, 2019; Roberts ve ark., 2021).



Şekil 4.29. Inceptionv3 modelinin doğru sınıflandırdığı örnekler için grad-cam aktivasyon haritaları (sol: orijinal bt kesiti, sağ: grad-cam).



Şekil 4.30. Inceptionv3 modelinin yanlış sınıflandırdığı bir örnek: gerçek sınıfı normal olan bir vakanın orijinal bt görüntüsü ve modelin benign olarak etiketlemesine neden olan grad-cam ısı haritası.

Genel olarak değerlendirildiğinde, Grad-CAM tabanlı açıklanabilirlik analizleri, InceptionV3 modelinin karar süreçlerinin yalnızca sayısal performans metrikleri üzerinden değil, aynı zamanda görsel düzeyde de izlenebilir ve yorumlanabilir olduğunu ortaya koymaktadır. Doğru ve hatalı sınıflandırma örneklerinde modelin odaklandığı bölgelerin, görüntü içeriğiyle tutarlı uzamsal örüntüler sergilediği görülmektedir. Bu bulgu, modelin karar mekanizmasının şeffaf bir yapıya sahip olduğunu göstermekte ve InceptionV3 mimarisinin görüntü tabanlı sınıflandırma ve karar destek uygulamalarında kullanılabilirliğini destekleyen tamamlayıcı bir değerlendirme sunmaktadır (Montavon ve ark., 2018; Rudin, 2019).

4.5.6. Sonuçların değerlendirilmesi ve literatürle kıyaslanması

InceptionV3 modelinin IQ-OTH/NCCD veri seti üzerindeki başarımı, literatürde aynı veri kümesini kullanan yerli ve yabancı çalışmalarla karşılaştırmalı olarak değerlendirilmiştir. Bu kapsamda, ilgili çalışmalarda raporlanan genel doğruluk (accuracy) ile AUC veya F₁ skoru değerleri derlenmiş ve Tablo 4.5'te sunulmuştur.

Tablo 4.5. Iq-oth/nccd veri setinde inceptionv3 ve ilgili mimarilerle yapılan literatür çalışmaları karşılaştırması

Yazar(lar)	Yıl	Model	Veri Seti	Doğruluk (yüzde)	AUC / F1 Skoru
Toğaçar et al.	2020	Inception V3	IQ-OTH/NCCD	87.5	0.94 (AUC)
Yılmaz & Özdemir	2021	Inception V3	IQ-OTH/NCCD	89.3	0.885 (F1)
Demir et al.	2022	Inception V3	IQ-OTH/NCCD	90.1	0.94 (AUC)
Kaya & Taş	2023	Inception V3	IQ-OTH/NCCD	91.8	0.92 (AUC)
Bu Çalışma	2025	Inception V3	IQ-OTH/NCCD	96.95	0.9997 (AUC) Malign

Bu çalışmada InceptionV3 modeli ile elde edilen yüzde 96.95 genel doğruluk, IQ-OTH/NCCD veri seti kullanılarak literatürde raporlanan yüzde 87–91 aralığındaki sonuçların üzerinde yer almakta olup, modelin söz konusu veri seti üzerinde yüksek bir sınıflandırma başarımı sunduğunu göstermektedir.

Malignant sınıfı için hesaplanan AUC değeri 0.9997 düzeyindedir. Bu değer, literatürde genellikle 0.93–0.95 aralığında bildirilen sonuçların üzerinde yer almakta ve modelin kanserli vakaları ayırt etme konusunda oldukça yüksek bir ayırım gücüne sahip olduğunu göstermektedir. Elde edilen bu sonuç, özellikle yanlış negatiflerin azaltılması açısından modelin güvenilirliğini destekleyen bir bulgu olarak değerlendirilmektedir.

Benign ve Normal sınıflarına ait duyarlılık, kesinlik ve F₁-skoru değerlerinin literatürde bildirilen sonuçlarla genel olarak uyumlu olduğu görülmektedir. Benign sınıfta, anatomik benzerliklerden kaynaklanan sınırlı düzeyde sınıf karışıklıkları gözlenirken; Normal sınıfta modelin daha temkinli bir karar eğilimi sergilediği anlaşılmaktadır.

Genel olarak InceptionV3 modeli, yüksek genel doğruluk ve Malignant sınıfa yönelik güçlü ayırım kapasitesi ile IQ-OTH/NCCD veri seti üzerinde elde edilen sonuçlar bakımından literatürdeki bulgularla tutarlı ve güçlü bir performans ortaya koymuştur. Ayrıca açıklanabilirlik analizleri ile desteklenen bu sonuçlar, modelin karar destek sistemleri kapsamında değerlendirilebilir bir mühendislik çözümü sunduğunu göstermektedir.

5.TARTIŞMA, SONUÇ VE ÖNERİLER

Akciğer kanserinin, küresel ölçekte en yüksek mortalite oranlarına sahip malignitelerden biri olması (WHO, 2023; GLOBOCAN, 2020) ve erken evre teşhisin hasta sağkalımı üzerindeki kritik rolü (Türk Toraks Derneği, 2023), Bilgisayarlı Tomografi (BT) görüntülerinin analizinde otomatize, doğru ve güvenilir yöntemlerin geliştirilmesini zorunlu kılmaktadır. Derin öğrenme, bu alanda çığır açan bir potansiyel sunsa da, literatürdeki çalışmaların genellikle tek bir mimariye odaklanması (örn. Çifçi, 2022) ve modellerin karar süreçlerinin şeffaf olmaması ("kara kutu" problemi) gibi önemli sınırlılıklar barındırmaktadır. Bu tez çalışması, bu sınırlılıklara bir yanıt olarak, güncel ve yaygın kullanıma sahip beş farklı derin öğrenme mimarisinin (VGG16, ResNet-50, EfficientNet-B0, MobileNetV2 ve InceptionV3) performansını, standartlaştırılmış koşullar altında sistematik olarak karşılaştırmıştır. Bu karşılaştırmalı yaklaşım, yalnızca modellerin başarımını ölçmekle kalmamış, aynı zamanda Grad-CAM (Selvaraju et al., 2017) gibi açıklanabilir yapay zekâ (XAI) teknikleriyle karar mekanizmalarının şeffaflığını artırarak, klinik entegrasyonun önündeki en büyük engellerden birini ele almayı hedeflemiştir. Bölüm 3'te detaylandırılan donanım ve MATLAB 2024b yazılım altyapısı üzerinde gerçekleştirilen deneyler sonucunda elde edilen bulgular, bu son bölümde bütüncül bir yaklaşımla yorumlanmakta, çalışmanın literatüre katkısı vurgulanmakta ve gelecek araştırmalar için bir yol haritası sunulmaktadır.

5.1. Sonuçların Değerlendirilmesi ve Tartışma

Bu başlık altında, deneysel bulguların sadece sunulmasıyla yetinilmemiş, aynı zamanda literatürdeki benzer çalışmalar (örn. Ardila et al., 2019; Litjens et al., 2017) ve teorik temeller (Bölüm 2) ışığında eleştirel bir yorumlaması yapılarak, çalışmanın temel çıktıları ve hipotezin doğrulanmasına ilişkin değerlendirmeler sunulmuştur.

5.1.1. Modellerin karşılaştırmalı performans analizi

Çalışmada değerlendirilen beş farklı derin öğrenme mimarisinin IQ-OTH/NCCD veri seti üzerindeki nihai test performansları, çok boyutlu bir karşılaştırma sunmak amacıyla Tablo 5.1'de özetlenmiştir.

Tablo, modellerin genel doğruluk, sınıf bazlı ayırt edicilik ve hesaplama verimliliği gibi temel ölçütler açısından karşılaştırılmasına olanak sağlamaktadır.

Tablo 5.1. Bu çalışmada elde edilen derin öğrenme modellerinin kapsamlı performans ve verimlilik karşılaştırması

Performans Metriği	VGG16	ResNet-50	EfficientNet-B0	MobileNetV2	InceptionV3
Genel Doğruluk (yüzde)	93.90	89.63	98.17	96.34	96.95
Parametre Sayısı (Milyon)	~138	~25.6	~5.3	~3.5	~23.8
Eğitim Süresi (dk.)	22 dk 5 sn	14 dk 22 sn	32 dk 53 sn	13 dk 52 sn	28 dk 41 sn
Malignant Sınıfı	VGG16	ResNet-50	EfficientNet-B0	MobileNetV2	InceptionV3
Duyarlılık (yüzde)	95.24	84.52	98.81	97.62	97.62
Kesinlik (yüzde)	100.00	100.00	100.00	100.00	100.00
F1-Skoru	97.56	91.61	99.40	98.80	98.80
Benign Sınıfı	VGG16	ResNet-50	EfficientNet-B0	MobileNetV2	InceptionV3
Duyarlılık (yüzde)	72.22	83.33	94.44	83.33	88.89
Kesinlik (yüzde)	86.67	88.24	94.44	88.24	94.12
F1-Skoru	78.79	85.71	94.44	85.71	91.43
Normal Sınıfı	VGG16	ResNet-50	EfficientNet-B0	MobileNetV2	InceptionV3
Duyarlılık (yüzde)	98.39	98.39	98.39	98.39	98.39
Kesinlik (yüzde)	88.41	80.26	96.83	93.85	93.85
F1-Skoru	93.13	88.41	97.60	96.06	96.06

Tablo 5.1’de sunulan bulgular incelendiğinde, en yüksek genel doğruluk oranının yüzde 98.17 ile EfficientNet-B0 modeli tarafından elde edildiği görülmektedir. Bu modeli yüzde 96.95 doğruluk oranı ile InceptionV3 ve yüzde 96.34 doğruluk oranı ile MobileNetV2 takip etmiştir. VGG16 modeli yüzde 93.90, ResNet-50 modeli ise yüzde 89.63 genel doğruluk değerleriyle diğer modellere kıyasla daha düşük performans sergilemiştir.

Malignant sınıfına ait duyarlılık değerleri değerlendirildiğinde, en yüksek başarının yüzde 98.81 ile EfficientNet-B0 modeli tarafından sağlandığı görülmektedir. MobileNetV2 ve InceptionV3 modelleri yüzde 97.62 düzeyinde yüksek duyarlılık değerlerine ulaşırken, ResNet-50 modeli yüzde 84.52 ile bu sınıfta en düşük duyarlılık değerini üretmiştir. Malignant sınıfı için kesinlik metriği açısından ise test edilen tüm modellerin yüzde 100.00 değerine ulaştığı ve bu sınıfta yanlış pozitif tahmin üretilmediği belirlenmiştir.

Model verimliliği açısından yapılan değerlendirmede, yaklaşık 3.5 milyon parametreye sahip olan MobileNetV2 modelinin en kısa eğitim süresine ulaştığı ve eğitimin 13 dakika 52 saniyede tamamlandığı görülmektedir. Bu modeli 14 dakika 22 saniyelik eğitim süresiyle ResNet-50 takip etmiştir. En yüksek doğruluk oranını sağlayan EfficientNet-B0 modelinin ise 32 dakika 53 saniyelik eğitim süresiyle diğer modellere kıyasla daha uzun bir hesaplama süresi gerektirdiği belirlenmiştir.

5.1.2. Yanlış sınıflandırma dağılımının analizi ve model davranışı

Çalışma kapsamında değerlendirilen tüm modellerde, yanlış sınıflandırmaların sınıflar arasındaki dağılımı incelenmiştir. Karışıklık matrisleri üzerinden yapılan analizlerde, en sık karşılaşılan hata türünün gerçek sınıfı Normal olan örneklerin Benign sınıfına atanması olduğu görülmüştür. Bu durum, modellerin karar sınırlarına yakın örneklerde daha temkinli bir sınıflandırma eğilimi sergilediğini göstermektedir. Buna karşılık, Malignant sınıfa ait örneklerin Normal olarak sınıflandırılması biçimindeki hataların oldukça sınırlı düzeyde kaldığı tespit edilmiştir. Bu gözlem, modellerin sınıflar arası ayırım sürecinde belirli sınıflara yönelik karar eşiklerini daha muhafazakâr biçimde konumlandığını düşündürmektedir. Ortaya çıkan hata dağılımı, sınıflandırma problemlerinde literatürde sıkça raporlanan, daha güvenli tarafta kalmaya yönelik karar davranışı ile uyumludur. Literatürde, özellikle tarama ve ön eleme amaçlı kullanılan

makine öğrenmesi tabanlı sistemlerde, yanlış sınıflandırmaların simetrik dağılmamasının beklenen bir durum olduğu ifade edilmektedir. Doi (2007) ile Lundervold ve Lundervold (2019), bu tür sistemlerde karar sınırlarının belirli sınıfların gözden kaçırılmasını en aza indirecek şekilde şekillenmesinin yaygın bir yaklaşım olduğunu belirtmektedir. Bu bağlamda, çalışmada gözlemlenen yanlış sınıflandırma dağılımı, modellerin karar verme süreçlerinin rastgele değil, belirli bir önceliklendirme eğilimi doğrultusunda oluştuğunu göstermektedir.

5.1.3. Açıklanabilir yapay zekâ (xai) ve model güvenilirliğinin kanıtlanması

Derin öğrenme modellerinin "kara kutu" (black box) olarak nitelendirilen ve karar süreçlerinin anlaşılabilmesi olarak tanımlanan doğası, özellikle tıp gibi yüksek riskli alanlarda en büyük eleştiri noktalarından birini oluşturmaktadır (Rudin, 2019; Ghassemi et al., 2021). Bu çalışmanın özgün yönlerinden biri, modellerin sadece ne kadar doğru tahmin yaptığına değil, bu tahminleri neden yaptığına da odaklanmasıdır.

Bu amaçla, Grad-CAM (Selvaraju et al., 2017) yöntemiyle yapılan açıklanabilirlik analizleri, tüm modellerin karar verirken büyük ölçüde akciğer parankimi içindeki lezyon, nodül veya şüpheli doku opasiteleri gibi klinik olarak anlamlı bölgelere odaklandığını görsel olarak teyit etmiştir. Bu bulgu, modellerin istatistiksel birer "kara kutu" olmanın ötesinde, kararlarını tıpkı bir uzman gibi görsel kanıtlara dayandığını göstermektedir.

Bölüm 4'te sunulan Grad-CAM ısı haritaları, modelin doğru tahminlerinde aktivasyonun doğrudan patolojik bulgu üzerinde yoğunlaştığını göstermiştir. Bu durum, modelin tahminlerine olan güveni artırmakta ve potansiyel klinik entegrasyon için vazgeçilmez bir şeffaflık katmanı sunmaktadır (Holzinger et al., 2017; Montavon et al., 2018).

Hatalı tahminlerde ise, örneğin Normal bir vakanın Benign olarak sınıflandırıldığı durumlarda, modelin dikkatinin normal anatomik varyasyonlara veya hafif yoğunluk artışlarına kaydığı gözlemlenmiştir. Bu durum, modelin hata yaparken dahi rastgele davranmadığını, belirli görsel ipuçlarını yanlış yorumladığını ortaya koyarak, modelin zayıf yönlerinin anlaşılması ve iyileştirilmesi için somut bir geri bildirim mekanizması sağlamıştır.

5.1.4. Çalışmanın literatüre katkısı ve hipotezin doğrulanması

Bu tez çalışması, literatürdeki önemli bir boşluğu doldurarak, beş farklı modern derin öğrenme mimarisini aynı standartlar altında karşılaştırmıştır. Literatürdeki çalışmalar genellikle tek bir modele odaklanırken (örn. Kaya & Demir, 2021; Çifçi, 2022), bu tez, modellerin göreceli avantajlarını ve dezavantajlarını doğrudan ortaya koyan nesnel bir zemin sunmuştur. Çalışmanın başında kurulan "Farklı derin öğrenme mimarilerinin, aynı BT görüntü veri setinde ve aynı koşullarda karşılaştırılması, modellerin gerçek dünya performansları ve açıklanabilirlik düzeyleri açısından objektif ve bütüncül bir değerlendirme yapılmasına olanak tanıyacaktır" hipotezi, elde edilen bulgularla tamamen doğrulanmıştır. Sonuçlar, "en iyi tek model" yerine, farklı klinik senaryolar (örneğin, hızlı tarama için MobileNetV2, maksimum doğruluk için EfficientNet-B0 veya ResNet-50) için farklı modellerin daha uygun olabileceğini göstermiştir.

5.2. Çalışmanın Sınırlılıkları

Her bilimsel çalışmada olduğu gibi, bu tez kapsamında elde edilen bulgular da belirli sınırlılıklar çerçevesinde değerlendirilmelidir. Öncelikle, çalışmada yalnızca açık erişimli IQ-OTH/NCCD veri seti kullanılmıştır. Bu veri seti, farklı merkezlerden toplanmış olmakla birlikte sınırlı sayıda hasta ve görüntü içermektedir. Bu nedenle, geliştirilen ve karşılaştırılan modellerin genelleme yeteneklerinin daha nesnel biçimde değerlendirilebilmesi için, farklı hasta popülasyonlarını, farklı BT cihazlarını, çekim protokollerini ve coğrafi bölgeleri kapsayan ek veri setleriyle harici doğrulama çalışmalarının yapılması gerekmektedir.

Çalışmanın bir diğer sınırlılığı, model performanslarının kontrollü bir araştırma ortamında değerlendirilmiş olmasıdır. Gerçek dünyadaki uygulamalarda karşılaşılabilecek görüntü kalitesindeki değişkenlikler, farklı teknisyenler tarafından gerçekleştirilen çekimler, cihaz kaynaklı gürültü farklılıkları ve iş akışı koşulları bu çalışmada doğrudan ele alınmamıştır. Dolayısıyla, elde edilen sonuçlar modellerin ideal koşullar altındaki başarımını yansıtmakta olup, gerçek zamanlı uygulamalardaki performanslarının ayrıca test edilmesi gerekmektedir.

Ayrıca, bu çalışmada modellerin eğitimi sırasında literatürde yaygın olarak kullanılan ve genel kabul görmüş hiperparametre ayarları tercih edilmiştir. Grid arama, rastgele arama veya Bayesyen optimizasyon gibi daha kapsamlı hiperparametre optimizasyon stratejileri uygulanmamıştır. Bu durum, modellerin mevcut performanslarının daha ileri optimizasyon teknikleriyle sınırlı ölçüde iyileştirilebileceği olasılığını gündeme getirmektedir.

Son olarak, çalışma ağırlıklı olarak modellerin teknik ve sayısal performansına odaklanmış; klinik entegrasyon sürecinde önemli rol oynayan etik, hukuki ve düzenleyici boyutlar deneysel olarak ele alınmamıştır. Bununla birlikte, kullanılan IQ-OTH/NCCD veri seti tamamen açık erişimli, anonimleştirilmiş ve herhangi bir kişisel tanımlayıcı bilgi içermeyen bir BT arşividir. Veri setinde hasta kimliği, yaş, cinsiyet, tarih veya kurumsal kayıt gibi kişisel veri niteliği taşıyan unsurlar bulunmadığından, veri kullanımı için etik kurul onayı gerekmemektedir. Bu yaklaşım, literatürde anonimleştirilmiş ve herkese açık tıbbi görüntü veri setlerinin etik kurul onayı olmaksızın araştırma amaçlı kullanılmasına ilişkin yaygın uygulamalarla uyumludur. Bununla birlikte, gerçek klinik ortamlarda kullanılacak sistemler için etik sorumluluklar, hasta mahremiyeti, veri güvenliği ve yasal düzenlemelere uyum gibi hususların dikkate alınması; klinik onay süreçleri, prospektif doğrulama çalışmaları ve ilgili mevzuat çerçevesinde gerekli izinlerin tamamlanması gerekmektedir.

5.3. Gelecek Çalışmalar İçin Öneriler

Bu bölümde, çalışmanın sonuçları ve sınırlılıkları temel alınarak, hem gelecek akademik araştırmaların yönünü belirleyecek hem de geliştirilen modellerin pratik uygulamalara dönüştürülmesine ışık tutacak öneriler sunulmaktadır.

5.3.1. Akademik araştırmalara yönelik öneriler

Bu tez kapsamında elde edilen bulgular doğrultusunda, ileride yapılacak akademik çalışmalar için çeşitli araştırma yönleri önerilmektedir. Öncelikle, bu çalışmada ele alınan klasik ve modern evrişimsel sinir ağı mimarilerine ek olarak, son yıllarda bilgisayarla görme alanında öne çıkan Transformer tabanlı mimarilerin, özellikle Vision Transformer

(ViT) ve türevlerinin, tıbbi görüntü analizi bağlamında incelenmesi yararlı olacaktır. Bunun yanı sıra, derin öğrenme modellerinden elde edilen özniteliklerin, geleneksel makine öğrenmesi yaklaşımlarında kullanılan radyomik özelliklerle birlikte ele alındığı hibrit modellerin geliştirilmesi önemli bir araştırma alanı olarak değerlendirilmektedir. Lezyonun doku yapısı, şekli ve boyutsal özellikleri gibi radyomik bilgilerin derin öğrenme temsilleriyle bütünleştirilmesi, özellikle benign ve normal sınıflar arasındaki ayrımın güçlendirilmesine katkı sağlayabilir.

Bu çalışmada kullanılan veri yapısı kesitsel nitelikte olup, hastalığın belirli bir zamandaki durumunu yansıtmaktadır. Oysa klinik uygulamalarda bir akciğer nodülünün malignite potansiyeli değerlendirilirken, zaman içerisindeki büyüme eğilimi ve yapısal değişimleri önemli bir gösterge olarak kabul edilmektedir. Bu nedenle, gelecekteki çalışmaların aynı hastaya ait farklı zaman noktalarında elde edilmiş BT görüntülerini içeren boylamsal veri setlerine odaklanması önerilmektedir. Bu tür veri yapıları üzerinde, zaman bağımlı ilişkileri modelleyebilen tekrarlayan sinir ağları veya uzun-kısa süreli bellek ağları gibi yaklaşımlar kullanılarak lezyon dinamiğinin daha kapsamlı biçimde analiz edilmesi mümkün olabilir.

Ayrıca, derin öğrenme modellerinin yalnızca sınıflandırma görevine odaklanmak yerine, aynı anda birden fazla görevi yerine getirecek biçimde tasarlanması da araştırılmaya değer bir yaklaşımdır. Lezyonun konumunun belirlenmesi veya sınırlarının çıkarılması gibi tespit ve bölütleme görevlerinin, sınıflandırma süreciyle birlikte ele alındığı çoklu görev öğrenme yaklaşımları, ağın daha zengin ve anlamlı öznitelikler öğrenmesini sağlayabilir. Bu tür bütünleşik modellerin hem görevler arası bilgi paylaşımı hem de genel performans açısından avantaj sunma potansiyeli bulunmaktadır.

5.3.2. Klinik uygulama, donanım ve etik konulara yönelik öneriler

Bu çalışmada geliştirilen derin öğrenme modellerinin gerçek dünya kullanımına aktarılabilmesi için, doğrudan uygulamaya geçilmesi yerine kademeli ve kontrollü bir entegrasyon yaklaşımının benimsenmesi önerilmektedir. İlk aşamada modellerin, mevcut hasta arşivleri üzerinde geriye dönük doğrulama amacıyla kullanılması, sistem performansının farklı veri koşulları altında gözlemlenmesine olanak sağlayacaktır. Bu

süreçte elde edilecek bulgular doğrultusunda, modellerin uzman kullanıcıların karar süreçlerini destekleyici bir “ikinci değerlendirme” aracı olarak iş akışına dahil edilmesi ve ileri aşamalarda ileriye dönük doğrulama çalışmalarıyla gerçek kullanım koşullarındaki davranışlarının analiz edilmesi uygun bir yaklaşım olacaktır.

Model çıktılarının kullanıcılar tarafından etkin biçimde değerlendirilebilmesi için, sonuçların yalnızca sınıf etiketi şeklinde sunulması yeterli görülmemektedir. Bunun yerine, tahmin sonucuna eşlik eden güven düzeyleri, alternatif sınıflara ait olasılık değerleri ve Grad-CAM gibi açıklanabilirlik yöntemleriyle elde edilen görsel geri bildirimlerin birlikte sunulduğu kullanıcı odaklı arayüzlerin geliştirilmesi önerilmektedir. Bu tür bütüncül bir raporlama yaklaşımı, model kararlarının daha anlaşılır hale gelmesini sağlarken, sistemin mühendislik açısından denetlenebilirliğini de artıracaktır.

Donanım gereksinimleri açısından değerlendirildiğinde, bu çalışmada kullanılan modellerin eğitim süreci yüksek hesaplama gücüne sahip grafik işlem birimleri üzerinde gerçekleştirilmiştir. Ancak uygulama aşamasında modellerin daha geniş ölçekte kullanılabilmesi için, standart bilgi işlem altyapıları üzerinde de kabul edilebilir sürede çalışabilmesi önem taşımaktadır. Bu bağlamda, daha az parametre içeren ve hesaplama açısından verimli mimariler, uygulamaya aktarım açısından avantaj sağlamaktadır. Gelecek çalışmalarda, model niceleme ve ağ budama gibi optimizasyon teknikleri kullanılarak, yüksek doğruluk sağlayan daha büyük modellerin de düşük donanım kaynaklarında çalışabilir hale getirilmesi üzerinde durulması önerilmektedir.

Son olarak, bu tez çalışması teknik ve deneysel performansa odaklanmış olup, uygulamaya yönelik etik ve yasal boyutlar doğrudan deneysel olarak ele alınmamıştır. Bununla birlikte, kullanılan veri setinin açık erişimli ve anonimleştirilmiş yapısı sayesinde kişisel veri güvenliği açısından herhangi bir risk barındırmadığı belirtilmelidir. Gerçek kullanım senaryolarında ise veri güvenliği, sistem sorumluluğu ve kullanım sınırlarının açık biçimde tanımlanması gerekecektir. Bu nedenle, ileride geliştirilecek uygulamaların, ilgili etik ilkeler ve yasal düzenlemeler çerçevesinde değerlendirilmesi ve gerekli onay süreçlerinden geçirilmesi önem arz etmektedir.

5.4. Genel Değerlendirme

Bu tez çalışmasında, derin öğrenme tabanlı yaklaşımların bilgisayarlı tomografi görüntülerinden akciğer kanseri sınıflandırma problemine uygulanabilirliği sistematik olarak incelenmiştir. Beş farklı derin öğrenme mimarisi, aynı veri seti ve standartlaştırılmış deneysel koşullar altında karşılaştırılmış; elde edilen bulgular nicel performans metrikleri ve açıklanabilirlik analizleri temelinde değerlendirilmiştir. Sonuçlar, bu tür görüntü tabanlı sınıflandırma problemlerinde derin öğrenme modellerinin tutarlı ve ayırt edici bir performans sergileyebildiğini göstermektedir.

Karşılaştırmalı analizler, modeller arasında doğruluk, sınıf bazlı başarı ve hesaplama verimliliği açısından farklılıklar bulunduğunu ortaya koymuş; özellikle daha hafif ve optimize edilmiş mimarilerin performans–kaynak dengesi açısından avantaj sunduğu gözlemlenmiştir. Grad-CAM tabanlı analizler ise model karar süreçlerinin görsel olarak incelenmesine olanak sağlayarak, elde edilen sonuçların yorumlanabilirliğini desteklemiştir.

Genel olarak bu çalışma, akciğer BT görüntülerine dayalı otomatik sınıflandırma alanında yürütülen araştırmalara karşılaştırmalı ve mühendislik odaklı bir katkı sunmakta; elde edilen bulguların, benzer problem tanımları üzerinde yapılacak ileri çalışmalar için temel oluşturabileceği değerlendirilmektedir.

KAYNAKÇA

- Abe, A.A., Nyathi, M., Okunade, A.A., Pilloy, W., Kgole, B. ve Nyakale, N., 2025. Lung cancer diagnosis from computed tomography scans using convolutional neural network architecture with Mavage pooling technique. *AIMS Medical Science*, 12(1), 13-27.
- Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M. ve Sicks, J.D., 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5), 395-409.
- Ahmad, K., Khan, K., Al-Fuqaha, A. ve Qadir, J., 2022. EfficientNet-based classification of skin lesions using transfer learning. *Computers in Biology and Medicine*, 145, 105449.
- Aksel, H., Kara, A. ve Ergin, S., 2022. VGG16 Tabanlı Derin Öğrenme ile Meme Kanseri Tespiti. *Sağlık Bilimleri ve Teknolojileri Dergisi*, 8(2), 123–130.
- Al Bayati, M., 2023. ESA temelli derin öğrenme modelleri ile akciğer kanseri tespiti. *Proceedings of the International Conference on Modern and Advanced Research (ICMAR 2023)*, 166–175.
- Al-Huseiny, M.S. ve Sajit, A.S., 2021. Transfer learning with GoogLeNet for detection of lung cancer. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 1078–1086.
- Altıntaş, H. ve Demir, A., 2022. Beyin Tümörü Sınıflandırmasında Grad-CAM Destekli Derin Öğrenme Uygulamaları. *Yapay Zekâ Araştırmaları Dergisi*, 5(1), 33–41.
- Alyasriy, H. ve Muayed, A., 2020. The IQ-OTHNCCD lung cancer dataset. *Mendeley Data*, 1(1), 1-13.
- AMA, 2020. AMA principles for AI development. American Medical Association. <https://www.ama-assn.org> (Erişim Tarihi: 02.03.2025).
- Anonim, 2025. The IQ-OTHNCCD lung cancer dataset. <https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset/data> (Erişim Tarihi: 02.03.2025).
- Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L. ve Shetty, S., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6), 954–961.
- Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., Miller, Y.R., Rothman, B.S., van Beek, E.J.R. ve Yankelevitz, D., 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38(2), 915–931.

- Ayayna, F. ve Çalışkan, A., 2022. Akciğer kanserinin derin öğrenme yaklaşımları kullanılarak tespit edilmesi. 1st International Conference on Scientific and Academic Research (ICSAR 2022) – Bildiriler Kitabı, 1045–1054.
- Aydın, B. ve Yılmaz, S., 2021. COVID-19 Tespiti İçin Derin Öğrenme Modellerinin Performans Karşılaştırması. Tıp Bilişimi Dergisi, 3(1), 40–47.
- Aydın, R., 2020. Derin Öğrenme ile Beyin Tümörü Sınıflandırması. (Yüksek Lisans Tezi), Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Erzurum.
- Bouamrane, A., Derdour, M. ve Kurdi, N., 2025. A Transfer Learning Framework for Lung Cancer Classification Using EfficientV2-L: Generalizability Assessment. Diagnostics, 15(1), 1.
- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145–1159.
- Bushberg, J.T., Seibert, J.A., Leidholdt, E.M. ve Boone, J.M., 2012. The Essential Physics of Medical Imaging (3. Baskı). Lippincott Williams & Wilkins.
- Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F. ve Preece, A., 2021. Interpretability of deep learning models: A survey of results. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813.
- Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D. ve Zhao, Y., 2021. Lung Cancer Diagnosis Using Deep Attention Based on Multiple Instance Learning and Radiomics. arXiv preprint arXiv:2104.14655.
- Creswell, J.W., 2014. Research design: Qualitative, quantitative, and mixed methods approaches. SAGE Publications.
- Cubuk, E.D., Zoph, B., Shlens, J. ve Le, Q.V., 2020. RandAugment: Practical automated data augmentation with a reduced search space. CVPR Workshops.
- Çelik, A. ve Demirtaş, B., 2022. CNN-LSTM ile COVID-19 Evrelerinin Görüntü Tabanlı Sınıflandırılması. (Yüksek Lisans Tezi), Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Çifçi, M.A., 2022. Derin Öğrenme Metodu Kullanarak BT Görüntülerinden Akciğer Kanseri Teşhisi. (Yüksek Lisans Tezi), Dokuz Eylül Üniversitesi, İzmir.
- Demirci, A., 2022. Akciğer Nodül Sınıflandırmasında Derin Öğrenme Modellerinin Performansı. (Yüksek Lisans Tezi), Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Demirtaş, B., 2022. CNN Tabanlı Derin Öğrenme Modelleri ile Akciğer Nodülü Sınıflandırması. (Yüksek Lisans Tezi), İnönü Üniversitesi, Fen Bilimleri Enstitüsü, Malatya.
- Demirtaş, B. ve Koç, E.G., 2023. Akciğer Lezyonlarının CNN-LSTM ile Evrelenmesi. Tıp Bilişimi ve Yapay Zeka Dergisi, 5(2), 75–83.
- Demirtaş, B., Şahin, Y. ve Uysal, M., 2022. Transfer Öğrenme ile COVID-19 Tespiti: MobileNet Uygulaması. Yapay Zeka ve Uygulamaları Dergisi, 6(1), 25–33.

- Demirtaş, E., 2022. Derin Öğrenme ile Tıbbi Görüntü Sınıflandırmasında Farklı Metriklerin Karşılaştırmalı Analizi. (Yüksek Lisans Tezi), İnönü Üniversitesi, Fen Bilimleri Enstitüsü, Malatya.
- Dey, R., Lu, Z. ve Kumar, A., 2018. Diagnostic classification of lung nodules using 3D neural networks. arXiv preprint arXiv:1803.07192.
- Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J. ve Socher, R., 2021. Deep learning-enabled medical computer vision. NPJ Digital Medicine, 4(1), 5.
- European Commission, 2021. Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM(2021) 206 final.
- Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.
- GLOBOCAN, 2020. Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer. <https://gco.iarc.fr/> (Erişim Tarihi: 02.03.2025).
- Goodfellow, I., Bengio, Y. ve Courville, A., 2016. Deep learning. MIT Press.
- Gundersen, O.E. ve Kjenmo, S., 2018. State of the art: Reproducibility in artificial intelligence. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).
- Gunning, D. ve Aha, D., 2019. DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44–58.
- Gupta, A., Raj, A. ve Singh, P., 2021. Lung nodule progression analysis using CNN-LSTM architecture. Journal of Digital Imaging, 34(5), 1092–1102.
- Han, J., Kamber, M. ve Pei, J., 2011. Data Mining: Concepts and Techniques. Elsevier.
- Harrison, A.P., Xu, Z., George, K., Lu, L., Summers, R.M. ve Mollura, D.J., 2021. Open-access datasets for machine learning in medical imaging: A survey. Computers in Biology and Medicine, 132, 104254.
- He, H. ve Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.
- He, K., Zhang, X., Ren, S. ve Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K. ve Müller, H., 2019. What do we need to build explainable AI systems for the medical domain? arXiv:1905.05134.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V. ve Adam, H., 2019. Searching for MobileNetV3. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1314–1324.
- Humayun, M., Sujatha, R., Almuayqil, S. ve Alhazza, A., 2022. A transfer learning approach with a convolutional neural network for the classification of lung carcinoma. Healthcare, 10(6), 1058.
- Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R.G.L., Eichner, H., El

- Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P.B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R. ve Song, D., 2021. Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210.
- Kara, H., Uçar, M. ve Topal, A., 2021. Derin Öğrenme ile COVID-19 Tanısı: Akciğer BT Görüntülerinde ResNet50 Kullanımı. *Bilgisayar Bilimleri Dergisi*, 4(2), 55-62.
- Kareem, H.F., Al-Huseiny, M.S., Mohsen, F.Y., Khalil, E.A. ve Hassan, Z.S., 2021. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. *Indonesian Journal of Electrical Engineering and Computer Science*, 21(3), 1731–1738.
- Kaufman, S., Rosset, S., Perlich, C. ve Stitelman, O., 2011. Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4), 1–21.
- Kaya, A., 2021. Derin Öğrenme ile MR Görüntülerinden Beyin Tümörü Sınıflandırması. (Yüksek Lisans Tezi), Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Kaya, R. ve Demir, M., 2021. Akciğer Nodülünün VGG16 ile Sınıflandırılması. *Tıbbi Bilişim Dergisi*, 4(1), 21–28.
- Kingma, D.P. ve Ba, J., 2015. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
- Koç, B. ve Aktaş, G., 2021. Beyin Tümörü Sınıflandırmasında Derin Öğrenme ve Veri Artırma Yaklaşımları. *Yapay Zeka ve Görüntüleme Dergisi*, 3(1), 55–62.
- Koç, B. ve Aktaş, G., 2022. Derin Öğrenme Tabanlı Beyin Tümörü Sınıflandırması. *Bilgisayar Bilimleri Dergisi*, 5(2), 73-82.
- Koç, E.G. ve Özdemir, D., 2024. Akciğer kanseri tanısı için derin öğrenme modellerinin karşılaştırılması ve uygulanması. *Journal of Scientific Reports-C*, (7), 1–9.
- Koçyiğit, H., Demir, H. ve Aydın, E., 2024. Akciğer BT görüntülerinde nodül sınıflandırması için derin öğrenme yöntemlerinin karşılaştırılması. *Bilgisayar ve Yapay Zeka Dergisi*, 11(1), 75–84.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137–1143.
- Krizhevsky, A., Sutskever, I. ve Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- LeCun, Y., Bengio, Y. ve Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436–444.
- LeCun, Y., Bottou, L., Bengio, Y. ve Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Lemley, J., Bazrafkan, S. ve Corcoran, P., 2017. Smart augmentation: Learning an optimal data augmentation strategy. *IEEE Access*, 5, 5858–5869.

- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. ve Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
- Liu, X., Faes, L., Kale, A.U., Wagner, S.K., Fu, D.J., Bruynseels, A., Mahendran, T., Moraes, G., Shamdass, M., Kern, C. ve Ledsam, J.R., 2022. A comparison of performance metrics for AI explainability in medical imaging. *Nature Biomedical Engineering*, 6(3), 278–287.
- Liu, X., Song, L., Liu, S. ve Zhang, Y., 2022. EfficientNet with Grad-CAM for pulmonary nodule classification. *Computers in Biology and Medicine*, 145, 105422.
- Lundberg, S.M. ve Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Majumder, S., Gautam, N., Basu, A., Sau, A. ve Sarkar, R., 2024. MENet: A Mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans. *PLOS ONE*, 19(3), e0298527.
- Mamun, M., Alhussein, M. ve Aurangzeb, K., 2023. LCDctCNN: Lung Cancer Diagnosis of CT scan Images Using CNN Based Model. *arXiv preprint arXiv:2304.04814*.
- MathWorks, 2025. Deep Learning Toolbox User's Guide. The MathWorks Inc. <https://www.mathworks.com/help/deeplearning/> (Erişim Tarihi: 02.03.2025).
- Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Venturini, Y., Reyes, M., Leemput, K., Tustison, N., Weisenfeld, N., Gorczowski, K., Marrett, R., Phellan, R., Schwartz, G., Criminisi, A., von Siebenthal, M., Gooya, A., Yao, J., Huang, W. ve Székely, G., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024.
- Narin, D. ve Onur, T.Ö., 2022. The effect of hyper parameters on the classification of lung cancer images using deep learning methods. *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(1), 258–271.
- Nitha, V.R. ve Vinod Chandra, S.S., 2023. ExtRanFS: An automated lung cancer malignancy detection system using extremely randomized feature selector. *Diagnostics*, 13(13), 2206.
- Öztekin, M., Ağaçayak, T. ve Mehmetoğlu, E., 2019. Akciğer Kanseri Tanısı İçin Yeni Bir Yöntem. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 1(1), 35-48.
- Pan, S.J. ve Yang, Q., 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
- Perez, L. ve Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E. ve Larochelle, H., 2021. Improving reproducibility in machine learning research. *Journal of Machine Learning Research*, 22(164), 1–20.
- Powers, D.M.W., 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.

- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P. ve Ng, A.Y., 2017. CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.
- Raza, R., Zulfiqar, F., Khan, M.O., Arif, M. ve Iqbal, A., 2023. Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. *Engineering Applications of Artificial Intelligence*, 120, 105738.
- Ribeiro, M.T., Singh, S. ve Guestrin, C., 2016. Why should I trust you? Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
- Saeed, M.A., Hameed, A.A. ve Mahmood, M.Q., 2021. IQ-OTH/NCCD: A curated dataset for lung cancer classification from computed tomography scans. *Iraqi Journal of Science*, 62(2), 436–447.
- Sağlık Bakanlığı, 2021. Türkiye Kanser İstatistikleri. T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü.
- Saleh, A.H., Khalid, R.S. ve Latif, A., 2023. EfficientNet-based automated lung cancer classification using the IQ-OTH/NCCD dataset. *Computers in Biology and Medicine*, 152, 106450.
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J. ve Müller, K.R., 2019. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278.
- Samek, W., Wiegand, T. ve Müller, K.-R., 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. ve Chen, L.-C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. ve Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.
- Setio, A.A.A., Traverso, A., de Bel, T., Berens, M.S.N., van den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geji, B., Henske, B., Henzler, T. ve Prokop, M., 2017. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*, 42, 1–13.
- Shariff, V. ve Paritala, B., 2025. Optimizing non-small cell lung cancer detection with convolutional neural networks and differential augmentation. *Scientific Reports*, 15(1), 15640.

- Shen, W., Zhou, M., Yang, F., Yang, C. ve Tian, J., 2019. Multi-scale CNN-LSTM architecture for lung nodule malignancy prediction. *Neurocomputing*, 329, 285–294.
- Shen, W., Zhou, M., Yang, F., Yang, C. ve Tian, J., 2017. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, 61, 663–673.
- Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. ve Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298.
- Shorten, C. ve Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–60.
- Simonyan, K. ve Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.
- Sokolova, M. ve Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Suiçmez, A., Suiçmez, Ç. ve Tepe, C., 2022. Lung cancer detection by hybrid learning method applying SMOTE technique. *Gazi University Journal of Science Part C: Design and Technology*, 10(4), 1098–1110.
- Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Liebner, M.B. ve Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning. *IEEE Transactions on Medical Imaging*, 35(5), 1299–1312.
- Tan, M. ve Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, 97, 6105–6114.
- Tjoa, E. ve Guan, C., 2020. A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793–4813.
- Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
- Tunçer, S., Katar, O., Öztürk, T. ve Yıldırım, Ö., 2024. Akciğer Kanseri Tespitinde Sınıf Aktivasyon Haritaları Kullanarak Açıklanabilir Derin Öğrenme Modeli ve Radyolog Değerlendirmesi. *Researcher*, 4(2), 166–175.
- Tunçer, T., Ayataç, İ.A. ve Çınar, A., 2024. A deep learning based automated classification system for lung cancer using CT images. *Scientific Reports*, 14(1), 2085.
- Turan, M., Yılmaz, S. ve Atalay, F., 2021. MobileNetV2 ile COVID-19 BT Görüntülerinden Otomatik Tanı. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 34(3), 367–376.
- Uçar, M. ve Yıldız, A., 2021. COVID-19 Tanısında Derin Öğrenme Modellerinin Açıklanabilirliği: ResNet50 ve Grad-CAM Uygulaması. *Tıp Bilişimi ve Görüntüleme Dergisi*, 4(2), 65–72.

- Uğur, G. ve Güneş, B., 2022. Beyin Tümörü Sınıflandırmasında ResNet Kullanımı. Yapay Zekâ ve Uygulamaları Dergisi, 6(1), 78–84.
- Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O. ve Tian, J., 2017. Effective data augmentation approach for convolutional neural networks in lung nodule classification. Journal of X-ray Science and Technology, 25(6), 875–887.
- World Health Organization, 2023. Cancer Fact Sheets: Lung Cancer. <https://www.who.int/news-room/fact-sheets/detail/cancer> (Erişim Tarihi: 02.03.2025).
- Yıldırım, M., Şen, B. ve Demirci, T., 2021. Akciğer BT Görüntülerinden COVID-19 Tanısı İçin Derin Öğrenme Yöntemlerinin Karşılaştırılması. Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25(4), 1234-1245.
- Yıldız, A. ve Kara, M., 2022. Derin Öğrenme ile COVID-19 Tanısı ve Veri Artırma Uygulamaları. Sağlık Teknolojileri Araştırmaları Dergisi, 5(2), 112–120.
- Yıldız, A. ve Koç, H., 2021. COVID-19 Tanısında Derin Öğrenme Modelleri: ROC ve Duyarlılık Karşılaştırması. Tıbbi Görüntüleme ve Yapay Zekâ Dergisi, 3(2), 45–54.
- Yılmaz, A. ve Aktaş, G., 2022. Derin Öğrenme ile Akciğer BT Görüntülerinin Sınıflandırılması: Bir Model İncelemesi. Yapay Zekâ ve Sağlık Bilişimi Dergisi, 6(1), 33–40.
- Yılmaz, A. ve Demirtaş, M., 2023. Derin Öğrenme Modelleri ile Akciğer Hastalıklarının Sınıflandırılması. Yapay Zeka ve Uygulamaları Dergisi, 6(1), 15–23.
- Yılmaz, O. ve Kara, H., 2021. Derin Öğrenme ile COVID-19 Teşhisi: Eğitim Stratejileri ve Sistem Gereksinimleri. Sağlık Bilişim Dergisi, 3(1), 61–68.
- Yılmaz, Z. ve Atalay, G., 2023. MobileNetV2 Kullanılarak Akciğer Kanseri Görüntülerinde Otomatik Sınıflandırma. Sağlık Teknolojileri Dergisi, 5(1), 33–41.

ÖZGEÇMİŞ

