

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAPAY ZEKA VE ROBOTİK ANABİLİM DALI

DERİN ÖĞRENME YÖNTEMLERİYLE MANYETİK REZONANS
GÖRÜNTÜLERİNDEN BEYİN TÜMÖRLERİNİN SINIFLANDIRILMASI

Erdoğan KÖKÜ

Danışman: Doç. Dr. İsmail AKGÜL

TEZ JÜRİ ÜYELERİ
Doç. Dr. Durmuş ÖZDEMİR
Doç. Dr. İsmail AKGÜL
Dr. Öğr. Üyesi Zafer DURAKLI

YÜKSEK LİSANS TEZİ
ERZİNCAN, 2026

© 2026 [Erdoğan KÖKÜ]. Tüm hakları saklıdır.

ÖZET

DERİN ÖĞRENME YÖNTEMLERİYLE MANYETİK REZONANS GÖRÜNTÜLERİNDEN BEYİN TÜMÖRLERİNİN SINIFLANDIRILMASI

Erdoğan KÖKÜ

Yüksek Lisans Tezi

Erzincan Binali Yıldırım Üniversitesi, Fen Bilimleri Enstitüsü,

Yapay Zeka ve Robotik Anabilim Dalı

Danışman: Doç. Dr. İsmail AKGÜL

2026, 84 sayfa

Bu tez çalışmasında, manyetik rezonans görüntülerinden (MRI) beyin tümörlerinin otomatik olarak sınıflandırılmasına yönelik derin öğrenme tabanlı bir tanı destek sistemi geliştirilmiş ve farklı evrişimli sinir ağı (CNN) mimarilerinin performansları karşılaştırmalı olarak değerlendirilmiştir. Çalışmada Kaggle platformunda paylaşılan Brain Tumor MRI Dataset veri seti kullanılmış olup glioma, meningioma, pituitary ve no tumor olmak üzere dört sınıfa ait toplam 7.023 görüntüden yararlanılmıştır. Araştırma kapsamında Keras Applications kütüphanesinde yer alan MobileNetV2, EfficientNetB0, VGG16, ResNet50 ve DenseNet121 modelleri transfer öğrenme yöntemiyle eğitilmiştir. Ayrıca, tez kapsamında özgün bir CNN mimarisi tasarlanmış ve MobileNetV2 ile bu Özgün CNN modelinin özellik çıkarım yeteneklerini birleştiren hibrit bir mimari geliştirilmiştir. Tüm modeller aynı ön işleme adımları, veri artırma teknikleri ve eğitim parametreleri kullanılarak adil bir biçimde karşılaştırılmıştır. Normal eğitim sonuçlarına göre en yüksek doğruluk %90,31 ile Özgün CNN modeli tarafından elde edilirken, hibrit model %85,51 doğruluk ile ikinci sırada yer almıştır. ResNet50, MobileNetV2 ve DenseNet121 modelleri sırasıyla %83,79, %80,40 ve %79,63 doğruluk değerleri üretmiştir. VGG16 modeli %73 doğruluk sağlarken, EfficientNetB0 yalnızca %30,89 doğruluk ile en düşük performansı göstermiştir. Model genellenebilirliğini değerlendirmek amacıyla beş katlı çapraz doğrulama yöntemi uygulanmıştır. Bu sonuçlara göre hibrit model %93,15 ortalama doğruluk ve %93,08 F1-skoru ile tüm modeller arasında en başarılı yapı olmuştur. Özgün CNN modeli ise %92,72 ortalama doğruluk değeri ile ikinci sırada yer almıştır. ResNet50, MobileNetV2, DenseNet121 ve VGG16 modelleri

sırasıyla %91,20, %89,12, %87,74 ve %82,28 ortalama doğruluk deęerleri elde etmiştir. EfficientNetB0 modeli apraz doęrulamada da yalnızca %27,45 doęruluk saęlayabilmiştir. Bu durum, ilgili mimarinin bu veri setindeki dokusal zellikleri yakalamada yaşadıęı yakınsama (convergence) sorununa işaret etmektedir. Elde edilen bulgular, zgün CNN ve zellikle hibrit mimarinin, MRI tabanlı beyin tmr sınıflandırma problemlerinde yksek doęruluk, dşk varyans ve gl genelleme kabiliyeti sunduęunu ortaya koymaktadır. Bu alıřma, klinik karar destek sistemlerinin geliřtirilmesi iin gvenilir bir temel sunmaktadır.

Anahtar Kelimeler: Beyin tmr, Manyetik rezonans grntleme, Derin ęrenme, Evriřimli sinir aęları, Transfer ęrenme, K-Fold apraz doęrulama, Tıbbi grnt sınıflandırma.



ABSTRACT

CLASSIFICATION OF BRAIN TUMORS FROM MAGNETIC RESONANCE IMAGES WITH DEEP LEARNING METHODS

Erdoğan KÖKÜ

Master's Thesis

Erzincan Binali Yıldırım University, Graduate School of Natural and Applied Sciences,

Department of Artificial Intelligence and Robotics

Advisor: Assoc. Prof. Dr. İsmail AKGÜL

2026, 84 pages

In this thesis, a deep learning–based diagnostic support system for the automatic classification of brain tumors from magnetic resonance imaging (MRI) images was developed, and the performances of different convolutional neural network (CNN) architectures were comparatively evaluated. The study employed the Brain Tumor MRI Dataset published on the Kaggle platform, which contains a total of 7,023 images belonging to four classes: glioma, meningioma, pituitary, and no tumor. Within the scope of the research, the MobileNetV2, EfficientNetB0, VGG16, ResNet50, and DenseNet121 models included in the Keras Applications library were trained using the transfer learning method. In addition, an original CNN architecture was designed as part of the thesis, and a hybrid architecture was developed by combining the feature extraction capabilities of MobileNetV2 with those of this original CNN model. All models were compared in a fair manner by employing the same preprocessing steps, data augmentation techniques, and training parameters. According to the standard training results, the highest accuracy was obtained by the original CNN model with 90.31%, while the hybrid model ranked second with an accuracy of 85.51%. The ResNet50, MobileNetV2, and DenseNet121 models produced accuracy values of 83.79%, 80.40%, and 79.63%, respectively. The VGG16 model achieved 73% accuracy, whereas the EfficientNetB0 model exhibited the lowest performance with only 30.89% accuracy. To evaluate model generalizability, five-fold cross-validation was applied. Based on these results, the hybrid model became the most successful structure among all models with an average accuracy of 93.15% and an F1-score of 93.08%. The original CNN model ranked second with an average

accuracy of 92.72%. The ResNet50, MobileNetV2, DenseNet121, and VGG16 models achieved average accuracy values of 91.20%, 89.12%, 87.74%, and 82.28%, respectively. The EfficientNetB0 model achieved only 27.45% accuracy in cross-validation as well. This indicates significant convergence difficulties of the corresponding architecture in capturing the specific textural characteristics of the brain tumor dataset. The obtained findings reveal that the original CNN and especially the hybrid architecture offer high accuracy, low variance, and strong generalization capability in MRI-based brain tumor classification problems. This study provides a reliable basis for the development of clinical decision support systems.

Keywords: Brain tumor, Magnetic resonance imaging, Deep learning, Convolutional neural networks, Transfer learning, K-Fold cross-validation, Medical image classification.



TEŐEKKÜR

Yüksek lisans tez çalışmamın her aşamasında bilgi, deneyim ve yönlendirmeleriyle bana yol gösteren, akademik birikimi ve sabrıyla çalışmamın şekillenmesine önemli katkılar sağlayan değerli danışmanım Doç. Dr. İsmail AKGÜL'e en içten teşekkürlerimi sunarım.

Çalışmam süresince karşılaştığım teknik ve bilimsel sorunların çözümünde görüşleriyle destek olan, bilgi ve tecrübelerini benimle paylaşan tüm hocalarıma teşekkür ederim.

Bu sürecin her aşamasında maddi ve manevi desteklerini hiçbir zaman esirgemeyen, beni daima motive eden ve yanımda olan sevgili aileme minnettarım. Özellikle yoğun ve zorlu çalışma dönemlerinde gösterdikleri anlayış ve sabır, bu tezin tamamlanmasında en büyük dayanağım olmuştur.

Erdoğan KÖKÜ

Ocak, 2026

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ	ix
SİMGELER VE KISALTMALAR DİZİNİ.....	x
1.GİRİŞ.....	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	2
1.3. Araştırmanın Kapsamı	3
1.4. Araştırmanın Önemi	4
1.5. Varsayımlar ve Sınırlılıklar	5
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR	7
2.1. Beyin Tümörleri ve Manyetik Rezonans Görüntüleme	7
2.2. Bilgisayar Destekli Tanı ve Yapay Zeka Yaklaşımları	8
2.3. Evrişimli Sinir Ağları ve Görüntü Sınıflandırma	10
2.4. Transfer Öğrenme	12
2.5. Veri Artırma ve Düzenleme Teknikleri	14
2.6. Model Değerlendirme Stratejilerine Kuramsal Bakış	16
2.7. Beyin Tümörü Sınıflandırması Üzerine Yapılan Çalışmalar	17
2.8. Literatür Değerlendirmesi ve Boşluk Analizi	20
3. MATERYAL VE YÖNTEM.....	23
3.1. Araştırma Modeli ve Yaklaşım	23
3.2. Veri Seti	23
3.3. Veri Ön İşleme ve Artırma Teknikleri	24
3.4. Kullanılan Derin Öğrenme Modelleri	25
3.4.1. MobileNetV2.....	25
3.4.2. EfficientNetB0	26
3.4.3. VGG16	26
3.4.4. ResNet50	26
3.4.5. DenseNet121	26

3.4.6. Özgün CNN modeli	26
3.4.7. Hibrit model	27
3.5. Model Eğitimi ve Optimizasyon	28
3.6. K-Fold Çapraz Doğrulama Yöntemi	29
3.7. Performans Ölçütleri	30
3.8. Deneysel Ortam	31
4. BULGULAR	33
4.1. Modellerin Normal Eğitim Başarı Analizleri	33
4.1.1. Özgün CNN modelinin normal eğitim performans analizi	33
4.1.2. Hibrit modelin normal eğitim performans analizi	36
4.1.3. ResNet50 modelinin normal eğitim performans analizi.....	38
4.1.4. MobileNetV2 modelinin normal eğitim performans analizi	41
4.1.5. DenseNet121 modelinin normal eğitim performans analizi.....	44
4.1.6. VGG16 modelinin normal eğitim performans analizi	46
4.1.7. EfficientNetB0 modelinin normal eğitim performans analizi	49
4.2. Modellerin Çapraz Doğrulama Başarı Analizleri	51
4.2.1. Özgün CNN modelinin çapraz doğrulama performans analizi	53
4.2.2. Hibrit modelin çapraz doğrulama performans analizi	53
4.2.3. ResNet50 modelinin çapraz doğrulama performans analizi	54
4.2.4. MobileNetV2 modelinin çapraz doğrulama performans analizi	55
4.2.5. DenseNet121 modelinin çapraz doğrulama performans analizi	56
4.2.6. VGG16 modelinin çapraz doğrulama performans analizi	57
4.2.7. EfficientNetB0 modelinin çapraz doğrulama performans analizi	58
5. TARTIŞMA VE SONUÇ.....	60
5.1. Normal Eğitim Sonuçlarının Tartışılması	60
5.2. K-Fold Çapraz Doğrulama Sonuçlarının Tartışılması	62
5.3. Sonuç	63
6. ÖNERİLER	65
KAYNAKÇA	66

TABLolar DİZİNİ

Tablo 1. Veri seti sınıf dağılımı ve görüntü sayıları.....	24
Tablo 2. Deneysel altyapı ve yazılım özellikleri	32
Tablo 3. Modellerin normal eğitim sonucu sayısal veriler	33
Tablo 4. Özgün cnn modelinin sınıflandırma performansı.....	34
Tablo 5. Hibrit modelin sınıflandırma performansı	36
Tablo 6. ResNet50 modelinin sınıflandırma performansı	38
Tablo 7. MobileNetV2 modelinin sınıflandırma performansı	41
Tablo 8. DenseNet121 modelinin sınıflandırma performansı	44
Tablo 9. VGG16 modelinin sınıflandırma performansı	46
Tablo 10. EfficientNetB0 modelinin sınıflandırma performansı	49
Tablo 11. Beş katlı çapraz doğrulama sayısal verileri	52

ŞEKİLLER DİZİNİ

Şekil 1. Veri seti görüntüleri	24
Şekil 2. Özgün CNN Model Diagramı	27
Şekil 3. Hibrit Model Diagramı	28
Şekil 4. Özgün cnn modelinin normal eğitim grafiği	34
Şekil 5. Özgün cnn modelinin karışıklık matrisi	35
Şekil 6. Hibrit modelin normal eğitim grafiği	37
Şekil 7. Hibrit modelin karışıklık matrisi	37
Şekil 8. ResNet50 modelinin normal eğitim grafiği	39
Şekil 9. ResNet50 modelinin karışıklık matrisi	40
Şekil 10. MobileNetV2 modelinin normal eğitim grafiği	42
Şekil 11. MobileNetV2 modelinin karışıklık matrisi	43
Şekil 12. DenseNet121 modelinin normal eğitim grafiği	45
Şekil 13. DenseNet121 modelinin karışıklık matrisi	45
Şekil 14. VGG16 modelinin normal eğitim grafiği	47
Şekil 15. VGG16 modelinin karışıklık matrisi	48
Şekil 16. EfficientNetB0 modelinin normal eğitim grafiği	50
Şekil 17. EfficientNetB0 modelinin karışıklık matrisi	50
Şekil 18. Özgün cnn modelinin k-fold performans grafiği	53
Şekil 19. Hibrit modelin k-fold performans grafiği	54
Şekil 20. ResNet50 modelinin k-fold performans grafiği	55
Şekil 21. MobileNetV2 modelinin k-fold performans grafiği	56
Şekil 22. DenseNet121 modelinin k-fold performans grafiği	57
Şekil 23. VGG16 modelinin k-fold performans grafiği	58
Şekil 24. EfficientNetB0 modelinin k-fold performans grafiği	59

SİMGELER VE KISALTMALAR DİZİNİ

BT	Computed Tomography – Bilgisayarlı Tomografi
CNN	Convolutional Neural Network – Evrişimli Sinir Ağı
CAD	Computer Aided Diagnosis – Bilgisayar Destekli Tanı
DL	Deep Learning – Derin Öğrenme
F1-Score	F1 Başarım Ölçütü
FN	False Negative – Yanlış Negatif
FP	False Positive – Yanlış Pozitif
GPU	Graphics Processing Unit – Grafik İşlem Birimi
IARC	International Agency for Research on Cancer – Uluslararası Kanser Araştırmaları Ajansı
K-Fold	K-Katlı Çapraz Doğrulama
MRI	Magnetic Resonance Imaging – Manyetik Rezonans Görüntüleme
PET	Positron Emission Tomography – Pozitron Emisyon Tomografisi
ReLU	Rectified Linear Unit – Doğrultulmuş Doğrusal Birim
TP	True Positive – Doğru Pozitif
TN	True Negative – Doğru Negatif

1. GİRİŞ

1.1. Problem Durumu

Kanser, dünya genelinde morbidite ve mortalite oranları yüksek olan başlıca sağlık sorunlarından biri olmaya devam etmektedir. Uluslararası Kanser Araştırmaları Ajansı (International Agency for Research on Cancer – IARC) tarafından yayımlanan güncel verilere göre, küresel ölçekte kanser insidansı ve buna bağlı ölümler her geçen yıl artış göstermekte olup, erken tanı ve doğru sınıflandırma süreçlerinin önemi giderek artmaktadır (IARC, 2022; Khazaei vd., 2020).

Manyetik Rezonans Görüntüleme (MRI), beyin tümörlerinin tanı ve değerlendirilmesinde yüksek yumuşak doku kontrastı sunması nedeniyle klinik uygulamalarda yaygın olarak kullanılmaktadır. Ancak MRI görüntülerinin manuel olarak değerlendirilmesi; zaman alıcı olması, gözlemciye bağlı farklılıklar içermesi ve uzmanlık gerektirmesi gibi nedenlerle sınırlılıklar barındırmaktadır. Bu durum, otomatik ve güvenilir sınıflandırma yöntemlerine olan ihtiyacı artırmaktadır (Litjens vd., 2017; Tiwari vd., 2019).

Son yıllarda derin öğrenme tabanlı yöntemler, özellikle Evrişimsel Sinir Ağları (Convolutional Neural Networks – CNN), tıbbi görüntü analizi alanında dikkat çekici başarılar elde etmiştir. CNN tabanlı modeller, görüntülerden ayırt edici özellikleri otomatik olarak öğrenebilme yetenekleri sayesinde, beyin tümörlerinin sınıflandırılmasında geleneksel makine öğrenmesi yöntemlerine kıyasla daha yüksek doğruluk oranları sunabilmektedir (Alzubaidi vd., 2021; Ahmad vd., 2022).

Bununla birlikte, CNN tabanlı modellerin performansı; kullanılan veri setinin yapısı, eğitim stratejileri ve doğrulama yöntemleri gibi birçok faktörden etkilenmektedir. Özellikle sınırlı veri setleriyle çalışılan tıbbi görüntüleme problemlerinde, modelin genellenebilirliğinin doğru biçimde değerlendirilmesi kritik öneme sahiptir. Bu bağlamda, yalnızca tek bir eğitim-test ayırımına dayalı değerlendirmeler, model performansının gerçekte olduğundan daha iyimser tahmin edilmesine yol açabilmektedir (McDermott vd., 2021).

Model performansının daha güvenilir biçimde değerlendirilmesi amacıyla kullanılan K-Fold Çapraz Doğrulama yöntemi, veri setinin farklı alt kümeler üzerinde tekrar tekrar eğitilip test

edilmesine olanak tanıyarak genellenebilirlik analizini güçlendirmektedir. Yapılan çalışmalar, K-Fold çapraz doğrulama yaklaşımının, özellikle tıbbi görüntü sınıflandırma problemlerinde model performansının istatistiksel olarak daha tutarlı biçimde raporlanmasını sağladığını göstermektedir (Yadav ve Shukla, 2016; Li vd., 2021).

Literatürde beyin tümörü sınıflandırmasına yönelik çok sayıda çalışma bulunmasına rağmen, farklı CNN mimarilerinin aynı veri seti üzerinde K-Fold çapraz doğrulama kullanılarak sistematik biçimde karşılaştırıldığı çalışmaların sınırlı olduğu görülmektedir. Bu durum, hangi mimarinin daha kararlı ve genellenebilir sonuçlar sunduğunun net biçimde ortaya konmasını zorlaştırmaktadır (Wadhwa vd., 2019).

Bu tez çalışmasında ele alınan problem, MRI görüntüleri kullanılarak beyin tümörlerinin CNN tabanlı modeller aracılığıyla sınıflandırılması ve elde edilen performans sonuçlarının K-Fold çapraz doğrulama yöntemi ile güvenilir biçimde değerlendirilmesidir. Böylece, derin öğrenme tabanlı yaklaşımların beyin tümörü sınıflandırmasındaki etkinliği, genellenebilirlik açısından bütüncül bir çerçevede analiz edilmektedir.

1.2. Araştırmanın Amacı

Bu çalışmanın temel amacı, derin öğrenme yöntemleri kullanılarak MRI beyin tümörlerinin (glioma, meningioma, pituitary ve tümör içermeyen görüntüler) yüksek doğrulukla otomatik olarak sınıflandırılmasını sağlamak ve farklı Keras tabanlı CNN modellerinin performanslarını karşılaştırmalı olarak analiz etmektir. Son yıllarda derin öğrenme tabanlı yaklaşımların tıbbi görüntü analizinde sağladığı önemli ilerlemeler, bu yöntemlerin beyin tümörü tanısında klinik karar destek sistemleri açısından uygulanabilirliğini artırmıştır (Aurna vd., 2022; Amin vd., 2022).

Bu kapsamda çalışmada, Keras derin öğrenme kütüphanesinin Applications modülünde yer alan ve literatürde yaygın biçimde kullanılan EfficientNetB0 (Tan ve Le, 2020), MobileNetV2 (Sandler vd., 2018), VGG16 (Simonyan ve Zisserman, 2015), ResNet50 (He vd., 2016) ve DenseNet121 (Huang vd., 2017) mimarileri kullanılarak karşılaştırmalı bir performans değerlendirmesi gerçekleştirilmiştir. Söz konusu mimariler, farklı ağ derinlikleri ve mimari tasarım yaklaşımları sunmakta olup, özellikle sınırlı veri kümeleriyle çalışılan tıbbi görüntü

sınıflandırma problemlerinde yaygın biçimde tercih edilmektedir (Alzubaidi vd., 2021; Zulfiqar vd., 2023).

Çalışmanın bir diğer amacı, yalnızca önceden eğitilmiş transfer öğrenme modelleriyle sınırlı kalmayarak, özgün olarak tasarlanmış bir CNN mimarisi ile bu mimarinin MobileNetV2 tabanlı özellik çıkarım katmanlarıyla birleştirildiği hibrit bir yapının geliştirilmesidir. Önerilen bu hibrit yaklaşımın, hesaplama verimliliği ve temsil gücü açısından daha dengeli bir yapı sunmasına rağmen, sınıflandırma başarımı bakımından derin ve karmaşık modellerle rekabet edip edemeyeceği araştırılmıştır. Güncel çalışmalar, hafif ve hibrit CNN mimarilerinin özellikle klinik uygulamalara yakın senaryolarda umut verici sonuçlar sunduğunu ortaya koymaktadır (Batool ve Byun, 2025; Wong vd., 2025).

1.3. Araştırmanın Kapsamı

Bu araştırma, yalnızca MRI görüntülerini temel almakta ve dört farklı sınıfı kapsamaktadır: glioma, meningioma, pituitary ve tümör içermeyen (no tumor) görüntüler. Çalışmada kullanılan veri seti, literatürde yaygın biçimde tercih edilen ve açık erişimli olarak Kaggle platformunda sunulan *Brain Tumor MRI Dataset*'tir (Kaggle, 2022). Güncel çalışmalarda da benzer biçimde açık erişimli MRI veri setlerinin, derin öğrenme tabanlı beyin tümörü sınıflandırma problemlerinde karşılaştırmalı analizler için uygun bir temel sunduğu ifade edilmektedir (Aurna vd., 2022; Sultan vd., 2023).

Bu çalışma kapsamında problem, iki boyutlu (2D) MRI görüntülerinin sınıflandırılması ile sınırlandırılmıştır. Tümör segmentasyonu, tümör hacminin nicel olarak hesaplanması, evreleme, çoklu modalite (T1, T2, FLAIR vb.) entegrasyonu ve üç boyutlu (3D) derin öğrenme yaklaşımları araştırma kapsamı dışında bırakılmıştır. Bu tercihin temel nedeni, sınıflandırma performansına odaklanarak farklı CNN mimarilerinin karşılaştırmalı analizinin daha kontrollü ve yorumlanabilir biçimde gerçekleştirilmesidir. Literatürde, sınıflandırma ve segmentasyon problemlerinin metodolojik olarak ayrı değerlendirilmesinin daha sağlıklı sonuçlar ürettiği vurgulanmaktadır (Wadhwa vd., 2019; Bakas vd., 2018).

Önerilen ve karşılaştırılan modeller yalnızca dijital ortamda test edilmiş olup, klinik doğrulama, hasta verileri üzerinde prospektif değerlendirme ve klinisyen geri bildirim bu çalışmanın kapsamı dışında tutulmuştur. Eğitim, doğrulama ve test süreçleri bulut tabanlı Google Colab

ortamında yürütülmüş; tüm deneysel çalışmalar TensorFlow/Keras derin öğrenme kütüphaneleri kullanılarak gerçekleştirilmiştir. Güncel çalışmalarda, bulut tabanlı platformların derin öğrenme deneyleri için tekrarlanabilirlik ve donanım erişimi açısından yaygın biçimde kullanıldığı bildirilmektedir (Xiong vd., 2021; Alzubaidi vd., 2021).

Bu kapsamda çalışma, derin öğrenme tabanlı beyin tümörü sınıflandırma problemini algoritmik performans ve genellenebilirlik açısından ele almakta olup, klinik karar süreçlerine doğrudan müdahale eden veya tedavi planlamasına yönelik uygulamaları içermemektedir. Elde edilen bulguların, klinik ortamlarda kullanılacak karar destek sistemlerine temel oluşturabilecek yönetsel bir çerçeve sunması amaçlanmaktadır.

1.4. Araştırmanın Önemi

Beyin tümörlerinin erken ve doğru teşhisi, hastaların yaşam süresinin uzatılması, tedaviye yanıtın artırılması ve uygun klinik müdahale stratejilerinin belirlenmesi açısından kritik bir öneme sahiptir. MRI tekniklerinin yaygınlaşmasıyla birlikte klinik ortamlarda üretilen görüntü sayısındaki hızlı artış, radyologlar üzerindeki iş yükünü önemli ölçüde artırmakta ve bu durum tanı sürecinde gözden kaçma ya da yorum farklılıkları gibi insan kaynaklı hataların ortaya çıkma riskini yükseltebilmektedir (Pereira vd., 2016). Bu bağlamda, tanı sürecini destekleyici, otomatik ve nesnel yöntemlere duyulan ihtiyaç giderek daha belirgin hale gelmektedir.

Son yıllarda yapay zeka ve özellikle derin öğrenme tabanlı yöntemler, tıbbi görüntü analizinde önemli bir dönüşüm sağlamış; karmaşık görüntü örüntülerinin öğrenilmesi, sınıflandırma doğruluğunun artırılması ve daha tutarlı sonuçların elde edilmesi açısından güçlü bir araç haline gelmiştir. Derin öğrenme temelli yaklaşımlar, tanı süresinin kısaltılması, gözlemciye bağlı değişkenliğin azaltılması ve tekrarlanabilir sonuçların elde edilmesi gibi avantajlar sunarak klinik karar destek sistemlerinin etkinliğini artırma potansiyeline sahiptir (Litjens vd., 2017; Sadr vd., 2025).

Bu tez çalışmasının önemi, birden fazla CNN mimarisinin karşılaştırmalı olarak değerlendirilmesi yoluyla, yüksek sınıflandırma başarımı sunarken aynı zamanda hesaplama maliyeti açısından verimli olan modellerin ortaya konulmasına katkı sağlamasından kaynaklanmaktadır. Özellikle MobileNetV2 ve EfficientNet gibi hafif ve ölçeklenebilir mimarilerin incelenmesi, sınırlı donanım kaynaklarına sahip sistemlerde dahi derin öğrenme

tabanlı tanı yaklaşımlarının uygulanabilirliğini artırma açısından önem taşımaktadır. MobileNetV2 mimarisi, ters artık (inverted residual) yapıları ve doğrusal dar boğaz (linear bottleneck) katmanları sayesinde düşük hesaplama maliyetiyle etkili özellik çıkarımı sağlarken (Sandler vd., 2018), EfficientNet mimarisi dengeli model ölçekleme yaklaşımıyla doğruluk ve verimlilik arasında güçlü bir denge sunmaktadır (Tan ve Le, 2019; Zulfiqar vd., 2023).

Ayrıca bu çalışma, K-Fold çapraz doğrulama yöntemi ile veri artırma tekniklerini birlikte kullanarak geliştirilen modellerin genellenebilirliğini güçlendirmeyi amaçlamaktadır. Bu yaklaşım, yalnızca tek bir eğitim–test ayırımına dayanan çalışmalara kıyasla daha güvenilir, istikrarlı ve istatistiksel açıdan anlamlı performans değerlendirmelerinin yapılmasına olanak tanımaktadır (McDermott vd., 2021; Yadav ve Shukla, 2016). Bu yönüyle tez çalışması, elde edilen bulguların belirli bir veri kümesine özgü kalmamasını ve farklı hasta gruplarına ait MRI görüntüleri üzerinde de benzer performans gösterebilecek yapay zeka tabanlı karar destek sistemlerinin geliştirilmesine yöntemsel bir katkı sunmayı hedeflemektedir.

1.5. Varsayımlar ve Sınırlılıklar

Varsayımlar

- Çalışmada kullanılan MRI görüntülerinin doğru ve güvenilir biçimde etiketlendiği,
- Veri setinde yer alan glioma, meningioma, pituitary ve no tumor sınıflarının birbirinden ayırt edilebilir görsel özellikler içerdiği,
- Eğitim ve test kümeleri arasında herhangi bir veri örtüşmesi bulunmadığı,
- Kullanılan Keras tabanlı derin öğrenme modellerinin uygun hiperparametreler ile optimize edildiği ve eğitim süreçlerinin yeterli epoch sayısı ile gerçekleştirildiği,
- Veri ön işleme ve artırma adımlarının model performansını olumsuz etkileyecek yanlılıklara neden olmadığı,
- Veri artırma tekniklerinin, orijinal görüntülerin patolojik özelliklerini bozmadan örneklem çeşitliliğini temsil ettiği varsayılmıştır.

Sınırlılıklar

- Çalışmada yalnızca MRI görüntüleri kullanılmış; Bilgisayarlı Tomografi (BT) veya Pozitron Emisyon Tomografisi (PET) gibi diğer görüntüleme modaliteleri kapsam dışı bırakılmıştır.

- Arařtırma yalnızca sınıflandırma problemine odaklanmış olup, tümör segmentasyonu, hacim ölçümü veya evreleme gibi işlemler gerçekleştirilmemiştir.
- Kullanılan veri setinin tek bir açık erişimli kaynaktan elde edilmiş olması nedeniyle, farklı merkezlerden toplanan verilerde model performansının değişebileceği öngörülmektedir.
- Geliştirilen modeller klinik ortamda gerçek hasta verileri üzerinde test edilmemiş ve klinik doğrulama süreci bu çalışmaya dahil edilmemiştir.



2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

2.1. Beyin Tümörleri ve Manyetik Rezonans Görüntüleme

Beyin tümörleri, merkezi sinir sisteminde anormal hücrel proliferasyon sonucu oluşan yapılar olarak tanımlanmaktadır ve klinik olarak malign veya benign karakter gösterebilmektedir (Louis vd., 2016). Histopatolojik sınıflandırma, tümörün hücrel kökenine, genetik profiline ve mikroskopik görünümüne dayalı olarak yapılmaktadır. En yaygın primer beyin tümörü alt tipleri arasında gliomalar, meningiomalar ve pituiter adenomalar yer almaktadır. Gliomalar glial hücre kökenli olup, genellikle daha invaziv ve malign seyirli tümörlerdir; özellikle glioblastoma multiforme alt grubu, invaziv büyüme ve kötü prognozla ilişkilidir (Goodenberger ve Jenkins, 2012; Louis vd., 2016). Meningiomalar meninkslerden türeyen genellikle benign lezyonlardır, fakat büyük hacimli olduklarında lokal kompresyona neden olabilirler. Pituiter adenomalar ise hipofiz bezinde ortaya çıkar ve endokrin dengesizliklere yol açarak klinik belirtiler oluşturabilir (Rogers vd., 2015).

Tıbbi görüntüleme yöntemleri, beyindeki bu tür patolojilerin erken dönemde tespiti, sınıflandırılması ve tedavi planlaması için hayati öneme sahiptir. Bu bağlamda, MRI özellikle nörolojik patolojilerin değerlendirilmesinde tercih edilen modalitedir (Wadhwa, Bhardwaj ve Verma, 2019). MRI, vücut dokularını güçlü manyetik alan ve radyo frekans (RF) darbeleri aracılığıyla görüntüleyerek yüksek yumuşak doku kontrastı sağlar ve iyonlaştırıcı radyasyon içermez. Bu özellik, özellikle beyin gibi hassas organların non-invaziv incelenmesinde önemli avantajlar sunar.

MRI protokolleri; T1-ağırlıklı, T2-ağırlıklı ve FLAIR (Fluid Attenuated Inversion Recovery) gibi sekansları içermektedir. T1 sekansları anatomi ve lezyon sınırlarının net gösterimi için kullanılırken, T2 ve FLAIR sekansları ödem, sıvı bileşenleri ve tümör infiltrasyonunun daha ayrıntılı değerlendirilmesine olanak tanır (Louis vd., 2016; Wadhwa vd., 2019). Özellikle FLAIR sekansı, serebrospinal sıvı sinyallerini baskılayarak periventriküler bölgelerdeki patolojik hiperintensiteleri daha belirgin hale getirir; bu durum glioma gibi infiltratif tümörlerin sınırlanmasında klinik olarak yüksek değere sahiptir.

Beyin tümörlerinin MRI ile değerlendirilmesi, uygun tedavi stratejilerinin belirlenmesinde kritik bir basamaktır. Tedavi kararı genellikle multidisipliner ekiplerce verilir; cerrahi,

radyoterapi ve kemoterapi gibi modaliteler tümörün tipi, yerleşimi ve biyolojik davranışına göre planlanır. Ancak MRI görüntülerinin manuel değerlendirilmesi, radyologun deneyimine dayalıdır ve bu süreç zaman alıcı olabilmektedir. Manuel değerlendirme, subjektif yoruma açık olduğundan tanı varyabilitesine yol açabilmekte ve özellikle yoğun klinik yük altında hata riskini artırabilmektedir (Litjens vd., 2017).

Bu bağlamda klinik uygulamalarda bilgisayar destekli tanı (Computer-Aided Diagnosis – CAD) sistemlerine olan talep her geçen yıl artmaktadır. CAD, görüntü tabanlı tanı süreçlerini otomatikleştirmek ve radyolog performansını artırmak amacıyla tasarlanmıştır. Derin öğrenme ve makine öğrenmesi temelli CAD yaklaşımları, MRI görüntülerinde tümör tespit ve sınıflandırma performansını artırmada önemli ilerlemeler göstermiştir. Örneğin, son dönemde geliştirilen CNN ve hibrit derin öğrenme modelleri, glioma, meningioma ve pituitar tümörleri yüksek doğrulukla sınıflandırabilmektedir (Alzubaidi vd., 2021; Batool ve Byun, 2025; Wong vd., 2025). Bu sistemler, radyologların klinik karar süreçlerine destek sağlamanın yanında tanı sürecini hızlandırmak ve tekrar edilebilirlik düzeyini artırmak için kullanılmaktadır.

Derin öğrenme temelli CAD sistemlerinin performansı, görüntü ön işleme teknikleri (normalizasyon, veri çoğaltma), model mimarisi ve eğitim stratejilerine bağlı olarak değişmektedir. Ayrıca, açıklanabilir yapay zeka yaklaşımları, model kararlarının klinik yorumlanabilirliğini artırmak için kullanılmaktadır (Aurna vd., 2022; Sadr vd., 2025). Bu eğilim, yapay zeka destekli sistemlerin klinik kabulünde önemli bir rol oynamaktadır, çünkü radyologların model çıktılarının başarısının ve sınırlamalarının daha iyi anlaşılmasına olanak tanır.

MRI beyin tümörlerinin tanı ve sınıflandırılmasında altın standart görüntüleme yöntemi olmaya devam etmektedir. Bununla birlikte, CAD ve yapay zeka destekli otomatik değerlendirme yöntemleri, klinik tanı süreçlerinin hızını, doğruluğunu ve tekrarlanabilirliğini artırmak üzere nöro-onkolojik görüntüleme entegre çözümler olarak literatürde geniş yer bulmaktadır.

2.2. Bilgisayar Destekli Tanı ve Yapay Zeka Yaklaşımları

CAD sistemleri, medikal görüntülerden elde edilen sayısal verilerin analiz edilerek hastalıkların otomatik ya da yarı otomatik biçimde tespit edilmesini, sınıflandırılmasını ve yorumlanmasını amaçlayan sistemlerdir. CAD yaklaşımları, özellikle radyolojik görüntüleme alanında klinik

karar verme sürecini desteklemek üzere geliştirilmiş olup, tanı doğruluğunu artırmayı ve yorumlayıcıya bağlı değişkenliği azaltmayı hedeflemektedir (Kumar vd., 2017; Litjens vd., 2017). Klasik bir CAD sisteminin temel bileşenleri; görüntü ön işleme, öznitelik (özellik) çıkarımı ve sınıflandırma aşamalarından oluşmaktadır.

Geleneksel CAD sistemlerinde, görüntülerden çıkarılan öznitelikler genellikle istatistiksel, şekil tabanlı, doku (texture) tabanlı veya frekans alanı özelliklerinden oluşmaktadır. Bu özellikler daha sonra Destek Vektör Makineleri (Support Vector Machines – SVM), Rastgele Orman (Random Forest – RF), K-En Yakın Komşu (K-Nearest Neighbor – KNN) ve Naive Bayes gibi klasik makine öğrenmesi algoritmaları ile sınıflandırılmaktadır (Kumar vd., 2017). Bu yöntemler, sınırlı veri setlerinde ve iyi tanımlanmış problemler için başarılı sonuçlar verebilmekle birlikte, elle öznitelik çıkarımına olan bağımlılık önemli bir dezavantaj olarak öne çıkmaktadır. Özellik çıkarım süreci, alan uzmanlığı gerektirmekte ve seçilen özelliklerin genellenebilirliği sınırlı kalabilmektedir.

Medikal görüntülerin karmaşık yapısı, geleneksel makine öğrenmesi yöntemlerinin yüksek boyutlu veriler üzerinde sınırlı performans sergilemesine neden olmuştur. Bu durum, özellikle beyin MRI görüntülerinde tümör sınırlarının belirsizliği, doku heterojenliği ve sınıflar arası benzerlik gibi faktörlerden kaynaklanmaktadır. Bu bağlamda, elle tasarlanan özelliklerin, tümörün biyolojik ve yapısal varyasyonlarını yeterince temsil edemediği literatürde sıklıkla vurgulanmaktadır (Litjens vd., 2017; Pereira vd., 2016).

Son yıllarda yapay zeka alanında yaşanan gelişmeler, CAD sistemlerinde derin öğrenme tabanlı yaklaşımların yaygın olarak kullanılmasına olanak sağlamıştır. Derin öğrenme, çok katmanlı yapay sinir ağları aracılığıyla verinin hiyerarşik temsillerini öğrenebilen bir makine öğrenmesi alt alanıdır. Bu yaklaşım, özellikle büyük veri kümeleri üzerinde karmaşık örüntülerin otomatik olarak keşfedilmesini mümkün kılmaktadır. Derin öğrenme modelleri, öznitelik çıkarımı ve sınıflandırma adımlarını tek bir uçtan uca (end-to-end) yapı içerisinde birleştirerek geleneksel CAD yaklaşımlarına kıyasla daha yüksek performans sunmaktadır.

Bu kapsamda, CNN, görüntü tabanlı tanı problemlerinde en yaygın kullanılan derin öğrenme mimarileri arasında yer almaktadır. CNN'ler, evrişim (convolution), havuzlama (pooling) ve tam bağlı (fully connected) katmanlar aracılığıyla görüntülerdeki uzamsal ilişkileri ve yerel örüntüleri etkin biçimde öğrenebilmektedir (Krizhevsky vd., 2012). Özellikle beyin MRI

görüntülerinde CNN tabanlı modeller, tümörlü ve sağlıklı dokular arasındaki karmaşık farkları insan müdahalesine gerek duymadan ayırt edebilme yeteneği göstermektedir.

CNN mimarilerinin en önemli avantajlarından biri, özneliklerin otomatik olarak öğrenilmesi ve modelin eğitim süreci boyunca bu özellikleri optimize edebilmesidir. Bu durum, manuel özellik mühendisliğine olan ihtiyacı büyük ölçüde ortadan kaldırmakta ve CAD sistemlerinin genellenebilirliğini artırmaktadır (Zeineldin vd., 2020). Ayrıca, son yıllarda geliştirilen derin CNN mimarileri (örneğin ResNet, DenseNet, EfficientNet) ve transfer öğrenme yaklaşımları, sınırlı sayıda medikal görüntü içeren veri setlerinde dahi yüksek sınıflandırma başarımı elde edilmesini sağlamıştır (Alzubaidi vd., 2021; Batool ve Byun, 2025; Wong vd., 2025).

Derin öğrenme tabanlı CAD sistemleri, yalnızca tümör tespiti ve sınıflandırmasıyla sınırlı kalmayıp, aynı zamanda segmentasyon, tümör derecelendirme ve tedavi yanıtının öngörülmesi gibi ileri klinik görevlerde de kullanılmaktadır. Bununla birlikte, bu sistemlerin klinik ortamlarda yaygın olarak benimsenebilmesi için model çıktılarının yorumlanabilirliği ve güvenilirliği büyük önem taşımaktadır. Bu nedenle son dönemde açıklanabilir yapay zeka (Explainable AI – XAI) yaklaşımları, derin öğrenme tabanlı CAD sistemlerine entegre edilerek klinisyenlerin model kararlarını daha iyi anlayabilmesi hedeflenmektedir (Aurna vd., 2022; Sadr vd., 2025).

CAD sistemleri medikal görüntü analizinde önemli bir dönüşüm geçirmiş; geleneksel makine öğrenmesi yaklaşımlarından derin öğrenme tabanlı, uçtan uca öğrenme sistemlerine doğru evrilmiştir. Özellikle CNN tabanlı yapay zeka modelleri, beyin tümörü gibi karmaşık patolojilerin analizinde yüksek doğruluk, azaltılmış insan bağımlılığı ve artan klinik verimlilik sağlaması nedeniyle literatürde geniş kabul görmektedir.

2.3. Evrişimli Sinir Ağları ve Görüntü Sınıflandırma

CNN, özellikle görüntü verilerinin analizi için tasarlanmış, çok katmanlı derin öğrenme mimarileridir. Bu ağlar temel olarak evrişim (convolution), havuzlama (pooling) ve tam bağlantılı (fully connected) katmanlardan oluşmaktadır. CNN mimarilerinin temel amacı, giriş görüntüsündeki uzamsal ilişkileri koruyarak hiyerarşik özellik temsilleri öğrenmektir. Evrişim katmanları, görüntü üzerinde kayan filtreler aracılığıyla kenar, köşe, doku ve şekil gibi düşük seviyeli görsel özellikleri otomatik olarak öğrenirken; ağ derinliği arttıkça bu özellikler daha

soyut, karmaşık ve sınıflandırıcı nitelikte temsillere dönüşmektedir (Albawi vd., 2017; Goodfellow vd., 2016).

Havuzlama katmanları, evrişim katmanlarından elde edilen öznelik haritalarının boyutunu azaltarak hesaplama maliyetini düşürmekte ve modele çeviri değişmezliği (translation invariance) kazandırmaktadır. Tam bağlantılı katmanlar ise öğrenilen yüksek seviyeli özellikleri kullanarak sınıflandırma kararını vermektedir. Bu mimari yapı sayesinde CNN modelleri, tıbbi görüntülerde sıkça karşılaşılan doku heterojenliği, sınıf içi varyasyon ve düşük kontrastlı sınırlar gibi zorlukların üstesinden gelmede yüksek başarı göstermektedir (Pereira vd., 2016; Zeineldin vd., 2020).

CNN tabanlı görüntü sınıflandırmanın gelişiminde önemli bir dönüm noktası, Krizhevsky ve diğerleri tarafından 2012 yılında geliştirilen AlexNet modeli olmuştur. AlexNet, ImageNet Large Scale Visual Recognition Challenge (ILSVRC) yarışmasında elde ettiği üstün performansla, derin öğrenmenin görüntü sınıflandırma alanında yaygın biçimde benimsenmesini sağlamıştır (Krizhevsky vd., 2012). Bu çalışmada ReLU aktivasyon fonksiyonunun kullanımı, veri çoğaltma (data augmentation) ve GPU tabanlı eğitim gibi yenilikler, derin CNN'lerin etkin biçimde eğitilebileceğini göstermiştir.

AlexNet'i takiben geliştirilen VGG16 mimarisi, daha derin ağların küçük boyutlu (3×3) evrişim filtreleri kullanılarak oluşturulabileceğini göstermiştir (Simonyan ve Zisserman, 2015). VGG tabanlı modeller, mimari sadeliği ve yüksek temsil gücü sayesinde tıbbi görüntü sınıflandırma çalışmalarında sıklıkla tercih edilmiştir. Bununla birlikte, artan katman sayısı hesaplama maliyetini ve aşırı uyum (overfitting) riskini artırmıştır.

Bu sınırlamaların aşılması amacıyla geliştirilen ResNet mimarisi, artık (residual) bağlantılar kullanarak derin ağların daha kararlı biçimde eğitilmesini sağlamıştır (He vd., 2016). ResNet50 gibi modeller, gradyan kaybolması problemini büyük ölçüde ortadan kaldırarak yüzlerce katman derinliğine sahip ağların eğitilmesine olanak tanımıştır. Bu özellik, özellikle beyin MRI görüntülerinde tümörlü ve tümörsüz dokular arasındaki ince yapısal farkların öğrenilmesinde önemli avantajlar sunmaktadır.

Benzer şekilde, DenseNet mimarisi, her katmanın kendisinden önceki tüm katmanlarla bağlantılı olduğu yoğun bağlantı (dense connectivity) yapısını kullanarak özellik tekrarını

azaltmış ve bilgi akışını güçlendirmiştir (Huang vd., 2017). DenseNet121 gibi modeller, daha az parametre ile yüksek sınıflandırma performansı sunması nedeniyle medikal görüntü analizinde yaygın biçimde kullanılmaktadır.

Hesaplama verimliliği ve taşınabilir uygulamalar açısından geliştirilen MobileNetV2, derinlik ayırık evrişimler (depthwise separable convolutions) ve ters artık bloklar (inverted residuals) kullanarak düşük donanım gereksinimiyle etkili performans elde etmeyi amaçlamaktadır (Sandler vd., 2018). Bu mimari, klinik karar destek sistemlerinin gerçek zamanlı uygulamalarında önemli bir avantaj sağlamaktadır.

Daha yakın dönemde önerilen EfficientNet ailesi ise ağ derinliği, genişliği ve giriş çözünürlüğünü dengeli biçimde ölçeklendiren bileşik ölçekleme (compound scaling) yaklaşımıyla dikkat çekmiştir (Tan ve Le, 2019). EfficientNetB0 gibi modeller, daha az parametre kullanarak yüksek doğruluk elde edebilmesi sayesinde, sınırlı veri setlerine sahip tıbbi görüntüleme problemlerinde etkin çözümler sunmaktadır.

CNN tabanlı görüntü sınıflandırma yaklaşımları, beyin tümörü MRI görüntülerinde tümör tipi sınıflandırma, malignite derecesi belirleme ve çok sınıflı tanı problemleri için yaygın biçimde kullanılmaktadır. Literatürde, CNN modellerinin geleneksel makine öğrenmesi yöntemlerine kıyasla daha yüksek doğruluk, duyarlılık ve F1-skoru değerleri sunduğu çok sayıda çalışma ile gösterilmiştir (Alzubaidi vd., 2021; Batoool ve Byun, 2025; Wong vd., 2025). Bununla birlikte, veri dengesizliği, sınırlı örnek sayısı ve model yorumlanabilirliği gibi sorunlar, güncel araştırmaların odak noktaları arasında yer almaktadır.

CNN, otomatik öznitelik öğrenme yetenekleri ve yüksek temsil gücü sayesinde görüntü sınıflandırma problemlerinde güçlü bir araç olarak öne çıkmaktadır. Özellikle beyin MRI görüntülerinde CNN tabanlı yaklaşımlar, CAD sistemlerinin temel bileşeni haline gelmiş ve klinik karar süreçlerine entegrasyon potansiyeli göstermektedir.

2.4. Transfer Öğrenme

Transfer öğrenme (transfer learning), bir problem alanında öğrenilen bilgi ve temsilin, farklı fakat ilişkili bir problem alanına aktarılması prensibine dayanan bir derin öğrenme yaklaşımıdır. Bu yöntemde, genellikle ImageNet gibi büyük ölçekli ve çeşitli veri kümeleri

üzerinde önceden eğitilmiş CNN modelleri, hedef probleme uyarlanarak yeniden eğitilmektedir. Transfer öğrenme, özellikle modelin erken katmanlarında öğrenilen genel görsel özniteliklerin (kenar, doku, temel şekiller) farklı görüntüleme problemleri arasında paylaşılabilir olmasına dayanmaktadır (LeCun vd., 2015; Yosinski vd., 2014).

Bu yaklaşımın en önemli avantajları arasında eğitim süresinin kısılması, hesaplama maliyetinin azalması ve sınırlı veri setlerinde dahi yüksek genelleme performansı elde edilmesi yer almaktadır. Tıbbi görüntüleme alanında kullanılan veri setleri, çoğu zaman etik, maliyet ve hasta gizliliği gibi nedenlerle sınırlı sayıda örnek içermektedir. Bu durum, derin öğrenme modellerinin sıfırdan (from scratch) eğitilmesini zorlaştırmakta ve aşırı uyum riskini artırmaktadır. Transfer öğrenme, bu temel problemi etkin biçimde ele alarak medikal görüntü analizinde yaygın biçimde benimsenmiştir (Rehman vd., 2021; Sajjad vd., 2019).

Transfer öğrenme uygulamalarında yaygın olarak iki temel strateji kullanılmaktadır: özellik çıkarıcı (feature extractor) yaklaşımı ve ince ayar (fine-tuning) yaklaşımı. Özellik çıkarıcı yöntemde, önceden eğitilmiş modelin evrişim katmanları sabitlenmekte ve yalnızca sınıflandırma katmanları hedef veri setine göre yeniden eğitilmektedir. İnce ayar yaklaşımında ise modelin üst katmanları kısmen veya tamamen yeniden eğitilerek, öğrenilen temsillerin hedef probleme daha iyi uyum sağlaması amaçlanmaktadır (Yosinski vd., 2014). Literatürde, beyin MRI görüntüleri gibi yüksek düzeyde alan özgüllüğü içeren veriler için ince ayar stratejisinin genellikle daha yüksek performans sunduğu bildirilmektedir (Swati vd., 2019).

Beyin tümörü sınıflandırma çalışmalarında transfer öğrenme yaklaşımının etkinliği çok sayıda çalışma ile gösterilmiştir. Sajjad ve diğerleri (2019), VGG16 mimarisini kullanarak MRI görüntülerinden glioma, meningioma ve pituitar tümörleri sınıflandırmış ve yaklaşık %94 doğruluk elde ettiklerini rapor etmiştir. Bu çalışma, derin CNN modellerinin transfer öğrenme ile medikal görüntülerde yüksek ayırt edicilik sağlayabildiğini ortaya koymuştur. Benzer biçimde, Swati ve diğerleri (2019) tarafından gerçekleştirilen çalışmada, ResNet50 tabanlı bir transfer öğrenme modeli kullanılarak beyin tümörü sınıflandırma probleminde %96'nın üzerinde doğruluk elde edilmiştir. ResNet mimarisinin artık bağlantılar sayesinde derin ağların kararlı biçimde eğitilmesine olanak tanınması, bu başarının temel nedenlerinden biri olarak gösterilmektedir.

Daha yakın dönemde geliştirilen EfficientNet mimarisi, ağ derinliği, genişliği ve giriş çözünürlüğünü dengeli biçimde ölçeklendiren bileşik ölçekleme (compound scaling) yaklaşımı ile dikkat çekmiştir (Tan ve Le, 2019). EfficientNet ailesi, daha az parametre kullanarak yüksek doğruluk elde edebilmesi sayesinde, sınırlı örnek sayısına sahip tıbbi görüntü veri setleri için özellikle uygun bir yapı sunmaktadır. Rehman ve diğerleri (2021), EfficientNet tabanlı transfer öğrenme yaklaşımını MRI görüntülerine uygulamış ve bu mimarinin beyin tümörü sınıflandırma görevlerinde hem yüksek doğruluk hem de hesaplama verimliliği sağladığını göstermiştir.

Literatürde ayrıca transfer öğrenmenin, veri çoğaltma (data augmentation), sınıf dengeleme ve düzenleme (regularization) teknikleriyle birlikte kullanıldığında performans artışının daha da belirginleştiği bildirilmektedir. Bununla birlikte, önceden eğitilmiş modellerin doğal görüntüler (natural images) üzerinde öğrenilmiş temsillere dayanması, tıbbi görüntülerle olan alan farkı (domain shift) nedeniyle bazı sınırlamalar da doğurabilmektedir. Bu nedenle güncel çalışmalarda, alan uyarlama (domain adaptation) ve özelleştirilmiş ince ayar stratejileri, transfer öğrenmenin etkinliğini artırmak amacıyla araştırılmaktadır (Batool ve Byun, 2025;Sadr vd., 2025).

Transfer öğrenme yaklaşımı, beyin tümörü MRI görüntülerinin sınıflandırılmasında yüksek doğruluk, düşük eğitim maliyeti ve sınırlı veri koşullarında güçlü genelleme yeteneği sunması nedeniyle derin öğrenme tabanlı CAD sistemlerinin temel bileşenlerinden biri haline gelmiştir. Bu yöntem, klinik uygulamalara yönelik yapay zeka modellerinin geliştirilmesinde pratik ve etkili bir potansiyel çözüm olarak literatürde geniş kabul görmektedir.

2.5. Veri Artırma ve Düzenleme Teknikleri

Derin öğrenme tabanlı modellerde karşılaşılan en temel sorunlardan biri, özellikle sınırlı veri setleriyle çalışıldığında ortaya çıkan aşırı öğrenme (overfitting) problemidir. Aşırı öğrenme, modelin eğitim verisine özgü örüntüleri ezberlemesi ve daha önce görmediği veriler üzerinde düşük genelleme performansı sergilemesi şeklinde tanımlanmaktadır. Tıbbi görüntüleme alanında, veri toplama sürecinin maliyetli ve kısıtlı olması nedeniyle bu problem daha belirgin hale gelmektedir (McDermott vd., 2021; Shorten ve Khoshgoftaar, 2019).

Bu bağlamda, veri artırma (data augmentation) yöntemleri, aşırı öğrenmenin önüne geçmek için yaygın biçimde kullanılan stratejiler arasında yer almaktadır. Veri artırma, mevcut eğitim görüntülerine çeşitli geometrik ve fotometrik dönüşümler uygulanarak, modelin eğitim sürecinde daha fazla ve çeşitli örnekle karşılaşmasını sağlamayı amaçlamaktadır. Döndürme (rotation), ölçekleme (scaling), kaydırma (translation), yatay/dikey ayna çevirme (flipping), kırpma (cropping) ve gürültü ekleme gibi işlemler, görüntülerin temel yapısal özelliklerini koruyarak veri çeşitliliğini artırmaktadır (Perez ve Wang, 2017; Shorten ve Khoshgoftaar, 2019).

Veri artırma teknikleri, özellikle CNN tabanlı modellerde konum ve yön değişmezliği (invariance) kazandırarak, öğrenilen özneliklerin daha genellenebilir hale gelmesine katkı sağlamaktadır. Beyin MRI görüntülerinde tümörlerin farklı konum, boyut ve şekillerde ortaya çıkabilmesi, bu tür dönüşümlerin etkinliğini daha da artırmaktadır. Literatürde, uygun veri artırma stratejilerinin kullanılmasıyla CNN modellerinin sınıflandırma doğruluğunda anlamlı artışlar elde edildiği birçok çalışma ile gösterilmiştir (Perez ve Wang, 2017; Rehman vd., 2020).

Veri artırmaya ek olarak, düzenleme (regularization) teknikleri, derin öğrenme modellerinin daha dengeli ve kararlı biçimde öğrenmesine katkı sağlamaktadır. Bu teknikler, modelin karmaşıklığını kontrol altına alarak aşırı uyumu azaltmayı ve genelleme performansını artırmayı hedeflemektedir. En yaygın kullanılan düzenleme yöntemleri arasında Dropout ve Batch Normalization yer almaktadır.

Dropout yöntemi, eğitim süreci sırasında ağ içerisindeki nöronların belirli bir oranının rastgele olarak devre dışı bırakılması prensibine dayanmaktadır. Bu yaklaşım, ağın belirli nöronlara aşırı bağımlı hale gelmesini engelleyerek, daha dayanıklı ve genellenebilir temsil öğrenimini teşvik etmektedir (Srivastava vd., 2014). Dropout yöntemi özellikle tam bağlantılı katmanlarda yaygın olarak kullanılmakla birlikte, güncel CNN mimarilerinde evrişim katmanlarında da uygulanabilmektedir.

Batch Normalization ise ağ içerisindeki aktivasyonların dağılımını normalize ederek, katmanlar arası dağılım kaymasını (internal covariate shift) azaltmayı amaçlamaktadır. Bu yöntem, öğrenme sürecini hızlandırmakta, daha yüksek öğrenme oranlarının kullanılmasına olanak tanımakta ve modelin eğitim sürecini daha kararlı hale getirmektedir (Ioffe ve Szegedy, 2015).

Batch Normalization, aynı zamanda düzenleyici bir etki göstererek, aşırı öğrenmenin azaltılmasına dolaylı olarak katkı sağlamaktadır.

Güncel literatürde, veri artırma ve düzenleme tekniklerinin birlikte kullanılmasının, tek başına uygulanan yöntemlere kıyasla daha yüksek performans sağladığı vurgulanmaktadır. Özellikle transfer öğrenme tabanlı CNN modellerinde, veri artırma ile zenginleştirilen eğitim setleri ve Dropout–Batch Normalization kombinasyonu, sınırlı MRI veri kümelerinde yüksek doğruluk, duyarlılık ve F1-skoru değerleri elde edilmesine olanak tanımaktadır (Rehman vd., 2021; Sadr vd., 2025).

Bu tez kapsamında geliştirilen derin öğrenme modellerinde, veri artırma ve düzenleme tekniklerinin tamamı sistematik biçimde uygulanmıştır. Böylece modellerin aşırı öğrenmeye karşı dayanıklılığı artırılmış, eğitim ve test performansları arasındaki fark minimize edilerek daha güvenilir ve genellenebilir sonuçlar elde edilmesi hedeflenmiştir.

2.6. Model Değerlendirme Stratejilerine Kuramsal Bakış

Derin öğrenme tabanlı sınıflandırma modellerinin başarımının değerlendirilmesi, tıbbi görüntü analizi gibi hata toleransının düşük olduğu alanlarda kritik bir aşama olarak kabul edilmektedir. Model performansının yalnızca doğruluk gibi tekil metrikler üzerinden raporlanması, özellikle sınıf dengesizliğinin bulunduğu veri setlerinde yanıltıcı sonuçlara yol açabilmektedir. Bu nedenle literatürde, model değerlendirme sürecinin çok boyutlu ele alınması ve farklı performans ölçütleri ile desteklenmesi gerektiği vurgulanmaktadır (Powers, 2009; Chicco ve Jurman, 2020).

Geleneksel değerlendirme yaklaşımlarında sıklıkla kullanılan tek seferlik eğitim–test bölünmesine dayalı stratejiler, modelin belirli bir veri alt kümesine aşırı uyum sağlaması riskini beraberinde getirmektedir. Özellikle veri setinin görece sınırlı olduğu tıbbi görüntüleme çalışmalarında, bu tür değerlendirme yaklaşımlarının modelin gerçek genellenebilirliğini yansıtmakta yetersiz kaldığı bildirilmektedir (McDermott vd., 2021; Wang vd., 2023). Bu durum, modelin farklı hasta grupları veya farklı görüntüleme protokolleri karşısındaki performansının öngörülememesine neden olabilmektedir.

Literatürde, model değerlendirme stratejilerinin yalnızca performans ölçüm aracı olarak değil, aynı zamanda farklı mimarilerin karşılaştırılabilirliğini sağlayan metodolojik bir bileşen olarak ele alınması gerektiği belirtilmektedir. Özellikle derin öğrenme tabanlı çalışmalarda, farklı ağ mimarilerinin ve eğitim stratejilerinin karşılaştırıldığı araştırmalarda, değerlendirme yaklaşımının tutarlılığı sonuçların güvenilirliği açısından belirleyici olmaktadır (Mohan vd., 2021; Yang vd., 2022).

Bununla birlikte, tıbbi görüntü sınıflandırma problemlerinde kullanılan değerlendirme stratejilerinin klinik gerçeklik ile uyumlu olması da önem taşımaktadır. Klinik uygulamalarda modellerin karşılaşacağı veri çeşitliliği göz önüne alındığında, değerlendirme sürecinin modelin kararlılığını ve istikrarını ortaya koyacak biçimde tasarlanması gerekmektedir. Aksi takdirde, deneysel ortamda yüksek başarı elde eden modellerin gerçek klinik senaryolarda benzer performans sergileyemediği durumlarla karşılaşılabilir (Esteva vd., 2021).

Bu bağlamda model değerlendirme stratejileri, yalnızca bir performans raporlama aşaması değil; geliştirilen yapay zeka modelinin güvenilirliği, genellenebilirliği ve klinik uygulanabilirliği açısından temel bir metodolojik unsur olarak değerlendirilmektedir. Bu nedenle, kullanılan değerlendirme yaklaşımının kuramsal temellerinin açık biçimde ortaya konması, çalışmanın bilimsel geçerliliğini güçlendiren önemli bir unsur olarak öne çıkmaktadır.

2.7. Beyin Tümörü Sınıflandırması Üzerine Yapılan Çalışmalar

Beyin tümörlerinin otomatik olarak sınıflandırılması, son yıllarda CAD sistemleri ile derin öğrenme tabanlı yaklaşımlardaki gelişmelere paralel olarak yoğun biçimde araştırılan bir konu haline gelmiştir. MRI yüksek yumuşak doku kontrastı sunması, bu görüntülerin CNN ile etkin biçimde analiz edilmesini mümkün kılmış; literatürde beyin tümörü sınıflandırmasına yönelik çok sayıda yüksek doğruluklu çalışma rapor edilmiştir.

Sajjad ve diğerleri (2019), VGG19 tabanlı bir CNN mimarisini transfer öğrenme yaklaşımıyla kullanarak beyin tümörü sınıflandırma probleminde yaklaşık %94 doğruluk elde etmiş; çalışmada veri artırma ve ince ayar (fine-tuning) stratejilerinin model performansı üzerindeki olumlu etkileri ayrıntılı biçimde ortaya konulmuştur. Benzer şekilde, Swati ve diğerleri (2019), ResNet50 mimarisini kullanarak glioma, meningioma ve pitüiter tümörlerin sınıflandırılmasında %97'ye varan başarı oranları raporlamış; artık bağlantıların (residual

connections) derin ağların eğitilebilirliğini ve genelleme kabiliyetini önemli ölçüde artırdığını göstermiştir.

Geleneksel CNN mimarilerine alternatif olarak geliştirilen kapsül sinir ağları (Capsule Networks – CapsNet) da beyin tümörü sınıflandırma çalışmalarında değerlendirilmiştir. Afshar ve diğerleri (2020), CapsNet tabanlı bir yaklaşım kullanarak %90'ın üzerinde doğruluk elde etmiş ve bu mimarinin uzamsal hiyerarşileri daha etkili biçimde modelleyebilmesi sayesinde özellikle sınırlı veri setlerinde avantaj sağlayabileceğini belirtmiştir. Bununla birlikte, CapsNet modellerinin yüksek hesaplama maliyeti ve eğitim sürecinin karmaşıklığı, CNN tabanlı modellere kıyasla yaygın kullanımını sınırlayan temel etkenler arasında yer almaktadır.

Literatürde, probleme özgü olarak tasarlanmış CNN mimarilerinin de önemli bir yer tuttuğu görülmektedir. Anaraki ve diğerleri (2019), 11 katmanlı özgün bir CNN mimarisi önererek beyin tümörü sınıflandırma probleminde yaklaşık %95 doğruluk elde etmiş; bu çalışma, görev odaklı mimari tasarımın hazır mimarilere alternatif olarak rekabetçi sonuçlar üretebileceğini ortaya koymuştur.

Sınıflandırma performansını artırmaya yönelik bir diğer önemli yaklaşım, ön işleme ve bölütleme (segmentation) aşamalarının sınıflandırma sürecine entegre edilmesidir. Zeineldin ve diğerleri (2020), FLAIR MRI görüntülerinde çok ölçekli CNN tabanlı bir segmentasyon yöntemi önererek tümör bölgelerinin otomatik ve başarılı biçimde ayrıştırılabileceğini göstermiştir. Bu çalışma, sınıflandırma öncesinde tümör bölgesinin doğru biçimde izole edilmesinin model performansını anlamlı ölçüde artırdığını ortaya koymaktadır. Literatürde, segmentasyon destekli sınıflandırma yaklaşımlarının özellikle heterojen tümör yapılarında daha tutarlı sonuçlar verdiği bildirilmektedir.

Rehman ve diğerleri (2021), farklı derin CNN mimarilerini karşılaştırmalı olarak incelemiş ve MRI görüntülerinden beyin tümörü sınıflandırmasında %95'in üzerinde doğruluk elde edilebildiğini raporlamıştır. Çalışmada ayrıca transfer öğrenme, veri artırma ve düzenleme tekniklerinin birlikte kullanımının performans üzerindeki belirleyici rolü vurgulanmış; EfficientNet mimarisinin daha az parametre ile yüksek doğruluk sunması nedeniyle tıbbi görüntüleme problemleri için uygun bir alternatif olduğu ifade edilmiştir.

Tan ve Le (2019) tarafından önerilen EfficientNet mimarisi, ağ derinliği, genişliği ve giriş çözünürlüğünü dengeli biçimde ölçekleyen bileşik ölçekleme yaklaşımıyla literatürde önemli bir yer edinmiştir. Bu mimarinin tıbbi görüntü sınıflandırma problemlerine uygulanabilirliği sonraki çalışmalarla doğrulanmış; EfficientNet tabanlı modellerin beyin tümörü sınıflandırmasında yüksek performans sergilediği gösterilmiştir (Rehman vd., 2020). Bu bulgular, hesaplama verimliliği ile sınıflandırma başarımı arasında etkili bir denge kurulabileceğini ortaya koymaktadır.

Model değerlendirme ve kararlılık analizleri açısından K-Fold çapraz doğrulama yönteminin kullanımı da literatürde yaygınlaşmıştır. Li ve diğerleri (2021), çapraz doğrulama temelli değerlendirme stratejilerinin özellikle sınırlı veri setlerinde model performansının daha güvenilir biçimde raporlanmasını sağladığını vurgulamış; bu yaklaşımın farklı veri bölümlerine karşı modelin tutarlılığının analiz edilmesine olanak tanıdığını belirtmiştir.

Son yıllarda MobileNetV2 ve EfficientNetB0 gibi hafif (lightweight) CNN mimarilerinin tıbbi görüntü sınıflandırma problemlerinde kullanımında belirgin bir artış gözlenmektedir. Bu mimariler, sınırlı donanım kaynaklarına sahip mobil ve gömülü sistemlerde dahi yüksek doğruluk oranları sunabilmeleri nedeniyle dikkat çekmektedir (Sandler vd., 2018; Tan ve Le, 2019). Klinik karar destek sistemlerinin gerçek zamanlı ve taşınabilir platformlara entegrasyonu açısından bu tür mimariler önemli bir araştırma alanı oluşturmaktadır.

2024 yılında yayımlanan bir çalışmada, Res-BRNet adlı derin kalıntı ve bölgesel CNN mimarisi kullanılarak beyin tümörü MRI sınıflandırmasında doğruluk, duyarlılık ve F1-skoru gibi performans ölçütlerinin yüksek değerlerle rapor edildiği bildirilmiştir. Bu modelin, yerel ve küresel doku özelliklerini birlikte öğrenmesi sayesinde sınıflandırma başarısını artırdığı belirtilmiştir (Hussain vd., 2024).

Aynı yıl yayımlanan bir diğer çalışmada ise birden fazla CNN modelinin hibrit bir yapı altında birleştirildiği çok sınıflı beyin tümörü sınıflandırma yaklaşımı ile yüksek doğruluk oranları elde edildiği raporlanmıştır; büyük ölçekli klinik veri setleri üzerinde modelin genellenebilirliği ayrıntılı biçimde analiz edilmiştir (Celik ve Inik, 2024).

Son yıllarda beyin tümörü sınıflandırmasına yönelik çalışmaların önemli bir bölümü, daha hafif ve optimize edilmiş CNN mimarilerine odaklanmaktadır. Wong ve diğerleri (2025) farklı derin

öğrenme modellerini MRI görüntülerine uygulayarak yüksek doğruluk oranları elde etmiş ve transfer öğrenmenin sınırlı veri koşullarında etkin bir çözüm sunduğunu göstermiştir. Benzer şekilde, Batoool ve Byun (2025) tarafından önerilen çok yollu hafif CNN mimarisi, optimal özellik seçimi ile çok sınıflı beyin tümörü sınıflandırmasında yüksek performans sağlamıştır. Ayrıca Sadr ve diğerleri (2025), derin öğrenme tabanlı bir yaklaşım ile MRI görüntülerinde tümör sınıflandırma doğruluğunu artıran gelişmiş bir model önermiştir.

Derin öğrenme tabanlı yaklaşımlar, sadece tümör tespiti değil, aynı zamanda tümörün evrelerinin belirlenmesinde de kritik bir rol oynamaya başlamıştır. Literatürdeki en güncel çalışmalardan biri olan Yang ve diğerleri (2025), geliştirdikleri 'BrainCNN' adlı özelleştirilmiş mimari ile beyin tümörlerinin otomatik olarak derecelendirilmesinde (grading) yüksek başarı elde etmişlerdir. Bu çalışma, karmaşık MRI görüntülerinin analizinde modele özgü iyileştirmelerin teşhis doğruluğunu doğrudan artırdığını kanıtlamaktadır.

Genel olarak değerlendirildiğinde, CNN ve transfer öğrenme tabanlı yaklaşımların beyin tümörü sınıflandırma problemlerinde yüksek doğruluk ve güçlü genelleme yeteneği sunduğu görülmektedir. Bununla birlikte, veri dengesizliği, model açıklanabilirliği ve klinik güvenilirlik gibi konular güncel çalışmaların odak noktaları arasında yer almaya devam etmektedir (Li vd., 2021; McDermott vd., 2021). Bu tez çalışması, literatürde rapor edilen yaklaşımlar doğrultusunda farklı CNN mimarilerinin performansını sistematik ve karşılaştırmalı bir biçimde değerlendirerek alana katkı sunmayı amaçlamaktadır.

2.8. Literatür Değerlendirmesi ve Boşluk Analizi

Beyin tümörü sınıflandırmasına yönelik literatür incelendiğinde, derin öğrenme tabanlı yaklaşımların yüksek doğruluk oranları sunduğu görülmekle birlikte, yöntemsel açıdan bazı sınırlılıkların devam ettiği dikkat çekmektedir. Özellikle çok sayıda çalışmada, sınırlı büyüklükteki veri kümeleri üzerinde yalnızca tek bir eğitim–test bölünmesine dayalı değerlendirmelerin yapıldığı rapor edilmektedir. Bu durum, elde edilen performans sonuçlarının veri setine özgü kalmasına ve modelin gerçek klinik veriler üzerindeki genellenebilirliğinin yeterince test edilememesine yol açabilmektedir (McDermott vd., 2021; Mohan vd., 2021; Wang vd., 2023).

Tek bölünmeye dayalı değerlendirme stratejileri, özellikle sınıf dağılımının dengesiz olduğu tıbbi görüntü veri setlerinde, performans ölçütlerinin olduğundan daha yüksek veya düşük raporlanmasına neden olabilmektedir. Bu bağlamda, modelin farklı veri alt kümeleri üzerindeki tutarlılığını ve kararlılığını değerlendirebilecek yöntemlerin kullanılması gerekliliği literatürde açık biçimde vurgulanmaktadır. K-Fold çapraz doğrulama yöntemi, her veri örneğinin hem eğitim hem de test sürecinde yer almasını sağlayarak bu soruna etkin bir çözüm sunmaktadır ve son yıllarda yapılan çalışmalarda giderek daha fazla benimsenmektedir (Kohavi, 1995; Rehman vd., 2021).

Literatürde dikkat çeken bir diğer sınırlılık, birçok çalışmanın yalnızca tek bir CNN mimarisine odaklanması ve alternatif mimarilerle sistematik bir karşılaştırma sunmamasıdır. Farklı mimarilerin aynı veri seti ve aynı değerlendirme protokolü altında karşılaştırılmaması, mimariler arası performans farklarının nesnel biçimde yorumlanmasını güçleştirmektedir. Ayrıca, hazır ve önceden eğitilmiş modellerin yaygın kullanımına karşın, özgün olarak tasarlanmış CNN mimarilerinin veya hibrit yapıların sınırlı sayıda çalışmada ele alındığı görülmektedir (Anaraki vd., 2019; Mohan vd., 2021).

Bunun yanı sıra, literatürde rapor edilen yüksek doğruluk oranlarına rağmen, bazı çalışmaların model kararlılığı, veri dengesizliği ve değerlendirme stratejileri açısından yeterli ayrıntıyı sunmadığı dikkat çekmektedir. Bu durum, farklı çalışmaların sonuçlarının doğrudan karşılaştırılmasını zorlaştırmakta ve alandaki bilgi birikiminin sistematik biçimde ilerlemesini sınırlamaktadır. Güncel araştırmalar, bu eksikliklerin giderilmesi için standartlaştırılmış değerlendirme protokollerinin ve çoklu doğrulama stratejilerinin önemine işaret etmektedir.

Bu tez çalışması, literatürde tanımlanan yöntemsel boşlukları ele almayı amaçlamaktadır. Çalışma kapsamında, farklı CNN tabanlı mimariler aynı veri seti üzerinde, aynı ön işleme ve değerlendirme koşulları altında karşılaştırmalı olarak eğitilmiş ve test edilmiştir. Model performansının daha güvenilir ve genellenebilir biçimde değerlendirilmesi amacıyla beş katlı çapraz doğrulama yöntemi kullanılmıştır. Bu yaklaşım, elde edilen sonuçların tek bir veri bölünmesine bağlı kalmasını engelleyerek istatistiksel açıdan daha sağlam çıkarımlar yapılmasına olanak tanımaktadır.

Ayrıca bu çalışmada, literatürde yaygın olarak kullanılan hazır mimarilere ek olarak, özgün olarak tasarlanan bir CNN modeli ve hibrit bir yapı da önerilerek performans iyileştirmesi

hedeflenmiştir. Bu yönüyle çalışma, yalnızca mevcut yaklaşımları uygulamakla kalmayıp, mimari tasarım ve yöntemsel değerlendirme açısından da literatüre katkı sunmaktadır. Elde edilen bulguların, beyin tümörü sınıflandırmasına yönelik CAD sistemlerinin geliştirilmesine hem teorik hem de uygulamalı düzeyde katkı sağlaması beklenmektedir.

Sonuç olarak, bu tez çalışması; kapsamlı literatür değerlendirmesi, çoklu CNN mimarilerinin karşılaştırmalı analizi, K-Fold çapraz doğrulama temelli değerlendirme stratejisi ve özgün model önerileri ile mevcut çalışmalarda belirlenen yöntemsel boşlukları ele almakta ve alandaki güncel araştırmalara sistematik ve güvenilir bir katkı sağlaması beklenmektedir (Batool ve Byun, 2025; Sadr vd., 2025).



3. MATERYAL VE YÖNTEM

3.1. Araştırma Modeli ve Yaklaşım

Bu çalışma, deneysel nitelikte bir araştırma olup, beyin tümörlerinin MRI görüntülerinden derin öğrenme yöntemleri kullanılarak sınıflandırılmasına yöneliktir. Araştırma modeli, nicel veri analizi ve denetimli öğrenme (supervised learning) yaklaşımına dayanmaktadır.

Çalışmada MobileNetV2, EfficientNetB0, VGG16, ResNet50 ve DenseNet121 gibi Keras Applications tabanlı derin öğrenme mimarilerinin yanı sıra, özgün olarak tasarlanan bir CNN modeli ile bu modelin MobileNetV2 tabanlı özellik çıkarım katmanlarıyla birleştirildiği hafif bir hibrit yapının performansları karşılaştırılmıştır. Tüm modeller aynı veri seti ve aynı ön işleme adımları kullanılarak eğitilmiş, böylece modeller arası karşılaştırmanın nesnel biçimde yapılması sağlanmıştır (Mohan vd., 2021).

Araştırmanın genel süreci aşağıdaki adımlardan oluşmaktadır:

1. Veri setinin hazırlanması
2. Görsellerin yeniden boyutlandırılması ve normalizasyonu
3. Veri artırma işlemlerinin uygulanması
4. Modellerin eğitilmesi
5. Model performanslarının değerlendirilmesi
6. Modeller arası karşılaştırma yapılması

Bu süreç, tıbbi görüntü sınıflandırma çalışmalarında yaygın olarak benimsenen derin öğrenme tabanlı modelleme adımlarına dayanmaktadır (Alzubaidi vd., 2021; Litjens vd., 2017).

3.2. Veri Seti

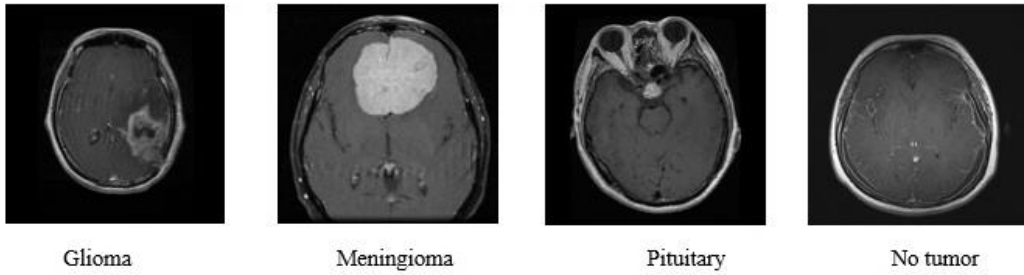
Bu çalışmada Kaggle platformunda paylaşılan Brain Tumor MRI Dataset (Kaggle, 2022) veri kümesinden yararlanılmıştır. Veri seti; glioma, meningioma, pituitary ve no tumor olmak üzere dört farklı sınıf altında toplanmış toplam 7.023 adet görüntüden oluşmaktadır (Şekil 1). Tüm görüntüler 512×512 piksel çözünürlüğe sahip olup, farklı düzlemlerden ve çekim protokollerinden elde edilen MRI kesitlerini içermektedir. Sınıflar arası dağılım nispeten dengeli olup her bir tümör türü için yaklaşık 1.700 görüntü bulunmaktadır (Tablo 1).

Tablo 1. Veri seti sınıf dağılımı ve görüntü sayıları

Sınıf Etiketi	Görüntü Sayısı	Açıklama
Glioma	1621	Malign glial hücre kökenli tümörler
Meningioma	1645	Meninks tabakasından köken alan tümörler
Pituitary	1757	Hipofiz bezi kaynaklı adenomalar
No Tumor	2000	Tümör bulgusu içermeyen normal kesitler
Toplam	7023	

Veri seti %80 eğitim ve %20 test olacak şekilde iki alt kümeye ayrılmıştır. Eğitim kümesinin %20'lik bölümü ayrıca doğrulama (validation) alt kümesi olarak kullanılmıştır. Bu yaklaşım sayesinde modelin aşırı öğrenme eğilimi azaltılmıştır.

Benzer veri kümeleri üzerinde gerçekleştirilen çalışmalar, bu tür bir sınıf dağılımının modelin farklı tümör tipleri arasında ayırt edici özellikleri öğrenebilmesi için yeterli çeşitlilik sağladığını göstermektedir (Sajjad vd., 2019; Swati vd., 2019).

**Şekil 1.** Veri seti görüntüleri

3.3. Veri Ön İşleme ve Artırma Teknikleri

MRI görüntüleri üzerinde öncelikle normalizasyon işlemi uygulanarak piksel değerleri 0–1 aralığına dönüştürülmüştür (rescale = 1/255). Bu işlem, ağız eğitim sürecinde daha kararlı biçimde öğrenmesini sağlamaktadır (Albawi vd., 2017).

Modelin genellenebilirliğini artırmak ve aşırı öğrenmeyi önlemek amacıyla veri artırma (data augmentation) tekniklerinden yararlanılmıştır. Bu işlemler Keras kütüphanesinde yer alan

ImageDataGenerator sınıfı kullanılarak gerçekleştirilmiştir. Uygulanan ön işleme ve artırma adımları aşağıda özetlenmiştir:

Boyutlandırma: Hazır mimariler (EfficientNet, VGG16 vb.) ImageNet standartlarına uygun olarak 224×224 piksel; özgün olarak tasarlanan CNN modeli ise hesaplama maliyetini optimize etmek amacıyla 128×128 piksel giriş boyutuyla yapılandırılmıştır.

Normalizasyon: Piksel değerleri 0–255 aralığından 0–1 aralığına normalize edilmiştir.

Veri artırma: Aşırı öğrenmeyi azaltmak amacıyla rastgele yatay ve dikey döndürme, yakınlaştırma, kaydırma ve parlaklık değişimleri gibi dönüşümler uygulanmıştır.

Bu tür veri artırma ve düzenleme yaklaşımlarının CNN tabanlı modellerin sınıflandırma performansını anlamlı biçimde artırdığı literatürde de vurgulanmaktadır (Perez ve Wang, 2017; Shorten ve Khoshgoftaar, 2019).

3.4. Kullanılan Derin Öğrenme Modelleri

Bu çalışmada farklı derin öğrenme mimarileri aynı veri seti üzerinde eğitilerek performansları karşılaştırılmıştır. Tüm modeller Keras Applications kütüphanesinden alınmış ve transfer öğrenme yaklaşımı kullanılarak yeniden eğitilmiştir.

3.4.1. MobileNetV2

MobileNetV2, düşük parametre sayısı ile mobil ve gömülü sistemler için optimize edilmiş bir CNN mimarisidir (Sandler vd., 2018). Derinlik ayrılabilir evrişim yapısı sayesinde hem hız hem de doğruluk açısından avantaj sağlamaktadır. Çalışmada, modelin son katmanı çıkarılarak yerine GlobalAveragePooling2D, Dropout (0.5) ve dört sınıflı Softmax çıkışa sahip Dense katmanı eklenmiştir.

3.4.2. EfficientNetB0

EfficientNet mimarisi, ağ derinliği, genişliği ve giriş çözünürlüğünü dengeli biçimde ölçeklendiren bir yaklaşıma sahiptir (Tan ve Le, 2019). Daha az parametreyle yüksek doğruluk sağlaması nedeniyle tıbbi görüntü uygulamalarında sıklıkla tercih edilmektedir. Bu çalışmada

EfficientNetB0 modeli weights="imagenet" parametresi kullanılarak transfer öğrenme ile eğitilmiştir.

3.4.3. VGG16

VGG16, toplam 16 katmandan oluşan klasik bir CNN mimarisidir (Simonyan ve Zisserman, 2015). Her evrişim bloğundan sonra ReLU aktivasyonu kullanılmış olup, basit ve düzenli yapısı nedeniyle birçok tıbbi görüntü uygulamasında referans model olarak değerlendirilmektedir (Sajjad vd., 2019).

3.4.4. ResNet50

ResNet50, artık bağlantılar (residual connections) sayesinde derin ağlarda karşılaşılan gradyan sönümlenmesi problemini azaltmaktadır (He vd., 2016). Bu özellik, karmaşık medikal görüntü sınıflandırma problemlerinde yüksek performans elde edilmesine olanak sağlamaktadır (Swati vd., 2019).

3.4.5. DenseNet121

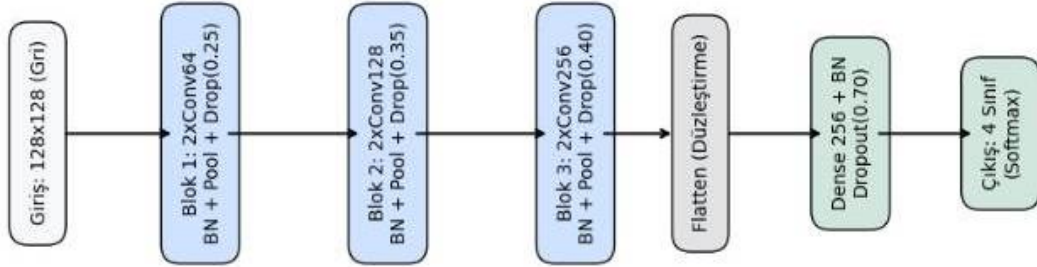
DenseNet mimarisi, katmanlar arasında yoğun bağlantılar kurarak bilgi kaybını azaltmakta ve özelliklerin yeniden kullanılmasını mümkün kılmaktadır (Huang vd., 2017). DenseNet121 modeli, bu avantajları sayesinde MRI görüntülerinin sınıflandırılmasında yaygın olarak tercih edilmektedir.

3.4.6. Özgün CNN Modeli

Bu tez kapsamında ayrıca özgün bir CNN mimarisi tasarlanmıştır. Model; üç evrişim bloğu, dropout katmanları ve tam bağlantılı (dense) katmanlardan oluşan optimize edilmiş bir yapıya sahiptir (Şekil 2). Amaç, yüksek doğruluk oranını korurken parametre sayısını azaltarak hesaplama verimliliğini artırmaktır.

Tüm modeller ReLU aktivasyon fonksiyonu, Softmax çıkış katmanı ve categorical_crossentropy kayıp fonksiyonu kullanılarak derlenmiştir. Öğrenme oranı 0.0001 ve

batch size değeri 32 olarak belirlenmiştir. Özgün CNN mimarilerinin beyin tümörü sınıflandırmasında etkili olduğu literatürde de vurgulanmaktadır (Anaraki vd., 2019).



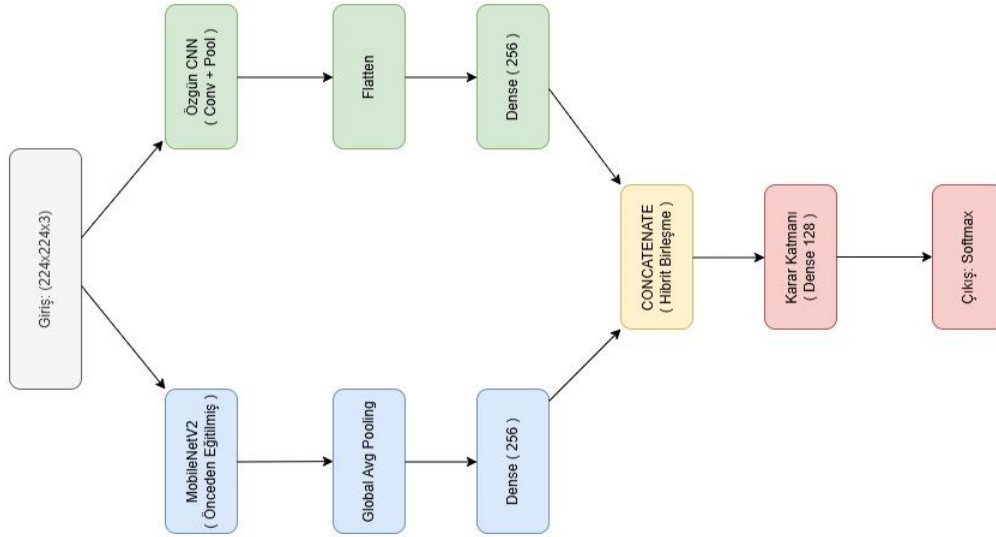
Şekil 2. Özgün CNN Model Diagramı

3.4.7. Hibrit Model

Bu çalışmada yalnızca tekil derin öğrenme mimarileri değil, farklı ağ yapılarını birleştiren hibrit bir yaklaşımın performansı da değerlendirilmiştir. Geliştirilen hibrit model, önceden eğitilmiş MobileNetV2 mimarisi ile çalışmada Özgün CNN modelinin özellik çıkarım yeteneklerini bir araya getirmektedir.

Hibrit mimarinin temel amacı, MobileNetV2'nin düşük parametre sayısı ile sağladığı genellenebilir özellik temsillerini, Özgün CNN modelinin yerel doku örüntülerini yakalama yeteneği ile birleştirerek daha ayırt edici bir özellik uzayı oluşturmaktır. Bu doğrultuda MobileNetV2'nin son evrişim katmanından elde edilen özellik haritaları ile Özgün CNN modelinin son evrişim bloğundan elde edilen çıktılar birleştirme (concatenation) işlemi ile tek bir vektörde toplanmıştır.

Elde edilen bu birleşik özellik vektörü; tam bağlantılı (Dense) katman, %50 Dropout katmanı, ReLU aktivasyon fonksiyonu ve Softmax çıkış katmanından oluşan sınıflandırma bloğuna aktarılmıştır (Şekil 3). Hibrit modelin eğitiminde Adam optimizasyon algoritması, categorical_crossentropy kayıp fonksiyonu ve 0.0001 öğrenme oranı kullanılmıştır. Bu yapı sayesinde hibrit mimari, hem düşük hesaplama maliyetine sahip olmuş hem de tekil modellere kıyasla daha kararlı ve yüksek doğruluklu bir sınıflandırma performansı sergilemiştir.



Şekil 3. Hibrit Model Diagramı

3.5. Model Eğitimi ve Optimizasyon

Tüm eğitim işlemleri Google Colab ortamında TensorFlow 2.15 ve Keras kütüphaneleri kullanılarak gerçekleştirilmiştir. Modellerin eğitiminde kullanılan hiperparametreler aşağıdaki şekilde sabit tutulmuştur: öğrenme oranı (learning rate) 0.0001, optimizasyon algoritması Adam, kayıp fonksiyonu categorical cross-entropy, epoch sayısı 20 ve batch size değeri 32 olarak belirlenmiştir.

Eğitim sürecinde ReduceLROnPlateau ve ModelCheckpoint geri çağırma (callback) fonksiyonları kullanılmıştır. Bu yöntemler, doğrulama kaybının durağanlaştığı durumlarda öğrenme oranını otomatik olarak düşürmek ve en iyi performansı gösteren model ağırlıklarını kaydetmek amacıyla tercih edilmiştir (Kingma ve Ba, 2015).

3.6. K-Fold Çapraz Doğrulama Yöntemi

K-Fold çapraz doğrulama, makine öğrenmesi ve derin öğrenme modellerinin performansını daha nesnel ve genellenebilir biçimde değerlendirmek amacıyla kullanılan istatistiksel bir model doğrulama tekniğidir. Bu yöntemde, mevcut veri seti rastgele ve genellikle dengeli olacak şekilde k eşit alt kümeye (fold) ayrılmakta; her iterasyonda bir alt küme test seti olarak ayrılırken, kalan k-1 alt küme eğitim sürecinde kullanılmaktadır. Bu işlem k kez tekrar edilerek her bir alt küme bir kez test seti olarak değerlendirilmiş olmaktadır (Kohavi, 1995).

K-Fold apraz doęrulamanın temel avantajı, model performansının tek bir eęitim-test blnmesine baęımlı kalmadan deęerlendirilmesine olanak saęlamasıdır. Bu yaklařım, zellikle veri setinin grece kk olduęu durumlarda, performans metriklerinin daha gvenilir ve istikrarlı biimde hesaplanmasını mmkn kılmaktadır. Tıbbi grnt sınıflandırma problemlerinde sıka karřılařılan rnek sayısının sınırlı olması, sınıf dengesizlięi ve yksek varyans riski, apraz doęrulama yntemlerinin kullanımını daha da nemli hale getirmektedir (Yao vd., 2022).

Derin ęrenme tabanlı modellerde, tek bir veri blnmesine dayalı deęerlendirmeler, modelin belirli bir veri daęılımına ařırı uyum saęlaması nedeniyle yanıltıcı sonular retebilmektedir. K-Fold apraz doęrulama ise her veri rneęinin hem eęitim hem de test srecine farklı iterasyonlarda dahil edilmesini saęlayarak, modelin genelleme yeteneęi hakkında daha kapsamlı bir deęerlendirme sunmaktadır. Bu zellik, zellikle klinik karar destek sistemlerinde kullanılacak modeller iin byk nem tařımaktadır; zira klinik ortamda modelin daha nce hi grmedięi veriler zerinde tutarlı performans gstermesi beklenmektedir.

Literatrde, K deęerinin seimi performans deęerlendirmesi zerinde doęrudan etkili bir faktr olarak ele alınmaktadır. Dřk K deęerleri (rneęin 3 veya 5), hesaplama maliyetini azaltırken; yksek K deęerleri (rneęin 10) daha ayrıntılı bir performans deęerlendirmesi sunabilmektedir. Ancak derin ęrenme modellerinde eęitim sresinin uzun olması nedeniyle, beř katlı (5-Fold) apraz doęrulama, hesaplama maliyeti ile deęerlendirme gvenilirlięi arasında dengeli bir zm olarak sıklıkla tercih edilmektedir (Sajjad vd., 2019; Swati vd., 2019).

Bu alıřmada, beř katlı apraz doęrulama yntemi kullanılarak her bir derin ęrenme modelinin performansı deęerlendirilmiřtir. K-Fold apraz doęrulama yalnızca eęitim veri seti zerinde uygulanmıř, baęımsız test veri seti nihai deęerlendirme iin kullanılmıřtır. Her bir katmanda elde edilen doęruluk, duyarlılık, kesinlik ve F1-skoru gibi performans ltleri hesaplanmıř; nihai sonular bu deęerlerin ortalaması alınarak raporlanmıřtır. Bu yaklařım sayesinde, geliřtirilen modellerin farklı veri blnmelerine karřı gsterdięi tutarlılık ve kararlılık ayrıntılı biimde analiz edilmiřtir.

Ayrıca K-Fold apraz doęrulama, farklı model mimarilerinin karřılařtırılmasında da adil bir deęerlendirme ortamı sunmaktadır. Tm modellerin aynı veri blnmeleri zerinde test

edilmesi, mimariler arası performans farklarının istatistiksel olarak daha anlamlı biçimde yorumlanmasına olanak tanımaktadır. Bu yönüyle çapraz doğrulama, yalnızca performans ölçümü değil, aynı zamanda model seçimi sürecinin de önemli bir bileşeni olarak değerlendirilmektedir.

K-Fold çapraz doğrulama yöntemi, özellikle sınırlı ve dengesiz tıbbi görüntü veri setleriyle çalışılan derin öğrenme tabanlı sınıflandırma problemlerinde, model performansının güvenilir ve genellenebilir biçimde değerlendirilmesini sağlayan güçlü bir yöntemdir. Bu tez kapsamında tercih edilen beş katlı çapraz doğrulama yaklaşımı, elde edilen sonuçların istatistiksel geçerliliğini güçlendirmekte ve geliştirilen modellerin klinik uygulamalara aktarılabilirliğini desteklemektedir.

3.7. Performans Ölçütleri

Model performansının değerlendirilmesinde doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru gibi sınıflandırma tabanlı ölçütler kullanılmıştır. Bu ölçütler, karışıklık matrisi (confusion matrix) bileşenleri olan doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) değerleri üzerinden hesaplanmaktadır (Powers, 2011; Sokolova ve Lapalme, 2009).

1. Doğruluk (Accuracy):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(Powers, 2011)

2. Kesinlik (Precision):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(Sokolova ve Lapalme, 2009)

3. Duyarlılık (Recall):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(Sokolova ve Lapalme, 2009)

4. F1-Skoru:

$$\text{F1 Skoru} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(Sokolova ve Lapalme, 2009; Powers, 2011)

Çok sınıflı sınıflandırma problemlerinde bu ölçütler her sınıf için ayrı ayrı hesaplanmış ve makro ortalama (macro average) yöntemi kullanılarak genel performans değerleri elde edilmiştir.

Buna ek olarak karışıklık matrisi kullanılarak modelin hangi sınıflarda hataya daha yatkın olduğu analiz edilmiştir. Bu ölçütlerin birlikte değerlendirilmesi, modelin doğruluk, hassasiyet ve genellenebilirlik düzeyinin kapsamlı biçimde ortaya konulmasını sağlamaktadır (Sokolova ve Lapalme, 2009).

3.8. Deneysel Ortam

Bu çalışmada gerçekleştirilen tüm deneyler, yüksek hesaplama gücü gerektiren derin öğrenme süreçlerini optimize etmek amacıyla bulut tabanlı Google Colab ortamında yürütülmüştür. Deneylerde kullanılan donanım ve yazılım altyapısı Tablo 2’de özetlenmiştir.

Tablo 2. Deneysel altyapı ve yazılım özellikleri

Bileşen / Yazılım	Özellikler
GPU (Grafik İşlemci)	NVIDIA Tesla T4 (16 GB GDDR6 VRAM)
Sistem Belleği (RAM)	12 GB
İşletim Sistemi	Ubuntu 22.04 LTS (Colab Sanal Makinesi)
Programlama Dili	Python 3.10
Derin Öğrenme Kütüphaneleri	TensorFlow 2.15, Keras
Veri İşleme ve Analiz	NumPy, Pandas, scikit-learn
Görüntü İşleme	OpenCV (cv2)
Görselleştirme	Matplotlib, Seaborn

Eđitim ve test ařamalarından elde edilen model ıktıları (kayıp deđerleri, dođruluk oranları ve metrikler) sistematik olarak CSV dosyalarına kaydedilmiřtir.

Elde edilen bu veriler; eđitim/dođrulama grafiklerinin oluřturulması, karıřıklık matrislerinin izilmesi ve modeller arası istatistiksel karıřılařtırmaların yapılması amacıyla kullanılmıřtır. GPU tabanlı bu altyapı, zellikle beř katlı (5-Fold) apraz dođrulama srecindeki yođun hesaplama yknn makul srelerde tamamlanmasına olanak sađlamıřtır.



4. BULGULAR

4.1. Modellerin Normal Eğitim Başarı Analizleri

Bu bölümde, beyin tümörlerinin sınıflandırılması amacıyla eğitilen derin öğrenme modellerinin (Özgün CNN Model, Hibrit Model, ResNet50, MobileNetV2, DenseNet121, VGG16 ve EfficientNetB0) performansları; standart eğitim verileri, sınıflandırma raporları, eğitim-test grafikleri ve karışıklık matrisleri üzerinden detaylı bir şekilde analiz edilmiştir.

Tablo 3'te tüm modellerin normal eğitim süreci sonucunda elde edilen sayısal performans verileri toplu olarak sunulmuştur. İlgili tabloda her bir modele ait doğruluk, duyarlılık ve diğer değerlendirme metrikleri karşılaştırmalı biçimde yer almaktadır. Modellerin performanslarına ilişkin ayrıntılı analiz ve yorumlara ise çalışmanın ilerleyen bölümlerinde detaylı olarak değinilmiştir.

Tablo 3. Modellerin normal eğitim sonucu sayısal veriler

Model	Accuracy	Precision	Recall	F1-Score
Özgün CNN Model	0.9031	0.9038	0.9031	0.9008
Hibrit Model	0.8551	0.8579	0.8551	0.8488
ResNet50	0.8379	0.8526	0.8379	0.8402
MobileNetV2	0.8040	0.8158	0.8040	0.7959
DenseNet121	0.7963	0.8094	0.7963	0.7904
VGG16	0.7300	0.7273	0.7300	0.7051
EfficientNetB0	0.3089	0.0954	0.3089	0.1458

4.1.1. Özgün CNN modelinin normal eğitim performans analizi

Çalışma kapsamında geliştirilen Özgün CNN modeli, Tablo 3'de görüldüğü üzere normal eğitim kapsamında %90,31 doğruluk oranı elde etmiştir. En yüksek başarıyı sergileyen mimari olduğu görülmektedir. Modelin genel kesinlik oranı %90,38 ve duyarlılık oranı %90,31 olarak kaydedilmiştir. Tablo 4'te Özgün CNN modelinin sınıflandırma performansına ilişkin bulgular verilmiştir.

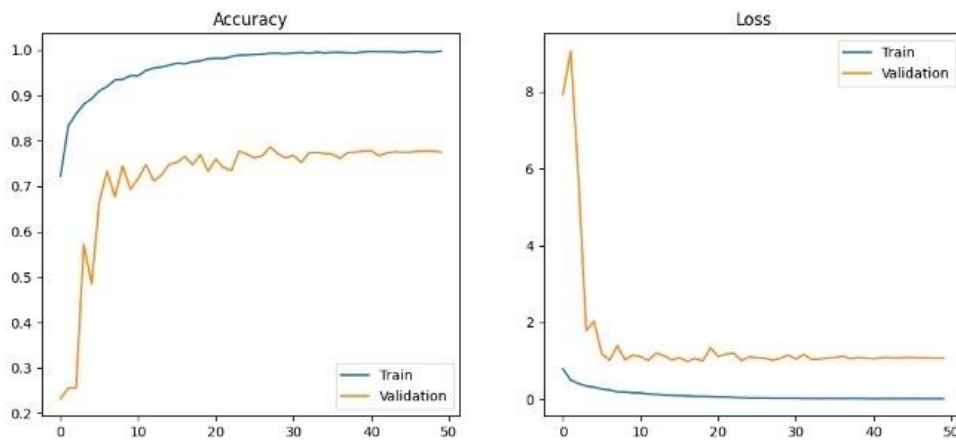
Tablo 4. Özgün cnn modelinin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.93	0.93	0.93	300
Meningioma	0.89	0.74	0.81	306
Notumor	0.92	0.94	0.93	405
Pituitary	0.87	1.00	0.93	300
Accuracy			0.90	1311
Macro avg	0.90	0.90	0.90	1311
Weighted avg	0.90	0.90	0.90	1311

Tablo 4'teki Özgün cnn modelinin sınıfsal bazlı verileri incelendiğinde, modelin Glioma sınıfı için model %0.93 kesinlik, %0.93 duyarlılık ve %0.93 F1-skoru değerlerine ulaşmıştır. Bu sonuç, glioma vakalarının neredeyse tamamının doğru biçimde tespit edildiğini göstermektedir. Meningioma sınıfı, %0.89 precision değerine sahip olmasına rağmen recall değerinin %0.74'de kalması, bazı meningioma örneklerinin diğer sınıflarla karıştığını göstermektedir. Notumor sınıfı, %0.92 kesinlik ve %0.94 duyarlılık değerleriyle en yüksek performans sergilenen sınıflardan biri olmuştur. Pituitary sınıfı için %1.00 recall değeri elde edilmesi, bu tümör türüne ait tüm örneklerin doğru sınıflandırıldığını ortaya koymaktadır.

Genel doğruluk değeri %90,31 olarak hesaplanmış olup, macro ve weighted ortalamalarda da %90 civarında dengeli bir performans elde edilmiştir.

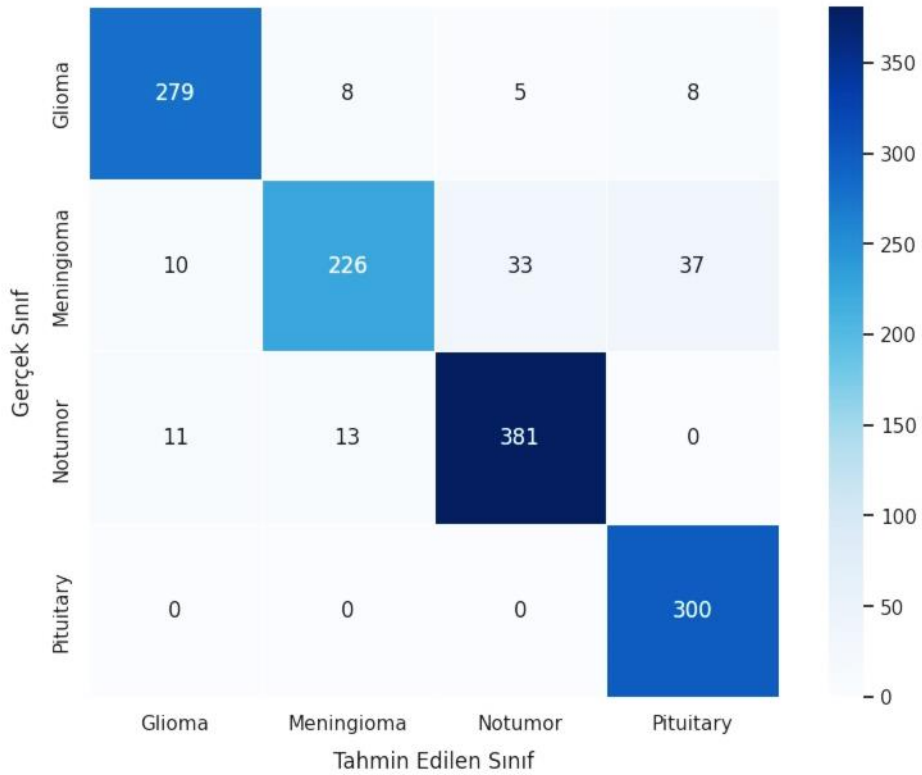
Şekil 4'te sunulan Özgün CNN modelinin normal eğitim ve doğrulama grafikleri, modelin kaybının (loss) istikrarlı bir şekilde düştüğünü ve aşırı öğrenme (overfitting) belirtisi göstermeden yakınsadığını kanıtlamaktadır.

**Şekil 4.** Özgün cnn modelinin normal eğitim grafiği

Şekil 5’te sunulan Özgün CNN modelinin karışıklık matrisi incelendiğinde, modelin her bir tümör sınıfı için ne kadar isabetli tahminler yaptığını ve hangi sınıfları birbirleriyle karıştırdığını göstermektedir. Matris üzerindeki koyu mavi diyagonal (köşegen) hattı, doğru tahmin edilen (True Positive) örnek sayılarını temsil etmektedir. Modelin en başarılı olduğu sınıf Pituitary’dir. 300 görüntünün tamamı hatasız olarak doğru tahmin edilmiştir. Bu sınıfta ne bir yanlış negatif (tümörü atlama) ne de bir yanlış pozitif (başka sınıfı pituitary sanma) hatası yapılmıştır.

405 sağlıklı doku görüntüsünden 381’i doğru sınıflandırılmıştır. 11 örnek Glioma, 13 örnek ise Meningioma olarak yanlış tahmin edilmiştir. Sağlıklı dokunun tümör olarak etiketlenmesi klinik açıdan "yanlış alarm" olarak değerlendirilir. 300 Glioma vakasından 279’u başarıyla tespit edilmiştir. Glioma sınıfının en çok Meningioma (8) ve Pituitary (8) sınıflarıyla karıştırıldığı görülmektedir.

Modelin en çok zorlandığı sınıf ise Meningioma’dır. 306 gerçek Meningioma vakasının 226’sı doğru tahmin edilirken; 37 vakası Pituitary, 33 vakası ise Notumor (sağlıklı) olarak yanlış sınıflandırılmıştır.



Şekil 5. Özgün cnn modelinin karışıklık matrisi

4.1.2. Hibrit modelin normal eğitim performans analizi

Tablo 3'teki genel performans verileri incelendiğinde; Hibrit modelin %85,51 doğruluk (accuracy) oranı ile Özgün CNN modelinden sonra en yüksek başarıyı sergileyen ikinci mimari olduğu görülmektedir. Modelin genel kesinlik oranı %85,79 ve duyarlılık oranı %85,51 olarak kaydedilmiştir. Tablo 5'te hibrit modelin sınıflandırma performansına ait bulgular verilmiştir.

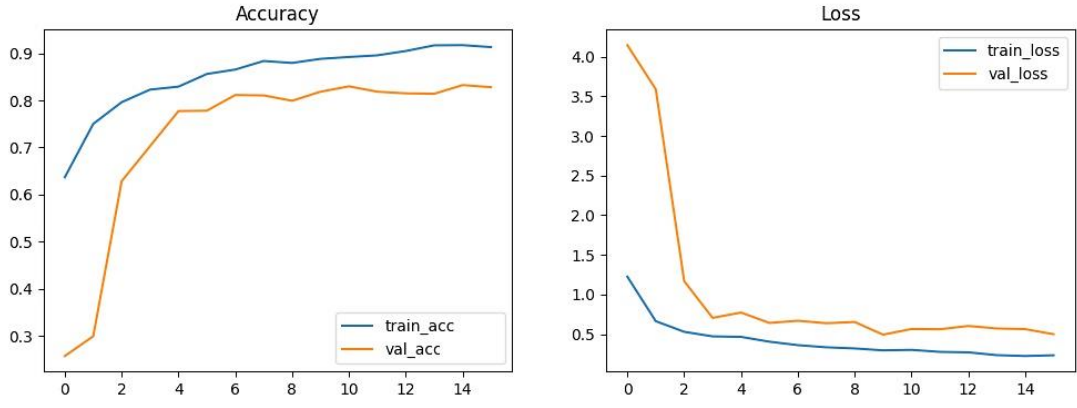
Tablo 5. Hibrit modelin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.97	0.79	0.87	300
Meningioma	0.79	0.62	0.69	306
Notumor	0.85	0.99	0.92	405
Pituitary	0.82	0.99	0.90	300
Accuracy			0.86	1311
Macro avg	0.86	0.84	0.84	1311
Weighted avg	0.86	0.86	0.85	1311

Tablo 5'te sunulan Hibrit modelin sınıfsal bazlı verileri incelendiğinde, Glioma sınıfı için Precision=0.97 ve Recall=0.79 elde edilmiş, bu da modelin glioma vakalarını yüksek doğrulukla tanıyabildiğini, ancak bazı glioma örneklerini kaçırabildiğini göstermektedir. Meningioma sınıfında Recall değerinin 0.62 olması, bu sınıfa ait örneklerin yaklaşık üçte birinin yanlış sınıflandırıldığını ortaya koymaktadır. Bu durum, meningioma ile diğer tümör tipleri arasındaki görsel benzerliklerden kaynaklanmış olabileceği düşünülmektedir.

Notumor sınıfı, Precision=0.85 ve Recall=0.99 değerleri ile en başarılı sınıf olarak öne çıkmıştır. Bu sonuç, hibrit modelin sağlıklı beyin dokularını patolojik örneklerden ayırmada oldukça güvenilir olduğunu göstermektedir. Pituitary sınıfında Recall=0.99 değeri elde edilmesi, bu tümör tipine ait tüm örneklerin doğru şekilde yakalandığını göstermektedir. Ancak Precision=0.82 değeri, başka sınıflara ait bazı örneklerin yanlışlıkla pituitary olarak etiketlendiğini ortaya koymaktadır.

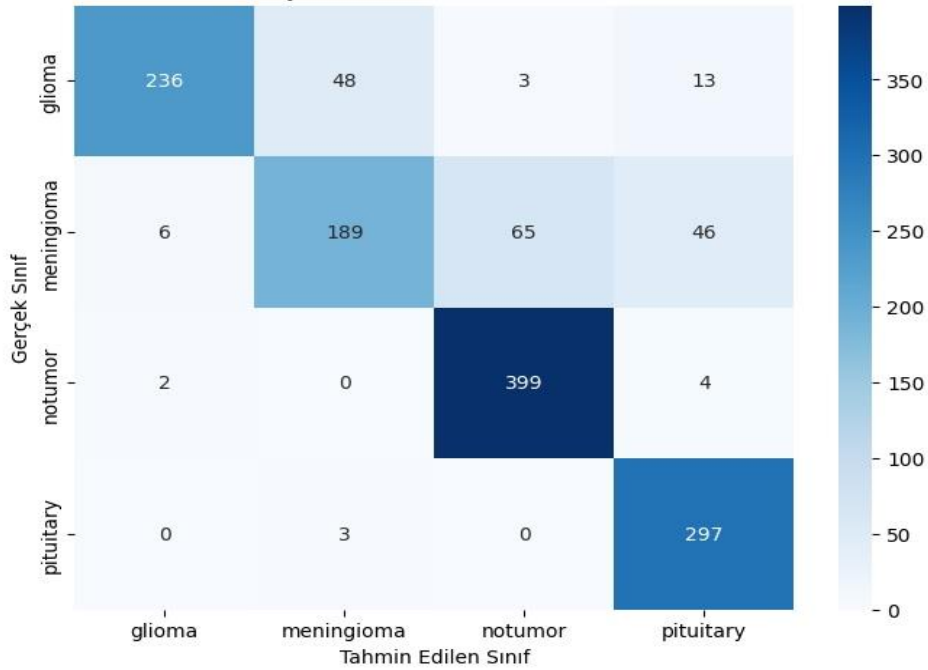
Genel doğruluk %86 olarak raporlanmış ve makro ortalama F1-skoru 0.84 seviyesinde gerçekleşmiştir. Bu bulgular, hibrit modelin özellikle “no tumor” ve “pituitary” sınıflarında güçlü olduğunu, ancak “meningioma” sınıfında sınırlı kaldığını göstermektedir.



Şekil 6. Hibrit modelin normal eğitim grafiği

Şekil 6’da sunulan Hibrit modelin normal eğitim grafiği incelendiğinde; eğitim doğruluğunun epoch ilerledikçe düzenli biçimde arttığı, doğrulama doğruluğunun ise belirli bir noktadan sonra plato çizdiği ve bazı epochlarda dalgalanmalar sergilediği görülmektedir.

Bu durum, hibrit modelin eğitim verisini oldukça iyi öğrendiğini ancak doğrulama verisi üzerinde zaman zaman aşırı öğrenme eğilimi gösterdiğini düşündürmektedir. Eğitim kaybı azalırken doğrulama kaybının bazı epochlarda artış göstermesi, modelin genelleme kabiliyetinin sınırlı kaldığı epoch aralıklarının bulunduğunu ortaya koymaktadır.



Şekil 7. Hibrit modelin karışıklık matrisi

Şekil 7’de Hibrit modelin test veri seti üzerindeki hata dağılımı gösteren karışıklık matrisi görselleştirilmiştir. 306 Meningioma vakasından 189’u doğru tahmin edilirken; 46 vakanın "Pituitary", 65 vakanın "Notumor" ve 6 vakanın "Glioma" olarak yanlış sınıflandırıldığı görülmektedir. Bu durum, Meningioma sınıfının morfolojik özelliklerinin model tarafından Pituitary sınıfı ile karıştırılmaya açık olduğunu göstermektedir.

"Notumor" (399/405 doğru) ve "Pituitary" (297/300 doğru) sınıflarındaki yüksek isabet oranları, modelin bu teşhis gruplarında klinik olarak yüksek güvenilirlik sunduğunu kanıtlamaktadır.

4.1.3. Resnet50 modelinin normal eğitim performans analizi

Tablo 3’teki genel performans verilerine göre; ResNet50 modeli %83,79 doğruluk (accuracy) oranı elde ederek, Özgün CNN ve Hibrit modellerin ardından üçüncü en başarılı mimari olmuştur. Modelin genel kesinlik değeri %85,26 ve F1-skoru %84,02 olarak hesaplanmıştır. Tablo 6’da ResNet50 modelinin sınıflandırma performansına ait bulgular verilmiştir.

Tablo 6. ResNet50 modelinin sınıflandırma performansı

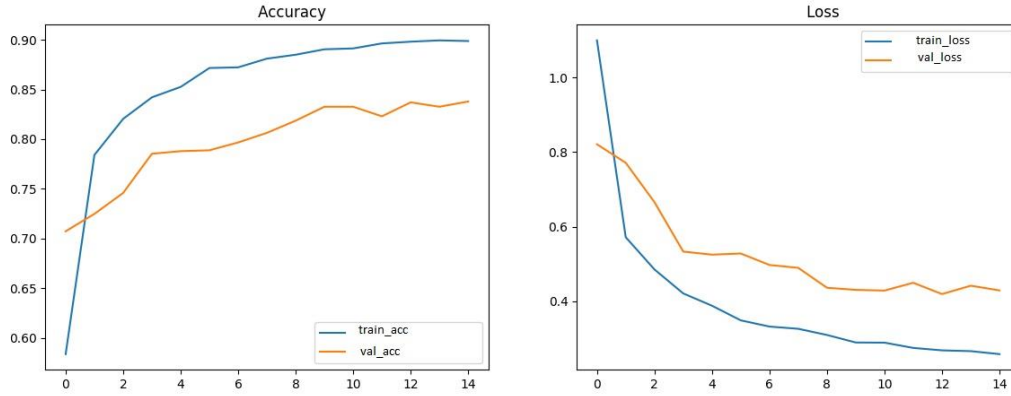
Class	Precision	Recall	F1-Score	Support
Glioma	0.96	0.84	0.90	264
Meningioma	0.68	0.77	0.73	267
Notumor	0.95	0.79	0.86	319
Pituitary	0.80	0.95	0.87	291
Accuracy			0.84	1141
Macro avg	0.85	0.84	0.84	1141
Weighted avg	0.85	0.84	0.84	1141

Tablo 6’da sunulan ResNet50 modelinin sınıfsal bazlı performans metrikleri incelendiğinde, Glioma sınıfı için %96 kesinlik ve %84 duyarlılık elde edilmiş, F1-skoru %90 olarak hesaplanmıştır. Bu durum modelin glioma vakalarını yüksek doğrulukla ayırt edebildiğini göstermektedir.

Meningioma sınıfında kesinlik %68, duyarlılık %77 olarak bulunmuş ve F1-skoru %73 seviyesinde kalmıştır. Bu sınıfın diğer sınıflarla daha fazla örtüşen özelliklere sahip olması, modelin karışıklık yaşamaya neden olmuş olabilir. Notumor sınıfında %95 kesinlik ve %79 duyarlılık elde edilmiş, F1-skoru %86 olarak raporlanmıştır.

Bu sonuç, modelin sağlıklı beyin görüntülerini başarılı bir şekilde tanımlayabildiğini göstermektedir. Pituitary sınıfı için %80 kesinlik ve %95 duyarlılık değerleri elde edilmiş, F1-skoru %87 olarak hesaplanmıştır. Pituitary tümörlerinin özgün yapısal özellikleri, modelin bu sınıfı yüksek oranda doğru tanımasına olanak sağlamıştır.

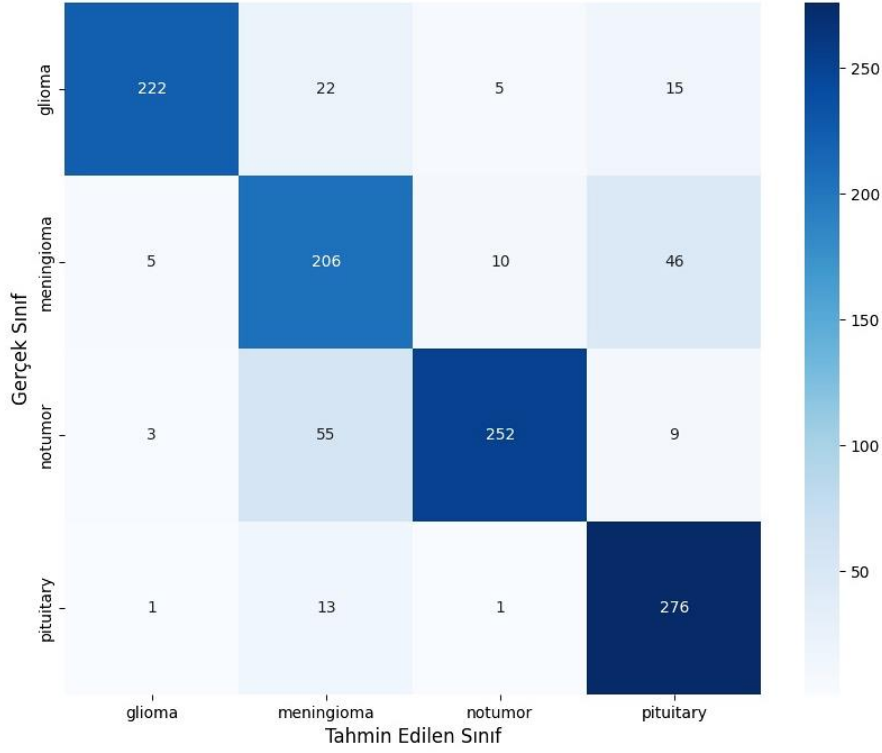
Genel doğruluk değeri %84 olup, ağırlıklı ortalama F1-skoru da %84 olarak raporlanmıştır.



Şekil 8. ResNet50 modelinin normal eğitim grafiği

Şekil 8'deki ResNet50 modeline ait eğitim ve doğrulama grafikleri incelendiğinde, modelin öğrenme sürecini yansıtmaktadır. Grafiklerde eğitim kaybının istikrarlı bir şekilde düştüğü, ancak doğrulama kaybında (validation loss) belirli aralıklarla dalgalanmalar yaşandığı görülmektedir.

Bu durum, modelin derin katman yapısının bu veri seti üzerinde optimize edilmesi sırasında yer yer kararsızlıklar yaşadığını göstermektedir.



Şekil 9. ResNet50 modelinin karışıklık matrisi

Şekil 9’da sunulan ResNet50 modeline ait karışıklık matrisi incelendiğinde, modelin glioma sınıfına ait 264 örneğin 222’sini doğru şekilde sınıflandırdığı görülmektedir. Bununla birlikte 22 örnek meningioma, 5 örnek notumor ve 15 örnek ise pituitary sınıfı olarak yanlış etiketlenmiştir. Bu sonuçlar, glioma sınıfının genel olarak yüksek doğrulukla tahmin edildiğini, ancak özellikle meningioma ile belirli ölçüde karışıklık yaşandığını göstermektedir.

Meningioma sınıfında 267 örneğin 206’sı doğru sınıflandırılmıştır. Bununla birlikte 46 örneğin pituitary olarak tahmin edilmesi dikkat çekmektedir. Bu hata oranı, meningioma ve pituitary tümörlerinin MR görüntülerinde benzer morfolojik özellikler gösterebilmesiyle açıklanabilir. Notumor sınıfına ait 319 görüntünün 252’si doğru biçimde sınıflandırılmıştır. Ancak 55 örneğin meningioma olarak tahmin edilmesi, sağlıklı beyin görüntülerinin bazı tümör tipleriyle karıştırılabildiğini ortaya koymaktadır. Bu durum, özellikle sınır bölgelerdeki doku yoğunluklarının ayırt edilmesinin model açısından zor olduğunu göstermektedir.

Pituitary sınıfında ise 291 örneğin 276’sı doğru tahmin edilmiştir. Sadece 13 örneğin meningioma, 1’er örneğin ise glioma ve notumor olarak yanlış sınıflandırıldığı görülmektedir. Bu sınıf için elde edilen yüksek doğru sınıflandırma sayısı, pituitary tümörlerinin yapısal olarak

daha belirgin özellikler taşıdığını ve ResNet50 modelinin bu özellikleri etkili bir şekilde öğrenebildiğini göstermektedir.

Genel olarak değerlendirildiğinde, ResNet50 mimarisinin özellikle pituitary ve glioma sınıflarında oldukça başarılı sonuçlar verdiği, buna karşın meningioma-notumor ve meningioma-pituitary sınıfları arasında görece daha fazla karışıklık yaşadığı görülmektedir.

4.1.4. MobileNetV2 modelinin normal eğitim performans analizi

Verimli işlem kapasitesi için optimize edilmiş olan MobileNetV2 modelinin beyin tümörü sınıflandırma performansı, normal eğitim stratejisi altında test edilmiş ve elde edilen veriler aşağıda detaylandırılmıştır.

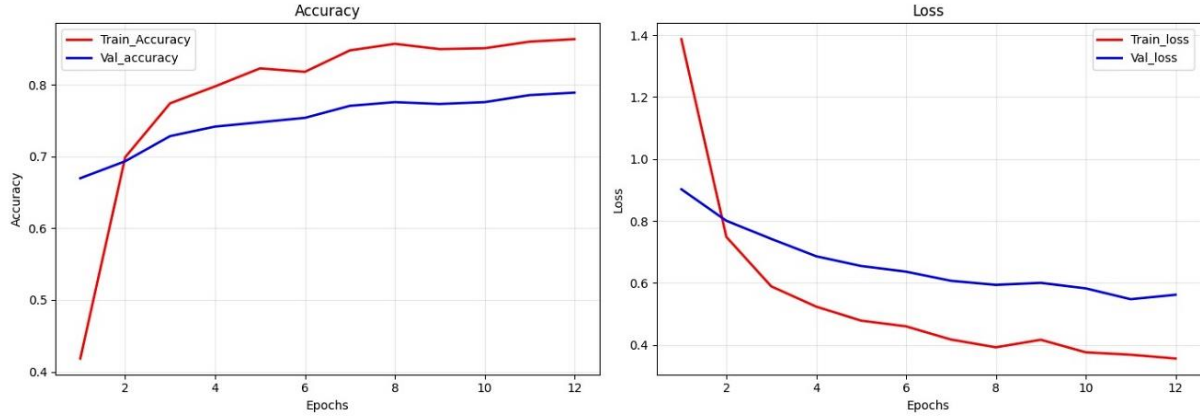
Tablo 3'teki genel performans tablosuna göre; MobileNetV2 modeli %80,40 doğruluk (accuracy) oranı elde etmiştir. Modelin genel kesinlik değeri %81,58 ve F1-skoru %79,59 olarak kaydedilmiştir. Bu sonuçlar, modelin ResNet50 gibi daha derin mimarilerin gerisinde kaldığını göstermektedir. Tablo 7'de MobileNetV2 modelinin sınıflandırma performansına ait bulgular verilmiştir.

Tablo 7. MobileNetV2 modelinin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.95	0.70	0.81	300
Meningioma	0.70	0.52	0.60	306
Notumor	0.89	0.95	0.92	405
Pituitary	0.69	1.00	0.82	300
Accuracy			0.80	1311
Macro avg	0.81	0.79	0.79	1311
Weighted avg	0.82	0.80	0.80	1311

Tablo 7'de yer alan MobileNetV2 modeline ait sınıfsal performans metrikleri incelendiğinde Pituitary; Model, bu sınıfta %1,00 duyarlılık (recall) değerine ulaşarak hipofiz tümörü vakalarının tamamını doğru tespit etmiştir. Ancak %69'luk kesinlik (precision) oranı, modelin diğer sınıflardan çok sayıda vakayı yanlışlıkla bu sınıfa dahil ettiğini (yalancı pozitif) göstermektedir.

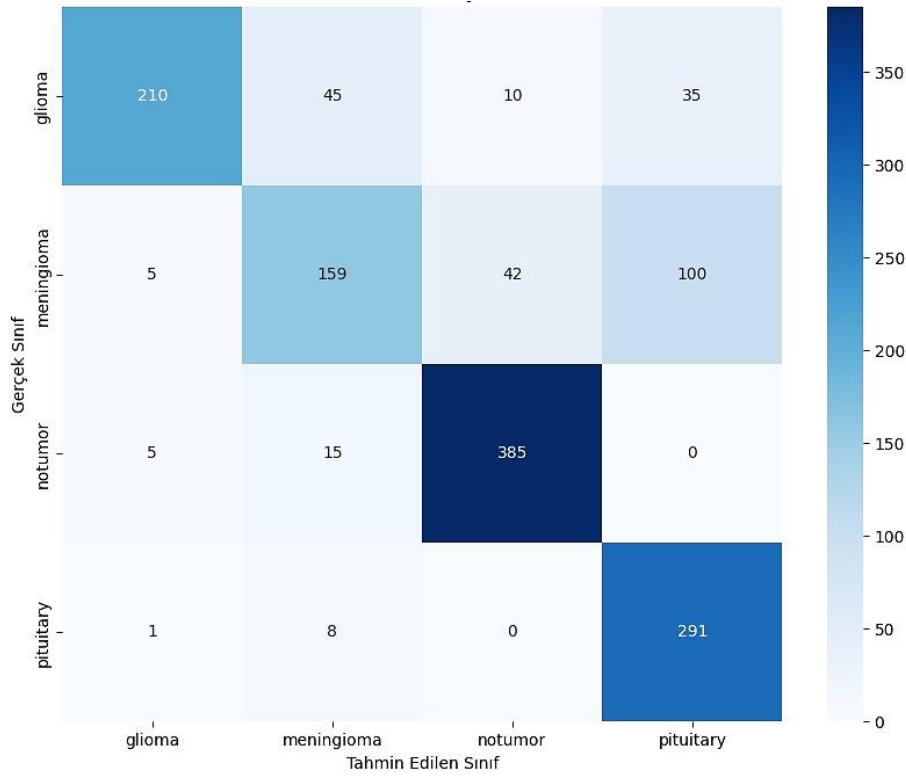
Modelin en çok zorlandığı sınıf %52 duyarlılık ve 0,60 F1-skoru ile Meningioma olmuştur. Bu durum, modelin meningeoma vakalarının yaklaşık yarısını teşhis edemediğini kanıtlamaktadır. Sağlıklı dokuların tespitinde %95 duyarlılık ve %89 kesinlik ile nispeten dengeli ve başarılı bir performans sergilenmiştir. Glioma; %95 kesinlik oranı ile modelin en güvenilir tahminleri Glioma sınıfında yaptığı görülmektedir.



Şekil 10. MobileNetV2 modelinin normal eğitim grafiği

Şekil 10'da sunulan MobileNetV2 modeline ait normal eğitim ve doğrulama grafikleri incelendiğinde, modelin öğrenme sürecindeki karakteristik özelliklerini yansıtmaktadır.

Eğitim doğruluğu %90 seviyelerine yaklaşırken, doğrulama doğruluğunun %80 civarında kalması, modelin eğitim verisine aşırı uyum sağlama (overfitting) eğiliminde olduğunu veya test verisi üzerinde genelleme yapmakta zorlandığını göstermektedir. Benzer şekilde, doğrulama kaybının eğitim kaybından daha yüksek bir seviyede stabilize olması bu durumu desteklemektedir.



Şekil 11. MobileNetV2 modelinin karışıklık matrisi

Şekil 11'deki MobileNetV2 modeline ait karışıklık matrisi incelendiğinde, gerçekte Meningioma olan 306 vakanın 159'u doğru tahmin edilirken, modelin bu sınıftaki en büyük hatayı 100 vakayı "Pituitary" olarak sınıflandırarak yaptığı görülmektedir. Bu 100 vakalık hata, Meningioma sınıfının duyarlılığının (Recall) neden düşük olduğunu ve Pituitary sınıfına neden çok fazla yanlış veri aktığını (düşük Precision) sayısal olarak kanıtlamaktadır.

300 Glioma vakasından 210'u doğru tahmin edilmiştir. Model, Glioma vakalarının 45'ini Meningioma ve 35'ini Pituitary olarak yanlış sınıflandırmıştır. 10 vakanın ise "Notumor" olarak sınıflandırılması, modelin tümör varlığını teşhis etmede küçük bir yanılma payı olduğunu göstermektedir.

Modelin en başarılı olduğu alanlardan biri sağlıklı doku teşhisidir. 405 Notumor vakasından 385'i doğru tahmin edilmiş, sadece 20 vaka (15 Meningioma, 5 Glioma) tümürlü olarak karıştırılmıştır. Dikkat çekici bir şekilde, hiçbir sağlıklı doku Pituitary olarak sınıflandırılmamıştır.

300 Pituitary vakasının 291'i doğru tahmin edilerek bu sınıfta %97 gibi yüksek bir iç başarı yakalanmıştır. Sadece 9 vaka (8 Meningioma, 1 Glioma) hatalı sınıflandırılmıştır.

4.1.5. Densenet121 modelinin normal eğitim performans analizi

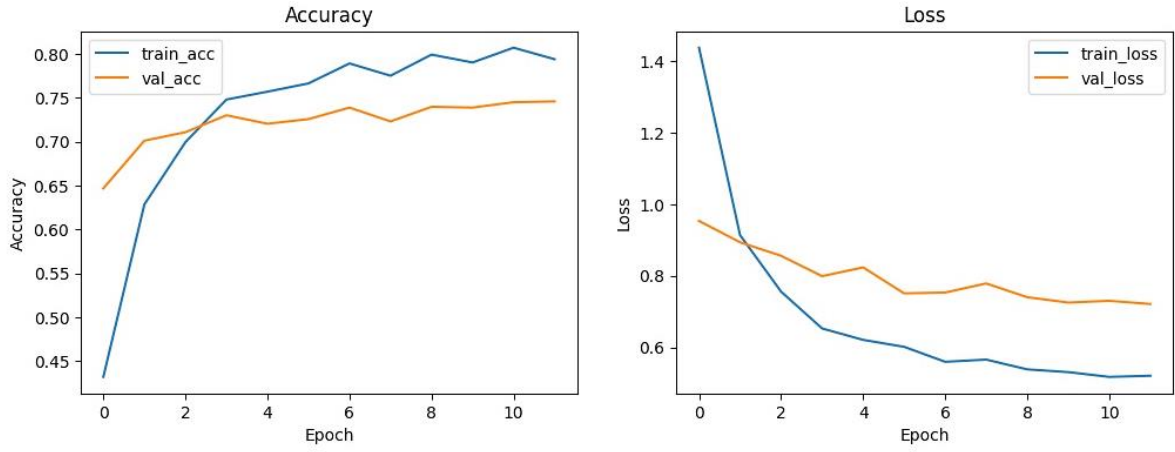
Tablo 3'teki genel performans verilerine göre; DenseNet121 modeli %79,63 doğruluk (accuracy) oranı elde etmiştir. Modelin genel kesinlik değeri %80,94 ve F1-skoru %79,04 olarak kaydedilmiştir. Bu sonuçlar, DenseNet121'in MobileNetV2 ile benzer bir başarıyı sergilediğini ancak ana modellerin (Özgün CNN ve Hibrit) gerisinde kaldığını göstermektedir. Tablo 8'de DenseNet121 modelinin sınıflandırma performansına ait bulgular verilmiştir.

Tablo 8. DenseNet121 modelinin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.94	0.69	0.79	300
Meningioma	0.67	0.54	0.60	306
Notumor	0.91	0.93	0.92	405
Pituitary	0.69	0.99	0.81	300
Accuracy			0.80	1311
Macro avg	0.80	0.79	0.78	1311
Weighted avg	0.81	0.80	0.79	1311

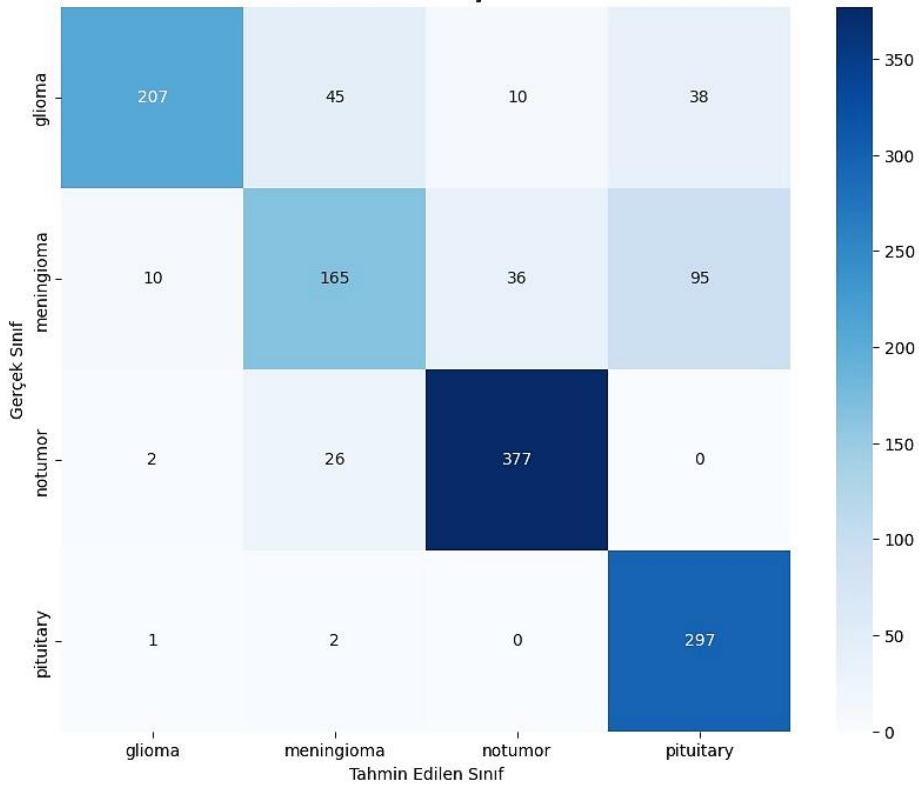
Tablo 8'de yer alan DenseNet121 modeline ait sınıfsal performans metrikleri incelendiğinde, Glioma; %94 kesinlik (precision) ve %69 duyarlılık (recall) oranı elde edilmiştir. Modelin Glioma teşhislerindeki doğruluğu yüksek olsa da, vakaların %31'ini kaçırdığı görülmektedir.

Modelin en çok zorlandığı sınıf %54 duyarlılık ve 0,60 F1-skoru ile Meningioma olmuştur. Sağlıklı dokuların tespitinde %91 kesinlik ve %93 duyarlılık ile yüksek ve dengeli bir başarı sergilenmiştir. Model bu sınıfta %99 duyarlılık değerine ulaşarak hipofiz tümörlerini yakalamada üstün bir performans göstermiştir; ancak %69'luk kesinlik oranı, diğer sınıflardan bu sınıfa önemli bir veri kayması olduğunu kanıtlamaktadır.



Şekil 12. DenseNet121 modelinin normal eğitim grafiği

Şekil 12'deki DenseNet121 modeline ait eğitim ve doğrulama grafikleri incelendiğinde, doğrulama kaybının (validation loss) eğitim kaybına göre daha yüksek bir seviyede dalgalanarak stabilize olduğu görülmektedir. Bu durum, modelin derin mimarisinin genelleme yaparken zorlandığını işaret etmektedir.



Şekil 13. DenseNet121 modelinin karışıklık matrisi

Şekil 13'teki DenseNet121 modeline ait karışıklık matrisi incelendiğinde, sınıfsal bazdaki performans düşüşlerinin nedenlerini sayısal olarak ortaya koymaktadır.

Gerçekte Meningioma olan 306 vakanın 165'i doğru tahmin edilirken; modelin en büyük hatayı 95 vakayı "Pituitary" olarak sınıflandırarak yaptığı görülmektedir. Bu bulgu, DenseNet121'in de tıpkı MobileNetV2 gibi Meningioma ve Pituitary sınıfları arasında morfolojik bir karışıklık yaşadığını teyit etmektedir.

300 Glioma vakasından 207'si doğru tahmin edilmiştir. Hataların 45'i Meningioma, 38'i ise Pituitary sınıflarına dağılmıştır. Modelin sadece 10 Glioma vakasına "Notumor" teşhisi koyması, tümör varlığını saptama konusundaki kararlılığını göstermektedir.

405 Notumor vakasından 377'si doğru tahmin edilmiş, hataların büyük çoğunluğu (26 vaka) Meningioma sınıfına gitmiştir. Pituitary sınıfında ise 300 vakadan 297'si doğru tahmin edilerek %99'luk bir iç başarı yakalanmıştır.

4.1.6. VGG16 modelinin normal eğitim performans analizi

Tablo 3'teki genel performans verilerine göre; VGG16 modeli %73,00 doğruluk (accuracy) oranı elde ederek, önceki bölümlerde incelenen ResNet50 ve DenseNet121 gibi modellerin gerisinde kalmıştır. Modelin genel kesinlik değeri %72,73 ve F1-skoru %70,51 olarak kaydedilmiştir. Tablo 9'da VGG16 modelinin sınıflandırma performansına ait bulgular verilmiştir.

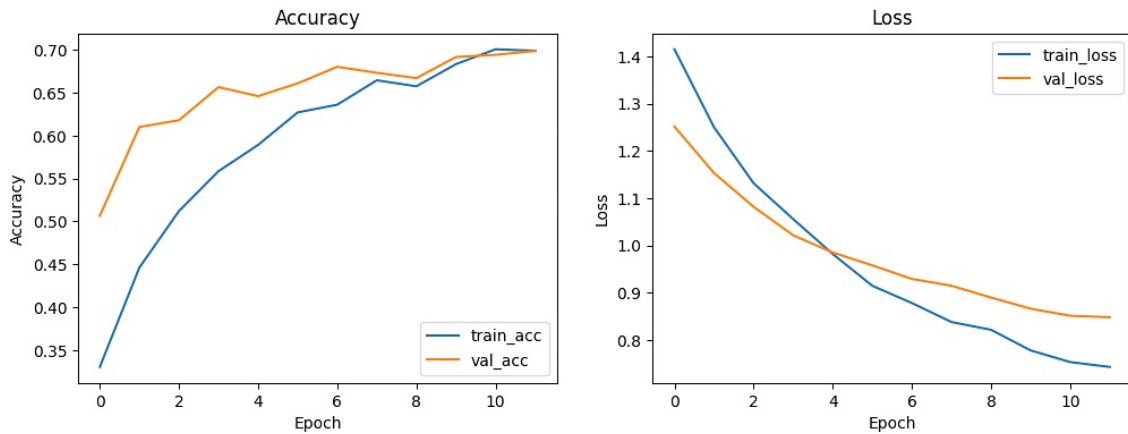
Tablo 9. VGG16 modelinin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.79	0.72	0.75	300
Meningioma	0.65	0.29	0.41	306
Notumor	0.80	0.93	0.86	405
Pituitary	0.64	0.92	0.75	300
Accuracy			0.73	1311
Macro avg	0.72	0.71	0.69	1311
Weighted avg	0.73	0.73	0.71	1311

Tablo 9’da yer alan VGG16 modeline ait sınıfsal performans metrikleri incelendiğinde modelin en başarısız olduğu sınıf %29 duyarlılık (recall) ve 0,41 F1-skoru ile Meningioma olmuştur. Bu değer, VGG16 modelinin Meningioma vakalarının büyük çoğunluğunu teşhis edemediğini ve farklı sınıflara atadığını göstermektedir.

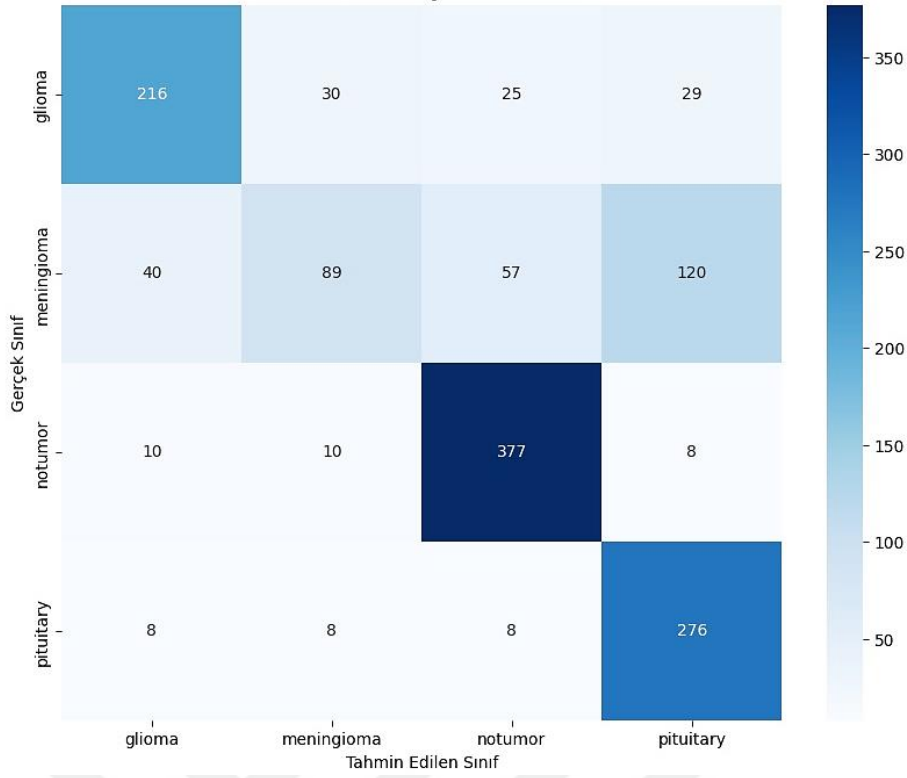
Model Pituitary sınıfında %92 duyarlılık değerine ulaşarak hipofiz tümörlerini yakalamada başarılı bir grafik çizmiştir. Ancak %64’lük kesinlik (precision) oranı, modele ait Pituitary tahminlerinin yaklaşık üçte birinin hatalı olduğunu kanıtlamaktadır.

Sağlıklı dokuların tespitinde %93 duyarlılık oranı ile yüksek bir yakalama başarısı gösterilse de, %80 kesinlik oranı bazı sağlıklı görüntülerin tümörlü olarak etiketlendiğine işaret etmektedir. Glioma sınıfında %79 kesinlik ve %72 duyarlılık ile ortalama bir performans sergilenmiştir.



Şekil 14. VGG16 modelinin normal eğitim grafiği

Şekil 14’teki VGG16 modeline ait eğitim ve doğrulama grafikleri incelendiğinde, kayıp (loss) değerinin diğer modellere oranla daha yüksek bir seviyede kaldığı ve öğrenme eğrisinin yavaş bir yakınsama gösterdiği saptanmıştır. Bu durum, modelin parametre yoğunluğunun mevcut veri seti için optimize edilmesinde güçlükler yaşandığını yansıtmaktadır.



Şekil 15. VGG16 modelinin karışıklık matrisi

Şekil 15'te sunulan VGG16 modeline ait karışıklık matrisi incelendiğinde, sınıfsal bazda incelendiğinde modelin teşhis kapasitesine dair kritik bulgular elde edilmiştir. Belirtilen matris değerleri, modelin neden diğer mimarilerin gerisinde kaldığını sayısal olarak açıklamaktadır.

Gerçekte Meningioma olan 306 vakanın sadece 89'u doğru tahmin edilebilmiştir. Modelin bu sınıftaki en büyük hatası, 120 vakayı Pituitary ve 57 vakayı Notumor olarak sınıflandırmasıdır. Özellikle tümörlü bir dokunun Notumor olarak etiketlenmesi, klinik risk açısından modelin en zayıf noktasını oluşturmaktadır.

Bu yüksek hata payı, Tablo 7'deki %29'luk düşük duyarlılık (recall) değerinin temel nedenidir. 300 Glioma vakasından 216'sı doğru tahmin edilmiştir. Hataların dağılımı incelendiğinde; 30 vakanın Meningioma, 25 vakanın Notumor ve 29 vakanın Pituitary olarak sınıflandırıldığı görülmektedir. Glioma sınıfındaki %72 duyarlılık oranı, modelin bu tümör tipinde Meningioma'ya göre daha kararlı olduğunu kanıtlamaktadır.

Model, 405 Notumor vakasından 377'sini doğru tahmin ederek sağlıklı dokuları ayırt etmede başarılı bir performans sergilemiştir. Sadece 28 vaka (10 Glioma, 10 Meningioma, 8 Pituitary)

hatalı şekilde tümörlü sınıflara dahil edilmiştir. 300 Pituitary vakasının 276'sı doğru sınıflandırılmış, hata payı sınıflar arasında (8 Glioma, 8 Meningioma, 8 Notumor) eşit dağılmıştır. Ancak Pituitary sütununa genel olarak bakıldığında, Meningioma'dan gelen 120 hatalı tahminin bu sınıfa yığılması, Pituitary sınıfının kesinlik (precision) oranını %64 seviyesine çekmiştir.

4.1.7. EfficientNetB0 modelinin normal eğitim performans analizi

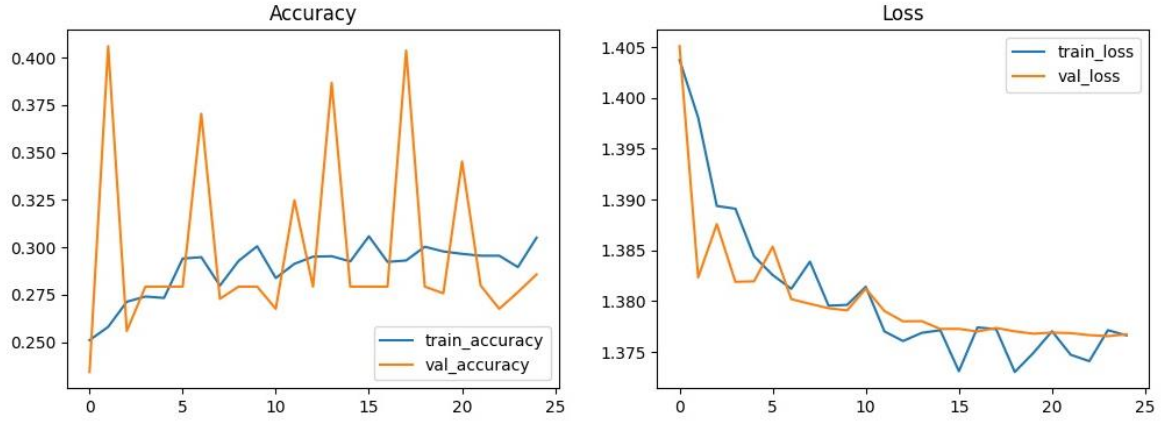
Tablo 3'teki genel performans verileri incelendiğinde; EfficientNetB0 modelinin %30,89 doğruluk (accuracy) oranı ile çalışmada test edilen modeller arasında en düşük performansı sergilediği görülmektedir. Modelin genel kesinlik değeri %0,09 ve F1-skoru %0,14 gibi oldukça düşük seviyelerde kalmıştır. Tablo 10'da EfficientNetB0 modelinin sınıflandırma performansına ait bulgulara verilmiştir.

Tablo 10. EfficientNetB0 modelinin sınıflandırma performansı

Class	Precision	Recall	F1-Score	Support
Glioma	0.00	0.00	0.00	300
Meningioma	0.00	0.00	0.00	306
Notumor	0.31	0.99	0.47	405
Pituitary	0.00	0.00	0.00	300
Accuracy			0.31	1311
Macro avg	0.08	0.25	0.12	1311
Weighted avg	0.09	0.31	0.14	1311

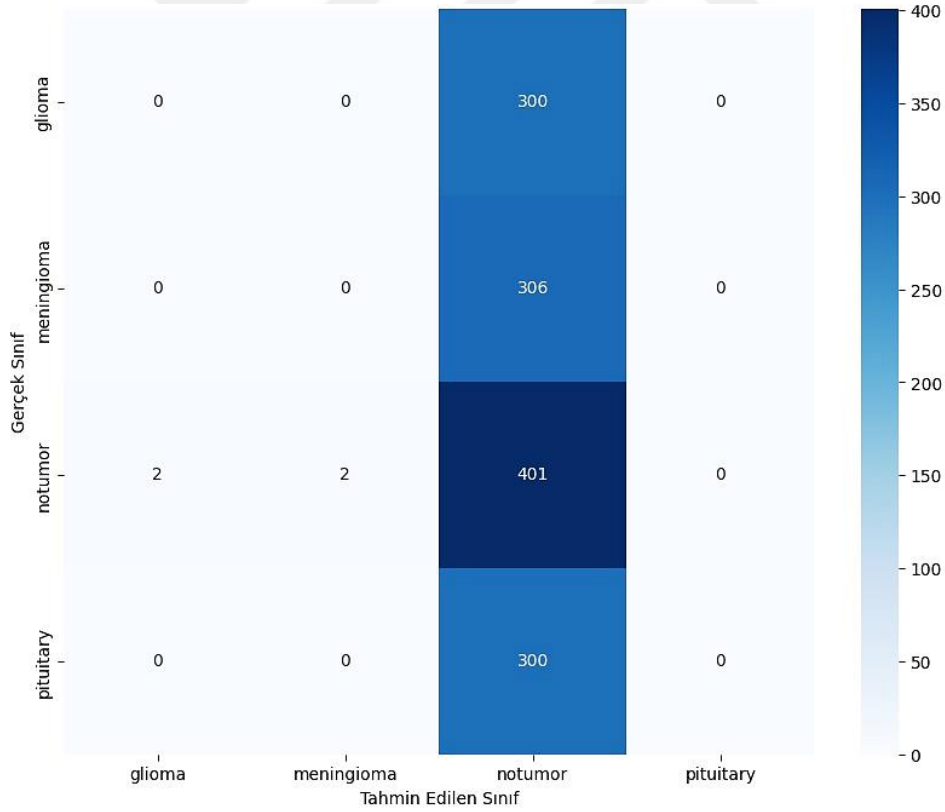
Tablo 10'da sunulan EfficientNetB0 modeline ait sınıfsal bazlı veriler, modelin sınıfları ayırt etme noktasında işlevsel bir sonuç üretmediğini kanıtlamaktadır.

Model, bu üç tümör sınıfında %0,00 kesinlik ve %0,00 duyarlılık oranları elde etmiştir. Bu durum, modelin tümörlü vakaların hiçbirini kendi sınıflarında doğru teşhis edemediğini göstermektedir. Modelin istatistiksel olarak başarılı görüldüğü tek alan %99 duyarlılık ile sağlıklı doku sınıfıdır. Ancak buradaki %31'lik kesinlik oranı, modelin aslında veri setindeki neredeyse tüm görüntüleri Notumor olarak sınıflandırma eğiliminde olduğunu ortaya koymaktadır.



Şekil 16. EfficientNetB0 modelinin normal eğitim grafiği

Şekil 16'daki EfficientNetB0 modeline ait eğitim ve doğrulama grafikleri incelendiğinde, doğruluk eğrisinin eğitim süreci boyunca yatay bir seyir izlediği ve kayıp (loss) değerinin anlamlı bir düşüş göstermediği saptanmıştır. Bu tablo, modelin öğrenememe (failure to converge) sorunu yaşadığını doğrulamaktadır.



Şekil 17. EfficientNetB0 modelinin karışıklık matrisi

Şekil 17’de sunulan EfficientNetB0 modeline ait karışıklık matrisi incelendiğinde; model, test veri setindeki vakaların ezici çoğunluğunu gerçek sınıfına bakmaksızın Notumor sütununa yığımıştır. 300 Glioma, 306 Meningioma ve 300 Pituitary vakasının tamamına yakını yanlışlıkla sağlıklı doku olarak etiketlenmiştir. Bu durum, modelin özellik çıkarımı aşamasında tümör morfolojisine dair karakteristik verileri yakalayamadığını kanıtlamaktadır.

EfficientNetB0 modeli, bu çalışma kapsamındaki beyin tümörü veri seti üzerinde öğrenme gerçekleştirememiş ve tüm vakaları tek bir sınıfa atama eğilimi göstererek klinik açıdan geçersiz sonuçlar üretmiştir.

4.2. Modellerin Çapraz Doğrulama Başarı Analizleri

Modellerin sadece tek bir eğitim-test ayrımıyla değil, verinin farklı bölümleriyle ne kadar tutarlı çalıştığını ölçmek amacıyla yapılan beş katlı çapraz doğrulama sonuçları Tablo 11’de sunulmuştur.

Tablo 11 incelendiğinde, modellerin K-Fold çapraz doğrulama sonuçlarına göre performanslarının farklılık gösterdiği görülmektedir. Ortalama doğruluk değerleri dikkate alındığında en yüksek performansın Hibrit Model tarafından elde edildiği %0,9315 belirlenmiştir. Hibrit modeli sırasıyla Özgün CNN Modeli %0,9272 ve ResNet50 %0,9120 takip etmektedir.

Orta düzey performans gösteren modeller arasında MobileNetV2 %0,8912 ve DenseNet121 %0,8774 yer alırken, VGG16 %0,8228 daha düşük doğruluk oranı sergilemiştir. En düşük performansın ise EfficientNetB0 modeline ait olduğu %0,2745 açıkça görülmektedir. Bu durum, ilgili modelin kullanılan veri kümesi üzerinde yeterli genelleme başarısı gösteremediğini ortaya koymaktadır.

Precision, Recall ve F1-Score ortalama değerleri de doğruluk oranlarıyla paralel bir dağılım göstermekte olup, en dengeli ve yüksek performansın yine Hibrit Model tarafından sağlandığı anlaşılmaktadır.

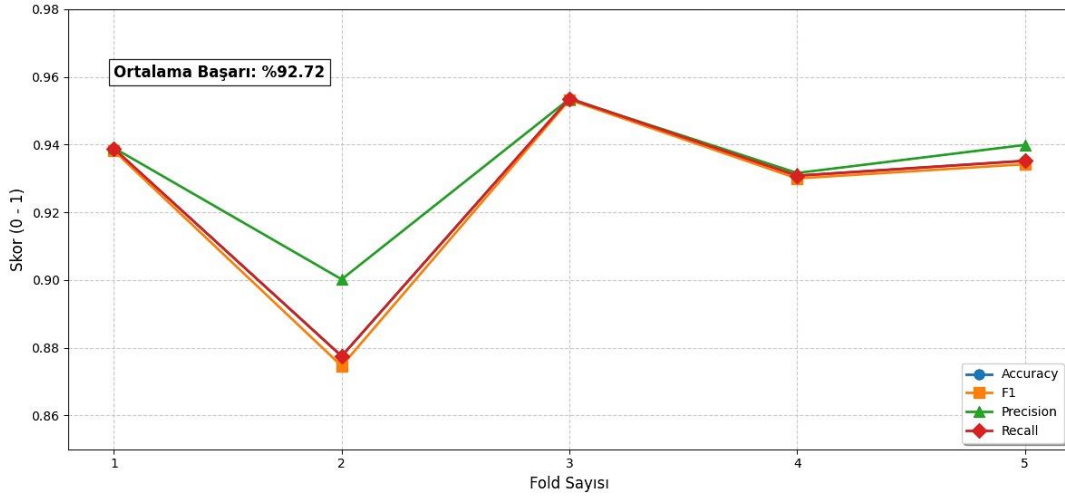
Her bir modele ait ayrıntılı performans analizleri çalışmanın ilerleyen bölümlerinde detaylı olarak ele alınmıştır.

Tablo 11. Beş katlı çapraz doğrulama sayısal verileri

Model	K-Fold	Accuracy	Precision	Recall	F1-Score
Özgün CNN Modeli	1	0.9388	0.9390	0.9388	0.9382
	2	0.8775	0.9002	0.8775	0.8746
	3	0.9536	0.9535	0.9536	0.9532
	4	0.9308	0.9316	0.9308	0.9300
	5	0.9352	0.9399	0.9352	0.9342
	Ortalama	0.9272	0.9328	0.9272	0.9261
Hibrit Model	1	0.9300	0.9290	0.9300	0.9291
	2	0.9291	0.9285	0.9291	0.9282
	3	0.9430	0.9428	0.9430	0.9425
	4	0.9316	0.9314	0.9316	0.9314
	5	0.9238	0.9234	0.9238	0.9229
	Ortalama	0.9315	0.9310	0.9315	0.9308
ResNet50	1	0.9081	0.9094	0.9081	0.9082
	2	0.9055	0.9099	0.9055	0.9055
	3	0.9168	0.9177	0.9168	0.9162
	4	0.9080	0.9126	0.9080	0.9087
	5	0.9220	0.9221	0.9220	0.9219
	Ortalama	0.9120	0.9143	0.9120	0.9121
MobileNetV2	1	0.8853	0.8854	0.8853	0.8835
	2	0.8941	0.8943	0.8941	0.8935
	3	0.8896	0.8884	0.8896	0.8887
	4	0.8966	0.8966	0.8966	0.8964
	5	0.8905	0.8899	0.8905	0.8882
	Ortalama	0.8912	0.8909	0.8912	0.8901
DenseNet121	1	0.8818	0.8871	0.8818	0.8816
	2	0.8573	0.8566	0.8573	0.8529
	3	0.8931	0.8938	0.8931	0.8923
	4	0.8782	0.8795	0.8782	0.8778
	5	0.8765	0.8804	0.8765	0.8750
	Ortalama	0.8774	0.8795	0.8774	0.8759
VGG16	1	0.8083	0.8134	0.8083	0.8068
	2	0.8416	0.8482	0.8416	0.8411
	3	0.8345	0.8349	0.8345	0.8321
	4	0.8152	0.8145	0.8152	0.8129
	5	0.8143	0.8211	0.8143	0.8126
	Ortalama	0.8228	0.8264	0.8228	0.8211
EfficientNetB0	1	0.2554	0.0652	0.2554	0.1039
	2	0.2790	0.0778	0.2790	0.1217
	3	0.2793	0.0780	0.2793	0.1219
	4	0.2793	0.0780	0.2793	0.1219
	5	0.2793	0.0780	0.2793	0.1219
	Ortalama	0.2745	0.0754	0.2745	0.1183

4.2.1. Özgün CNN modelinin çapraz doğrulama performans analizi

Özgün CNN modeli, çapraz doğrulama sürecinde oldukça yüksek bir kararlılık sergilemiştir. Tablo 11'e göre model, 3. katlamada (Fold 3) %95,36 ile en yüksek doğruluğuna ulaşırken, 5 katlamanın ortalamasında %92,72 doğruluk oranı elde etmiştir. Modelin her bir katlamada benzer yakınsama eğilimleri gösterdiği ve ortalama %93,28 kesinlik (precision) değeri ile teşhis güvenilirliğini kanıtladığı görülmektedir. Özgün CNN modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 18'de grafiksel olarak sunulmuştur.



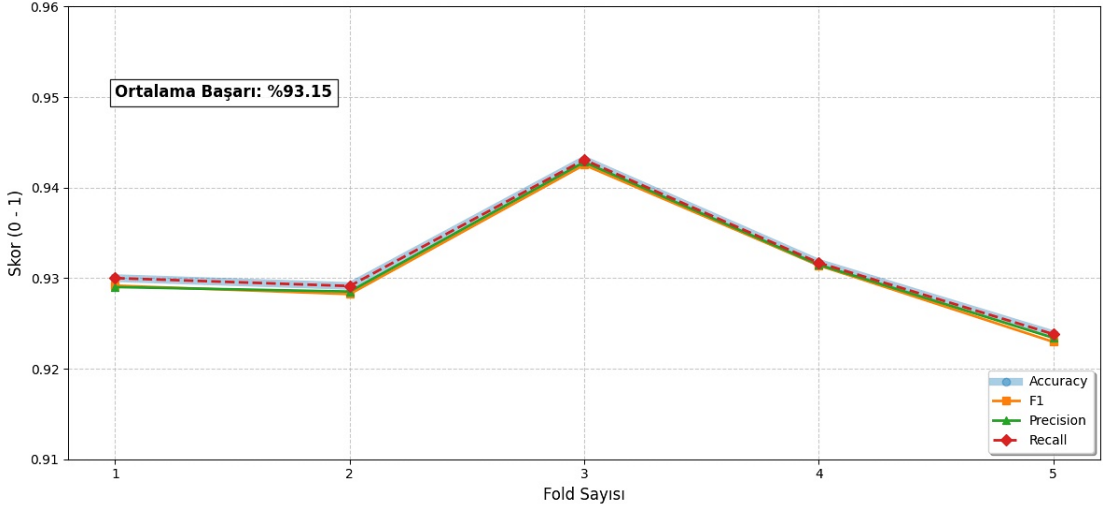
Şekil 18. Özgün cnn modelinin k-fold performans grafiği

Şekil 18'de sunulan Özgün cnn modeli k-fold performans grafiği incelendiğinde, tüm katlarda metriklerin birbirine oldukça yakın seyrettiği ve modelin farklı veri alt kümelerinde tutarlı sonuçlar ürettiği görülmektedir. Bu durum, modelin aşırı öğrenmeye yatkın olmadığını ve güçlü bir genelleme yeteneğine sahip olduğunu göstermektedir.

4.2.2. Hibrit modelin çapraz doğrulama performans analizi

Hibrit model, çapraz doğrulama sonuçlarında tüm modeller arasında en stabil ve en yüksek ortalamaya sahip model olarak öne çıkmaktadır. Tablo 11 verileri incelendiğinde, katlamalar arasındaki farkın çok düşük olduğu ve %93,15 ortalama doğruluk ile genel performansta Özgün CNN modelini geride bıraktığı görülmektedir.

Bu durum, Hibrit mimarinin farklı veri alt gruplarında genelleme yeteneğinin daha güçlü olduğunu ortaya koymaktadır. Hibrit modelin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 19’da grafiksel olarak sunulmuştur.



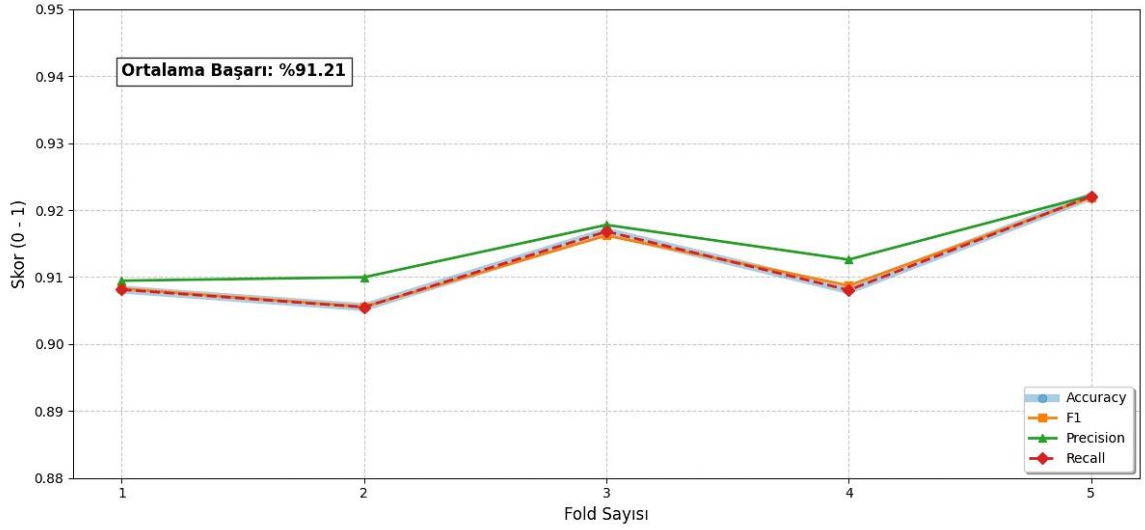
Şekil 19. Hibrit modelin k-fold performans grafiği

Şekil 19’da sunulan Hibrit model k-fold performans grafiği incelendiğinde, metriklerin tüm katlarda stabil seyrettiğini ve özellikle 3. katlamada en yüksek değerlere ulaşıldığını göstermektedir. Bu sonuçlar, MobileNetV2’nin özellik çıkarım kapasitesinin, Özgün CNN yapısı ile birleştirildiğinde daha etkili bir temsil gücü oluşturduğunu ortaya koymaktadır.

4.2.3. Resnet50 modelinin çapraz doğrulama performans analizi

ResNet50 modeli, çapraz doğrulama sürecinde %91,20 ortalama doğruluk oranı yakalamıştır. Tablo 11’de görüldüğü üzere modelin katlamaları %90 ile %92 arasında dar bir bantta seyretmektedir.

Bu istatistik, modelin derin mimarisinin veri setindeki özellikleri öğrenmede tutarlı olduğunu ancak Hibrit ve Özgün CNN modellerinin sunduğu hassasiyete tam olarak erişemediğini göstermektedir. ResNet50 modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 20’de grafiksel olarak sunulmuştur.



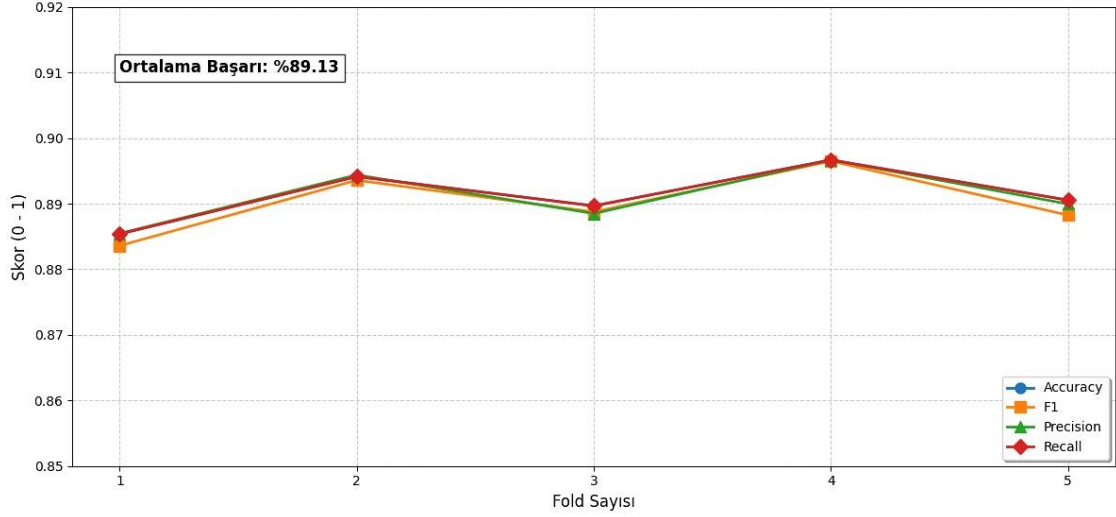
Şekil 20. ResNet50 modelinin k-fold performans grafiği

Şekil 20’de verilen ResNet50 k-fold performans grafiği değerlendirildiğinde, katlamalar arasında düşük varyans gözlenmiş olup ResNet50’nin tutarlı fakat sınıflandırma başarısı açısından optimize edilmiş mimarilere göre daha sınırlı kaldığı anlaşılmaktadır.

4.2.4. MobileNetV2 modelinin çapraz doğrulama performans analizi

Hafif mimarili MobileNetV2 modeli, %89,12 ortalama doğruluk oranı ile çapraz doğrulamayı tamamlamıştır. Tablo 11’deki değerler, modelin normal eğitimdeki %80’lik sonucuna göre çapraz doğrulamada daha iyi bir performans sergilediğini göstermektedir.

Karışıklık matrisindeki sınıfsal karışıklıklara rağmen, modelin genel veri dağılımında %89’luk bir başarıyı sürdürebilmesi dikkat çekicidir. MobileNetV2 modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 21’de grafiksel olarak sunulmuştur.



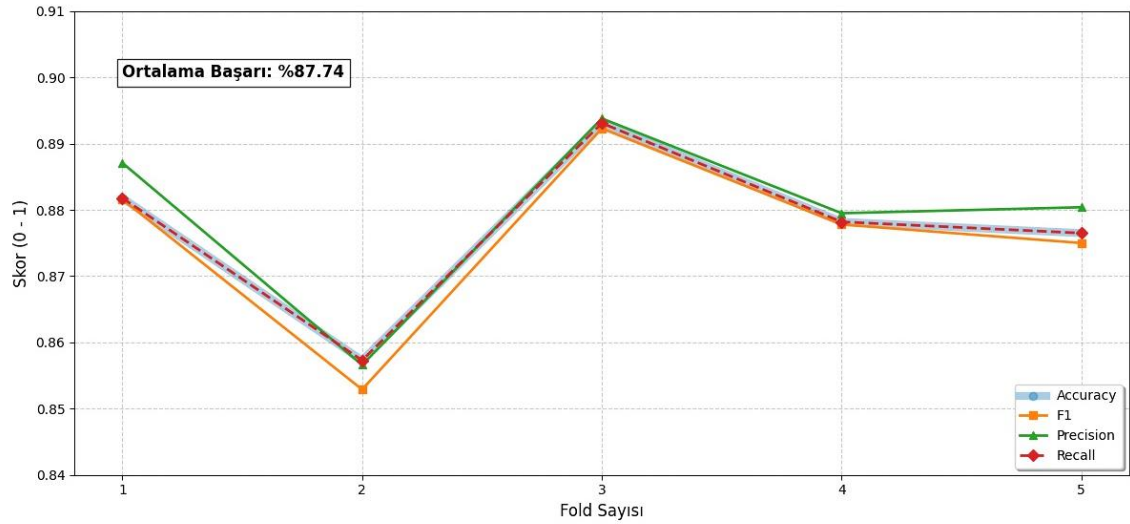
Şekil 21. MobileNetV2 modelinin k-fold performans grafiği

Şekil 21’de verilen MobileNetV2 k-fold performans grafiği incelendiğinde, modelin tüm katlarda birbirine yakın sonuçlar üretmesine rağmen mutlak başarı düzeyinin ResNet50 ve önerilen modellerden düşük olduğu görülmektedir. Bu durum, hafif mimarilerin hız avantajına karşılık sınıflandırma doğruluğu açısından bazı sınırlamalar içerdiğini göstermektedir.

4.2.5. DenseNet121 modelinin çapraz doğrulama performans analizi

DenseNet121 modeli, beş katlı çapraz doğrulama aşamasında %87,74 ortalama doğruluk elde etmiştir. Katlamalar bazında performans incelendiğinde; modelin en yüksek başarısını üçüncü katlamada %89,31 ile sergilediği, en düşük başarısını ise ikinci katlamada %85,73 ile aldığı görülmektedir. Modelin ortalama F1-skoru (%87,59), sınıflar arası genel bir denge sağlandığını gösterse de, katlamalar arasındaki yaklaşık %3,58’lik sapma, modelin veri setindeki bazı alt gruplara karşı diğer modellere oranla daha hassas olduğunu ortaya koymaktadır.

Bu sonuçlar, DenseNet121’in güçlü bir özellik iletim yapısına sahip olmasına rağmen, %93 bandına ulaşan Hibrit ve Özgün CNN modellerinin yaklaşık %5,41 gerisinde kaldığını doğrulamaktadır. DenseNet121 modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 22’de grafiksel olarak sunulmuştur.



Şekil 22. DenseNet121 modelinin k-fold performans grafiği

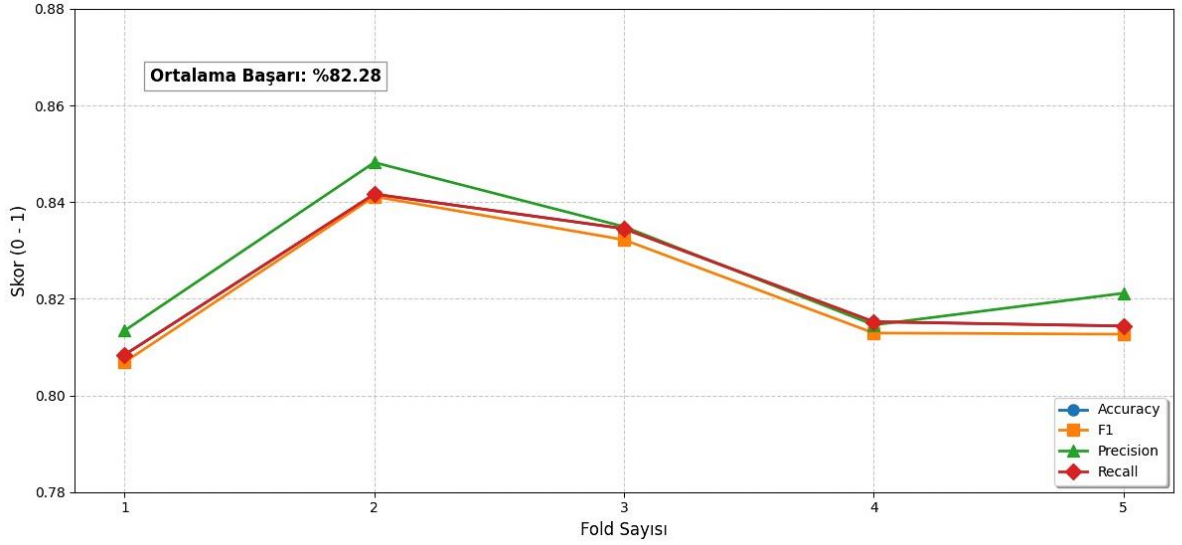
Şekil 22’de sunulan DenseNet121 k-fold performans grafiği değerlendirildiğinde, katlamalar arasında gözlemlenen belirgin dalgalanmaları (varyans) ortaya koymaktadır. Özellikle ikinci ve üçüncü katlamalar arasındaki performans farkı, DenseNet121 mimarisinin bu veri seti üzerindeki öğrenme kararlılığının düşük olduğunu ve dolayısıyla modelin genelleme kabiliyetinin sınırlı kaldığını doğrulamaktadır. Bu durum, yoğun bağlantılı yapının tıbbi görüntülerin karakteristik varyasyonlarına karşı Hibrit ve Özgün CNN modelleri kadar dirençli olmadığını göstermektedir.

4.2.6. VGG16 modelinin çapraz doğrulama performans analizi

VGG16 modeli, beş katlı çapraz doğrulama aşamasında %82,28 ortalama doğruluk oranı sergileyerek, güncel mimarilerin (ResNet, DenseNet, Hibrit) gerisinde kalmıştır. Tablo 11 verileri incelendiğinde; modelin en düşük performansının %80,83 (Fold 1), en yüksek performansının ise %84,16 (Fold 2) olduğu görülmektedir.

Modelin normal eğitim sürecinde elde ettiği yaklaşık %73’lük doğruluk oranına kıyasla çapraz doğrulama sonuçlarının daha yüksek çıkması, modelin veri setinin belirli alt kümelerinde (subsets) daha başarılı öz nitelik çıkarımı yapabildiğini göstermektedir. Ancak ortalama F1-skoru (%82,11) ve katlamalar arasındaki performans değişimi, VGG16’nın derin ve parametre yoğun yapısının, beyin tümörü görüntülerindeki karmaşık dokuları genelleştirmede Hibrit ve Özgün CNN modelleri kadar kararlı bir duruş sergileyemediğini ortaya koymaktadır.

VGG16 modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 23'te grafiksel olarak sunulmuştur.



Şekil 23. VGG16 modelinin k-fold performans grafiği

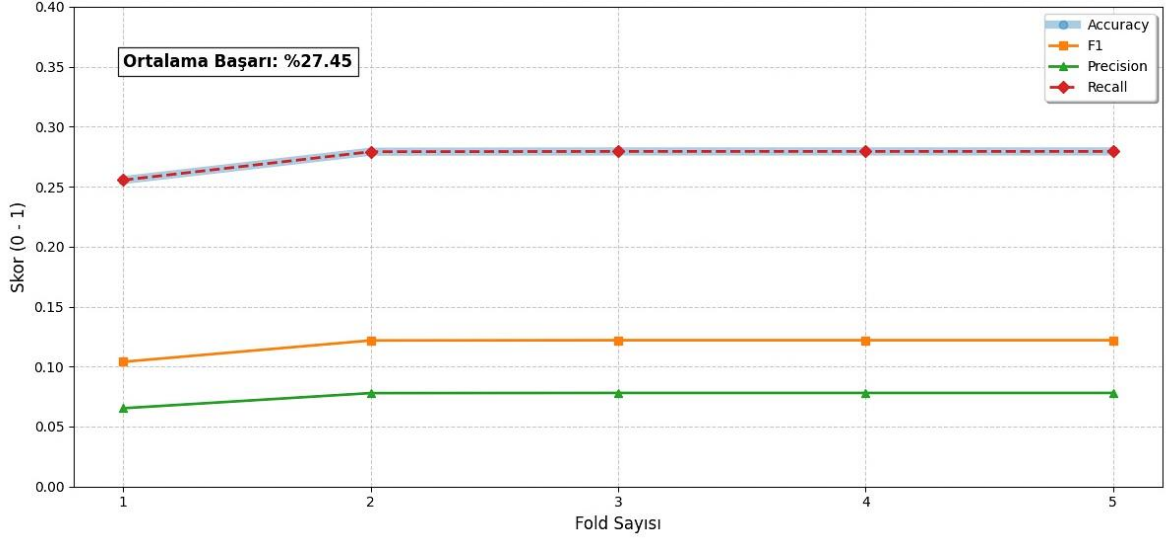
Şekil 23'te sunulan VGG16 k-fold performans grafiği değerlendirildiğinde, özellikle ilk ve son katlamalarda performansın belirgin biçimde düştüğü görülmekte olup, bu durum VGG16'nın parametre sayısının fazla olmasına rağmen bu veri seti için yeterince etkin temsil oluşturamadığını göstermektedir.

4.2.7. EfficientNetB0 modelinin çapraz doğrulama performans analizi

EfficientNetB0 modeli, çapraz doğrulama sürecinde de başarısızlığını sürdürerek %27,45 ortalama doğruluk oranında kalmıştır. Tablo 11 verilerine göre tüm katlamaların %25,54 ile %27,93 gibi dar bir bantta sıkışması, modelin veri setinden bağımsız olarak mimari bir öğrenme engeli yaşadığını ve hiçbir veri kombinasyonunda başarılı olamadığını kanıtlamaktadır.

Bu durum, EfficientNet mimarisinin temelini oluşturan bileşik ölçeklendirme (compound scaling) ve derinlik katsayılı evrişim (depthwise separable convolution) yapılarının, beyin tümörü görüntülerindeki spesifik doku özneliklerini yakalamada yetersiz kaldığını göstermektedir. Ayrıca, modelin hassasiyet (Precision) değerinin %0,0754 gibi çok düşük bir seviyede kalması, modelin sınıfları birbirinden ayırt edemediğini ve rastgele tahmin düzeyinin

dahi altında kaldığını doğrulamaktadır. EfficientNetB0 modelinin k-fold çapraz doğrulama sürecine ait performans değişimi Şekil 24'te grafiksel olarak sunulmuştur.



Şekil 24. EfficientNetB0 modelinin k-fold performans grafiği

Şekil 24'te sunulan EfficientNetB0 k-fold performans grafiği değerlendirildiğinde, metriklerin tüm katlarda oldukça düşük ve neredeyse sabit kaldığını göstermektedir. Bu durum, EfficientNetB0 mimarisinin bu veri seti üzerinde transfer öğrenme yoluyla etkin bir özellik çıkarımı gerçekleştiremediğini açıkça ortaya koymaktadır.

Çalışma kapsamında eğitilen yedi farklı modelin hem normal eğitim hem de beş katlı çapraz doğrulama sonuçları birlikte değerlendirildiğinde; Özgün CNN Modeli ve Hibrit Model'in beyin tümörü sınıflandırmasında açık ara en başarılı sistemler olduğu saptanmıştır. Özellikle Hibrit modelin çapraz doğrulamada ulaştığı %93,15 ortalama doğruluk, modelin klinik ortamlarda farklı hasta verileriyle karşılaştığında sergileyeceği yüksek başarının bir göstergesidir. Karışıklık matrisleri üzerinden yapılan analizler, modellerin çoğunluğunun "Notumor" ve "Pituitary" sınıflarında mükemmel yakın sonuçlar verdiğini; ancak "Meningioma" ve "Glioma" ayrımında derin mimarilerin daha hassas katmanlara ihtiyaç duyduğunu ortaya koymuştur.

Bu çalışma, Özgün CNN Model ve Hibrit yaklaşımlarının tıbbi görüntü analizinde hem doğruluk hem de güvenilirlik açısından geleneksel transfer öğrenme modellerine göre daha üstün bir alternatif sunduğunu kanıtlamaktadır.

5. TARTIŞMA VE SONUÇ

Bu çalışmada beyin tümörlerinin MRI görüntülerinden otomatik olarak sınıflandırılmasına yönelik olarak Özgün CNN, Hibrit, ResNet50, MobileNetV2, DenseNet121, VGG16 ve EfficientNetB0 modelleri karşılaştırmalı biçimde değerlendirilmiştir.

Normal eğitim sonuçları bağımsız test veri seti üzerinde elde edilmiş, çapraz doğrulama sonuçları ise yalnızca eğitim veri seti kullanılarak hesaplanmıştır. Modellerin performansları hem normal eğitim hem de beş katlı çapraz doğrulama stratejileri altında analiz edilmiştir.

5.1. Normal Eğitim Sonuçlarının Tartışılması

Tablo 3’de sunulan normal eğitim sonuçları incelendiğinde, en yüksek doğruluk değerinin %90.31 ile Özgün CNN modeli tarafından elde edildiği görülmektedir. Özgün CNN modelinin precision, recall ve F1-skoru metriklerinde de dengeli ve yüksek değerler üretmesi, önerilen mimarinin sınıflar arası ayırt edici özellikleri başarılı şekilde öğrendiğini göstermektedir. Literatürde, görev-özel olarak tasarlanan CNN mimarilerinin, özellikle medikal görüntüleme uygulamalarında genel amaçlı derin ağlara kıyasla daha etkili sonuçlar üretebildiği rapor edilmiştir (Khan vd., 2020; Singh ve Agarwal, 2023). Bu bağlamda elde edilen bulgular, literatürle uyumlu bir biçimde, özgün olarak tasarlanan CNN yapısının veri setine daha iyi adapte olduğunu ortaya koymaktadır.

Hibrit model, %85.51 doğruluk değeri ile normal eğitim senaryosunda ikinci en başarılı model olarak öne çıkmıştır. Özellikle sınıf bazlı sonuçlar incelendiğinde, “no tumor” sınıfında elde edilen %97 recall ve %95 F1-skoru, hibrit mimarinin sağlıklı beyin dokularını ayırt etmede yüksek güvenilirlik sunduğunu göstermektedir. Transfer öğrenme tabanlı mimarilerin medikal görüntülerde güçlü genellenebilir özellikler sunduğu, ancak görev-özel ince dokusal örüntülerin yakalanmasında sınırlı kalabildiği literatürde belirtilmektedir (Khan vd., 2020; Nazir vd., 2021).

Bu çalışmada hibrit mimarinin, MobileNetV2’nin genellenebilir özellikleri ile Özgün CNN modelinin yerel doku temsillerini birleştirmesi, bu sınırlılığı büyük ölçüde azaltmıştır. Normal eğitim sürecinde hibrit modelin Özgün CNN’in gerisinde kalması, önceden eğitilmiş ağırlıkların sınırlı epoch sayısı altında tam olarak adapte edilememesiyle açıklanabilir.

ResNet50 modeli %83.79 doğruluk değeri ile orta seviyede bir performans sergilemiştir. Derin artık (residual) bağlantıların eğitim kararlılığını artırdığı bilinmesine rağmen (He vd., 2016), bu çalışmada ResNet50'nin özellikle "meningioma" ve "glioma" sınıflarında daha düşük duyarlılık göstermesi, modelin yüksek derinliğinin veri setindeki sınırlı örnek çeşitliliğiyle uyumlu çalışmadığını düşündürmektedir. Benzer şekilde, literatürde derin mimarilerin küçük ve orta ölçekli medikal veri setlerinde aşırı karışıklık nedeniyle performans kaybı yaşayabildiği rapor edilmiştir (Nazir vd., 2021).

MobileNetV2 ve DenseNet121 modelleri sırasıyla %80.40 ve %79.63 doğruluk değerleri elde etmiştir. MobileNetV2'nin düşük parametre sayısına dayalı hafif mimarisi, hesaplama açısından avantaj sağlasa da, derin özellik temsillerinin sınırlı kalması özellikle "meningioma" sınıfında düşük recall değerlerine yol açmıştır. Benzer şekilde DenseNet121, yoğun bağlantı yapısı sayesinde özellik aktarımını güçlendirmesine rağmen, bu veri setinde beklenen performans artışını sağlayamamıştır. Bu durum, DenseNet mimarisinin yüksek bağlantı karmaşıklığının medikal MRI görüntülerinde her zaman avantaj sağlamadığını gösteren çalışmalarla örtüşmektedir (Khan vd., 2020; Nazir vd., 2021).

VGG16 modeli %73 doğruluk değeri ile nispeten zayıf bir performans sergilemiştir. VGG16'nın yaklaşık 138 milyon parametreye sahip olması, medikal görüntüler gibi sınırlı veri setlerinde aşırı öğrenmeye yatkın bir yapı oluşturabilmektedir. Bu durum, literatürde VGG tabanlı modellerin transfer öğrenme süreçlerinde medikal görüntüler için hantal kaldığını belirten çalışmalarla uyumludur (Khan vd., 2020; Nazir vd., 2021).

EfficientNetB0 modeli %30.89 doğruluk ile normal eğitim sürecinde en düşük performansı sergilemiştir. EfficientNet mimarisinin, ImageNet üzerinde doğal görüntüler için optimize edilmiş ölçekleme stratejilerine dayandığı bilinmektedir (Tan ve Le, 2019). Bu çalışmada elde edilen düşük performans, EfficientNetB0'nun derinlik ve genişlik katsayılarının MRI görüntülerindeki dokusal hiyerarşiyi yeterince temsil edemediğini ve medikal görüntülerde her zaman üstün performans garanti etmediğini göstermektedir. Benzer sonuçlar, bazı önceden eğitilmiş derin mimarilerin görev uyumsuzluğu nedeniyle medikal veri setlerinde başarısız olabildiğini rapor eden çalışmalarla örtüşmektedir (Nazir vd., 2021).

5.2. K-Fold Çapraz Doğrulama Sonuçlarının Tartışılması

Beş katlı çapraz doğrulama sonuçları incelendiğinde (Tablo 11), modellerin genelleme yetenekleri daha net biçimde ortaya konmuştur. Özgün CNN modeli, katlamalar arasında düşük varyansla %92.72 ortalama doğruluk değerine ulaşmış ve üçüncü katlamada %95.36 ile en yüksek performansını sergilemiştir. Bu durum, önerilen özgün mimarinin yalnızca tek bir eğitim-test bölünmesine değil, farklı veri dağılımlarına karşı da kararlı olduğunu göstermektedir. Literatürde, görev-özel CNN mimarilerinin çapraz doğrulama altında daha tutarlı performans sergilediği belirtilmiştir (Kumar vd., 2021; Singh ve Agarwal, 2023).

Hibrit model, %93.15 ortalama doğruluk değeri ile çapraz doğrulama sürecinde tüm modeller arasında en yüksek performansı elde etmiştir. Şekil 19'da sunulan K-Fold sonuçları, katlamalar arasındaki varyansın son derece düşük olduğunu ve hibrit mimarinin güçlü bir genelleme kapasitesine sahip olduğunu göstermektedir. Bu bulgu, transfer öğrenme tabanlı modellerin, Özel CNN yapılarıyla birleştirildiğinde daha ayırt edici ve kararlı özellik uzayları oluşturabildiğini rapor eden çalışmalarla uyumludur (Ali vd., 2022; Kumar vd., 2021). Normal eğitim sonuçlarına kıyasla hibrit modelin çapraz doğrulamada daha iyi performans göstermesi, modelin genelleme kabiliyetinin güçlendiğini ortaya koymaktadır.

ResNet50 modeli %91.20 ortalama doğruluk değeri ile tutarlı bir performans sergilemiş, ancak hibrit ve Özgün CNN modellerinin gerisinde kalmıştır. Şekil 20'de verilen katlama bazlı grafikler, modelin genel eğiliminin stabil olduğunu, ancak bazı sınıflarda hatalı sınıflandırmaların devam ettiğini göstermektedir. Bu durum, derin mimarilerin çapraz doğrulama altında daha stabil çalışmasına rağmen, veri setine özgü ince dokusal ayrımları yakalamada sınırlı kalabildiğini düşündürmektedir (He vd., 2016).

MobileNetV2 modeli, hafif mimarisine rağmen beş katlı çapraz doğrulama sonucunda %89.12 ortalama doğruluk elde ederek dengeli bir performans sergilemiştir. Katlamalar arasındaki düşük varyans, MobileNetV2'nin genellenebilir özellik temsili açısından dirençli bir yapı sunduğunu göstermektedir. Literatürde MobileNet tabanlı mimarilerin medikal görüntülerde düşük hesaplama maliyetiyle rekabetçi performans sunabildiği vurgulanmaktadır (Sandler vd., 2018; Swati vd., 2019). Ancak, tek başına kullanıldığında özellik haritalarının sınırlı derinliği nedeniyle, hibrit ve Özgün CNN modellerinin gerisinde kalmıştır.

DenseNet121 modeli %87.74 ortalama doğruluk elde etmiş ve katlamalar arasında MobileNetV2'ye kıyasla daha yüksek varyans göstermiştir. Yoğun bağlantı yapısının özellik aktarımını güçlendirmesine rağmen, bu mimarinin parametre karmaşıklığı nedeniyle önerilen modellere kıyasla daha sınırlı bir genelleme performansı sunduğu görülmektedir. Benzer gözlemler, DenseNet tabanlı mimarilerin bazı medikal veri setlerinde beklenen performans artışını sağlayamadığını bildiren çalışmalarla uyumludur (Huang vd., 2017).

VGG16 modeli %82.28 ortalama doğruluk ile çapraz doğrulama sürecinde de orta düzeyde bir performans sergilemiştir. EfficientNetB0 modeli ise %27.45 ortalama doğruluk ile çapraz doğrulamada da başarısız bir performans göstermiştir. Bu sonuçlar, ImageNet tabanlı ön eğitim stratejilerinin her medikal görüntüleme problemi için uygun olmayabileceğini ve model seçiminin veri setine özgü yapılması gerektiğini vurgulayan literatürle örtüşmektedir (Kumar vd., 2021; Nazir vd., 2021).

Genel olarak elde edilen bulgular, hem normal eğitim hem de çapraz doğrulama stratejileri altında Özgün CNN ve Hibrit modellerin diğer derin öğrenme mimarilerine kıyasla daha yüksek doğruluk, daha dengeli precision–recall değerleri ve daha güçlü genelleme yetenekleri sunduğunu göstermektedir. Özellikle hibrit mimarinin, MobileNetV2'nin düşük parametreliliği ve genellenebilir yapısı ile Özgün CNN modelinin yerel özellik çıkarım kabiliyetini birleştirmesi, literatürde rapor edilen yüksek başarı oranlarıyla uyumlu ve rekabetçi sonuçlar üretmiştir.

5.3. Sonuç

Bu çalışmada, beyin tümörlerinin MRI görüntülerinden otomatik olarak sınıflandırılması amacıyla Özgün CNN, Hibrit, ResNet tabanlı ResNet50, MobileNet tabanlı MobileNetV2, DenseNet tabanlı DenseNet121, VGG tabanlı VGG16 ve EfficientNet tabanlı EfficientNetB0 modelleri karşılaştırmalı olarak değerlendirilmiştir. Modeller hem standart eğitim-test bölünmesi hem de beş katlı çapraz doğrulama yöntemi kullanılarak analiz edilmiştir.

Elde edilen sonuçlar, görev-özel olarak tasarlanan Özgün CNN mimarisinin normal eğitim senaryosunda en yüksek doğruluk değerine ulaştığını göstermiştir. Buna karşılık, çapraz doğrulama sonuçları hibrit modelin farklı veri bölünmelerinde daha kararlı performans sergilediğini ve en yüksek ortalama doğruluk değerini elde ettiğini ortaya koymuştur. Bu bulgular, özgün olarak tasarlanan mimarilerin veri setine daha iyi uyum sağlayabildiğini,

transfer öğrenme ile desteklenen hibrit yaklaşımların ise genelleme performansını artırabildiğini göstermektedir.

Bu çalışmada kullanılan beş katlı çapraz doğrulama yöntemi yalnızca eğitim veri seti üzerinde uygulanmış, bağımsız test veri seti ise modellerin nihai performanslarının değerlendirilmesi amacıyla kullanılmıştır. Bu yaklaşım, model seçim sürecinde veri sızıntısını önleyerek elde edilen sonuçların daha güvenilir ve genellenebilir olmasını sağlamıştır.

Deneysel sonuçlar, derin ve yüksek parametrelili mimarilerin her zaman daha iyi performans üretmediğini göstermiştir. Özellikle VGG16 ve EfficientNetB0 modellerinin görece düşük performansı, model karmaşıklığının veri seti büyüklüğü ve özellik yapısı ile uyumlu olması gerektiğini ortaya koymaktadır. Buna karşılık, daha kompakt ve görev-özel mimariler ile hibrit yaklaşımlar daha başarılı sonuçlar üretmiştir.

Çalışmanın önemli katkılarından biri, özgün olarak tasarlanan CNN mimarisi ile transfer öğrenme tabanlı bir modelin birleştirilmesiyle oluşturulan hibrit yapının karşılaştırmalı olarak analiz edilmesidir. Elde edilen bulgular, hibrit mimarilerin medikal görüntü sınıflandırma problemlerinde güçlü bir alternatif oluşturduğunu göstermektedir.

Bununla birlikte çalışmanın bazı sınırlılıkları bulunmaktadır. Kullanılan veri setinin sınırlı büyüklüğü, özellikle derin mimarilerin potansiyel performansının tam olarak ortaya çıkmasını engellemiş olabilir. Ayrıca çalışma yalnızca MRI görüntüleri ile sınırlı tutulmuş olup farklı görüntüleme modaliteleri değerlendirilmemiştir.

Gelecek çalışmalarda daha büyük ve çok merkezli veri setlerinin kullanılması, veri artırma stratejilerinin genişletilmesi ve dikkat mekanizmaları (attention mechanisms) veya transformer tabanlı mimarilerin incelenmesi önerilmektedir. Ayrıca geliştirilen modellerin klinik karar destek sistemlerine entegre edilmesi, bu alandaki uygulamalı çalışmalar açısından önemli bir adım olacaktır. Genel olarak elde edilen sonuçlar, önerilen Özgün CNN ve hibrit modellerin beyin tümörü sınıflandırma probleminde etkili ve güvenilir yaklaşımlar sunduğunu göstermektedir. Bu modellerin CAD sistemlerinde kullanılabilecek önemli bir potansiyele sahip olduğu değerlendirilmektedir.

6. ÖNERİLER

Bu çalışma kapsamında elde edilen bulgular doğrultusunda aşağıdaki öneriler sunulmaktadır:

- Gelecek çalışmalarda veri seti daha geniş hasta popülasyonları ile zenginleştirilmeli ve farklı Manyetik Rezonans Görüntüleme sekansları (T1, T2, FLAIR) birlikte kullanılarak çok kanallı girdi yapısı oluşturulmalıdır.
- Önerilen hibrit mimarinin farklı önceden eğitilmiş ağlar ile entegrasyonu test edilerek mimarinin genellenebilirliği analiz edilmelidir.
- Klinik kullanıma yönelik olarak model çıktılarının, ağın odaklandığı bölgeleri görselleştiren ve karar verme sürecini şeffaf hale getiren Grad-CAM ve türevi yöntemlerle desteklenmesi önerilmektedir (Chattopadhyay vd., 2020).
- Çok merkezli veri setleriyle doğrulama yapılarak modelin farklı cihaz, çekim protokolleri ve merkezler arası varyasyonlara karşı dayanıklılığı araştırılmalıdır.
- Veri dengesizliği problemini azaltmak amacıyla SMOTE, focal loss veya sınıf ağırlıklandırma gibi yöntemlerin etkisi incelenmelidir.
- Model performansı yalnızca doğruluk temelli değil; ROC–AUC, sınıf bazlı F1-skorları ve klinik hata türlerine (özellikle yanlış negatifler) odaklanan metriklerle de değerlendirilmelidir.
- Modelin hafifletilmesi amacıyla bilgi damıtma (knowledge distillation) ve model sıkıştırma teknikleri uygulanarak gerçek zamanlı klinik kullanım için hesaplama maliyetleri düşürülmelidir.
- MRI görüntülerine ek olarak klinik demografik verilerin entegre edildiği çok kipli (multimodal) yaklaşımların tanı performansına etkisi araştırılmalıdır.
- Modelin klinik karar destek sistemlerine entegrasyonu için kullanıcı dostu grafik arayüzler geliştirilerek hekim geri bildirimleri doğrultusunda sistemin pratikteki kullanılabilirliği değerlendirilmelidir.

KAYNAKÇA

- Afshar, P., Plataniotis, K. N., & Mohammadi, A. (2020). Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1170–1174).
- Ahmad, S., Ullah, T., Ahmad, I., Al-Sharabi, A., Ullah, K., ... & Ali, S. (2022). A novel hybrid deep learning model for metastatic cancer detection. *Computational Intelligence and Neuroscience, 2022*, 8141530.
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *Proceedings of the 2017 International Conference on Engineering and Technology (ICET)*, 1–6.
- Ali, L., et al. (2022). Hybrid deep learning models for medical image classification using feature fusion. *Scientific Reports, 12*(1), 1-15.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., ... & Fadhel, M. A. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, 53.
- Amin, J., Sharif, M., Haldorai, A., et al (2022). Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex & Intelligent Systems, 8*, 3161–3183.
- Anaraki, A. K., Ayati, M., & Kazemian, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering, 39*(1), 63-74.
- Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. K. M., & Moni, M. A. (2022). A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Computers in Biology and Medicine, 146*, 105539.
- Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., ... & Crimi, A. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.
- Batool, A., & Byun, Y. C. (2025). A lightweight multi-path convolutional neural network architecture using optimal features selection for multiclass classification of brain tumor using magnetic resonance images. *Results in Engineering, 25*, 104327. doi:10.1016/j.rineng.2025.104327

- Celik, M., & Inik, O. (2024). Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification. *Expert Systems with Applications*, 238(Part E), 122159. <https://doi.org/10.1016/j.eswa.2023.122159>
- Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2020). Grad-CAM++: Generalized gradient-weighted class activation mapping for visual explanations. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 839–847).
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 1-13.
- Goodenberger, M. L., & Jenkins, R. B. (2012). Genetics of adult glioma. *Cancer Genetics*, 205(12), 613–621.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708.
- Hussain, M., et al. (2024). Res-BRNet: A deep residual and regional convolutional neural network for brain tumor classification from MRI. *Journal of Medical Imaging and Health Informatics*.
- International Agency for Research on Cancer. (2022). *The Global Cancer Observatory—All Cancers*. Lyon, France: International Agency for Research on Cancer—WHO.
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 448–456).
- Kaggle. (2022). Brain Tumor Classification (MRI). Erişim Tarihi: 18 Ocak 2025, <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>
- Khan, M. A., et al. (2020). Brain tumor classification using fine-tuning of deep convolutional neural networks. *IEEE Access*, 8, 103444-103454.

- Khazaei, Z., Goodarzi, E., Borhaninejad, V., Iranmanesh, F., ... Sarabi, R. E. (2020). The association between incidence and mortality of brain cancer and human development index (HDI): An ecological study. *BMC Public Health*, 20, 1696.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (San Diego, CA, USA).
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 14(2), 1137–1145.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 25 (pp. 1097–1105).
- Kumar, S., ... et al. (2021). Robust evaluation of CNN models for brain MRI classification with K-fold cross-validation. *Journal of Medical Systems*, 45(3), 1–12.
- Kumar, Y., Koul, A., Singh, R., & Pyari, M. (2017). Image segmentation and classification techniques for brain tumor detection: A review. *International Journal of Computer Science and Mobile Computing*, 6(11), 31–38.
- Kwan, R. K.-S., Evans, A. C., & Pike, G. B. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11), 1085–1097.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M. J., & Vercauteren, T. (2021). On the importance of effective validation in medical image classification. *Medical Image Analysis*, 65, 101766.
- Li, Z., et al. (2021). Robustness evaluation of deep learning models in medical image analysis. *IEEE Access*.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
- Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., ... & Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathologica*, 131, 803–820.
- McDermott, M. B. A., Wang, S., Marinsek, N., Ranganath, R., ... & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. *Science Translational Medicine*, 13(586), eabb1655.

- Mohan, P., Subramaniam, V., & Singh, J. P. (2021). Effective evaluation of deep learning models for medical image classification. *IEEE Access*, 9, 11234-11245.
- Nazir, M., et al. (2021). Deep learning for multi-class brain tumor classification: A review of state-of-the-art techniques. *Informatics in Medicine Unlocked*, 23, 100519.
- Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5), 1240–1251.
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- 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.
- Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2021). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing*, 40(1), 508–524.
- Rogers, L., Barani, I., Chamberlain, M., Kaley, T. J., ... & Vogelbaum, M. A. (2015). Meningiomas: Knowledge base, treatment outcomes, and uncertainties. A RANO review. *Journal of Neurosurgery*, 122, 4–23.
- Sadr, H., Nazari, M., Yousefzadeh-Chabok, S., Emami, H., Rabiei, R., & Ashraf, A. (2025). Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis. *Image and Vision Computing*, 159, 105555. doi:10.1016/j.imavis.2025.105555
- Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of Computational Science*, 30, 174–182.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510-4520.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA.

- Singh, S., & Agarwal, A. (2023). Custom vs. pre-trained CNN architectures for medical imaging: A comparative analysis. *International Journal of Computer Applications*, 185(12), 22-30.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Sultan, H., Owais, M., Nam, S. H., Haider, A., Akram, R., Usman, M., & Park, K. R. (2023). MDFU-net: multiscale dilated features up-sampling network for accurate segmentation of tumor from heterogeneous brain data. *Journal of King Saud University - Computer and Information Sciences*, 35(5), 101560.
- Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, 75, 34–46.
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 6105–6114.
- Tiwari, A., Srivastava, S., & Pant, M. (2019). Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern Recognition Letters*, 131, 244–260.
- Wadhwa, A., Bhardwaj, A., & Verma, V. S. (2019). A review on brain tumor segmentation of MRI images. *Magnetic Resonance Imaging*, 61, 247–259.
- Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2023). Robust evaluation of medical image segmentation and classification. *Medical Image Analysis*, 84, 102698.
- Wong Y, Su ELM, Yeong CF, Holderbaum W, Yang C (2025) Brain tumor classification using MRI images and deep learning techniques. *PLoS One* 20(5): e0322624. <https://doi.org/10.1371/journal.pone.0322624>.
- Xiong, S., Wu, G., Fan, X., et al (2021). MRI-based brain tumor segmentation using FPGA-accelerated neural network. *BMC Bioinformatics*, 22.
- Yadav, S., & Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In *Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 78–83.

- Yang, J., Siddique, M. A., Ullah, H., Gilanie, G., Por, L. Y., Alshathri, S., . . . Gadekallu, T. R. (2025). BrainCNN: Automated brain tumor grading from magnetic resonance images using a convolutional neural network-based customized model. *SLAS Technology*, *34*, 100334. doi:10.1016/j.slast.2025.100334
- Yang, X., Yu, Z., & Lu, S. (2022). Comprehensive performance evaluation of deep learning in brain tumor diagnosis. *Diagnostic Pathology*, *17*, 45.
- Yao, J., et al. (2022). Cross-validation strategies for deep learning in medical image analysis. *Medical Image Analysis*, *78*, 102400.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, *27*.
- Zeineldin, R. A., Karar, M. E., Cobca, J., Aradi, B., ... & Burgert, O. (2020). DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance images. *Access IEEE*, *8*, 75920–75939.
- Zulfiqar, F., Bajwa, U. I., & Mehmood, Y. (2023). Multi-class classification of brain tumor types from MR images using EfficientNets. *Biomedical Signal Processing and Control*, *84*, 104777.