



**MEDİKAL GÖRÜNTÜ ANALİZİNDE GÜRÜLTÜ SALDIRILARINA
KARŞI DERİN ÖĞRENME MODELLERİNİN PERFORMANSLARININ
KARŞILAŞTIRILMASI**

Gökçe OK

**YÜKSEK LİSANS TEZİ
BİLGİ GÜVENLİĞİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2024

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Gökçe OK

19/07/2024

MEDİKAL GÖRÜNTÜ ANALİZİNDE GÜRÜLTÜ SALDIRILARINA KARŞI DERİN ÖĞRENME MODELLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI

(Yüksek Lisans Tezi)

Gökçe OK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2024

ÖZET

Medikal görüntülerin analizi tıbbi tanı ve teşhis çalışmalarında önemli rol oynamaktadır. Teknolojinin gelişimiyle yapay zeka destekli medikal görüntü analiz sistemlerinin kullanımı yaygınlaşmaktadır. Birçok alanda yüksek başarı sağlayan derin öğrenme modelleri sağlık sektöründe de ilgi görmektedir. Medikal görüntü analizinde kullanılan derin öğrenme modelleri yüksek doğruluk ile görevlerini yerine getirmesine karşın birtakım saldırılara karşı savunmasızdır. Medikal görüntülere çok küçük düzeylerde gürültü eklenerek derin öğrenme modellerinin manipülasyonuna neden olan çekişmeli saldırılar dikkate alınması gereken tehlikelerin başında gelmektedir. Bu saldırıların doğasının anlaşılması ve medikal görüntü analizinde kullanılan derin öğrenme modellerinin hassasiyetlerinin keşfedilmesi, uygun savunma yöntemlerinin geliştirilmesi açısından önem arz etmektedir. Bu çalışmada, tümör tespitinde kullanılan manyetik rezonans (MR) ve bilgisayarlı tomografi (BT) görüntü setleri üzerinde iki gradyan tabanlı gürültü saldırısının etkinliği değerlendirilmektedir. Veri setlerine tek adımlı hızlı gradyan işareti yöntemi (FGSM) ve iteratif bir saldırı olan izdüşürülen gradyan inişi (PGD) uygulanmaktadır. Her bir veri seti için temiz test setleri ve saldırıların üç farklı pertürbasyon parametresi ile ürettiği çekişmeli örneklerden oluşan 7 senaryo ele alınmaktadır. Bu senaryolar 20 farklı ön eğitilmiş derin öğrenme modeli kullanılarak analiz edilmekte ve derinlemesine bir karşılaştırma gerçekleştirilmektedir. Aynı zamanda pertürbasyon parametresi, görüntü modalitesi, derin öğrenme algoritmaları ve saldırıların iteratif özellikleri gibi unsurların saldırı başarısı üzerindeki etkisi incelenmektedir. Ayrıca, FGSM ve PGD saldırıları ile üretilen örneklerin orijinal görüntüye benzerlikleri ve kaliteleri ölçülerek görüntülerdeki pertürbasyon derecesine bağlı bozulmanın saldırı başarısı ile ilişkisi araştırılmaktadır. Elde edilen deneysel sonuçlar ve bulgular kapsamlı olarak sunulmaktadır.

Bilim Kodu : 92403

Anahtar Kelimeler : Medikal görüntü analizi, gürültü saldırısı, derin öğrenme, transfer öğrenmesi, FGSM, sınıflandırma

Sayfa Adedi : 270

Danışman : Prof. Dr. Murat DENER

BENCHMARKING OF DEEP LEARNING MODELS AGAINST ADVERSARIAL ATTACKS IN MEDICAL IMAGE ANALYSIS

(M. Sc. Thesis)

Gökçe OK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2024

ABSTRACT

Medical image analysis plays a pivotal role in medical diagnostics and research studies. With the advent of technology, the proliferation of artificial intelligence-supported systems for medical image analysis is increasingly prevalent. Deep learning models, renowned for their high efficacy across various domains, are gaining prominence within the healthcare sector. Despite their commendable accuracy in medical image analysis tasks, deep learning models are susceptible to adversarial attacks. These attacks involve the manipulation of models by introducing imperceptible perturbations to medical images. Understanding the nature of such attacks and identifying the vulnerabilities of deep learning models in medical image analysis is paramount for devising effective defense mechanisms. In this study, we assess the efficacy of two gradient-based adversarial attacks, namely fast gradient signature method (FGSM) and projected gradient descent (PGD), on magnetic resonance (MR) and computed tomography (CT) image datasets utilized for tumor detection. We conduct experiments across seven scenarios for each dataset, comprising clean test sets and adversarial examples generated by attacks with varying perturbation parameters. Subsequently, we analyze these scenarios using twenty distinct pre-trained deep learning models, facilitating an extensive comparative analysis. Furthermore, we investigate the impact of factors such as perturbation parameters, image modalities, deep learning algorithms, and the iterative nature of attacks on attack success rates. Additionally, we assess the similarity and quality of samples generated by FGSM and PGD attacks concerning the original images, elucidating the relationship between image distortion due to perturbation and attack success. The experimental results and findings are comprehensively presented to provide insights into the effectiveness of adversarial attacks in medical image analysis.

Science Code : 92403

Key Words : Medical image analysis, adversarial attacks, deep learning, transfer learning, FGSM, classification

Page Number : 270

Supervisor : Prof. Dr. Murat DENER

TEŞEKKÜR

Bu tezi hazırlamama destek olan, bilgi ve tecrübelerini benimle paylaşan değerli danışman hocam Prof. Dr. Murat DENER' e teşekkürlerimi sunarım. Bu süreçte her zaman yanımda olan, bana her koşulda inanan ve desteklerini esirgemeyen çok sevgili aileme, karşılaştığım her zorlukta devam etmem için beni teşvik eden kıymetli arkadaşlarıma içten şükranlarımı sunarım. Ayrıca, Bilim İnsanı Destek Programları Başkanlığı (BİDEB) eğitim ve burs destekleri kapsamında beni destekleyen ve tezimin gerçekleştirilmesine katkıda bulunan TÜRKİYE BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA KURUMU (TÜBİTAK)'a minnettarlığımı sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	xii
ŞEKİLLERİN LİSTESİ.....	xiv
SİMGELER VE KISALTMALAR	xxviii
1. GİRİŞ.....	1
2. KAVRAMSAL ÇERÇEVE.....	5
2.1. Medikal Görüntü Analizi	5
2.1.1. Medikal görüntüleme	5
2.1.2. Medikal görüntü analizinde güvenlik.....	8
2.2. Çekişmeli Saldırıları	10
2.2.1. Çekişmeli örnek üretimi	12
2.2.2. Hızlı gradyan işaretli yöntemi (FGSM)	14
2.2.3. İzdüşürülen gradyan inişi (PGD)	14
2.3. Derin Öğrenme	15
2.3.1. Evrimsel sinir ağları	16
2.3.2. Evrimsel sinir ağ mimarileri	18
2.4. Değerlendirme Metrikleri	33
2.4.1. Model performans değerlendirme metrikleri	33
2.4.2. Görüntü kalite ve benzerlik değerlendirme metrikleri.....	36

	Sayfa
3. LİTERATÜR TARAMASI	41
3.1. Mevcut Saldırıların Medikal Görüntüler Üzerindeki Etkinliğine İlişkin Yapılan Çalışmalar	41
3.2. Medikal Görüntülere Yönelik Geliştirilen Çekişmeli Saldırıları	43
3.3. Medikal Görüntü Analizinde Çekişmeli Saldırıları Karşı Geliştirilen Savunma Yöntemleri	44
4. KULLANILAN ARAÇLAR VE VERİ SETLERİ	49
4.1. Kullanılan Araçlar	49
4.1.1. Google Colaboratory.....	49
4.1.2. Numpy	50
4.1.3. Scikit-learn	50
4.1.4. Tensorflow ve Keras	50
4.1.5. Matplotlib ve Seaborn.....	50
4.1.6. Sear	51
4.2. Veri Setleri	51
4.2.1. Beyin tümör MR görüntü seti.....	51
4.2.2. Böbrek BT görüntü seti	52
4.2.3. Verilerin ön işleme	52
5. SINIFLANDIRMA MODELLERİNİN EĞİTİMİ	55
5.1. Deney Ortamı	55
5.2. Deney Tasarımı.....	55
5.3. Hiper Parametre Değerlerinin Belirlenmesi	56
5.4. İnce Ayarlı Transfer Öğrenme	57
5.5. Modellerin Eğitimi.....	58

	Sayfa
5.6. Çekişmeli Test Örneklerinin Üretimi	60
6. DENEY SONUÇLARI VE BULGULAR	63
6.1. Beyin Tümör MR Görüntü Seti Deney Sonuçları	63
6.1.1. Xception.....	66
6.1.2. VGG16.....	70
6.1.3. VGG19	74
6.1.4. DenseNet121	78
6.1.5. DenseNet169	82
6.1.6. DenseNet201	86
6.1.7. EfficientNetV2B0	90
6.1.8. EfficientNetV2B1	94
6.1.9. EfficientNetV2B2	98
6.1.10. EfficientNetV2B3	102
6.1.11. EfficientNetV2S	106
6.1.12. InceptionResNetV2	110
6.1.13. InceptionV3.....	114
6.1.14. MobileNetV1	119
6.1.15. MobileNetV2	124
6.1.16. MobileNetV3Small	128
6.1.17. NASNetMobile	132
6.1.18. ResNet50V2	136
6.1.19. ResNet101V2	140
6.1.20. ResNet152V2	144

	Sayfa
6.2. Böbrek BT Görüntü Seti Deney Sonuçları	149
6.2.1. Xception	153
6.2.2. VGG16	157
6.2.3. VGG19	161
6.2.4. DenseNet121	165
6.2.5. DenseNet169	169
6.2.6. DenseNet201	173
6.2.7. EfficientNetV2B0	177
6.2.8. EfficientNetV2B1	181
6.2.9. EfficientNetV2B2	185
6.2.10. EfficientNetV2B3	189
6.2.11. EfficientNetV2S	193
6.2.12. InceptionResNetV2	197
6.2.13. InceptionV3	201
6.2.14. MobileNetV1	205
6.2.15. MobileNetV2	209
6.2.16. MobileNetV3Small	213
6.2.17. NASNetMobile	217
6.2.18. ResNet50V2	231
6.2.19. ResNet101V2	225
6.2.20. ResNet152V2	229
6.3. Görüntülerin Kalite ve Benzerlik Değerlendirmesi	233
6.3.1. Beyin Tümör MR görüntülerinin saldırılara karşı kalite ve benzerliklerinin değerlendirilmesi	233

Sayfa

6.3.2. Böbrek BT görüntülerinin saldırılara karşı kalite ve benzerliklerinin değerlendirilmesi.....	235
6.4. Veri Setlerinin Performans ve Benzerlik Karşılaştırması.....	237
7. SONUÇ VE ÖNERİLER	247
KAYNAKLAR	251
EKLER.....	259
EK-1. Beyin Tümör MR Veri Seti İçin Modellerin Eğitimi.....	260
EK-2. Böbrek BT Veri Seti İçin Modellerin Eğitimi.....	265
ÖZGEÇMİŞ	270

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. Kullanılan görüntü setlerinin ikili sınıflandırmaya göre örnek dağılımları	53
Çizelge 5.1. Beyin Tümör MR görselleri ile eğitilen modellerin eğitimlerinin sayısal verileri	59
Çizelge 5.2. Böbrek BT görselleri ile eğitilen modellerin eğitimlerinin sayısal verileri	60
Çizelge 5.3. Bir Böbrek BT görüntüsü üzerinde FGSM saldırısı uygulanması ile meydana gelen gürültü ve çekişmeli örnekler.....	61
Çizelge 5.4. Bir Beyin Tümör MR görüntüsü üzerinde PGD saldırısı uygulanması ile meydana gelen gürültü ve çekişmeli örnekler	62
Çizelge 6.1. Beyin Tümör MR test görüntüleri ile modellerin elde ettiği performans değerleri	63
Çizelge 6.2. Beyin Tümör MR görüntüleri üzerine uygulanan FGSM saldırıları sonucunda modellerin elde ettiği performans değerleri	64
Çizelge 6.3. Beyin Tümör MR görüntüleri üzerine uygulanan PGD saldırıları sonucunda modellerin elde ettiği performans değerleri.....	65
Çizelge 6.4. Böbrek BT test görüntüleri ile modellerin elde ettiği performans değerleri	150
Çizelge 6.5. Böbrek BT görüntüleri üzerine uygulanan FGSM saldırıları sonucunda modellerin elde ettiği performans değerleri	151
Çizelge 6.6. Böbrek BT görüntüleri üzerine uygulanan PGD saldırıları sonucunda modellerin elde ettiği performans değerleri	152
Çizelge 6.7. Beyin Tümör MR test görüntüleri üzerine uygulanan FGSM ve PGD saldırıları sonucunda elde edilen görüntülerin ortalama kalite ve benzerlik skorları	235
Çizelge 6.8. Böbrek BT test görüntüleri üzerine uygulanan FGSM ve PGD saldırıları sonucunda elde edilen görüntülerin ortalama kalite ve benzerlik skorları	237
Çizelge 6.9. Beyin Tümör MR görüntüleri üzerinde orijinal görüntülerin sınıflandırılma başarısı ile FGSM ve PGD saldırı senaryolarının başarıları sınıflandırma arasındaki doğruluk oranı farkları	240

Çizelge	Sayfa
Çizelge 6.10. Böbrek BT görüntüleri üzerinde orijinal görüntülerin sınıflandırılma başarısı ile FGSM ve PGD saldırı senaryolarının sınıflandırma başarıları arasındaki doğruluk oranı farkları.....	241

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Temel CNN mimarisi	16
Şekil 2.2. VGG16 modeline ait mimari gösterim	19
Şekil 2.3. VGG19 modeline ait mimari gösterim	20
Şekil 2.4. InceptionV1 modül gösterimi	21
Şekil 2.5. InceptionV2 modül gösterimi	22
Şekil 2.6. InceptionV3 modül gösterimi	22
Şekil 2.7. ResNetV1 ve ResNetV2 blok yapıları	24
Şekil 2.8. ResNet mimari gösterim	25
Şekil 2.9. InceptionResNetV2 blok yapıları	26
Şekil 2.10. DenseNet mimari gösterimi	27
Şekil 2.11. Xception mimari gösterimi	28
Şekil 2.12. MobileNetV1 mimari gösterimi	30
Şekil 2.13. MobileNetV2 mimari gösterimi	31
Şekil 2.14. MBConv ve Fused-MBConv blok yapıları	33
Şekil 2.15. Karmaşıklık matrisi	34
Şekil 4.1. Kullanılan araçların logoları	49
Şekil 4.2. Beyin Tümör MR görüntü örnekleri (soldaki tümör, sağdaki normal)	52
Şekil 4.3. Böbrek BT görüntü örnekleri (soldaki tümör, sağdaki normal)	52
Şekil 5.1. Çalışma boyunca takip edilen deney prosesi	55
Şekil 5.2. Çalışmada kullanılan genel model tasarımı	56
Şekil 6.1. Temiz Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi	66
Şekil 6.2. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi	67

Şekil	Sayfa
Şekil 6.3. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait ROC eğrileri.....	68
Şekil 6.4. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi.....	69
Şekil 6.5. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait ROC eğrileri.....	70
Şekil 6.6. Temiz Beyin Tümör MR görüntüleri üzerinde VGG16 modeline ait karmaşıklık matrisi.....	70
Şekil 6.7. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait karmaşıklık matrisi.....	71
Şekil 6.8. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait ROC eğrileri.....	72
Şekil 6.9. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait karmaşıklık matrisi.....	73
Şekil 6.10. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait ROC eğrileri.....	74
Şekil 6.11. Temiz Beyin Tümör MR görüntüleri üzerinde VGG19 modeline ait karmaşıklık matrisi.....	74
Şekil 6.12. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait karmaşıklık matrisi.....	75
Şekil 6.13. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait ROC eğrileri.....	76
Şekil 6.14. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait karmaşıklık matrisi.....	77
Şekil 6.15. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait ROC eğrileri.....	78
Şekil 6.16. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet121 modeline ait karmaşıklık matrisi.....	78
Şekil 6.17. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi.....	79
Şekil 6.18. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait ROC eğrileri	80

Şekil	Sayfa
Şekil 6.19. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi.....	81
Şekil 6.20. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait ROC eğrileri.....	82
Şekil 6.21. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet169 modeline ait karmaşıklık matrisi.....	82
Şekil 6.22. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi.....	83
Şekil 6.23. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait ROC eğrileri	84
Şekil 6.24. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi.....	85
Şekil 6.25. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait ROC eğrileri.....	86
Şekil 6.26. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet201 modeline ait karmaşıklık matrisi.....	86
Şekil 6.27. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi.....	87
Şekil 6.28. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait ROC eğrileri	88
Şekil 6.29. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi.....	89
Şekil 6.30. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait ROC eğrileri.....	90
Şekil 6.31. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi.....	90
Şekil 6.32. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi	91
Şekil 6.33. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri.....	92
Şekil 6.34. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi	93

Şekil	Sayfa
Şekil 6.35. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri.....	94
Şekil 6.36. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi.....	94
Şekil 6.37. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi	95
Şekil 6.38. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri.....	96
Şekil 6.39. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi	97
Şekil 6.40. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri.....	98
Şekil 6.41. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi.....	98
Şekil 6.42. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi	99
Şekil 6.43. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri.....	100
Şekil 6.44. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi	101
Şekil 6.45. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri.....	102
Şekil 6.46. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi.....	102
Şekil 6.47. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi	103
Şekil 6.48. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri.....	104
Şekil 6.49. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi	105
Şekil 6.50. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri.....	106

Şekil	Sayfa
Şekil 6.51. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	106
Şekil 6.52. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	107
Şekil 6.53. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait ROC eğrileri	108
Şekil 6.54. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	109
Şekil 6.55. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait ROC eğrileri	110
Şekil 6.56. Temiz Beyin Tümör MR görüntüleri üzerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	110
Şekil 6.57. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	111
Şekil 6.58. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri	112
Şekil 6.59. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	113
Şekil 6.60. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri	114
Şekil 6.61. Temiz Beyin Tümör MR görüntüleri üzerinde InceptionV3 modeline ait karmaşıklık matrisi.....	115
Şekil 6.62. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi.....	116
Şekil 6.63. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait ROC eğrileri.....	117
Şekil 6.64. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi.....	118
Şekil 6.65. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait ROC eğrileri.....	119
Şekil 6.66. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV1 modeline ait karmaşıklık matrisi.....	120

Şekil	Sayfa
Şekil 6.67. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi	121
Şekil 6.68. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait ROC eğrileri.....	122
Şekil 6.69. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi.....	123
Şekil 6.70. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait ROC eğrileri.....	124
Şekil 6.71. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV2 modeline ait karmaşıklık matrisi.....	124
Şekil 6.72. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi	125
Şekil 6.73. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait ROC eğrileri.....	126
Şekil 6.74. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi.....	127
Şekil 6.75. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait ROC eğrileri.....	128
Şekil 6.76. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV3Small modeline ait karmaşıklık matrisi.....	128
Şekil 6.77. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi	129
Şekil 6.78. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait ROC eğrileri	130
Şekil 6.79. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi	131
Şekil 6.80. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait ROC eğrileri	132
Şekil 6.81. Temiz Beyin Tümör MR görüntüleri üzerinde NASNetMobile modeline ait karmaşıklık matrisi.....	132
Şekil 6.82. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi	133

Şekil	Sayfa
Şekil 6.83. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait ROC eğrileri.....	134
Şekil 6.84. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi	135
Şekil 6.85. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait ROC eğrileri.....	136
Şekil 6.86. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet50V2 modeline ait karmaşıklık matrisi.....	136
Şekil 6.87. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi	137
Şekil 6.88. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait ROC eğrileri	138
Şekil 6.89. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi.....	139
Şekil 6.90. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait ROC eğrileri.....	140
Şekil 6.91. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet101V2 modeline ait karmaşıklık matrisi.....	140
Şekil 6.92. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi	141
Şekil 6.93. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde modeline ait ROC eğrileri.....	142
Şekil 6.94. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 ResNet101V2 modeline ait karmaşıklık matrisi	143
Şekil 6.95. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 ResNet101V2modeline ait ROC eğrileri	144
Şekil 6.96. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet152V2 modeline ait karmaşıklık matrisi.....	145
Şekil 6.97. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi	146
Şekil 6.98. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait ROC eğrileri	147

Şekil	Sayfa
Şekil 6.99. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi.....	148
Şekil 6.100. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait ROC eğrileri	149
Şekil 6.101. Temiz Böbrek BT görüntüleri üzerinde Xception modeline ait karmaşıklık matrisi.....	153
Şekil 6.102. FGSM saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait karmaşıklık matrisi.....	154
Şekil 6.103. FGSM saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait ROC eğrileri.....	155
Şekil 6.104. PGD saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait karmaşıklık matrisi.....	156
Şekil 6.105. PGD saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait ROC eğrileri.....	157
Şekil 6.106. Temiz Böbrek BT görüntüleri üzerinde VGG16 modeline ait karmaşıklık matrisi.....	157
Şekil 6.107. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait karmaşıklık matrisi.....	158
Şekil 6.108. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait ROC eğrileri.....	159
Şekil 6.109. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait karmaşıklık matrisi.....	160
Şekil 6.110. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait ROC eğrileri.....	161
Şekil 6.111. Temiz Böbrek BT görüntüleri üzerinde VGG19 modeline ait karmaşıklık matrisi.....	161
Şekil 6.112. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait karmaşıklık matrisi.....	162
Şekil 6.113. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait ROC eğrileri.....	163
Şekil 6.114. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait karmaşıklık matrisi.....	164

Şekil	Sayfa
Şekil 6.115. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait ROC eğrileri.....	165
Şekil 6.116. Temiz Böbrek BT görüntüleri üzerinde DenseNet121 modeline ait karmaşıklık matrisi.....	165
Şekil 6.117. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi.....	166
Şekil 6.118. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait ROC eğrileri.....	167
Şekil 6.119. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi.....	168
Şekil 6.120. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait ROC eğrileri.....	169
Şekil 6.121. Temiz Böbrek BT görüntüleri üzerinde DenseNet169 modeline ait karmaşıklık matrisi.....	169
Şekil 6.122. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi.....	170
Şekil 6.123. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait ROC eğrileri.....	171
Şekil 6.124. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi.....	172
Şekil 6.125. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait ROC eğrileri.....	173
Şekil 6.126. Temiz Böbrek BT görüntüleri üzerinde DenseNet201 modeline ait karmaşıklık matrisi.....	173
Şekil 6.127. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi.....	174
Şekil 6.128. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait ROC eğrileri.....	175
Şekil 6.129. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi.....	176
Şekil 6.130. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait ROC eğrileri.....	177

Şekil	Sayfa
Şekil 6.131. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi.....	177
Şekil 6.132. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi.....	178
Şekil 6.133. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri.....	179
Şekil 6.134. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi.....	180
Şekil 6.135. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri.....	181
Şekil 6.136. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi.....	181
Şekil 6.137. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi.....	182
Şekil 6.138. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri.....	183
Şekil 6.139. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi.....	184
Şekil 6.140. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri.....	185
Şekil 6.141. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi.....	185
Şekil 6.142. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi.....	186
Şekil 6.143. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri.....	187
Şekil 6.144. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi.....	188
Şekil 6.145. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri.....	189
Şekil 6.146. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi.....	189

Şekil	Sayfa
Şekil 6.147. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi.....	190
Şekil 6.148. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri.....	191
Şekil 6.149. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi.....	192
Şekil 6.150. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri.....	193
Şekil 6.151. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	193
Şekil 6.152. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	194
Şekil 6.153. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait ROC eğrileri.....	195
Şekil 6.154. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi.....	196
Şekil 6.155. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait ROC eğrileri.....	197
Şekil 6.156. Temiz Böbrek BT görüntüleri üzerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	197
Şekil 6.157. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	198
Şekil 6.158. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri	199
Şekil 6.159. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi.....	200
Şekil 6.160. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri	201
Şekil 6.161. Temiz Böbrek BT görüntüleri üzerinde InceptionV3 modeline ait karmaşıklık matrisi.....	201
Şekil 6.162. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi.....	202

Şekil	Sayfa
Şekil 6.163. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait ROC eğrileri.....	203
Şekil 6.164. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi.....	204
Şekil 6.165. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait ROC eğrileri.....	205
Şekil 6.166. Temiz Böbrek BT görüntüleri üzerinde MobileNetV1 modeline ait karmaşıklık matrisi.....	205
Şekil 6.167. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi.....	206
Şekil 6.168. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait ROC eğrileri.....	207
Şekil 6.169. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi.....	208
Şekil 6.170. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait ROC eğrileri.....	209
Şekil 6.171. Temiz Böbrek BT görüntüleri üzerinde MobileNetV2 modeline ait karmaşıklık matrisi.....	209
Şekil 6.172. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi.....	210
Şekil 6.173. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait ROC eğrileri.....	211
Şekil 6.174. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi.....	212
Şekil 6.175. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait ROC eğrileri.....	213
Şekil 6.176. Temiz Böbrek BT görüntüleri üzerinde MobileNetV3Small modeline ait karmaşıklık matrisi.....	213
Şekil 6.177. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi	214
Şekil 6.178. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait ROC eğrileri	215

Şekil	Sayfa
Şekil 6.179. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi.....	216
Şekil 6.180. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait ROC eğrileri.....	217
Şekil 6.181. Temiz Böbrek BT görüntüleri üzerinde NASNetMobile modeline ait karmaşıklık matrisi.....	217
Şekil 6.182. FGSM saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi.....	218
Şekil 6.183. FGSM saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait ROC eğrileri.....	219
Şekil 6.184. PGD saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi.....	220
Şekil 6.185. PGD saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait ROC eğrileri.....	221
Şekil 6.186. Temiz Böbrek BT görüntüleri üzerinde ResNet50V2 modeline ait karmaşıklık matrisi.....	221
Şekil 6.187. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi.....	222
Şekil 6.188. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait ROC eğrileri.....	223
Şekil 6.189. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi.....	224
Şekil 6.190. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait ROC eğrileri.....	225
Şekil 6.191. Temiz Böbrek BT görüntüleri üzerinde ResNet101V2 modeline ait karmaşıklık matrisi.....	225
Şekil 6.192. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi.....	226
Şekil 6.193. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait ROC eğrileri.....	227
Şekil 6.194. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi.....	228

Şekil	Sayfa
Şekil 6.195. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait ROC eğrileri.....	229
Şekil 6.196. Temiz Böbrek BT görüntüleri üzerinde ResNet152V2 modeline ait karmaşıklık matrisi.....	229
Şekil 6.197. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi.....	230
Şekil 6.198. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait ROC eğrileri.....	231
Şekil 6.199. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi.....	232
Şekil 6.200. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait ROC eğrileri.....	233
Şekil 6.201. Orijinal Beyin Tümör MR görüntü örneği ve farklı pertürbasyona sahip FGSM saldırısına bağlı olarak üretilen çekişmeli örnekler	234
Şekil 6.202. Orijinal Beyin Tümör MR görüntü örneği ve farklı pertürbasyona sahip PGD saldırısına bağlı olarak üretilen çekişmeli örnekler	234
Şekil 6.203. Orijinal Böbrek BT görüntü örneği ve farklı pertürbasyona sahip FGSM saldırısına bağlı olarak üretilen çekişmeli örnekler	236
Şekil 6.204. Orijinal Böbrek BT görüntü örneği ve farklı pertürbasyona sahip PGD saldırısına bağlı olarak üretilen çekişmeli örnekler	236
Şekil 6.205. FGSM saldırısı ile üretilen MR ve BT örneklerinin derin öğrenme modelleri bazında doğruluk performanslarının karşılaştırılması	243
Şekil 6.206. PGD saldırısı ile üretilen MR ve BT örneklerinin derin öğrenme modelleri bazında doğruluk performanslarının karşılaştırılması	244
Şekil 6.207. FGSM ve PGD saldırıları ile üretilen MR ve BT örneklerinin orijinal görüntülere kıyasla ortalama kalite ve benzerliklerinin karşılaştırılması ...	245

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

ϵ	pertürbasyon değeri
db	desibel
inf	infimum
sn	saniye

Kısaltmalar

Açıklamalar

AUC	Eğrinin Altındaki Alan
BIM	Temel İteratif Yöntem
BT	Bilgisayarlı Tomografi
CNN	Evrişimsel Sinir Ağı
C&W	Carlini & Wagner
DICOM	Tıpta Dijital Görüntüleme ve İletişim
FGSM	Hızlı Gradyan İşareti Yöntemi
Fused-MBConv	Birleştirilmiş Mobil Ters Darboğaz Evrişim Bloğu
MBConv	Mobil Ters Darboğaz Evrişim Bloğu
MR	Manyetik Rezonans
MSE	Ortalama Karesel Hata
MS-SSIM	Çok Ölçekli Yapısal Benzerlik İndeksi
PACS	Görüntü Saklama ve İletişim Sistemleri
PET	Pozitron Emisyon Tomografisi
PGD	İzdüşürülen Gradyan İnişi
RMSE	Kök Ortalama Karesel Hata
ROC	Alıcı İşlem Karakteristik
SPECT	Tek Foton Emisyon Tomografisi
SSIM	Yapısal Benzerlik İndeksi

Kısaltmalar**Açıklamalar****TL**

Transfer Öğrenme

PSNR

Tepe Sinyal-Gürültü Oranı

UAP

Evrensel Çekişmeli Pertürbasyon

UQI

Evrensel Kalite İndeksi

VIF

Görsel Bilgi Doğruluğu

VIFP

Piksel Tabanlı Görsel Bilgi Doğruluğu

1. GİRİŞ

Medikal görüntüler organlar, hücreler ve patolojik oluşumlar gibi çeşitli anatomik bilgiler sağlayan önemli veri kaynaklarıdır. Bu görüntüler, genel itibariyle, sağlık uzmanlarının tanılama ve tespit çalışmalarında hızlı ve verimli karar vermesini sağlamak amacıyla kullanılmaktadır. Günümüzde retinopati ve katarakt gibi göz kusurlarının tespiti, cilt lezyonlarının takibi, kanser ve tümör oluşumlarının teşhisi ve segmentasyonu, karaciğer ölçümü, akciğer hastalıklarının sınıflandırılması gibi çeşitli görevlerde medikal görüntülerden yararlanılmaktadır. Medikal görüntülerin analizi ile daha güvenilir ve isabetli kararlar alınması beklenmektedir. Ancak çeşitli prensiplerle elde edilen ve birbirinden farklı yapısal özellikler barındıran medikal görüntülerin sağlık uzmanları tarafından analiz edilmesi zaman ve maliyet açısından elverişli olmamaktadır. Bu durum medikal görüntü analizinde bilgisayar destekli tespit ve tanı sistemlerinden yararlanmasını kaçınılmaz hale getirmektedir [1-3].

Teknolojinin gelişimiyle birlikte, karar verme sürecini yüksek doğruluk ile otomatize eden yapay zeka tabanlı karar destek sistemlerinin kullanımı yaygınlaşmaktadır. Yapay zekanın alt dallarından biri olan derin öğrenme, kendini pek çok alanda kanıtlamakla birlikte sağlık uzmanlarının ve araştırmacıların dikkatini çekmeyi başarmıştır. Medikal görüntü analizinde derin öğrenme kullanımı, ABD Gıda ve İlaç İdaresi'nin (FDA) çalışmaları ile resmi olarak kabul görmüştür. 2018 yılında uzman yardımı olmaksızın gerçek zamanlı orta düzeyli diyabetik retinopati tespiti yapan IDX-DR yapay zeka algoritması onaylanmıştır. Ardından, 2020 yılında gerçek zamanlı diyabetik retinopati tespiti yapan yapay zeka destekli EyeArt platformu onaylanmıştır. Yapay zeka tabanlı sistemlerin belirli kısıtlamaları olduğu bilinmesine karşın çeşitli kurum ve kuruluşlar medikal görüntülerin analizi için yapay zeka tabanlı cihaz ve platformlar geliştirmek üzere çalışmalarına devam etmektedir. Literatürde, medikal görüntülerin sınıflandırılması ve segmentasyonu için özel olarak geliştirilen çeşitli derin öğrenme modelleri yer almaktadır.

Derin öğrenme modelleri, medikal görüntülerin analizinde yüksek başarı elde ederek hastalıkların teşhis ve tespit süreçlerine katkı sağlamasına rağmen güvenlik konusunda bazı soru işaretlerine neden olmaktadır. Son yıllarda medikal alanda yapılan bazı araştırmalar, medikal görüntü analizinde kullanılan derin öğrenme modellerinin çekişmeli

saldırı (adversarial attack) adı verilen manipölasyonlara karşı hassas ve savunmasız olduğunu göstermektedir. Bu saldırılar, yalnızca derin öğrenme algoritmasını etkileyebilecek düzeyde ufak bir manipölasyon gerçekleştirerek modeli yanlış karar vermeye zorlamaktadır. Çekişmeli saldırıların medikal görüntü analizinde başarılı olmasının nedenleri arasında medikal görüntü setlerinin sınırlı sayıda veri içermesi, derin öğrenme modellerinin yapısal özellikleri, medikal görüntüleri diğer görüntülerden ayırtıran karakteristik dokuların ve parametrelerin varlığı gibi çeşitli durumlar sıralanmaktadır. Bu durumlar dikkate alınarak çeşitli savunma yöntemleri üzerine çalışmakta ve çekişmeli saldırıların etkinliğinin azaltılması hedeflenmektedir. Ancak, sağlık alanında verilebilecek yanlış bir karar kaynak sarfiyatına ve ek maliyetlere neden olmasının yanı sıra bir insanın hayatına mal olabilecek potansiyel bir risk anlamı taşımaktadır. Bu nedenle çekişmeli saldırıların davranışlarını öğrenmek ve derin öğrenme modelleri üzerindeki etkisini anlamak kritik bir önem arz etmektedir. Tez çalışmasında, medikal görüntülerin analizinde kullanılan derin öğrenme yöntemlerinin gradyan tabanlı çekişmeli saldırılara karşı sağlamlığı kıyaslanmakta, medikal görüntü modaliteleri, saldırı parametreleri ve iterasyon/adım sayısının saldırı başarısına etkisi araştırılmaktadır.

Bu çalışmanın katkıları şu şekildedir;

- Medikal görüntülerin manipölasyonuna yönelik gerçekleştirilen çekişmeli saldırılar hakkında literatürdeki çalışmalar doğrultusunda kapsamlı bir değerlendirme sunulmaktadır.
- MR ve BT görüntülerinden oluşan iki açık kaynaklı veri seti üzerinde 20 ön eğitilmiş derin öğrenme modeli eğitilmekte ve sınıflandırma başarısı karşılaştırılmaktadır.
- Belirlenen 20 ön eğitilmiş derin öğrenme modelinin FGSM ve PGD saldırı senaryoları karşısındaki sağlamlıkları ve sınıflandırma kabiliyetleri karşılaştırılmaktadır.
- Tek adımlı FGSM ve çok adımlı PGD saldırılarının medikal görüntü verileri üzerindeki etkinliği karşılaştırılmaktadır. Saldırıları üç farklı pertürbasyon değeri ile gerçekleştirilerek saldırı parametrelerinin başarı üzerindeki etkisi incelenmektedir.
- MR ve BT görüntü modalitelerinin saldırılara karşı direnci kıyaslanmakta ve medikal görüntü modalitelerinin saldırı başarısı üzerindeki etkisi incelenmektedir.

- FGSM ve PGD saldırıları ile üretilen çekişmeli örneklerin kalitesi ve orijinal görüntüye benzerlik değerleri 7 farklı metrik ile ölçülmektedir. Görüntüler üzerindeki ortalama bozulma oranının saldırı başarısı ile ilişkisi değerlendirilmektedir.

Tez çalışması, kavramsal çerçeve, literatür taraması, kullanılan araçlar ve veri setleri, sınıflandırma modellerinin eğitimi, deneysel sonuçlar ve bulguların değerlendirilmesi, sonuç ve öneriler bölümlerinden oluşmaktadır. Kavramsal çerçeve bölümünde, çalışmanın anlaşılması için gerekli temel bilgiler açıklanmaktadır. Literatür taraması bölümünde, konu kapsamında geçmişte gerçekleştirilen araştırmalara yer verilmektedir. Kullanılan araçlar ve veri setleri bölümünde, kullanılan yazılım kütüphaneleri, çalışma ortamı ve çalışmada kullanılan veri setlerine ilişkin bilgiler sunulmaktadır. Sınıflandırma modellerinin eğitimi bölümünde, deney ortamı, deney tasarımı, kullanılan hiper parametre ayarları ve kullanılan ince ayarlı transfer öğrenme yöntemi açıklanmaktadır. Ayrıca, modellerin eğitim sürecinde elde edilen sayısal ölçümler sunulmakta ve test aşamasında kullanılan çekişmeli örneklerin üretim detaylarına yer verilmektedir. Deneysel sonuçlar ve bulgular bölümünde, her bir veri setine yönelik gerçekleştirilen saldırılar için tek tek modellerin performansları değerlendirilmektedir. Diğer yandan, üretilen çekişmeli görüntülerin kaliteleri ve benzerlikleri incelenmektedir. Son bölümde, tez çalışması sonucunda varılan nihai sonuçlar açıklanmaktadır.

2. KAVRAMSAL ÇERÇEVE

Bu bölümde, tez çalışması kapsamında ele alınan konunun temel çerçevesi sunulmaktadır. Medikal görüntü analizinde görüntüleme ve güvenlik riskleri, çekişmeli saldırılar, derin öğrenme ve evrimsel sinir ağ mimarileri, modellerin değerlendirmesinde kullanılan performans metrikleri ve görüntü benzerlik metrikleri açıklanmaktadır.

2.1. Medikal Görüntü Analizi

Medikal görüntü analizinde esas alınan nokta görüntü modaliteleridir. Medikal görüntü modalitelerinin analizinde güvenlik unsuru ise bu çalışmanın temelini oluşturmaktadır. Bu bölümde, başlıca medikal görüntüleme yöntemleri ve medikal görüntü analizinde karşılaşılan güvenlik riskleri açıklanmaktadır.

2.1.1. Medikal görüntüleme

Medikal görüntüler farklı teknik ve süreçlere sahip çeşitli medikal görüntüleme yöntemleri ile elde edilmektedir. Bir hastalığın teşhisinde kullanılabilecek en uygun görüntüleme yöntemi uzman kararı ile belirlenmektedir. Elde edilen tomografi, röntgen, ultrason gibi medikal görüntü türleri modalite olarak adlandırılmaktadır. Her bir modalite kendi doğasına uygun özelliklere, süreçlere ve çıktılara sahiptir. Medikal görüntü modalitelerinin büyük bir kısmı radyasyon ve ses dalgalarını kullanarak vücudun ilgili bölümlerinin incelenmesini sağlayan invazif olmayan yaklaşıma dayanmaktadır. Endoskopi gibi bazı modaliteler ise iğne ya da kesi yoluyla vücuda enjekte edilen bir cisim aracılığıyla ilgili dokuların incelenmesini sağlayan invazif yaklaşımları esas almaktadır [4]. Medikal görüntü denildiğinde akla ilk gelen radyoloji görüntüleridir. Radyoloji genel itibarıyla röntgen, bilgisayarlı tomografi (BT), manyetik rezonans (MR), ultrason ve mamografi görüntülerini kapsayan hastalıkların tanı ve tedavisinde yönlendirme amacıyla yaygın olarak kullanılan bir görüntüleme dalıdır. Radyolojik görüntüler genellikle invazif olmayan yöntemler ile elde edilmektedir ve elektromanyetik spektrumda frekans ve dalga boylarına dağılmaktadır [5]. Radyolojideki görüntüleme modaliteleri şunlardır:

- X-Ray: Wilhelm Röntgen tarafından bulunan X-Ray en eski medikal görüntüleme yöntemidir. Genellikle kemik dokularının kullanılan bir yöntemidir. Röntgende sert dokular beyaz görünürken, dokunun yoğunluğu azaldıkça renk siyaha yaklaşmaktadır. Örneğin, kemik dokusu beyaz, nispeten yumuşak olan yağ dokusu gri ve akciğerlerdeki hava ise siyah görüntülenmektedir.
- Bilgisayarlı Tomografi: X-Işınlarının farklı açı ve mesafelerden gönderilmesi suretiyle vücudun 3 boyutlu olarak taranmasını sağlayan medikal görüntü modalitesidir. Bu durum çözünürlüğü ve ayrıntı düzeyi yüksek görüntüler elde edilmesine olanak tanımaktadır. Bu sayede erken teşhis için uygun bir görüntüleme adayıdır. Bilgisayarlı tomografi taramalarında dokulardaki x-ışını yoğunluğunu ölçmek için Hounsfield skalası kullanılmaktadır. Hounsfield skalası bilgisayarlı tomografi için geliştirilen özel bir ölçü birimidir. Tomografi görüntülerini standardize etmekte ve böylece görüntünün nicel olarak değerlendirilmesine olanak tanımaktadır. Hounsfield skalasında hava negatif değer aralığındayken, dokular pozitif değer aralığındadır. Dokuların sertliği arttıkça değer artmaktadır.
- Manyetik Rezonans: Güçlü bir manyetik alan içerisinde manyetik gradyanlar ve radyofrekans dalgaları kullanılarak elde edilen bir medikal görüntü modalitesidir. Yüksek anatomik çözünürlük sağlayarak dokuların ayrıntılı görüntülerini sunmaktadır. Manyetik rezonans görüntüleme, vücutta en fazla bulunan atom olan Hidrojen (H) temel alınarak inceleme gerçekleştirilmektedir. MR görüntüleri teknik ve çevresel kaynaklı dış faktörlerden kolayca etkilenebilmektedir. Bu durum modalitenin standartlaştırılmasını zorlaştırmaktadır. Tomografi görüntülerine kıyasla yumuşak dokular hakkında daha ayrıntılı bilgi sağlamaktadır. Ancak görüntüleme için daha uzun bir zaman dilimine ihtiyaç duyulmaktadır.
- Ultrason: İnsan kulağıyla algılanamayacak kadar yüksek frekanslı ses dalgalarının yansımalarını kullanarak anatomik dokuları görüntüleyen bir modalitedir. Ses dalgalarının insan vücuduna bir yan etkisinin olmadığı bulgusunun elde edilmesi, gerçek zamanlı ve düşük maliyetli görüntüleme imkanı sunması nedeniyle özellikle ilk adım tanı aşamasında tercih edilmektedir. Ancak derin doku özelliklerinin tespit edilmesi ve görüntülenmesinde yetersiz kalabilmektedir.
- Mamografi: Meme hastalıklarının tarama ve teşhis sürecinde yaygın kullanılan bir görüntüleme modalitesidir. Düşük dozlu x-ışını göğüsten geçirilerek bir dedektöre gönderilir. Geleneksel mamografide film dedektörü kullanılmaktadır. Dijital

mamografide ise katı hal dedektörleri ile alınan ışın elektrik sinyallerine dönüştürülmektedir. Dijital mamografi minimum görüntü kaybı ile anomali tespitine imkan vermesi, görüntü üzerinde kontrast ayarlaması yapılabilmesi ve kolay dağıtılabilmesi ile ön plana çıkmaktadır. Mamografide dokunun sertliği arttıkça renk beyaza yaklaşmaktadır.

Bahsedilen radyolojik görüntüler dışında dermoskopi, endoskopi, oftalmoloji gibi çeşitli medikal görüntü modaliteleri bulunmaktadır. Diğer yandan, nükleer tıp uygulamalarında radyofarmasötik maddelerin vücuda enjeksiyonuyla yayılan radyasyonun kaydedilerek biyokimyasal değişikliklerin görüntülendiği PET ve SPECT modaliteleri bulunmaktadır. Bilinen başka bir modalite ise insan gözüyle tespit edilemeyen mikroorganizmaların ve cansız maddelerin yüksek uzamsal çözünürlükte görüntülenmesi için kullanılan mikroskopidir. Mikroskopisi ışık renk skalasının çeşitli fazlarından yararlanarak moleküler bazda doku özelliklerinin incelenmesine olanak tanımaktadır. Her bir modalite kendine has noktalara odaklanmakta ve bu doğrultuda sınırlı bilgi sunmaktadır. Görüntü füzyonu ile farklı modalitelere ait görüntülerin geometrik hizalanması sonucu hibrit görüntüler elde edilmektedir. Bu görüntülerin her iki görüntünün sahip olduğu bilgileri eksiksiz barındırması ve bu sınırları koruması gerekmektedir. Radyoloji ve nükleer tıp modalitelerinin füzyonu yaygın kullanılmaktadır. Radyoloji görüntüleri yüksek uzaysal çözünürlüklü anatomik bilgiler açısından zengindir. Bunun aksine PET ve SPECT düşük uzamsal çözünürlüklü metabolik ve fizyolojik bilgiler sağlamaktadır. Bu modalitelerin birleşimi ile oluşan füzyon görüntüler hastalıkların teşhis ve tedavi süreçlerinde daha ayrıntılı bilgi sağlamakta ve uzmanların karar sürecini kolaylaştırmaktadır. Yaygın kullanılan hibrit görüntü modalitelerinden bazıları SPECT-MR, PET-MR, PET-BT, SPECT-BT ve BT-MR' dir [4].

Medikal görüntüler farklı donanım ve prosedürler ile elde edilmektedir. Ancak, önemli ortak özelliklerinden biri dijital ortamlarda saklanması için yüksek kayıt kapasitesi gerektirmeleridir. Günümüzde bu görüntülerin sayısal olarak depolanması, erişilmesi ve iletilmesi için yüksek kayıt kapasitesi sunan PACS (Picture Archiving and Communication Systems) kullanılmaktadır [6, 7]. PACS medikal film arşivlerine olan ihtiyacı ortadan kaldırmakta ve tüm uzmanların sistem üzerinden kolaylıkla verilere ulaşmasına olanak tanımaktadır. PACS ile arşivleme sorunu ortadan kalkmıştır. Ancak, farklı donanımlar ile elde edilen görüntülerin uyumlu çalışabilmesi için standart dosya biçiminde tutulması

gerekmektedir. Amerikan Radyoloji Koleji (ACR, American College of Radiology) ve Ulusal Elektrik Üreticileri Derneği (NEMA, National Electrical Manufacturers Association) DICOM (Digital Imaging and Communications in Medicine) standardını geliştirerek ortak görüntü formatı probleminin üstesinden gelmiştir. Medikal görüntülerin standartlaştırılması ile farklı sağlık merkezlerinde farklı cihazlar ile elde edilen görüntülerin ortak kullanımı sağlanmaktadır. DICOM iki kısımdan oluşmaktadır. İlk kısımda hasta, sağlık merkezi, kullanılan cihaz, tanı, pozlama ve kesit bilgileri gibi ek metinsel bilgiler yer almaktadır. İkinci kısımda ise ham medikal görüntü bulunmaktadır. DICOM bu özelliği ile bir veri tabanını andırmaktadır [8]. DICOM verileri belirli sıkıştırma işlemleri uygulanarak yalnızca görüntü içeren JPEG, PNG gibi formatlara dönüştürülebilmektedir. Bu tez çalışmasında JPG formatına dönüştürülmüş açık kaynaklı MR ve BT medikal görüntüleri kullanılmaktadır.

2.1.2. Medikal görüntü analizinde güvenlik

Medikal görüntülerin analizinde yapay zeka tabanlı bilgisayarlı destek sistemlerinin yaygın kullanımı birtakım güvenlik endişelerini de beraberinde getirmektedir. Medikal görüntü analiz modellerinin medikal görüntüleme cihazlarından kaynaklı istenmeyen gürültülere karşı toleransı düşündürücüdür. Diğer yandan, saldırganlar tarafından yapay zeka modellerini yanlış karar vermeye sevk eden manipülasyon saldırılarının varlığı da bir başka tartışma konusudur. Görüntülerin manipülasyonu için gerçekleştirilen çekişmeli saldırılar üzerine yapılan araştırmalar daha uzun bir geçmişe dayanmasına rağmen medikal görüntüler üzerindeki etkinliklerinin değerlendirilmesine yönelik çalışmalar 2018 yılında başlamıştır. Bu çalışmalar, medikal görüntülerin analizinde yapay zeka tabanlı sistemlerin kullanımının yaygınlaşmasıyla ivme kazanmıştır.

Medikal görüntü analizinde çekişmeli saldırıların etkinliğinin ölçülmesi ve nedenlerinin anlaşılması gerekli önleyici yaklaşımların geliştirilmesi için kritik bir konudur. Çekişmeli saldırıların başarısının ardında yatan nedenler üzerine yapılan çalışmalar medikal görüntülerin karakteristikleri ve derin öğrenme modellerinin özellikleri üzerine yoğunlaşmaktadır. Medikal görüntülerde etiketli verilerin yetersizliği ve karmaşık biyolojik dokuların varlığı çekişmeli saldırıların etkinliğini artırmaktadır. Derin öğrenme modellerinin doğrusallığı, yetersiz genelleme kabiliyeti, modellerin sağlam olmayan

özellikleri ve aşırı parametrelendirilmesi gibi unsurlar ise çekişmeli saldırılara zemin hazırlamaktadır [9-12].

Medikal veriler dünya genelinden çeşitli ulusal ve uluslararası düzenlemelerle güvence altına alınan hassas verilerdir. Bir kişinin rızası dışında sağlık verilerine erişilmesi, dağıtılması ve kullanılması kısıtlanmaktadır. Aksi durum, kişisel verilerin gizliliğinin ihlali anlamına gelmektedir. Öte yandan, eldeki medikal görüntülerin doğru şekilde etiketlenmesi zaman ve uğraş gerektiren bir faaliyettir. Bu işlemin gerçekleştirilmesi için sağlık uzmanlarına ihtiyaç vardır. Ancak uzmanlar arasında değişen teşhis farklılıkları, yanlış tanılama, kaba taneli etiketler, sınıf dengesizliği ve etiket sızıntısı gibi süreci olumsuz etkileyecek sorunlar ortaya çıkabilmektedir. Doğru etiketlenmiş erişime açık medikal verilerin sınırlı olması analiz çalışmalarına zorluk ve sınırlama getirmektedir. Derin öğrenme modellerinin verimli çalışabilmesi için kullanılan veri setlerinin büyük olması beklenir. Etiketli medikal görüntü setlerinin küçük boyutlu olması derin öğrenme modellerinin performansını olumsuz etkilemektedir. Yetersiz veri derin öğrenme modellerinin aşırı öğrenmesi ve yetersiz genelleme kabiliyeti ile sonuçlanmaktadır.

Medikal görüntüler sahip oldukları birtakım yapısal özellikler ile diğer görüntülerden ayrılmaktadır. Medikal görüntüler biyolojik dokular ve nicel bilgiler içeren yüksek kaliteli görüntülerdir. Bu görüntülerde organların konumları, aydınlatma ve parlaklık parametreleri gibi karakteristik unsurlar önem kazanmaktadır [10, 13]. Medikal görüntülerin genel anatomik bir çerçeve sunması derin öğrenme modellerinin görüntüdeki önemsiz bölgelere yoğunlaşmasına neden olabilmektedir [10]. Farklı prosedürlere sahip cihazlar ile elde edilen medikal görüntü modalitelerinin saldırılar karşısında farklı duyarlılıkta olması da bir başka araştırma konusudur. Çalışma [14]' de histopatoloji görüntülerinin saldırılara karşı röntgen görüntülerinden daha dirençli olduğu ortaya konulmaktadır. Çalışma [15] ise röntgen görüntülerinin bilgisayarlı tomografi görüntülerinden daha dirençli olduğunu göstermektedir. Ancak veri seti boyutları, kullanılan derin öğrenme yaklaşımları ve saldırı türleri gibi dinamikler modalitelerin dirençleri hakkında net bir çıkarım yapmak için yeterli koşulları sağlamamaktadır.

Medikal görüntü analizinde kullanılan derin öğrenme yöntemleri doğal görüntüler kullanılarak geliştirilmiştir. Benzer şekilde, küçük boyutlu veri setlerinin neden olduğu dezavantajı gidermek için kullanılan transfer öğrenme yaklaşımında kullanılan ImageNet

gibi büyük ölçekli görüntü setleri doğal görüntülerden oluşmaktadır. Bu durum ön eğitilmiş derin öğrenme yaklaşımlarında kullanılan ağırlıkların doğal görüntüleri esas aldığı ve medikal görüntülerin özelliklerini yansıtmakta yetersiz kalabileceği anlamına gelmektedir. Modellerin küçülmesi saldırılara karşı direnci artıran bir faktördür. Basit ön eğitilmiş modeller sınıflandırma görevlerinde önemli özellikleri daha iyi kavramaktadır [16]. Bu durum büyük ve karmaşık modellerin saldırılara karşı daha savunmasız olduğu anlamına gelmektedir. Karmaşık modellerin sağlamlığını artırmak için başvurulmuş en yaygın yöntemlerin başında çekişmeli eğitim gelmektedir.

2.2. Çekişmeli Saldırıları

Çekişmeli saldırılar sınıflandırma, semantik segmentasyon ve nesne tespiti gibi görevlerde yapay zeka tabanlı modelleri yanıltarak modelin yanlış karar vermesini hedefleyen saldırılardır. Çekişmeli saldırılar, en basit haliyle, girdilere insan gözüyle algılanamayacak düzeyde küçük pertürbasyonlar ekleyerek manipüle edilmiş çıktılar üretmektedir. Gözle görülemeyen ancak makinelerin kararlarını değiştirmek için yeterli olan bu pertürbasyonlar gürültü olarak da adlandırılmaktadır. Eklenen pertürbasyonlar girdi örneklerinin karar sınırı boyunca hareket etmesine neden olmaktadır. Oluşan manipüle edilmiş yeni görüntülere ise çekişmeli örnekler denilmektedir. Çekişmeli örneklerin oluşumunun altında yatan faktörlere dair çeşitli varsayımlar üzerinde durulmaktadır. Szegedy ve arkadaşları [17] çekişmeli örneklerin oluşumunu derin öğrenme modellerinin düşük olasılıklı veri uzayındaki genelleme kabiliyetinin yetersizliği ile ilişkilendirmektedir. Goodfellow ve arkadaşları [12] modellerin yüksek boyutlu veri uzayında doğrusal davranış sergilediğini ve bunun sonucunda çekişmeli örneklerin meydana geldiğini ortaya koymaktadır. Bu görüşe karşın, Papernot ve arkadaşları [18] ise çekişmeli örneklerin oluşumunun derin ağ modellerinin aşırı doğrusal olmama durumundan kaynaklandığını ileri sürmektedir. Çekişmeli örnekler genellikle bir saldırgan tarafından bilinçli olarak gerçekleştirilmektedir. Ancak medikal görüntüleme ekipmanının doğasına, görüntü işleme tekniklerine veya verinin elde edildiği dış ortama bağlı olarak istenmeyen gürültüler meydana gelebilmektedir. Örneğin, bilgisayarlı tomografi gibi harekete duyarlı görüntüleme modalitelerinde, hastanın ya da cihazın işlem sırasındaki hareketlerine bağlı olarak istenmeyen gürültüler meydana gelebilmektedir. Hasta ve cihaz hareketleri, metalik cisimlerin varlığı gibi unsurlar sonucu meydana gelen gürültülü görüntülere tıp dilinde artefakt denilmektedir. Artefaktlar doğal koşullar altında meydana gelen çekişmeli

örneklerdir. Bu çalışmada bilinçli olarak gerçekleştirilen çekişmeli saldırılar ve çekişmeli örnek üretim yöntemleri üzerinde durulmaktadır.

Çekişmeli saldırılar modelin eğitim, test ve dağıtım aşamalarında gerçekleştirilebilmektedir. Eğitim aşamasında gerçekleştirilen çekişmeli saldırılar eğitim verilerinin değiştirilmesi, etiketlerin manipülasyonu, ilave çekişmeli veri enjeksiyonu, hedef modelin mantıksal manipülasyonu ile hedef modelin kararına müdahale etmektedir. Test aşamasında gerçekleştirilen saldırılar ise hedef modele doğrudan bir müdahalede bulunmak yerine ağ mimarisi ve parametreleri hakkında ilgi edinerek modelin sağlamlığını manipüle etmektedir. Çekişmeli saldırılar modelin hangi aşamasında gerçekleşirse gerçekleşsin benzer amaçlara sahiptir. Bu amaçlar doğru tahmin başarısını düşürerek modele güvenin azaltılması, girdinin herhangi bir yanlış tahmin ile sonuçlanması, tüm girdilerin spesifik olarak belirlenmiş yanlış bir tahmin ile sonuçlanması ve spesifik bir girdi sınıfının belirlenen spesifik bir yanlış tahmin sınıfı ile sonuçlanması olarak dört kategoride değerlendirilebilir [19].

Çekişmeli saldırılar benzer amaçlara sahip olmasına karşın çeşitli özelliklere sahiptir. Bu durum saldırıların farklı yaklaşımlara göre sınıflandırılmasına olanak tanımaktadır. Bu çalışmada çekişmeli saldırılar dört farklı çerçevede incelenmektedir: Saldırganın bilgisi, saldırının odağı, saldırının frekansı ve saldırıda kullanılan pertürbasyon.

Çekişmeli saldırılar saldırırganın edindiği bilgiye göre beyaz kutu saldırısı ve kara kutu saldırısı olarak iki sınıfa ayrılır:

- Beyaz kutu saldırısı: Saldırganın eğitim verileri, model mimarisi, gradyan bilgileri, ağırlık değerleri ve kullanılan hiper parametreler gibi modele ait tüm bilgilere sahip olduğu durumdur.
- Kara kutu saldırısı: Saldırganın modele erişemediği varsayılmaktadır. Bu nedenle, saldırırganın hedef modelin temel yapısı ve parametreleri hakkında herhangi bir bilgisi yoktur. Yalnızca orijinal verilere ve hedef modelin çıktı sonuçlarına erişimi bulunmaktadır.

Çekişmeli saldırıların odağına göre hedefli saldırı ve hedefsiz saldırı olarak iki sınıfa ayrılır:

- Hedefli saldırı: Hedef modelin, girdileri belirli bir sınıfa yönlendirerek yanlış sınıflandırma yapmasına yol açan saldırılardır. Bu tür saldırılar çok sınıflı sınıflandırma problemlerinde yaygın olarak kullanılmaktadır.
- Hedefsiz saldırı: Herhangi bir yanlış sınıf belirlenerek sınıflandırma sonucunun sınırlandırılmadığı saldırılardır. Girdilerin yanlış herhangi bir sınıfta değerlendirilmesi yeterlidir. Hedefli saldırılara kıyasla daha kolay uygulanmaktadır.

Çekişmeli saldırıların frekansına göre tek seferlik saldırı ve çok adımlı/iteratif saldırı olarak iki sınıfa ayrılır:

- Tek seferlik saldırı: Saldırı algoritması üzerinde yalnızca bir kez hesaplama gerçekleştirilerek çekişmeli örnek üretilen saldırılardır.
- İteratif saldırı: Saldırı algoritması üzerinde birçok kez hesaplama gerçekleştirilerek çekişmeli örnek üretilen saldırılardır. İteratif saldırılar, tek seferlik saldırılardan daha uzun zamana ve sınıflandırıcı etkileşimine ihtiyaç duymaktadır. Buna karşılık tek seferlik saldırılara kıyasla daha yüksek etkinlik göstermektedir.

Çekişmeli saldırıda kullanılan pertürbasyona göre evrensel pertürbasyon saldırısı ve spesifik pertürbasyon saldırısı olarak iki sınıfa ayrılır:

- Spesifik pertürbasyon saldırısı: Belirli bir hedef modeli yanıltmaya yönelik belirlenen bir pertürbasyon ile çekişmeli örneklerin üretildiği saldırılardır. Bir hedefe odaklandığı için genellikle ilgili modelin yapı ve parametrelerinin bilindiği beyaz kutu saldırılarıdır.
- Evrensel pertürbasyon saldırısı: Yalnızca bir modeli hedeflemek yerine birden çok modeli ya da model grubunu yanıltmak amacıyla çekişmeli örneklerin üretildiği saldırılardır. Saldırının aktarılabilirliği esasına dayanmaktadır. Farklı modellerin benzer nitelikleri ve zayıflıkları dikkate alınarak modelleri yanıltabilecek evrensel çekişmeli örnekler üretilmektedir [20].

2.2.1. Çekişmeli örnek üretimi

Çekişmeli saldırıların çekişmeli örnekler üretmek için kullandığı yaklaşımlar gradyan tabanlı, skor tabanlı ve karar tabanlı olmak üzere üçe ayrılmaktadır. Gradyan tabanlı

yöntemlerde çekişmeli örnekler üretmek için modelin gradyan bilgisinden yararlanılmaktadır. Kayıp fonksiyonunun gradyanı hesaplanarak modelin yanlış karar vermesi için gereken değişikliklerin hangi yönde ve hangi büyüklükte olacağı belirlenmektedir. Gradyan yönünde yapılan küçük değişiklikler ile modeli yanıltmak için gerekli olan en küçük pertürbasyon değerinin (ϵ) bulunması amaçlanır. Yapılan değişiklikler belirlenen giriş kısıtlarına uygun olmalıdır. Giriş kısıtları üretilen çelişkili örneklerin daha gerçekçi ve orijinalinden ayırt edilemeyecek düzeyde olması için önemlidir. Piksel, kontrast ve renk değişimi gibi girdi kısıtlarının yanı sıra girdide gerçekleştirilen değişikliklerin boyutunu ve yoğunluğunu sınırlamak için L-normları kullanılmaktadır. Skor tabanlı yöntemler, modelin çıktı değerlerini baz alarak çelişkili örnekler üretmektedir. Modelin çıktı değerleri skor olarak adlandırılmaktadır. Skor değerleri yanlış bir sınıfa yakınsanacak şekilde optimize edilmekte ve girdiler üzerinde skor değerinin manipülasyonunu destekleyecek değişiklikler yapılmaktadır. Skor tabanlı yöntemler modelin yapısı ve parametre bilgilerine ihtiyaç duymamaktadır. Hedefli saldırı senaryoları gerçekleştirmek için yaygın kullanılan yöntemlerdir. Karar tabanlı yöntemler, yalnızca model tarafından nihai çıktı olarak elde edilen sınıf etiketlerine erişebilmektedir. Hedef modelin softmax sınıflandırıcısı sorgulanarak yinelemeli bir reddetme örnekleme süreci takip edilmektedir. Bu süreçte amaç, hedef modelin karar mekanizmasını anlamak ve modeli yanıltmak için gerekli olan en küçük pertürbasyon değerini bulmaktır. Modelin yanlış karar vermesini sağlayacak en küçük değişiklik esas alınarak girdi verilerinden çekişmeli örnekler üretilmektedir. Gradyan tabanlı yöntemlerden daha az parametre değişikliği gerektirmektedir. Bu özelliği ile esnek ve basit bir çerçeve çizmektedir [21]. Her çekişmeli örnek üretim yönteminin kendine özgü avantajları ve dezavantajları bulunmaktadır. Çekişmeli örnek üretim yöntemleri farklı saldırı senaryoları altında ön plana çıkmaktadır. Saldırı senaryosuna göre özelleştirilebilmekte ya da hibrit olarak kullanılabilir. Bu çalışmada, gradyan tabanlı çekişmeli örnek üretme yöntemleri kullanılmaktadır. Gradyan tabanlı yöntemler uygulanması nispeten kolay, hızlı ve etkili olması nedeniyle en yaygın kullanılan beyaz kutu yaklaşımlarıdır. Çekişmeli saldırılara karşı geliştirilen birçok savunma yöntemine karşı kırılabilir olmasına karşı çekişmeli saldırı araştırmaları için önemli bir konumdadır.

2.2.2. Hızlı gradyan işareti yöntemi (FGSM)

Goodfellow ve arkadaşları [12] tarafından sunulan tek adımlı bir beyaz kutu saldırısıdır. Bu yöntem ile derin yapılı sinir ağlarındaki yüksek doğrusallığın çekişmeli örneklerle neden olduğunu gösterilmektedir. Tek adımlı bir işlem gerçekleştirdiği için hızlı ve hesaplama maliyeti düşük bir yöntemdir. FGSM’ de kayıp fonksiyonunun gradyanı hesaplanır ve gradyan işareti alınır. Her pikselde gradyanın yönü boyunca yalnızca bir hesaplama gerçekleştirilmektedir. Gradyan işareti epsilon ile çarpılarak gradyanın büyüklüğü belirlenir. Elde edilen değer orijinal girdiye eklenerek çekişmeli örnek üretilir. Hızlı gradyan işaret yöntemi gradyan değişimlerini genellikle L_∞ normu ile sınırlandırmaktadır. L_∞ normu gradyan değişimini girdinin piksellerindeki en büyük mutlak değer ile sınırlamaktadır. FGSM saldırısının formülasyonu:

$$x^* = x + \varepsilon \cdot \text{sign}(\nabla_x J(\theta, x, y)) \quad (2.1)$$

x girdi, x^* üretilen çekişmeli örnek, ε gradyanın büyüklüğünü belirleyen parametre, sign gradyanın yönünü gösteren gradyan işaretidir. $\nabla_x J(\theta, x, y)$ modelin kayıp fonksiyonunun x girdisine göre gradyanını belirtir. ∇ gradyan operatörüdür. J kayıp fonksiyonudur. θ modelin parametrelerini, y ise girdilerin doğru sınıf etiketini belirtir.

2.2.3. İzdüşürülen gradyan inisi (PGD)

Madry ve arkadaşları [22] tarafından sunulan çok adımlı bir saldırıdır. Küçük adım boyutları ile pertürbasyonlar oluşturarak FGSM’yi birden çok kez tekrarlandığı versiyonudur. PGD’ de pertürbasyon üretimi $[-\epsilon, \epsilon]$ aralığında rastgele bir değerle başlamaktadır. Süreç boyunca, piksel değeri gradyan yönünde α kadar ilerleyerek güncellenmektedir. PGD saldırısında genellikle L_∞ normu kullanılmaktadır. Gradyan sonrası elde edilen değerlerin ϵ sınırlarını ihlal etmesi durumunda izdüşüm işlemi gerçekleştirilir. Böylelikle, yeni girdinin belirlenen sınırlar içerisinde bulunan alt uzaya yansıtılması sağlanmaktadır.

$t = 0$ olmak üzere T adım için:

$$x = x^t$$

$$x^{t+1} = \Pi_{x,\varepsilon}(x^t + \alpha \cdot \text{sign}(\nabla_x J(\theta, x^t, y))) \quad (2.2)$$

x^t t adımının girdisi, x^{t+1} t adımında üretilen çekişmeli örnek, ε maksimum değişiklik miktarını belirten parametre, α iterasyonun adım boyutu yani saldırının yoğunluğunu belirleyen parametre, Π izdüşüm fonksiyonu, sign gradyanın yönünü gösteren gradyan işaretidir. $\nabla_x J(\theta, x^t, y)$ modelin kayıp fonksiyonunun x^t girdisine göre gradyanını belirtir. ∇ gradyan operatörüdür. J kayıp fonksiyonudur. θ modelin parametrelerini, y ise girdilerin doğru sınıf etiketini belirtir.

2.3. Derin Öğrenme

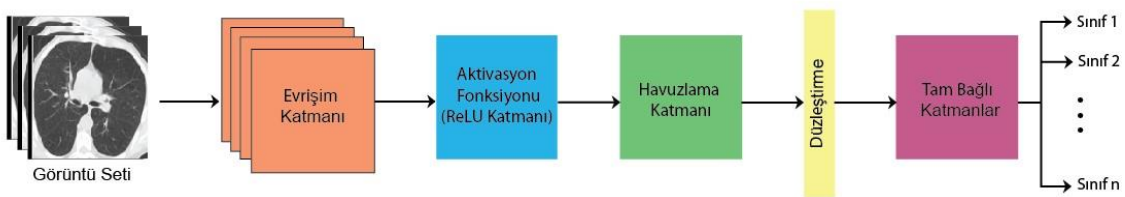
Derin öğrenme, insan beyninin öğrenme mekanizmasını taklit ederek verilerin işlenmesi temeline dayanan çok aşamalı makine öğrenme alt dalıdır. Derin ifadesi ile anlatılmak istenen birbiri ardına sıralanan çok sayıda katmanın varlığıdır. Derin öğrenme modelleri farklı katman ve düğümlerin kullanılabilmesi için ölçeklenebilmektedir. Derin öğrenme algoritmalarının eğitiminde önemli olan iki nokta bulunmaktadır. Bunlardan ilki katman yapısı sayesinde verilerin adım adım işlenerek daha anlamlı özniteliklerin elde edilebilmesidir. Bu sayede karmaşık problemlerin çözümü kolaylaşmaktadır. Oysa sığ yöntemler (bir veya iki katmandan oluşan) olarak da bilinen makine öğrenme modelleri karmaşık problemler karşısında her zaman yeterli olmamaktadır. İkincisi ise derin öğrenme katmanlarının kendinden önceki ve sonraki katmanlara göre kendini güncelleyebilme kapasitesidir. Geri yayılım (backpropagation) olarak adlandırılan bu güncelleme işlemi katmanlardaki parametrelerin düzenlenerek elde edilen çıktı ile hedef çıktının yakınsanmasını sağlamaktadır. Buna bir nevi hatadan öğrenmek denilebilir [23].

Derin öğrenme, akademik ve endüstriyel uygulama alanlarında geniş ölçekte başarılı sonuçlar elde eden evrensel bir yaklaşımdır. Derin öğrenmeyi ön plana çıkaran başlıca neden çeşitli alanlarda ele alınan birçok problemin çözümünde yüksek performans göstermesidir. Ancak tek nedenin performans olduğu söylenemez. Derin öğrenmeyi önemli kılan bir diğer faktör, ham verilerden öznitelikleri ve gösterimleri otomatik çıkarabilme kabiliyetidir. Optimize edilmiş öznitelikleri otomatik öğrenebilmesi bir modelin

sağlamlığını ve adaptasyon yeteneğini geliştirmektedir. Makine öğrenme yöntemlerinde öznitelik çıkarımı zaman ve emek isteyen bir iştir. Biraz daha açmak gerekirse, makine öğrenmesi ile öznitelik çıkarımı, çeşitli birçok yöntemin ve alanında yetkin uzmanların görüşlerinin bir arada kullanılmasını gerektiren işlemler bütünüdür. Derin öğrenme ise bu süreci uçtan uca tek sisteme indirgemektedir.

2.3.1. Evrişimsel sinir ağları

Evrişimsel sinir ağları (CNN) görüntü analizi ve görüntü işleme çalışmalarında yaygın bir şekilde yararlanılan geleneksel bir derin sinir ağı modelidir. Evrişimsel sinir ağları, biyolojik olarak, görsel korteksi oluşturan hücre dizisinin simülasyonu şeklinde tanımlanabilir [24]. Evrişimsel sinir ağları, genellikle, katmanları arasında az sayıda bağlantı bulunan ve verilerin uzamsal ilişkilerini korumayı hedefleyen yapılardır. Temel bir CNN mimarisi evrişim katmanı, havuzlama katmanı, aktivasyon katmanı ve nihai çıktıları hesaplayan tam bağlı katmandan oluşmaktadır. Bu bileşenler dışında bir de kayıp fonksiyonundan yararlanılmaktadır. Şekil 2.1’ de basit bir CNN mimarisi görülmektedir. CNN girdileri ızgara şeklindedir. Girdiler, giriş katmanları tarafından uzamsal ilişkiler muhafaza edilerek ağa alınmaktadır. Yani, giriş katmanındaki nöron sayısı görüntüdeki piksel sayısına eşittir. Her piksel ayrı bir nörona geçmektedir. Eğitim aşamasında geri yayılım ve gradyan iniş kullanılmaktadır.



Şekil 2.1. Temel CNN mimarisi

Temel evrişimsel sinir ağı bileşenleri:

- **Evrişim Katmanı:** Birbiri ardına sıralanmış evrişimli filtre koleksiyonundan oluşmaktadır. Çekirdek (kernel) bilinen bir diğer adıdır. CNN mimarisinde özellik çıkarma işleminin gerçekleştiği en karakteristik ve kritik bileşendir. Girdi görüntüsü ile filtreler birleşerek özellik haritası denilen bir matris oluşturulmaktadır. Evrişim katmanları boyunca, her bir katmanın oluşturduğu özellik haritası ardında yer alan bir

sonraki katmanın girdisidir. Mimaride ilk sıralarda yer alan evrişim katmanları çizgi, kare ve üçgen gibi temel geometrik özellikleri çıkarırken, ilerleyen katmanlarda daha kompleks özellikler çıkarılmaktadır.

- **Havuzlama Katmanı:** Özellik haritalarını alt örneklemeden sorumlu katmanlardır. Gereksiz katmanları yok sayarak özellik haritasının boyutunu azaltmaktadır. Havuzlama işleminde önemli noktalar ön plana çıkarılarak yüksek özellikler korunmaktadır. Minimum havuzlama, maksimum havuzlama ve global ortalama havuzlama gibi çeşitli havuzlama yöntemleri bulunmaktadır.
- **Aktivasyon Katmanı:** Bir nöronun gelen bilgiyi aktarıp aktarmayacağına karar veren doğrusal olmayan fonksiyondur. Evrişim katmanlarından sonra kullanılarak doğrusallığı engellemekte ve karmaşık bilgilerin öğrenilmesini sağlamaktadır. Eğitim sürecinde geri yayılım gerçekleştirdiği için farklılaşma kabiliyeti bulunmaktadır [24]. Sigmoid, tanh, ReLU, Swish gibi farklı aktivasyon fonksiyonları bulunmaktadır.
- **Düzleştirme Katmanı:** Havuzlama işleminden sonra elde edilen matrisleri tek boyutlu dizilere dönüştürmekten sorumlu katmandır. Tek boyutlu diziler tam bağlı katmanlara girdi olarak aktarılmaktadır.
- **Tam Bağlı Katman:** CNN modelinin sonunda bulunan sınıflandırıcı katmandır. Her nöron bir önceki katmanın tüm nöronlarına bağlıdır. Bu katmanın çıktısı CNN modelinin nihai çıktısını oluşturmaktadır.
- **Kayıp Fonksiyonları:** Modelin eğitim sürecinde gerçek çıktı ile elde edilen çıktı arasındaki hatayı hesaplamak için kullanılan fonksiyondur.

Görüntü işleme çalışmalarında sıkça karşılaşılan bir tanım ise transfer öğrenme (TL)' dir. Transfer öğrenimi, bir problem üzerinde eğitilen belirli bir modelin daha sonra farklı bir problemde referans noktası olarak kullanıldığı bir öğrenme türüdür. Model büyük bir veri seti ile bir ön eğitim işleminden geçer. Ön eğitim adımı ImageNet seti yaygın olarak kullanılmaktadır. Eğitilen model geniş çaplı bir öğrenme sürecinden geçtiği için öğrenilen özelliklerin bir başka probleme uyarlanması fayda sağlamaktadır. Transfer öğrenme, yetersiz eğitim verilerinin bulunduğu ve/veya temel probleme benzer süreç takip eden durumlarda verimli bir çözüm yolu olmaktadır [25]. Aynı zamanda, model ağırlıklarının baştan başlatılması gerekmediği için hesaplama yükü azalmakta ve işlem hızı artmaktadır. CNN mimarisi, ağırlık paylaşımı sayesinde, eğitilen ağ parametrelerinin sayısını azaltarak modelin genelleştirme yeteneğini artırabilir. Diğer yandan, hiper parametre ayarları, eğitim

süresinin ve model derinliğinin artırılması, düzenlileştirme gibi iyileştirici düzenlemeler gerçekleştirilebilmektedir [24].

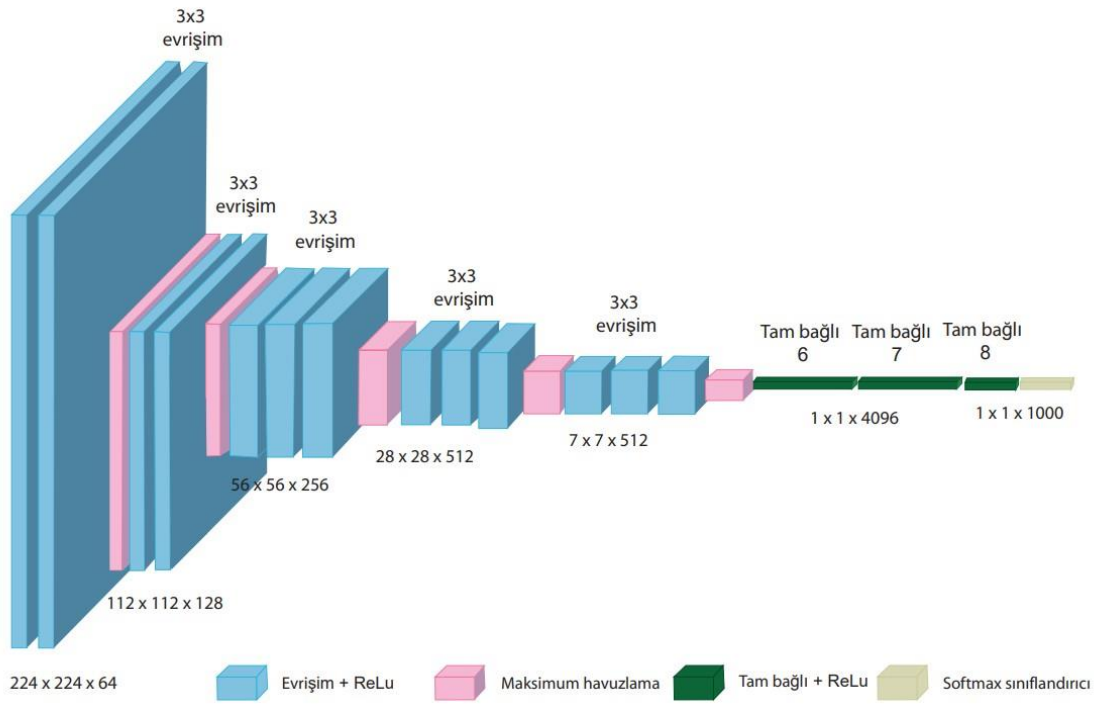
2.3.2. Evrişimsel sinir ağı mimarileri

Bu bölümde, görüntü işleme çalışmalarından öncü rol oynayan ve bu çalışmada faydalanılan CNN mimarileri açıklanmaktadır.

VGG mimarileri

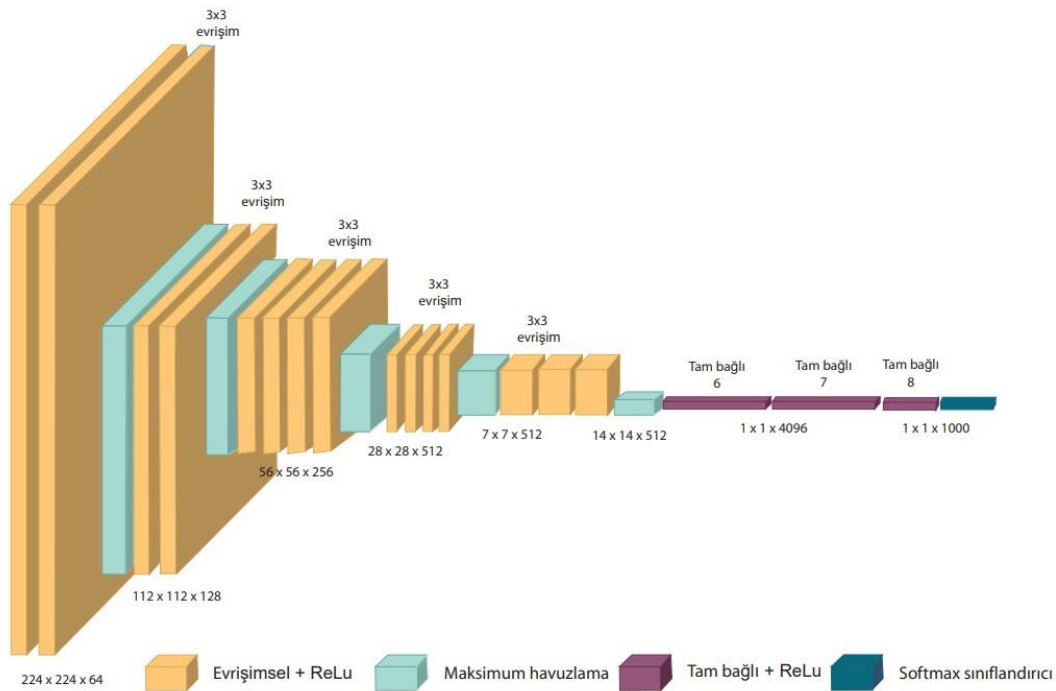
VGGNet olarak da bilinen VGG [26] mimarisinin tam açılımı Visual Geometry Group (Görsel Geometri Grubu)' dur. Mimari, Oxford Üniversitesinde geliştirilmiştir. VGG bir derin ağ yapısı olmasına karşın basit bir tasarıma sahiptir. Mimaride büyük boyutlu evrişim katmanları yerine 3x3 sabit çekirdek boyutlu birden çok evrişim katmanı kullanılmaktadır. Bu sayede, parametre sayında kayda değer azalma olmaktadır. Ayrıca, evrişim katmanlarının sayısındaki artış ile ağı öğrenme kabiliyetinin artırılmakta ve hesaplama miktarında azaltılmaktadır. VGG16 ve VGG19 olarak iki model tasarımı bulunmaktadır.

VGG16 modeli, 16 katmandan oluşan bir tasarıma sahiptir. İkili ve üçlü kombinasyonlardan oluşan on üç evrişim katmanı bulunmaktadır. Son üç katmanında tam bağlı katmanlar bulunmaktadır. Modelde aktivasyon işlemi için ReLU aktivasyon fonksiyonu kullanılmaktadır. Son katmanında ise Softmax aktivasyon fonksiyonundan yararlanılmaktadır. Katmanlar arasında boyut azaltma işlemini gerçekleştirmek için 2x2 çekirdek boyutlarına sahip iki kaydırmalı maksimum havuzlama katmanları kullanılmaktadır. VGG16 mimarisi Şekil 2.2'de sunulmaktadır.



Şekil 2.2. VGG16 modeline ait mimari gösterim

VGG19 modeli, 19 katmandan oluşan bir tasarıma sahiptir. İkili ve dörtlü kombinasyonlardan oluşan on altı evrişim katmanı bulunmaktadır. Son üç katmanında tam bağı katmanlar bulunmaktadır. Modelde aktivasyon işlemi için ReLU aktivasyon fonksiyonu kullanılmaktadır. Son katmanında ise Softmax aktivasyon fonksiyonundan yararlanılmaktadır. Katmanlar arasında boyut azaltma işlemini gerçekleştirmek için 2x2 çekirdek boyutlarına sahip iki kaydırmalı maksimum havuzlama katmanları kullanılmaktadır. VGG19 mimarisi Şekil 2.3'te sunulmaktadır.



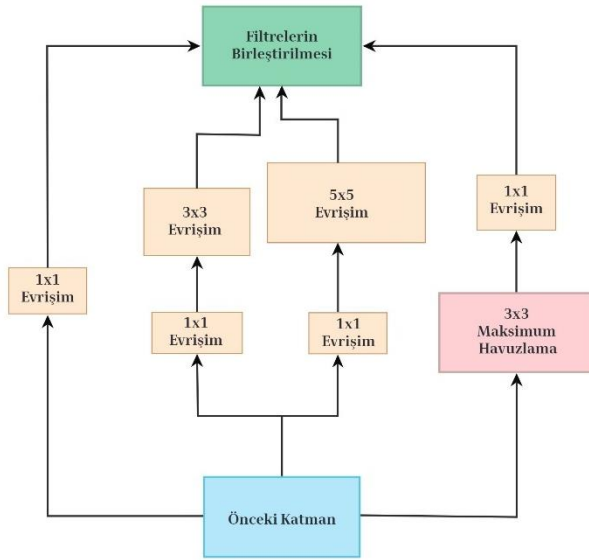
Şekil 2.3. VGG19 modeline ait mimari gösterim

Inception mimarileri

Inception mimarileri önceki araştırmalarda odaklanılan derin ağ mimarilerinin sorunlarına çözüm bulmak ve daha düşük hesaplama maliyetine sahip ağlar oluşturmak için tasarlanmıştır. Inception mimarisi daha az derinliğe sahip geniş yapılara odaklanmakta ve nispeten daha küçük boyutlu evrişim filtreleri kullanmaktadır. Inception [27] ilk olarak 2014 yılında Google tarafından tanıtılmıştır Zamanla çeşitli Inception versiyonları geliştirilmiş ve Inception model ailesi meydana gelmiştir. Bu versiyonlar gelişim sırasına göre versiyon numaraları ile adlandırılmıştır. InceptionV1, InceptionV2 [28], InceptionV3 [29] ve InceptionV4 [30] bilinen Inception mimarileridir. Bu çalışmada yalnızca InceptionV3 mimarisi kullanılmaktadır. Ancak InceptionV3 mimarisinin daha net anlaşılması için gelişim sürecinde temel oluşturan InceptionV1 ve InceptionV2 mimarileri de bu başlık altında açıklanmaktadır.

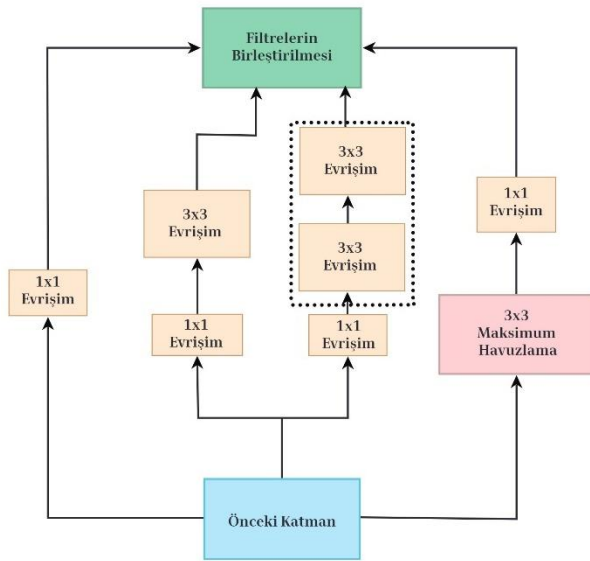
InceptionV1 diğer adıyla GoogleLeNet, doğrusal olarak sıralanmış dokuz Inception modülünden meydana gelmektedir. Inception modülü aynı seviyede farklı boyutlarda filtreler kullanarak evrişim işlemi ve maksimum havuzlama işlemi gerçekleştirir. Bu işlemler sonucunda eş zamanlı çıkarılan özellik haritaları birleştirilerek bir sonraki katmana aktarılmaktadır. Şekil 2.4’ te görülen InceptionV1 modülünde, önceki katman ile

filtre birleştirme katmanları arasında dört alt ağ bulunmaktadır. İlk alt ağda 1x1 evrişim işlemi gerçekleştirilmektedir. İkinci alt ağda 1x1 ve 3x3 evrişim işlemi gerçekleştirilmektedir. Üçüncü alt ağda 1x1 ve 5x5 evrişim işlemi gerçekleştirilmektedir. İkinci ve üçüncü alt ağlarda 1x1 evrişim uygulanması giriş kanal sayısının azaltılmasını sağlar. Dördüncü alt ağda 3x3 maksimum havuzlama işlemi gerçekleştirilmektedir. Bu ağda 1x1 evrişim işlemi havuzlama işleminde sonra gerçekleşir. Alt ağlarda farklı boyutlarda filtreler kullanılmasının amacı farklı ölçeklerdeki desenlerin paralel olarak çıkarılmasını sağlamaktır. Çıkarılan bu özellik haritaları filtre birleştirme kanalında birleştirmekte ve sonraki katmana aktarılmaktadır.



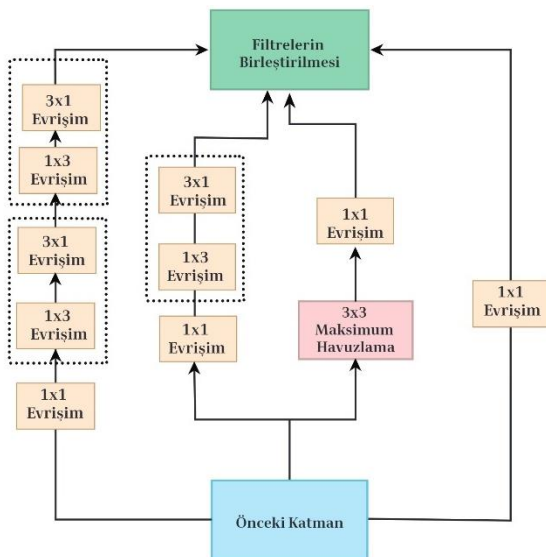
Şekil 2.4. InceptionV1 modül gösterimi

InceptionV1 modeline yığın normalizasyon katmanları eklenmesiyle BN-Inception olarak da bilinen InceptionV2 modeli ortaya çıkmaktadır. Yığın normalizasyonu, evrişim katmanları ile aktivasyon katmanları arasında yığın ortalama ve yığın varyansları üzerinden gerçekleştirilmektedir. Yığın normalizasyonu, başlangıç değerlerinin ve parametre ölçeklerinin gradyanlar üzerindeki etkisini indirgeyerek başlangıçtaki öğrenme oranının ve eğitim hızının optimizasyonunu sağlamaktadır. InceptionV2 mimarisini InceptionV1' den farklılaştıran bir diğer faktör modül tasarımındaki değişikliktir. Şekil 2.5' te görüldüğü gibi, InceptionV2 modülünde, InceptionV1' de bulunan 5x5 evrişim filtresi art arda konumlanan iki 3x3 evrişim filtresiyle değiştirilmiştir. Bu sayede evrişim işleminde gerçekleştirilen matris çarpımı azaltılarak hesaplama maliyeti iyileştirilmektedir.



Şekil 2.5. InceptionV2 modül gösterimi

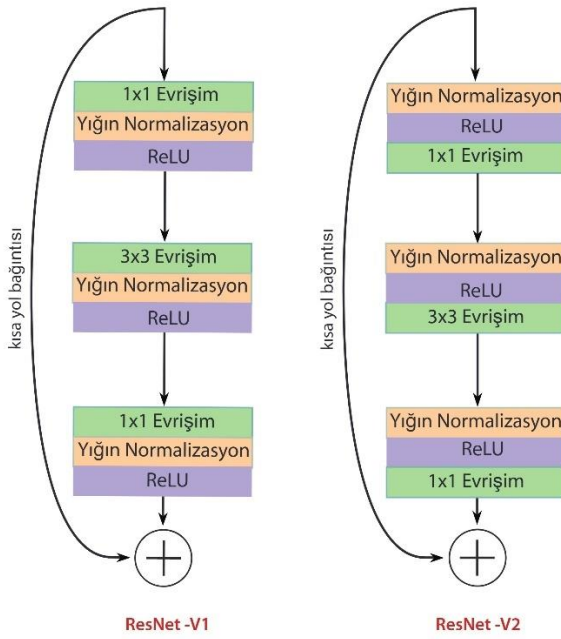
InceptionV3 ise InceptionV2’ deki çarpanlara ayırma fikrini bir adım öteye götürmektedir. Bu mimarinin tasarımında 3x3 evrişim filtresi 3x1 ve 1x3 şeklinde birbiri ardına uygulanan iki evrişim parçasına ayrılmaktadır (Şekil 2.6). Bu işlem faktörize edilmiş evrişim olarak adlandırılmaktadır. 3x1 evrişim filtresi yatay düzlemde özellikleri çıkarırken, 1x3 evrişim filtresi ise dikey düzlemde özellikleri çıkarmaktadır. Faktörize edilmiş evrişim sonucunda matris çarpımıyla orantılı olarak üçte bir oranında daha az hesaplama maliyeti ortaya çıkmaktadır. InceptionV3, ek olarak, 7x7 faktörize edilmiş evrişim, RMSProp optimizasyon algoritması, düzenleyici bir yöntem olan etiket yumuşatma ve yardımcı sınıflandırıcı yeniliklerini kapsamaktadır.



Şekil 2.6. InceptionV3 modül gösterimi

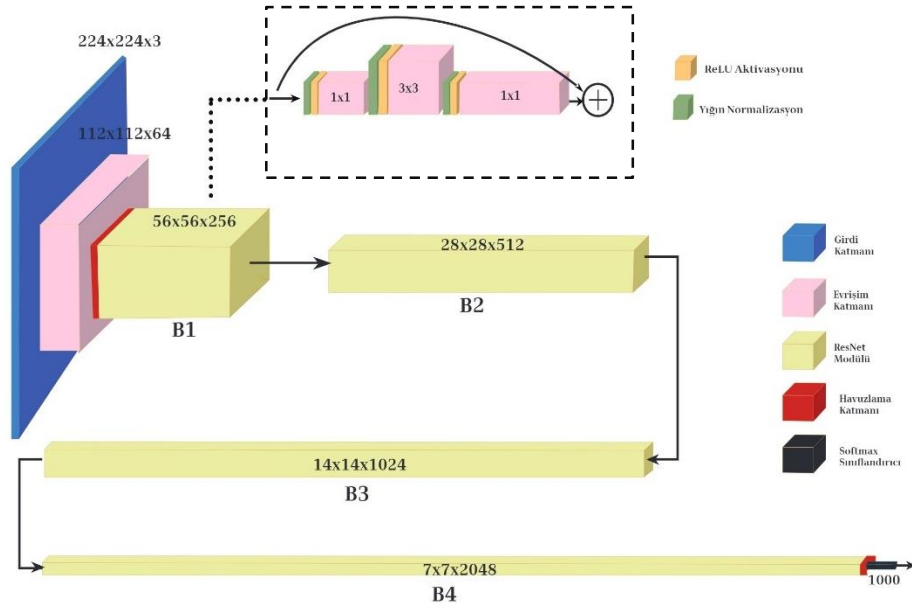
ResNet mimarileri

ResNet [31, 32] mimarisinin açılımı Residual Network (Artık Ağ) 'dür. ResNet, her blokta 2 ya da 3 evrişimin gerçekleştiği artık blok yapılarından meydana gelmektedir. Artık bloklarda evrişim katmanlarından sonra yığın normalizasyonu ve ReLU aktivasyonu yer almaktadır. Resnet, bir katmandaki girdi değerlerinin bir sonraki katmanın girdi değerlerine eklenmesiyle kaybolan gradyanların engellenmesini amaçlamaktadır. Bir katmanın girdisinin bir başka katmanın girdisine eklenmesini sağlayan doğrudan bağlantılara kısa yol bağlantısı denilmektedir. Bu bağlantılar sayesinde girdi ve çıktı arasındaki fark öğrenilerek kaybolan gradyan problemi önlenmektedir. Ayrıca, kısa yol bağlantısı ile ağdaki belirli katmanların çıktıları atlanarak derin ağlarda eğitim sürecinin daha verimli ve hızlı gerçekleştirilmesi sağlanır. Bu aktarım sırasında, girdi ve çıktı boyutların eşleştirilmesi işlemine kimlik eşleme denilmektedir. Kimlik eşleme, gradyanların kolay şekilde geri yayılmasını ve derin ağlarda eğitim sürecinin kolaylaşmasını sağlamaktadır. Artık bloklardaki karmaşıklığı azaltmak için bir darboğaz artık bloğu tasarlanmıştır. Darboğaz artık bloğu 1x1, 3x3, 1x1 evrişim katmanlarından oluşmaktadır. Birinci 1x1 katmanının girdi değerlerinin bir sonraki 1x1 katmanın çıktı değerlerine eklenmesiyle meydana gelmektedir. Birden fazla artık blok yığınlarından oluşan ResNet mimarisi ağın derinliğine katman sayılarına göre adlandırılmaktadır. En yaygın kullanılan ResNet50, ResNet101 ve ResNet152 modelleri darboğaz artık bloklarından oluşmaktadır. Bu çalışmada ResNet mimarisinin ikinci versiyonun olan ResNetV2 kullanılmaktadır. ResNetV2 mimarisinin bir önceki versiyondan temel farkı erken aktivasyondur (pre-activation). ResNetV2 mimarisinde evrişim katmanından önce yığın normalizasyonu ve ReLU aktivasyonu yer almaktadır. Bu yapı, ağda eğitim hızının artmasına ve ağın derinleşmesine destek olmaktadır. ResNetV1 ve ResNetV2 modellerine ait blok yapıları Şekil 2.7'de verilmektedir.



Şekil 2.7. ResNetV1 ve ResNetV2 blok yapıları

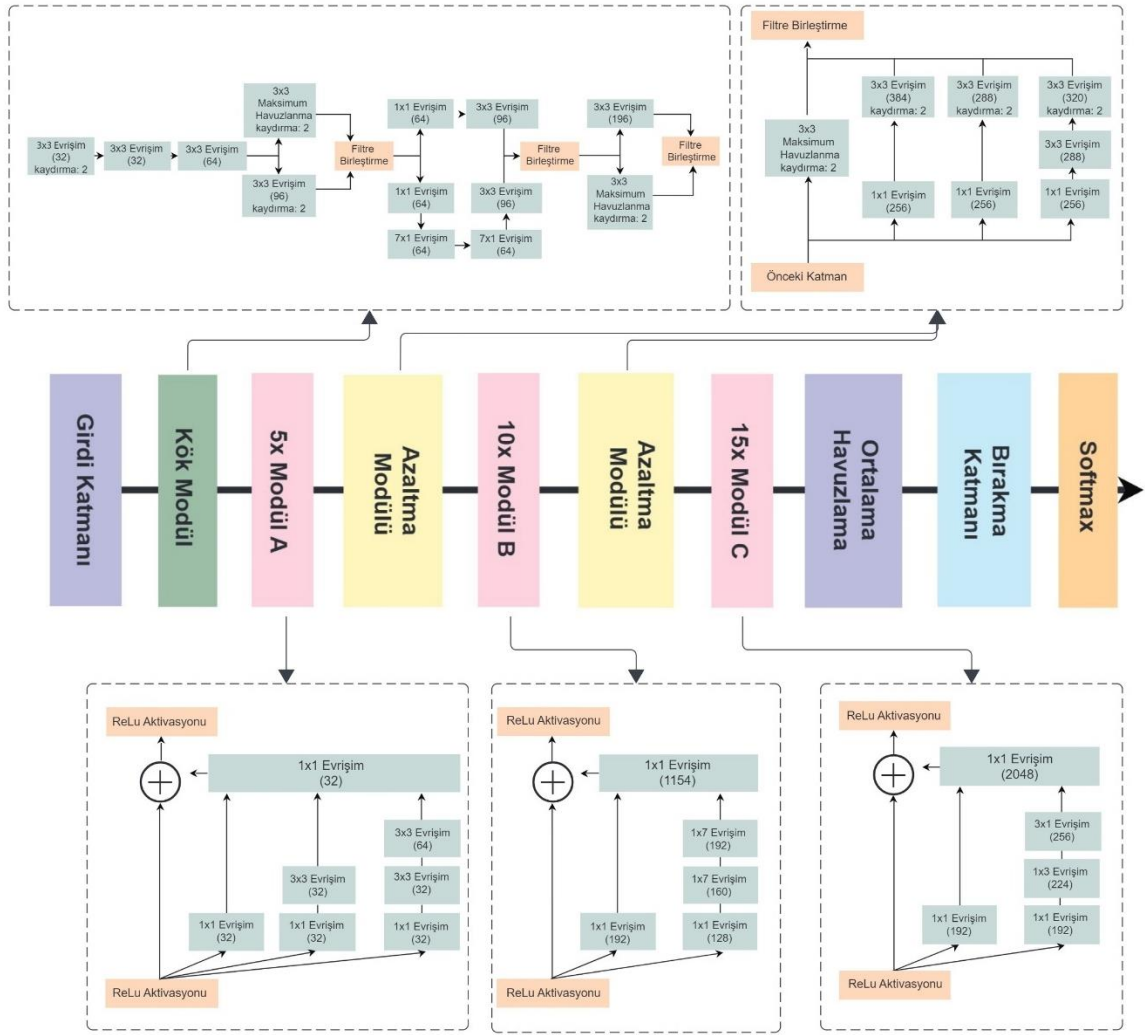
Şekil 2.8’ de temel Resnet mimarisi gösterilmektedir. ResNet, temel olarak, ardışık sıralanmış dört bloktan meydana gelmektedir. Artık bloklardan önce giriş katmanını takip eden 7x7 boyutlu evrişim katmanı yer almaktadır. Ardından gelen bloklar, katman sayısına bağlı olarak tekrar etmektedir. ResNetV2’ de darboğaz artık bloklar kullanılmaktadır. Her bir blokta 3 evrişim katmanı yer almaktadır. ResNet50V2 mimarisi 3 adet B1 bloğu, 4 adet B2 bloğu, 6 adet B3 bloğu ve 3 adet B4 bloğu olmak üzere 16 bloktan oluşmaktadır. ResNet101V2 mimarisi 3 adet B1 bloğu, 4 adet B2 bloğu, 23 adet B3 bloğu ve 3 adet B4 bloğu olmak üzere 33 bloktan oluşmaktadır. ResNet150V2 mimarisi 3 adet B1 bloğu, 8 adet B2 bloğu, 36 adet B3 bloğu ve 3 adet B4 bloğu olmak üzere 50 bloktan oluşmaktadır. Mimarinin son adımında 1000 nöron içeren tam bağlı katman bulunmaktadır. Bir bloktan diğerine geçişlerde evrişim katmanında iki kaydırma gerçekleştirilerek girdi boyutu yarıya düşürülür.



Şekil 2.8. ResNet mimari gösterim

InceptionResNetV2 mimarisi

Google tarafından 2016 yılında sunulan InceptionResNetV2 [30], Inception ve ResNet modellerinin tasarımsal kombinasyonu ile meydana gelen hibrit bir derin ağ mimarisidir. Inception modellerinin paralel işleme kabiliyeti ve düşük hesaplama maliyeti gibi yapısal avantajlarının yanı sıra ResNet mimarisinin artık bağlantılar ile daha derin ağlar oluşturma kapasitesinden yararlanmaktadır. InceptionResNetV2 mimarisi 5 modülden oluşmaktadır: kök modül, A modülü, B modülü, C modülü ve azaltma modülü. InceptionResNetV2 mimarisinde giriş katmanının ardından kök modülü yer almaktadır. Kök modülünde elde edilen düşük boyutlu özellik haritaları diğer modüllerine aktarılmaktadır. InceptionResNet modülleri paralel evrim katmanları ve bir kısa yol bağlantısı içermektedir. A modülü 5 kez, B modülü 10 kez ve C modülü 5 kez tekrar etmektedir. A, B ve C modülleri arasında azaltma modülü yer almaktadır. Azaltma modülü özellik haritalarının boyutlarını azaltarak hesaplama maliyetlerinin düşürülmesini sağlamaktadır. Çıktı değerleri ortama havuzlama işleminden geçmekte ve ardından bırakma katmanı ile bir kısmı atılmaktadır. Geri kalan çıktı değerleri softmax aktivasyon fonksiyonuna sahip sınıflandırma katmanında sınıflandırılmaktadır. InceptionResNet modülleri Şekil 2.9’da gösterilmektedir.

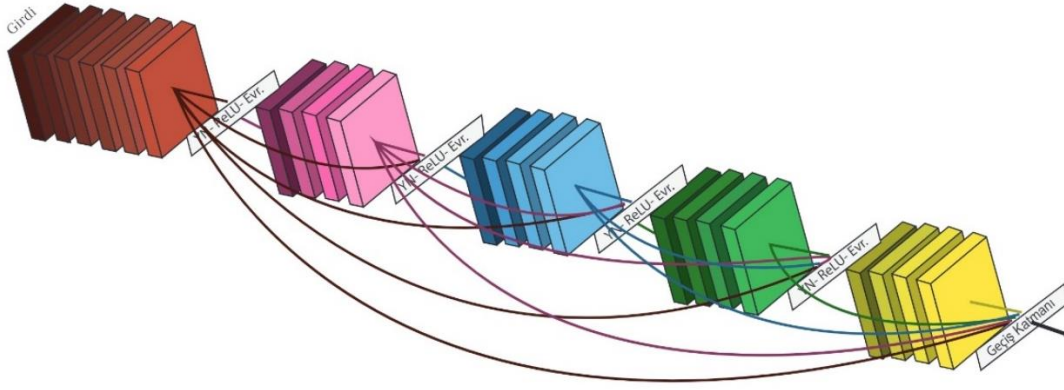


Şekil 2.9. InceptionResNetV2 blok yapıları

DenseNet mimarileri

DenseNet [33], ResNet mimarisindeki sinyallerin sonraki katmana aktarılması fikrine benzer bir yaklaşıma sahiptir. DenseNet mimarisinde her katman diğer tüm katmanlara ileri beslemeli şekilde bağlanarak maksimum bilgi akışı sağlamaktadır. Bir katmanın girdileri kendinden önceki tüm katmanların özellik haritalarından oluşturmaktadır. Aynı şekilde katmanın ortaya koyduğu özellik haritası sonraki tüm katmanlara aktarılmaktadır. Katmanların kendinden önceki katmanlara ait ortak bilgiyi işleyebilmesi zengin desenler oluşturulmasını sağlar ve gradyan yayılımını güçlendirir. Katmanlar arasındaki bu bağlantı yoğunluğu DenseNet mimarisinin adının kaynağıdır. Bu yoğun bağlantıları barındıran katmanlara, blok ismi verilmektedir. DenseNet mimarisinde yoğun bloklarda erken aktivasyon yaklaşımı kullanılmaktadır. Yoğun blok içindeki bir katmana aktarılan özellik

haritaları öncelikle yığın normalizasyon ve ReLU aktivasyon katmanından geçmektedir. Elde edilen çıktılar ilgili katmana aktarılmaktadır. DenseNet mimarisi ResNet'ten farklı olarak, özellik haritalarının aktarırken toplamayla birleştirmek yerine birbirine ekleme yöntemi kullanmaktadır. DenseNet mimarisinin 121, 160 ve 201 katman içeren üç versiyonu bulunmaktadır. DenseNet mimarisi Şekil 2.10'da gösterilmektedir.

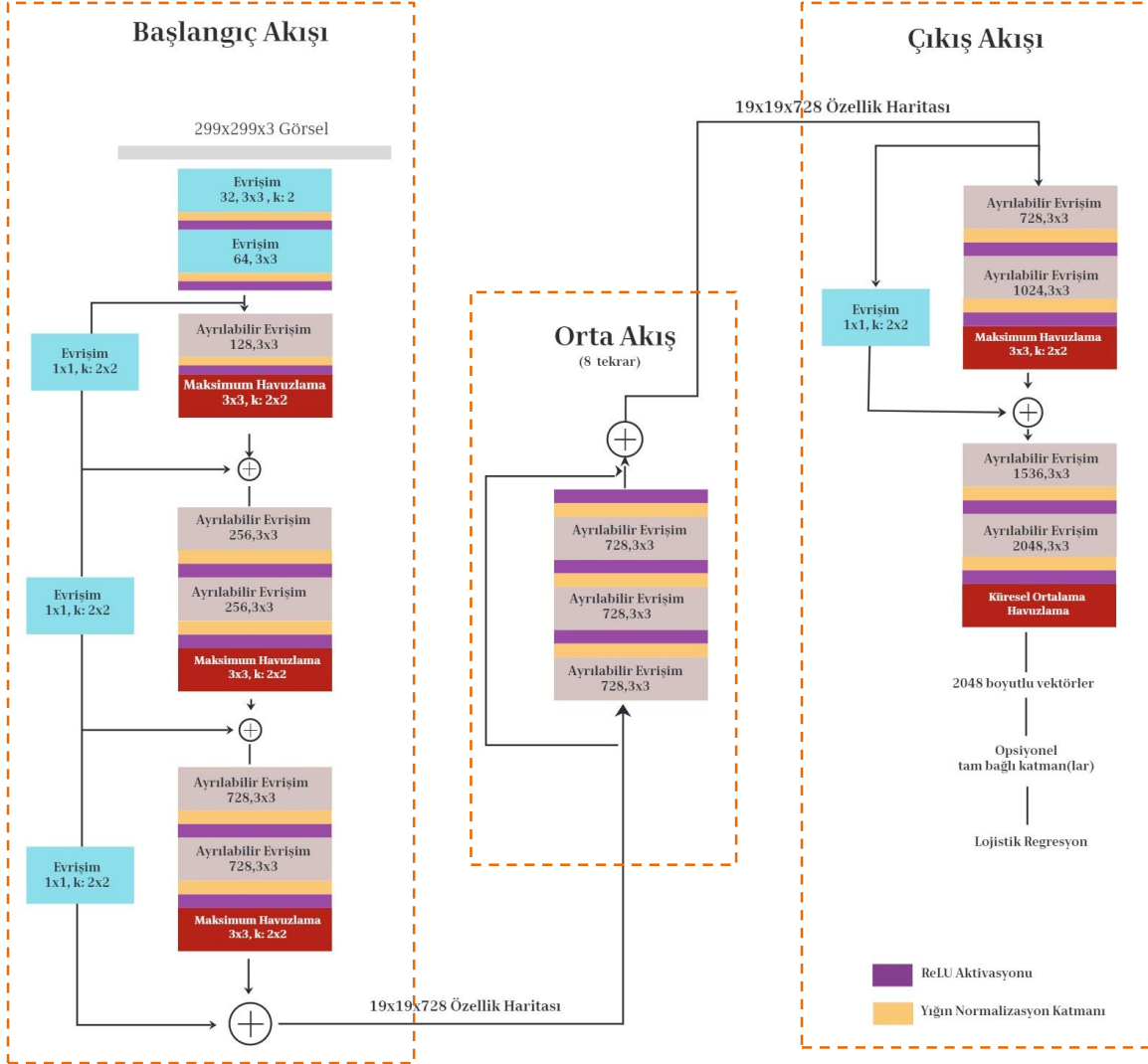


Şekil 2.10. DenseNet mimari gösterimi

Xception mimarisi

Inception mimarisinden esinlenerek geliştirilen Xception [34], adını İngilizce "Extreme Inception" kelimelerinin birleşiminden almaktadır. Xception, Inception modülleri yerine değiştirilmiş derinlemesine ayrılabilir evrişim katmanlarının uyarlandığı bir mimaridir. Xception mimarisi, kısaca, artık bağlantılar ile güçlendirilmiş değiştirilmiş derinlemesine ayrılabilir evrişim katmanlarının doğrusal bir yığını olarak tanımlanabilir. Mimaride, bilinen standart derinlemesine ayrılabilir evrişim ağlarından farklı olarak öncelikle piksel düzeyinde uzamsal/noktasal evrişim işlemi gerçekleştirilmektedir. Noktasal evrişim işleminden sonra elde edilen özellik haritaları üzerinde kanal boyunca derinlemesine/uzaysal evrişim işlemi gerçekleştirilir. Ayrıca Xception mimarisinin bir diğer farkı ReLU ativasyon fonksiyonunun kullanımına dayalı doğrusal olmama durumunun varlığıdır. Şekil 2.11' de gösterilen Xception mimarisi başlangıç, orta ve çıkış olmak üzere üç çeşit akıştan oluşmaktadır. Girdi, başlangıç akışından geçtikten sonra orta akışta sekiz yineleme yaparak çıkış akışına aktarılmaktadır. Mimari 36 evrişim katmanına sahiptir. Evrişim katmanlarının ilk ve son modüllerinde artık bağlantı yoktur. Geri kalan 14 modül doğrusal artık bağlantılara sahiptir. Her bir evrişim ve derinlemesine ayrılabilir evrişim katmanının sonunda yığın normalizasyonu ve ReLU aktivasyonu

gerçekleştirilmektedir. Xception mimarisinin orijinal sunumunda sınıflandırma için lojistik regresyon kullanılmış ve tam bağlı katman kullanımı opsiyonel bırakılmıştır. Bu tez kapsamında tam bağlı katmanlar kullanılmaktadır.



Şekil 2.11. Xception mimari gösterimi

NASNet mimarisi

NASNet [35] pekiştirmeli öğrenme arama yöntemi ile optimize edilen ayrılabilir evrişim ve havuzlama katmanlarından oluşan ölçeklenebilir bir mimaridir. Mimarinin geliştiricileri Sinir Mimarisi Arama (NAS) yaklaşımından esinlenerek modelin daha az maliyet ile ölçeklenebilirliği konusunda çalışmıştır. NAS yaklaşımı, tekrarlayan sinir ağlarını (RNN) kullanarak bir alt ağ uzayında optimal mimariyi arar. Temel evrişimli ağ mimarisi önceden manuel olarak belirlenmektedir. Arama uzayındaki evrişimsel sinir ağları, aynı

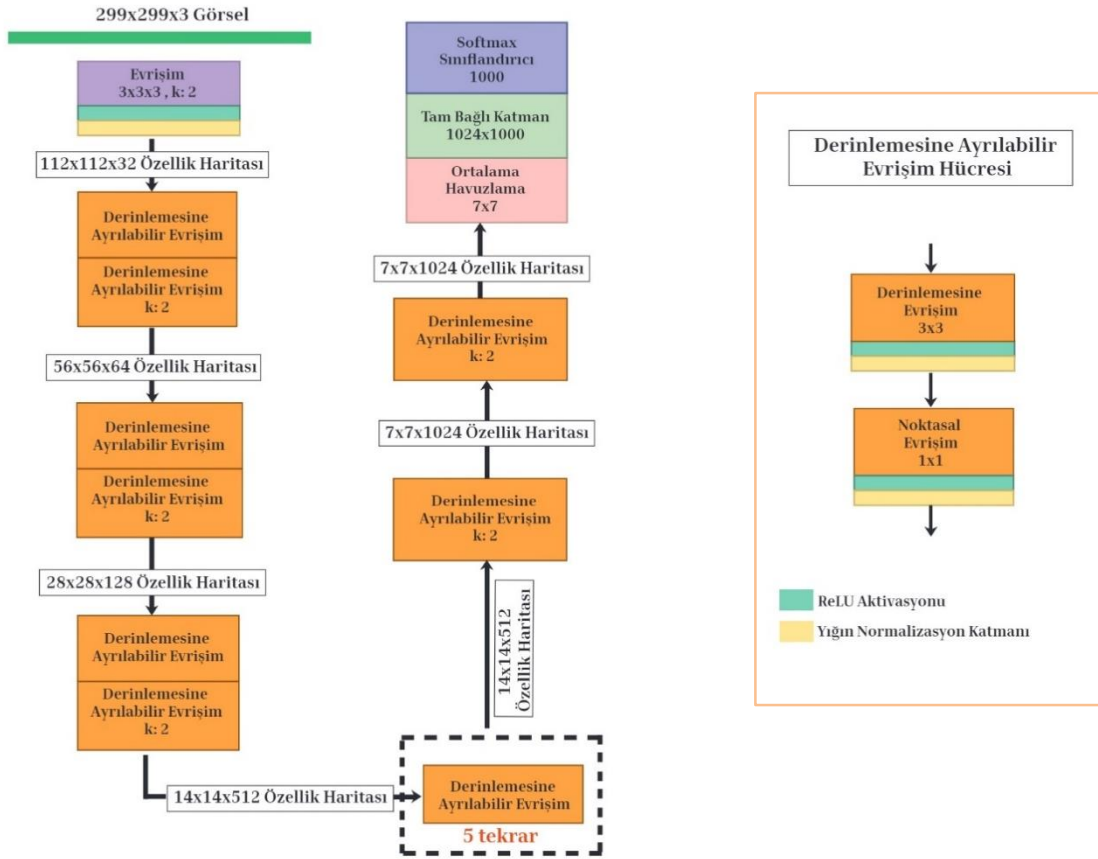
mimari tasarıma sahip olmasının yanı sıra farklı ağırlıklara sahip yinelenen katmanlardan oluşmaktadır. NASNet en iyi ağ mimarisini aramak yerine en iyi hücre yapısını aramaya odaklanmaktadır. Hücre arama havuzunda bulunan yapılardan bazıları birim aktarma, 1x1 evrişim, 3x3 evrişim, 3x3 genişlemiş evrişim, 1x3 sonra 3x1 evrişim, 1x7 sonra 7x1 evrişim, 3x3 ortalama havuzlama, 3x3 maksimum havuzlama, 5x5 maksimum havuzlama, 7x7 maksimum havuzlama, 3x3 derinlemesine ayrılabilir evrişim, 5x5 derinlemesine ayrılabilir evrişim, 7x7 derinlemesine ayrılabilir evrişim olarak sıralanabilir. Arama havuzunun hücre düzeyine indirgenmesi mimarinin hızını ve ağın aktarılabilme kabiliyeti artırmaktadır. NASNet mimarisinde normal hücre ve indirgenme hücresi olarak iki tür evrişim hücresi bulunmaktadır. Normal hücre, girdi olarak alınan bir özellik haritasından aynı boyutlara sahip bir özellik haritası oluşturmaktadır. İndirgeme hücresi ise büyük boyutlu bir özellik haritasının boyutlarını iki kat azaltmaktadır. Bu yapı sayesinde farklı boyuttaki görüntülerin kolayca ölçeklenebilmesi ve hesaplama yükünün azaltılması sağlanmaktadır. Evrişimli katman ve filtre sayıları esnektir. Yapılacak değişimler ile farklı hesaplama kabiliyetlerine sahip NASNet versiyonları oluşturulabilir. NASNet mimarisi NASNetMobile ve NASNetLarge olmak üzere iki tasarıma sahiptir. Bu tez çalışması kapsamında yalnızca NASNetMobile kullanılacaktır.

MobileNet mimarileri

MobileNet, mobil cihazlar ve gömülü sistemler gibi kaynak kısıtlı ortamlarda derin öğrenme modellerinin efektif şekilde kullanılabilmesi için tasarlanmış bir model ailesidir. Tasarım açısından diğer mimarilere kıyasla daha küçük ve hızlıdır. Seri her bir versiyonda performansı artırmak için çeşitli yenilikler ve iyileştirmeler denemektedir. Bu bölümde MobileNetV1, MobileNetV2 ve MobileNetV3 modelleri açıklanacaktır.

MobileNetV1 [36] ilk MobileNet versiyonudur. MobileNetV1 tasarımında en önemli yapıtaşını derinlemesine ayrılabilir evrişim katmanları oluşturmaktadır. Derinlemesine ayrılabilir evrişimi iki aşamalı olarak gerçekleştirilir: İlk aşamada giriş kanalı boyunca derinlemesine 3x3 boyutlu evrişim gerçekleştirilir. Bu aşamada her giriş kanalı için ayrı bir çıktı kanalı üretilmektedir. İkinci aşamada 1x1 noktasal evrişim işlemi gerçekleştirilir. Bir önceki aşamada üretilen çıktılar doğrusal olarak birleştirilir. Noktasal evrişim işlemi kanal boyutundaki değişimi kontrol etmek için esneklik sağlamaktadır. MobileNetV1, ek olarak, mimarinin boyutunu ve karmaşıklığını belirlemek için genişlik çarpanı (width multiplier)

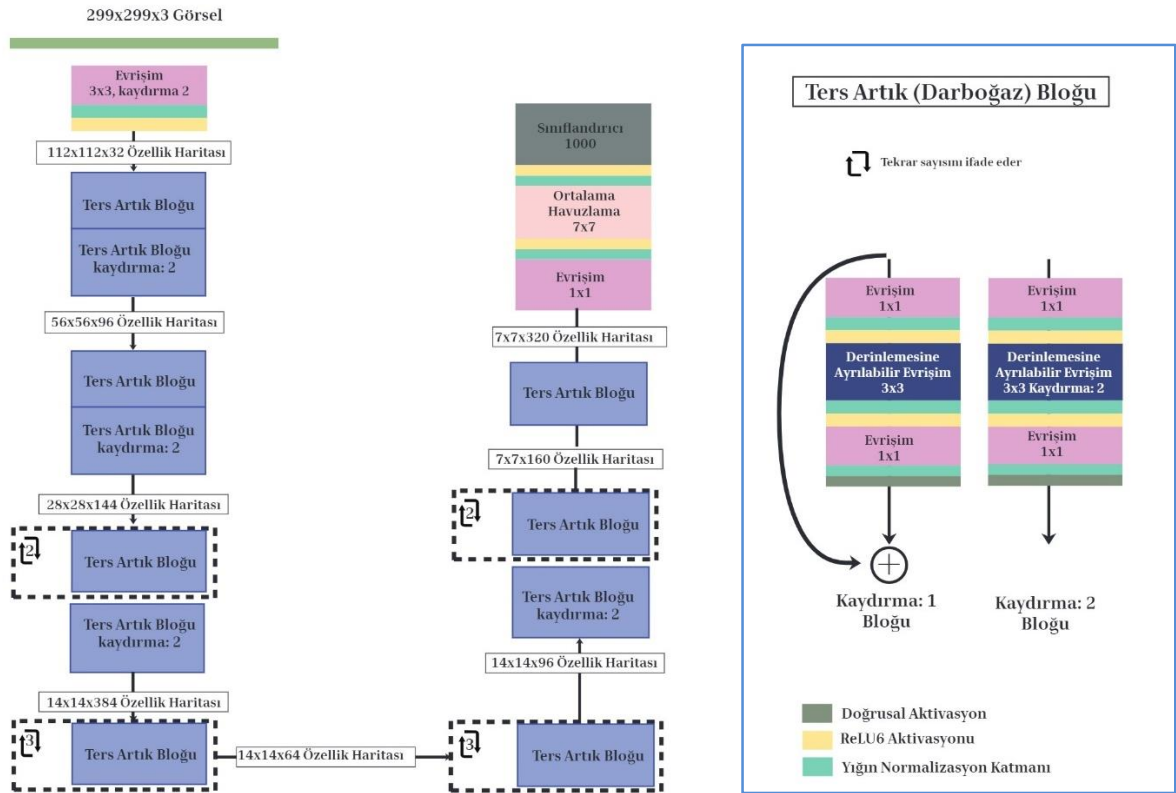
ve çözünürlük çarpanı (resolution multiplier) adı verilen iki hiper parametre kullanmaktadır. Genişlik çarpanı, kanal sayısını yani ağın genişliğini düzenlemektedir. Çözünürlük çarpanı ise girdinin boyutunun ayarlanmasından sorumludur. Şekil 2.12’ de gösterilen MobileNetV1 mimarisi, standart bir evrişim katmanı ve yinelenen derinlemesine ayrılabilir evrişim katmanlarından oluşmaktadır. Evrişim işlemleri kaydırmalı ve kaydırmaz olarak iki farklı şekilde gerçekleştirilmektedir. Kaydırmalı evrişim katmanlarında 2 kaydırma uygulanmadır. Kaydırma sonucunda girdi boyutu yarıya düşmektedir. Her bir evrişim işleminin ardından yığın normalizasyonu ve ReLU aktivasyonu gerçekleştirilmektedir. Evrişim katmanlarını takip eden bir ortalama havuzlama katmanı, tam bağlı katman ve 1000 nöronlu softmax sınıflandırıcısı yer almaktadır.



Şekil 2.12. MobileNetV1 mimari gösterimi

MobileNetV2 [37], bir önceki versiyonda kullanılan derinlemesine ayrılabilir evrişim katmanlarının yerine derinlemesine ayrılabilir evrişimi de kapsayan ters artık blokları kullanmaktadır. Ters artık blok yapısı ResNet mimarisindeki artık blokların ters işleyiş mekanizmasına dayanmaktadır. Artık bloklarda blok içindeki kanal sayısının azaltılması

esas alınırken, ters artık bloklarda kanal sayısı genişleme faktörü ile orantılı olarak artırılmaktadır. Ters artık blokları darboğaz katmanları ve doğrusal aktivasyondan oluşan doğrusal darboğazlar içermektedir. Doğrusal darboğazlar modelin genişletilmiş özellik haritalarından daha karmaşık özellikleri öğrenmesine olanak tanımaktadır. Ek olarak, ters artık bloklarda mobil cihazlarda kullanmak üzere tasarlanmış doğrusal olmayan ReLU6 aktivasyon fonksiyonu bulunmaktadır. ReLU6, ReLU fonksiyonunun sınırlandırılmış bir versiyonudur. ReLU6 aktivasyon fonksiyonu, girdi değerlerini 0 ile 6 arasında sınırlamakta ve modelin işlem yükünü azaltmaktadır. MobileNetV2 geliştiricileri, darboğazların gerekli bilgileri taşıdığı ve genişleme katmanının doğrusal olmama durumunu desteklemek için var olduğu sezgisine dayanarak darboğazlar arasına kısayol bağlantıları yerleştirmiştir. MobileNetV2 mimarisi Şekil 2.13’ te gösterilmektedir.



Şekil 2.13. MobileNetV2 mimari gösterimi

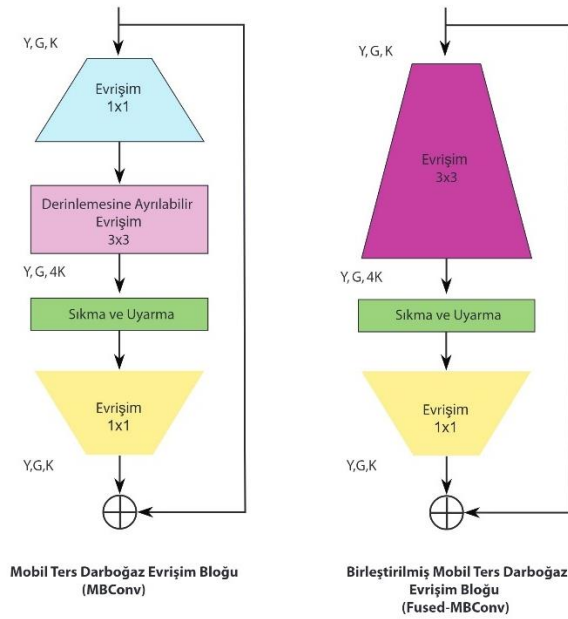
MobileNetV3 [38] otomatik sinir ağı arama algoritmalarıyla oluşturulmuş bir mimaridir. NAS kullanılarak gerçekleştirilen arama çalışmasında ödül fonksiyonu olarak doğruluk değerinin yanı sıra gecikme de dikkate alınmıştır. Ödül fonksiyonu doğruluk ve gecikmenin bir oranı olarak belirlenmiştir. MobileNetV3 mimarisi bir önceki versiyondan farklı olarak ters darboğaz bloklarıyla birlikte sıkma ve uyarma (squeeze and excitation)

yaklaşımından yararlanmaktadır. Sıkma bloğunda, girdi değerleri küresel ortalama havuzlama işleminden geçerek sayısal bir değere sıkıştırılmaktadır. Uyarma bloğunda tam bağlı iki katman bulunmaktadır. İlk katmanın nöron sayısı, filtre sayısı ve sıkma katsayısının çarpanı kadar azaltılmaktadır. Tam bağlı katmanlar arasında ReLU aktivasyon fonksiyonu ile doğrusal olmama durumu sağlanmaktadır. İkinci tam bağlı katmanın ardından daha yumuşak değerler elde etmek ve geçişi sağlamak için Sigmoid aktivasyonu kullanılmaktadır. MobileNetV3 mimarisinde Sigmoid yerine ReLu6 fonksiyonunun girdi değerini 3 artırarak ve çıkan sonucu 6'ya bölerek hesaplanan sert Sigmoid fonksiyonu kullanılmaktadır. Diğer yandan, MobileNetV3 mimarisinin daha sığ sayılabilecek ilk katmanlarında ReLU aktivasyon fonksiyonu kullanılırken, daha derin olarak nitelendirilen ileri katmanlarda sert Swish aktivasyonu kullanılmaktadır. Aktivasyon fonksiyonlarının “sert” versiyonlarının tercih edilmesinin nedeni standart aktivasyon fonksiyonlarından daha az hesaplama maliyeti gerektirmeleridir.

EfficientNet mimarileri

EfficientNet [39] derinlik-genişlik-çözünürlük ölçeklendirme parametreleri ile B0-7 aralığında sekiz farklı model versiyonu sunan bir CNN mimarisidir. Mimaride genişlik, derinlik, çözünürlük ve bırakma oranı olmak üzere beş farklı parametre bulunmaktadır. Bu parametreler belirli bir formül çerçevesinde sabitlenmiştir. Parametrelerden birinde yapılacak bir değişiklik tüm parametrelerin yeniden ölçeklendirilmesini gerektirmektedir. EfficientNet mimarisinde 7 adet Mobil Ters Darboğaz Evrişim Bloğu (MBConv) yer almaktadır. Birbirine bağlı bu bloklar derinlik katsayısına göre tekrar etmektedir. MBConv bloğunda 1x1 evrişimli genişleme katmanı, 3x3 derinlemesine evrişim katmanı, sıkıştırma ve uyarma katmanı ve 1x1 noktasal evrişim katmanı bulunmaktadır. Bloklardaki filtre sayısı genişleme oranı ile orantılı olarak artmaktadır. Modelde aktivasyon fonksiyonu olarak Swish aktivasyonu kullanılmaktadır. Sıkıştırma ve uyarma katmanlarında ise Sigmoid aktivasyonu bulunmaktadır. EfficientNetV2 [40], EfficientNet mimarisindeki sorunları çözerek hız ve parametre verimliliği sağlamak için geliştirilen ikinci versiyondur. EfficientNet’teki derinlemesine evrişimli katmanlar daha az parametreye sahip olmasına karşın oldukça yavaş çalışmaktadır. Yeni versiyonda MBConv ve Birleştirilmiş Mobil Ters Darboğaz Evrişim Bloğu (Fused-MBConv) kombinasyonu ile elde edilen bir mimari tasarlanmıştır. Şekil 2.14’te MBConv ve Fused-MBConv bloklarının tasarımı gösterilmektedir. Fused-MBConv bloğu genişletme katmanı ve derinliksel evrişim

katmanını birleştirerek tek bir 3x3 evrişim işlemi gerçekleştirmektedir. Fused-MBConv bloğunda parametre sayısında artış olmasına karşın hesaplama verimliği sağlanmaktadır. EfficientNetV2' de daha küçük bir genişleme oranı kullanılmaktadır. Ayrıca 5x5 evrişim filtreleri yerine 3x3 evrişim filtreleri kullanılmaktadır. EfficientNetV2-S modeli üzerine önceki versiyona benzer bir ölçeklendirme işlemi gerçekleştirilerek EfficientNetV2-M ve EfficientNetV2-L modelleri elde edilmektedir. Yeni versiyonda eğitim hızını artırmak için girdi boyutu 480x480 piksel ile sınırlandırılmaktadır.



Şekil 2.14. MBConv ve Fused-MBConv blok yapıları

2.4. Değerlendirme Metrikleri

Bu bölümde, modellerin başarısını değerlendirmek için kullanılan performans metriklerinden bahsedilmektedir. Ayrıca, orijinal ve çekişmeli görüntülerin kalite ve benzerlik değerlendirme metrikleri açıklanmaktadır.

2.4.1. Model performans değerlendirme metrikleri

Sınıflandırma modelleri iki veya daha fazla sınıfı tespit etmek için kullanılan algoritma yapılarıdır. Bu modellerin performanslarını ölçmek ve değerlendirmek için kullanılan doğruluk, kesinlik, F1 skoru gibi çeşitli metrikler bulunmaktadır. Bu metriklerin temelinde karmaşıklık matrisi denilen ölçüm tabloları yatmaktadır. Karmaşıklık matrisleri gerçek

değerler ile modelin ortaya koyduğu tahmini değerlerin çapraz olarak incelenmesine olanak tanıyan tablolardır. Şekil 2.15’ te iki sınıflı bir sınıflandırma işlemine ait karmaşıklık matrisi tablosu verilmektedir.

		Tahmin	
		Negatif	Pozitif
Gerçek	Negatif	GN	HP
	Pozitif	HN	GP

Şekil 2.15. Karmaşıklık matrisi

Karmaşıklık matrisi dört değer birleşiminden meydana gelmektedir. Bu değerler beyinde bir kitlenin olup olmadığını tespit eden bir yapay zeka örneği ile açıklanabilir.

- Gerçek Pozitif (GP): Kitle olarak sınıflandırılan dokunun gerçekte de kitle olduğu durumdur.
- Hatalı Pozitif (HP): Kitle olarak sınıflandırılan dokunun gerçekte sağlıklı olduğu durumdur.
- Gerçek Negatif (GN): Sağlıklı olarak sınıflandırılan bir dokunun gerçekte de sağlıklı olduğu durumdur.
- Hatalı Negatif (HN): Sağlıklı olarak sınıflandırılan bir dokunun gerçekte kitle dokusu olduğu durumdur.

Karmaşıklık matrisinde yer alan değerler ile hesaplanan Doğruluk (Accuracy), Kesinlik (Precision), Geri Çağırma (Recall), F1-Skoru, Özgüllük (Specificity/True Negative Rate) metrikleri de modellerin performanslarının değerlendirilmesinde kullanılmaktadır. Bu metriklerden kısaca bahsetmek gerekirse:

- Doğruluk: Modelin sınıflandırma performansı hakkında genel bilgiyi sağlayan metriktir.

$$\text{Doğruluk} = \frac{GP + GN}{GP + GN + HP + HN} \quad (2.3)$$

- Kesinlik: Gerçek pozitif olarak sınıflandırılan örneklerin tüm pozitif örneklerin ne kadarını oluşturduğunu gösteren metriktir.

$$\text{Kesinlik} = \frac{GP}{GP + HP} \quad (2.4)$$

- Geri Çağırma: Gerçek pozitif olarak sınıflandırılan örneklerin tüm gerçek pozitif örneklere oranını veren metriktir. Duyarlılık/ Gerçek Pozitif Oranı olarak da bilinmektedir.

$$\text{Geri Çağırma} = \frac{GP}{GP + HN} \quad (2.5)$$

- F1-Skoru: Geri Çağırma ve Kesinlik değerlerinin harmonik ortalaması ile elde edilen bir metriktir. Dengesiz veri setlerinin sınıflandırılmasında doğruluktan daha kullanışlı bir değerlendirme metriğidir.

$$F1 - \text{Skoru} = 2 * \frac{\text{Kesinlik} * \text{Geri Çağırma}}{\text{Kesinlik} + \text{Geri Çağırma}} \quad (2.6)$$

- Özgüllük: Gerçek negatif olarak sınıflandırılan örneklerin tüm gerçek negatif örneklere oranını veren metriktir. Gerçek Negatif Oranı olarak da bilinmektedir.

$$\text{Özgüllük} = \frac{GN}{GN + HP} \quad (2.7)$$

- Alıcı İşlem Karakteristik (ROC, Receiver Operating Characteristic) ve Eğrinin Altındaki Alan (AUC, Area Under the Curve): ROC eğrisi ikili sınıflamada sıklıkla tercih edilmektedir. ROC, gerçek pozitif oranı (geri çağırma) ve hatalı pozitif oranı (1 –

özgüllük) arasındaki ilişkiyi göstermektedir. Gerçek pozitif oranı y ekseninde, hatalı pozitif oranı x ekseninde konumlanmaktadır. İki değer dengeyi gözetmek için sınıflar arası ayrımı sağlayan bir kriter kullanılmaktadır. Bu kritere eşik değeri (threshold) denilmektedir. ROC eğrisini tek bir değer ile özetlemek için AUC metriği kullanılmaktadır. AUC, ROC eğrisinin altında kalan alana verilen addır. AUC değerinin 1'e ne kadar yakınsanması modelin sınıf ayrımını başarı ile gerçekleştirdiğini göstermektedir. AUC' un 0,5 olması ise modelin rastgele sınıflandırma yaptığını yani uydurduğunu göstermektedir.

2.4.2. Görüntü kalite ve benzerlik değerlendirme metrikleri

Farklı parametrelere sahip çekişmeli görüntüler ile orijinal görüntüler arasındaki kalite ve benzerliğin karşılaştırılması için çeşitli objektif ölçüm yöntemleri geliştirilmektedir. Bu metriklerden bir kısmı matematiksel hesaplamalar ile görüntü üzerindeki belli özellikleri, değişimleri, benzerlik ve hata oranlarını değerlendirmektedir. Ancak bu hesaplamalar insanın görsel algısını modellemede yeterince başarılı olmamaktadır. Bu nedenle daha karmaşık kalite ve benzerlik değerlendirmeleri için görüntünün kontrastı, renk uyumu ve görüntüler arasındaki yapısal benzerlikler gibi faktörleri dikkate alarak insanın görsel algısını modelleyebilen metrikler tercih edilmektedir. Bu çalışmada hata ölçüm metrikleri ve insan görsel algısına uyumlu benzerlik metrikleri birlikte kullanılmaktadır.

Ortalama karesel hata (MSE, Mean squared error)

MSE iki görüntünün piksel düzeyinde birbirine karşılık gelen değerlerinin farklarının karelerinin ortalaması ile hesaplanan bir hata ölçüm metriğidir. Basit bir matematiksel hesaplama olmasına karşın insanın görsel algısına duyarlı değildir. Görüntülerin pikselleri arasındaki bozulmayı hesaplarken düşük kontrastlı ve ayrıntıları değerlendirmek için yeterli olmayabilir. Bu nedenle iki görüntü arasındaki benzerliği değerlendirmek için diğer benzerlik ve kalite metrikleriyle birlikte kullanılmaktadır. MSE değeri 0' a yaklaştıkça görüntüler arasındaki bozulma oranı az olmaktadır. MSE metriğinin matematiksel formülü:

$$MSE(x, y) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [x(i, j) - y(i, j)]^2 \quad (2.8)$$

Burada x ve y görüntüleri, m ve n görüntülerin piksel bazlı yükseklik ve genişlik boyutlarını, $x(i, j)$ ve $y(i, j)$ ise (i, j) konumundaki pikselin değerini temsil etmektedir.

Kök ortalama karesel hata (RMSE, Root mean squared error)

RMSE, MSE' nin karekökü alınarak hesaplanan bir hata ölçüm metriğidir. Görüntüler arasında piksel düzeyinde birbirine karşılık gelen değerlerin ortalama karesel farkının standart sapmasını ifade etmektedir. Kolay yorumlanabilir bir metrik olmasına karşın insanın görsel algısına duyarlı değildir. Diğer görüntü kalite ve benzerlik metrikleri ile kullanılması daha kapsamlı bir değerlendirme sağlamaktadır. RMSE metriğinin matematiksel formülü:

$$RMSE = \sqrt{MSE} \quad (2.9)$$

Tepe sinyal-gürültü oranı (PSNR, Peak signal-to-noise ratio)

PSNR [41], görüntülerdeki piksel düzeyindeki farklılıkları dikkate alarak benzerlik hesaplayan bir metriktir. Maksimum sinyal değeri ile MSE değerlerini temel almaktadır. Maksimum sinyal değeri ile bozulmaya neden olan gürültü arasındaki oran hesaplanarak sinyal-gürültü ilişkisi değerlendirilmektedir. PSNR değeri genellikle dB ile ifade edilmektedir. PSNR değerinin yüksek olması, sinyal-gürültü oranının düşük olduğunu yani görüntüdeki hata oranının daha az olduğunu göstermektedir.

$$PSNR = 10 \cdot \log_{10} \left(\frac{Maksimum\ Sinyal\ Değeri^2}{MSE} \right) \quad (2.10)$$

Yapısal benzerlik indeksi (SSIM, Structural similarity index)

SSIM [41] iki görüntü arasındaki düşük ve yüksek seviyeli özelliklerin birlikte değerlendirilmesine olanak tanıyan benzerlik ve kalite metriğidir. İki görüntü arasındaki yapısal, kontrast ve luminans benzerliklerini değerlendiren algısal bir metriktir. Yapısal benzerliğin değerlendirilmesi için görüntülerin piksel yoğunlukları arasında istatistiksel

hesaplama gerçekleştirilmektedir. Luminans piksel değerlerinin ortalaması ile kontrast ise piksel değerlerinin standart sapması ile ölçülmektedir. Üç faktörün değeri birleştirilerek SSIM değeri elde edilir. SSIM değeri 0 ila 1 aralığında ondalık bir değer ile temsil edilmektedir. Değer 1' e yaklaştıkça görüntüler arasındaki benzerlik oranı yükselmektedir. SSIM değerinin matematiksel formülü:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.11)$$

Burada, μ_x ve μ_y değerleri x ve y görüntülerinin piksel değerlerinin ortalaması, σ_x^2 ve σ_y^2 , x ve y görüntülerinin piksel değerlerinin varyansını, σ_{xy} görüntülerin piksel değerlerinin kovaryansını temsil etmektedir. Sıfıra bölüm hatasını engellemek için c_1 ve c_2 düzenleme parametreleri kullanılmaktadır.

Çok ölçekli yapısal benzerlik indeksi (MS-SSIM, Multi-scale structural similarity index)

MS-SSIM [42], iki görüntü arasındaki yapısal benzerliği çok ölçekli olarak değerlendiren bir SSIM versiyonudur. Çok ölçekli durum ile anlatılmak istenen görüntülerin farklı çözünürlüklerde benzerliklerinin değerlendirilmesidir. Farklı çözünürlüklerdeki benzerlik düzeyleri farklı ayrıntılara sahiptir. MS-SSIM değerinin matematiksel formülü:

$$MS - SSIM(x, y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (2.12)$$

Burada, x ve y görselleri $\bar{x}$ ve $\bar{y}$ görüntülerin piksellerinin ortalama değerlerini, σ_x^2 ve σ_y^2 görsellerin piksellerinin varyanslarını, σ_{xy} ise görsellerin piksellerinin kovaryansını göstermektedir.

Piksel tabanlı görsel bilgi doğruluğu (VIFP, Pixel based visual information fidelity)

Görsel Bilgi Doğruluğu (VIF) [43] doğal sahne istatistiklerini (natural scene statistics) kullanarak görüntünün kalitesini değerlendirmeyi hedefleyen bir metriktir. Genel olarak görüntüler arasında paylaşılan bilginin ilişkisi dikkate alınmaktadır. VIF, Bilgi Doğruluğu

Kriteri (Information Fidelity Criterion) fikrine dayanmaktadır [44, 45]. Görüntüler belirli bölgelere ve bloklara ayrılarak özellikleri arasındaki benzerlik hesaplamaları gerçekleştirilmektedir. VIF değeri insan görsel algılamasını temsilen yapısal benzerlik, kontrast ve luminans değişimleri değerlendirilerek elde edilmektedir. VIFP ise VIF metriğinin piksel düzeyinde görüntünün yerel özelliklerinin değerlendirmesini sağlayan özelleştirilmiş bir versiyonudur. VIFP değeri 0 ila 1 aralığındandır ve 1'e yaklaştıkça görüntüler arasındaki benzerlik oranının yüksek olduğu kabul edilmektedir.

Evrensel kalite indeksi (UQI, Universal quality image index)

UQI [46] iki görüntü arasındaki yapısal benzerlikleri ve piksel düzeyindeki farklılıklarını ele alarak görüntünün kalitesini değerlendiren bir metriktir. Bir görüntüdeki bozulmayı belirlemek için korelasyon kaybı, luminans değişimi ve kontrast değişimi olmak üzere üç bileşen değerlendirilmektedir. Görüntülerin ortalama, varyans ve kovaryansı hesaplanarak bir kalite skoru elde edilmektedir. UQI değeri 1'e yaklaştıkça görüntüler arasındaki benzerlik oranı artmaktadır.

3. LİTERATÜR TARAMASI

Bu bölümde literatür araştırmaları, mevcut saldırı yöntemlerinin medikal görüntüler üzerindeki etkinliği, medikal görüntüleri hedef almak amacıyla geliştirilen saldırılar ve medikal görüntüleri saldırılara karşı korumak amacıyla kullanılan savunma yöntemleri olmak üzere üç başlık altında aktarılabacaktır. İlgili çalışmaların tespitinde Web Of Science veri tabanı kullanılmaktadır.

3.1. Mevcut Saldırıların Medikal Görüntüler Üzerindeki Etkinliğine İlişkin Yapılan Çalışmalar

Medikal görüntü analiz modellerine yönelik çekişmeli saldırıların varlığına dair ilk çalışma Paschali ve arkadaşları [47] tarafından gerçekleştirilmiştir. Çalışmada cilt lezyonlarının ön eğitilmiş modeller ile sınıflandırılmasında çekişmeli saldırıların etkinliği değerlendirilmiştir. Modellerin genelleştirme kabiliyeti ve sağlamlığı arasındaki ilişki ele alınmıştır. Finlayson ve arkadaşları [48] ise fundoskopi, dermoskopi ve röntgen görüntülerinin derin sinir ağları ile analizinde beyaz kutu ve kara kutu saldırılarının etkinliğini değerlendirmiştir. Shah ve arkadaşları [49] diyabetik retinopati teşhisinde kullanılan geleneksel evrimsel sinir ağlarının I-FGSM saldırısı karşısında savunmasız olduğunu ortaya koymuştur. Yoo ve Choi [50]'nin gerçekleştirdiği benzer bir araştırmada ise diyabetik retinopati teşhisinde kullanılan üç farklı ön eğitilmiş modelin FGSM saldırısına karşı hassasiyetlerini kıyaslamıştır. Anand ve arkadaşları [51] ise transfer öğrenme (TL) ve kendi kendine denetimli öğrenme (self-supervised learning) yöntemlerinin FGSM ve PGD beyaz kutu saldırılarına karşı sağlamlığını test etmiştir. Sınıflandırma ve segmentasyon görevleri üzerinde yapılan karşılaştırmada kendi kendine denetimli öğrenmenin daha sağlam özellikler öğrendiği görülmüştür. Kovalev ve arkadaşları [52], pulmonolojik ve onkolojik hastalıkların teşhisinde kullanılan dört görüntü modalitesinin ön eğitilmiş modeller ile sınıflandırılmasında C&W saldırısının etkinliğini değerlendirmiştir. InceptionV3 modelinin en hassas model olduğu ortaya konulmuştur. Yang ve arkadaşları [53], histopatoloji ve röntgen görüntüleri üzerine uygulanan HopSkipJump sınır tabanlı kara kutu saldırısı ile işaret aktivasyon ağı bozulmalarındaki artışın saldırıların tespitindeki önemini ve caydırıcılığını araştırmıştır. Allyn ve arkadaşları [54], derin öğrenme tabanlı dermoskopik tanılama sistemlerinde tüm görüntülerin melanositik nevüs olarak sınıflandırması için çekişmeli bir sinir ağı tasarlamıştır. Hedefli

saldırı sonucunda sınıflandırma başarısı %84'ten %67'ye düşmüştür. Gongye ve arkadaşları [55], COVID-19 teşhisinde FGSM ve PGD saldırılarının etkinliğini değerlendirmiştir. PGD saldırısının tüm görüntüleri yanlış sınıflandırdığı görülmüştür. Zhou ve arkadaşları [56], meme kanseri teşhisi için geliştirdiği VGG11 tabanlı AI-CAD sisteminin sağlamlığını değerlendirmek üzere bir dizi deney gerçekleştirmiştir. Üretici çekişmeli ağ ile üretilen örneklerin uzmanların deneyimleri doğrultusunda %29 ila %71 aralığında doğru sınıflandırıldığı görülmüştür. AI-CAD ise %30,9 doğru sınıflandırma gerçekleştirmiştir. Kovalev ve Voynov [14] görüntü modalitesinin, veri seti boyutunun ve kontrol parametreleri ile saldırı başarısı arasındaki ilişkiyi araştırmıştır. On üç veri seti üzerinde gerçekleştirilen sekiz sınıflandırma işlemi sonucunda pertürbasyon genliğindeki artışın saldırı başarısını artırdığı sonucuna ulaşılmıştır. Saldırının iterasyon sayısı da saldırının başarısını artıran bir diğer faktördür. Bortsova ve arkadaşları [57], kara kutu saldırısına karşı derin sinir ağlarının sağlamlığını etkileyen faktörler incelenmiştir. Pertürbasyon değerinin saldırı başarısı ve tespit edilebilirliği ile ilişkisi ortaya konulmuştur. Diğer yandan, ön eğitilmiş model başarısının aktarılabilirlikle doğrusal artış gösterdiği gözlemlenmiştir. Ma ve arkadaşları [10] fundoskopi, dermoskopi ve röntgen üzerine uygulanan FGSM, BIM, PGD ve C&W beyaz kutu saldırılarına karşı ResNet50 modelinin sağlamlığını değerlendirmiştir. Çalışmada medikal görüntülerin doğal görüntülerden daha savunmasız olduğu ve sınıf sayısındaki artışın saldırı başarısında rol oynadığı görülmüştür. Kong ve arkadaşları [58], yığın normalizasyon işleminin modellerin çekişmeli saldırılara karşı savunmasız olmasına neden olduğunu göstermiştir. Hirano ve arkadaşları [59], ön eğitilmiş derin öğrenme modellerinin hedefli ve hedefsiz Evrensel Çekişmeli Pertürbasyon (UAP) saldırılarına karşı direncini incelemiştir. UAP saldırılarının modeller üzerinde %80'den fazla başarı elde ettiği ve modeller arasında aktarılabilirliğinin sınırlı olduğu ortaya konulmuştur. Minagi ve arkadaşları [60], mevcut ön eğitilmiş derin sinir ağlarının doğal görüntüler ile oluşturulan UAP örnekleri ile kara kutu saldırıları gerçekleştirmiştir. Rastgele başlatmanın saldırı performansını düşürmesine karşın medikal teşhis sistemlerinde güvenilirliğinin sorgulanması önerilmiştir. Diyasa ve arkadaşları [61], kolanjokarsima sınıflandırmasında dokuz farklı saldırının etkinliğini değerlendirmiştir. Veri seti dengesizliğinin saldırı başarısına etkisinin değişken olduğu görülmüştür. Rodriguez ve arkadaşları [16], medikal görüntü analizinde kullanılan modellerin karmaşıklık düzeylerinin saldırı başarısındaki etkisi araştırılmıştır. Basit modellerin saldırılara karşı daha dirençli olduğu görülmüştür. Ancak çekişmeli eğitimin karmaşık modellerin direnç düzeylerini maksimize ettiği ortaya konulmuştur. Rahman ve arkadaşları

[62], IoT ile toplanan medikal veriler üzerinde dokuz farklı saldırı metodu test etmiştir. Bunun sonucunda COVID-19 verilerinin çekişmeli saldırılarında dahil olduğu çeşitli saldırılara karşı savunmasız olduğu sonucuna ulaşılmıştır. Li ve Liu [63] ise COVID-19 görüntü tespit sistemi üzerinde FGSM saldırısı uygulayarak sınıflandırma performansının %80' den %0'a düştüğünü göstermiştir. Aynı zamanda daha sağlam bir derin öğrenme tabanlı COVID-19 sınıflandırma modeli tasarlamak için mevcut savunma yöntemlerine bir bakış getirmiştir. Pal ve arkadaşları [15], COVID-19 tespitinde röntgen ve tomografi görüntülerinin sağlamlığını araştırmıştır. Tomografi görüntüleri daha kaliteli ve ayrıntılı görüntü sağlamasına karşın, her iki görüntü seti de küçük pertürbasyon değerine sahip bir FGSM saldırısına savunmasızdır. Joel ve arkadaşlarının [64] gradyan tabanlı FGSM, BIM ve PGD saldırıları kullanarak gerçekleştirdiği çalışmada, pertürbasyon değerinin, eşik değerinin üzerinde olduğu durumlarda çekişmeli örneğin tespitinin kolaylaştığı sonucuna ulaşılmıştır. Pertürbasyon değeri küçüldükçe tespit edilme olasılığı da düşmektedir. Pertürbasyon eşiği ise modelin eğitiminde kullanılan çekişmeli örneklerin üretildiği pertürbasyon boyutu ile ilişkilidir.

3.2. Medikal Görüntülere Yönelik Geliştirilen Çekişmeli Saldırıları

Byra ve arkadaşları [65], ultrason görüntü modalitesine özgü çekişmeli bir saldırı tasarlamıştır. Karaciğer yağlanması teşhisinde kullanılan ultrason görüntüleri RF sinyalleri ile yeniden yapılandırılmıştır. Zayıflanma katsayısı ve sıkıştırma eşiği üzerinde gerçekleştirilen modifikasyon neticesinde piksel yoğunluğunu hedef alan kara kutu çekişmeli saldırısı meydana gelmiştir. Mangaokar ve arkadaşları [66], CycleGAN stil aktarım yöntemine dayalı bir saldırı tasarlamıştır. Jekyll adı verilen saldırı fundoskopi ve röntgen görüntüleri üzerinde iki makine öğrenme odaklı algılama şeması ile araştırılmıştır. Bu şemalar kör tespit (blind detection) ve denetimli tespittir (supervised detection). Denetimli tespit şeması saldırganın orijinal ve sahte görüntü setlerine erişimi olduğunu varsayarken, kör tespit ise saldırganın görüntü setlerine veya eğitim modeline erişimi olmadığı varsayımına dayanmaktadır. Denetimli tespit şemasının saldırıya karşı daha umut verici sonuçlar elde ettiği görülmüştür. Yao ve arkadaşları [67], medikal görüntülerden üretilen çekişmeli örneklerin tespit edilmesinin doğal görüntülere kıyasla daha kolay olduğu fikrini savunmaktadır. Bu nedenle, medikal görüntülerdeki çekişmeli özellikleri gizlemek için Hiyerarşik Özellik Kısıtlaması (Hierarchical Feature Constraint) adı verilen bir yöntem geliştirmiştir. Bu yöntem beyaz kutu saldırılarına entegre edilerek

kullanılmaktadır. BIM, PGD ve C&W beyaz kutu saldırıları ile kıyaslandığında daha başarılı olduğu görülmüştür. Kulkarni ve Bhambani [68] önerilen bir ilgi alanına (region of interest) dayalı hedefsiz bir beyaz kutu saldırısı geliştirmiştir. Kryptonite adı verilen bu saldırının amacı saldırının etkinliğini yerelleştirerek iyileştirebilmektir. İlgi alanı, piksel yoğunluklarına bağlı olarak görüntüyü ikili forma getiren ve pikselleri ayıklayan OTSU Eşikleme yöntemi ile belirlenir. Böylece gürültü görüntünün belirli bir alanıyla sınırlandırılır. Alan içindeki piksel değişimlerini kontrol etmek için ise pertürbasyon değeri sınırlandırılır. Apostolidis ve Papakostas [69] kimlik doğrulama, içerik doğrulama ve görüntü entegrasyonu gibi uygulamalar ile verinin gizliliğinin, bütünlüğünün ve güvenliğinin sağlanmasında kullanılan filigran ekleme uygulamasının yaratacağı bozulmanın derin sinir ağlarındaki çekişmeli etkisini araştırmıştır. Bu çalışmanın sonucunda filigran ekleme saldırısı adı verilen kara kutu saldırı yöntemi önermiştir. Geliştirilen yöntem ile sınıflandırma başarısında dikkat çekici bir azalma olmuştur. Bu durum dijital filigran ekleme uygulamasının truva atı (trojan) potansiyeli taşıdığını göstermektedir. Koga ve Takemoto [70], küçük veri setleri kara kutu UAP saldırısı üretme fikrini ileri sürmüştür. Derin sinir ağı çıktılarını esas alan Tepe Tırmanma Araması (Hill Climbing Search) kullanarak hedefli ve hedefsiz saldırı yapabilen girdiden bağımsız bir saldırı tasarlamıştır. Saldırı sorgu verimli SimBA saldırısı temel almaktadır. Kara kutu saldırısının %40 ila %90 aralığında başarı gösterdiği ve beyaz kutu saldırılarından daha etkili olduğu belirtilmiştir. Maliamanis ve arkadaşları [71], görüntülere yapısal bir ekleme yapmadan gerçekleştirilen hedefsiz bir kara kutu saldırısı önermiştir. Mb-AdA adı verilen saldırı ortagonal görüntü momentlerine dayanır ve belirli görüntü bilgilerini kaldırarak modelin kararını etkilemeyi amaçlar. Momentlerin sırası ilerledikçe taşıdıkları özelliklerin anlamsal değeri artmaktadır. Bundan dolayı son sıralarda yer alan momentler kaldırılması tercih edilmektedir. Önerilen yöntemin sınıflandırma görevindeki performansı FGSM, PGD ve Kare saldırısı (Square attack) ile kıyaslandığında daha yüksek SSIM değeri elde ettiği görülmektedir.

3.3. Medikal Görüntü Analizinde Çekişmeli Saldırlara Karşı Kullanılan Savunma Yöntemleri

Vatian ve arkadaşları [72] iki soruya cevap aramaktadır. Bu sorulardan ilki orijinal medikal görüntü örnekleri analizde çelişkili örnek davranışı sergileyebilir mi? Diğer soru ise doğal görüntülerde kullanılan temel savunma yöntemleri medikal görüntüleri savunmak için

yeterli midir? Çekişmeli örneklerin eğitim modelinin türüne bağlı olarak farklı derecelerde tezahür edebilir. Ayrıca çekişmeli eğitimin ters edebileceği ortaya konulmuştur. Liu ve arkadaşları [73], akciğer tomografileri üzerinde nodül tespit sistemlerinde veri büyütme uygulamasının PGD saldırısının etkilerini hafifletebildiğini göstermiştir. Taghanaki ve arkadaşları [74], Mahalanobis mesafe fonksiyonu ile doğrusal olmayan radyal tabanlı evrişimsel özellik eşleme yöntemi kullanarak pertürbasyonların etkisini azaltan bir savunma yöntemi önermiştir. Yöntem aktivasyon fonksiyonu değiştirmekte ve böylece modelin karmaşıklığı artmadan doğrusallığından kaynaklı dezavantaj giderilmektedir. Xue ve arkadaşları [75], mevcut gürültü giderici ve CNN modelleriyle birleştirilebilen bir gömülü otomatik kodlayıcı geliştirerek modellerin sağlamlığını artırmayı hedeflemiştir. Geliştirilen gömülü otomatik kodlayıcı FGSM, I-FGSM ve C&W saldırılarına karşı CNN modellerinin direncini artırmaktadır. Daanouni ve arkadaşları [76] Sinir Yapısı Öğrenme (Neural Structure Learning) ve Çok Kafalı Dikkat (Multi-Head Attention) yapılarının kombinasyonundan oluşan NSL-MHA-CNN yöntemini sunmuştur. NSL-MHA-CNN modelin pertürbasyon üreten çekişmeli komşularla uçtan uca eğitimini esas alan savunma yöntemidir. Önerilen yöntem FGSM saldırılarına karşı ek eğitim maliyeti gerektirmeden model sağlamlığını iyileştirmektedir. Li ve Zhu [77], bir sınıflandırıcıda elde edilen özelliklerin gürültü kirliliğindeki artışın çekişmeli örneklerin tespitini kolaylaştırdığını göstermiştir. Li ve arkadaşları [78], yarı denetimli çekişmeli eğitim ve denetimsiz çekişmeli tespit küçük veri setleri ile gelişmiş çekişmeli saldırılara karşı savunma sağlamaktadır. Rizky ve Suyanto [79] çekişmeli eğitim ve derin k en yakın komşu yöntemlerinden oluşan hibrit bir savunma yaklaşımı önermiştir. Optimum k değeri bulunduğunda, fundoskopi görüntülerine karşı gerçekleştirilen FGSM, PGD ve BIM saldırılarına karşı hafifletme gerçekleştirilmiştir. Xu ve arkadaşları [80], önceden belirlenen sabit pertürbasyon değerlerinin çekişmeli eğitim sürecinde meydana getirdiği problemleri çözmek için çok pertürbasyonlu çekişmeli eğitim (MPAdvT, multi-perturbations adversarial training) önermiştir. Diğer taraftan, yanlış sınıflandırılan örneklerin model sağlamlığı üzerinde daha çok söz sahibi olduğu göz önüne alınarak, kayıp fonksiyonuna yanlış sınıflandırmaya duyarlı bir düzenlileştirme eklenmiştir. Bu yöntem ise yanlış sınıflandırma farkındalıklı çekişmeli eğitim (MAAdvT, misclassification-aware adversarial training) olarak adlandırılmıştır. Yu ve Zhou [81], uzamsal uyarlanabilirliği olan düşük geçişli filtreleri öğrenmek için geliştirilen Konuma Duyarlı Kenar Yumuşatma Filtreleri (Position-Aware Anti-Aliasing Filters) yaklaşımının çekişmeli saldırılara karşı modellerin direncini artırdığını göstermiştir. Chen ve arkadaşları

[82], budama ve dikkat mekanizmasına dayalı model tabanlı bir savunma yaklaşımı sunmaktadır. Dikkat katmanı ile modelin dikkat aralığı küçültürerek önemli bölgelere odaklanılmasını sağlamaktadır. Budama mekanizması ise çalışma [10]'de bahsedilen medikal görüntülerin biyolojik dokularından ve derin öğrenme modellerinin aşırı parametrelendirilmesinden kaynaklanan dezavantajlara çözüm sunmaktadır. Ma ve arkadaşları [10] ise çekişmeli örnekleri tespit etmek için çekirdek yoğunluğu (KD, kernel density), yerel içsel boyutsallık (LID, local intrinsic dimensionality), derin özellikler (DFeat, deep features) ve sayısal olarak belirlenmiş özellikler (QFeat, quantized features) olmak üzere dört dedektör sunmuştur. He ve arkadaşları [83], yüksek kaliteli hatalı negatif örnek üretmek için bir dedektör, bir üretici ve bir veri artırma çerçevesi önermiştir. Eğitim verilerinin dağılımı genişlemesi ve veriler arasındaki boşluğun artması sağlanarak yeniden eğitilen dedektörlerin sağlamlığının geliştirilmektedir. Ahmed ve arkadaşları [84], önerdikleri merkezi aktif öğrenme yöntemi (centroid active learning method) adı verilen denetimsiz savunma yaklaşımının FGSM saldırısına karşı direnci artırdığını göstermiştir. Yang ve arkadaşları [85], stokastik iniş işareti aktivasyon ağı (stochastic descent sign activation network) çekişmeli kara kutu saldırılara karşı yüksek direnç ve aktarılabilirlik sağladığını ortaya koymuştur. Sun ve arkadaşları [86], çok örnekli sağlam kendi kendine eğitim (MIRST, multi-instance robust self-training) adı verilen bir savunma yaklaşımı benimsemiştir. Bu yaklaşımda çok adımlı bir saldırı ile üretilen çekişmeli örnekler kullanılmaktadır. Ayrıca İlk havuzlama katmanı maksimum düşüş katmanı ile değiştirilerek eğitimin ilk adımında kararsız özellikler arındırılarak sağlamlık artırılmaktadır. MIRST yöntemi FGSM, PGD ve C&W saldırılarını hafifletmiştir. Kansal ve arkadaşları [87], çekişmeli eğitim sürecinde meydana gelen etiket sızıntısına karşı üst düzey temsil güdümlü gürültü giderici adı verilen bir filtre geliştirmiştir. Etiket sızıntısından kastedilen çekişmeli örneklerin sınıflandırma başarısı artarken orijinal görüntülerin sınıflandırma başarısının düşmesidir. Bu durum modellerin sağlamlığı karşısında güvensizlik meydana getirmektedir. Filtre beyaz kutu saldırılarına karşı başarılı olmuştur. Maliamanis ve arkadaşları [71] geliştirdiği Mb-AdA saldırısına karşı yine aynı saldırı ile üretilen çekişmeli örnekler ile eğitim verilerin artırılmasının model direncini artırdığını göstermektedir. Xu ve arkadaşları [88], ön eğitilmiş modellerin çıkarım süresine bağlı medikal sağlam ve yeniden eğitim gerektirmeyen teşhis (medical robust and retrain-less diagnostic) yöntemi geliştirmiştir. Orijinal test girdisi için birçok çekişmeli kopya oluşturularak etiketlenir. Bu kopyaların etiketlerine dayalı olarak çoğunluk oylaması gerçekleştirilir ve karar belirlenir. Shi ve arkadaşları [89], ön işleme ile görüntülerdeki

gürültülerin azaltılmasının CNN mimarilerinin direncini artıracak fikrine dayanarak seyreklik giderici operatörler (sparsity denoising operators) önermiştir. Bu operatör, CNN katmanları arasına konumlandırılarak örneğin içindeki ve örneklerin arasındaki gürültülerin giderilmesinden sorumludur. Operatörün beyaz kutu ve kara kutu saldırılarına karşı yüksek düzeyde başarı sağladığı görülmüştür.

4. KULLANILAN ARAÇLAR VE VERİ SETLERİ

Bu bölümde, tez çalışmasında kullanılan araçlar, yazılım kütüphaneleri, veri setleri ve veri setlerine uygulanan ön işlemler açıklanmaktadır.

4.1. Kullanılan Araçlar



Şekil 4.1. Kullanılan araçların logoları

Çalışmada Google Colaboratory geliştirme ortamı, Numpy matematik kütüphanesi, Scikit-learn makine öğrenme kütüphanesi, Tensorflow çerçevesi, Keras derin öğrenme kütüphanesi, Matplotlib ve Seaborn görselleştirme kütüphaneleri ve Sesar görüntü benzerlik karşılaştırma kütüphanesi kullanılmaktadır (Şekil 4.1).

4.1.1. Google Colaboratory

Google Colaboratory, bilinen adıyla Google Colab yapay zeka ve makine öğrenme projeleri geliştirmek üzere sunulmuş web tarayıcısı üzerinde kullanılan bulut tabanlı bir Python kodlama ortamıdır. Jupiter Notebook benzeri bir çalışma arayüzü sunmaktadır. Google Colab’ da birçok Python kütüphanesi ön yüklü gelmektedir. Diğer kütüphaneler linux komutları ile yüklenebilmektedir. Google Colab, Github ile entegre edilebilmekte ve ortak çalışma olanağı tanımaktadır. Google Cloud altyapısı ile çeşitli donanımsal özelliklere sahip merkezi işlem birimi (CPU), grafik işleme birimi (GPU) ve tensör işleme birimi (TPU) kullanıma sunulmaktadır. Ücretli ve ücretsiz versiyonları bulunmaktadır.

4.1.2. Numpy

Sayısal Python anlamına gelen “Numerical Python” ifadesinin kısaltılmasıyla oluşmuştur. Büyük matrisler ve çok boyutlu diziler gibi matematiksel işlemlerin gerçekleştirilmesinde yaygın kullanılan temel bir python kütüphanesidir. Numpy bir dizi arayüzü sunmaktadır. Numpy, dizileri Python listelerine kıyasla daha hızlı ve kolay hesaplama olanağı sunmaktadır. Bilimsel ve matematiksel hesaplamaların gerçekleştirilmesinde Tensorflow ve Theano gibi çeşitli kütüphaneler ile bütünleşmiş şekilde kullanılabilir.

4.1.3. Scikit-learn

Scikit-learn, sınıflandırma, regresyon ve kümeleme gibi makine öğrenme görevlerinde kullanılmak üzere tasarlanmış açık kaynaklı bir kütüphanedir. Python dili ile yazılmıştır. Birçok denetimli ve denetimsiz makine öğrenme algoritmasını desteklemektedir. Ayrıca veri ön işleme, model seçimi ve veri setinin veri bölünmesi için kullanılmaktadır.

4.1.4. Tensorflow ve Keras

Tensorflow, Google Beyin takımı tarafından geliştirilen açık kaynaklı bir yapay zeka çerçevesidir. Makine öğrenme ve derin öğrenme modellerinin geliştirilmesi ve eğitilmesi için sıklıkla tercih edilmektedir. Tensorflow, Python dışında R, Java, C++, Julia gibi birçok programlama dilini desteklemektedir. Tensorflow nispeten düşük seviyeli bir uygulama programlama arayüzüne (API) sahiptir ve doğrudan bir model geliştirmek zordur. Keras ise derin öğrenme modelleri tasarlamak için geliştirilen ve üst düzey API sunan bir kütüphanedir. Keras, 2017 yılında Tensorflow 2 üzerine entegre edilmiş ve Tensorflow’un çekirdek kütüphanesinde yerini almıştır. Böylece Tensorflow üzerinde derin öğrenme modellerinin geliştirilmesi için uygun bir ortam sağlanmıştır. Sade ve anlaşılabilir yapısı gereği geliştiriciler için Keras’ın öğrenilmesi ve uygulanması kolaydır. Tensorflow ve Keras Google Colab ortamında ön yüklü olarak gelmektedir.

4.1.5. Matplotlib ve Seaborn

Matplotlib ve Seaborn makine öğrenme projelerinde verilerin keşfini ve sunumunu kolaylaştırmak için sıklıkla tercih edilen veri görselleştirme kütüphaneleridir. Matplotlib

çizgi, dağılım ve çubuk grafikleri gibi düşük seviyeli görselleştirme işlemlerinin gerçekleştirildiği temel bir Python kütüphanesidir. Seaborn, Matplotlib kütüphanesi üzerine inşa edilmiş bir görselleştirme kütüphanesidir. Seaborn karmaşık verilerin görselleştirilmesi için yüksek seviyeli bir arayüz sunmaktadır. Matplotlib kütüphanesinden farklı olarak Pandas veri çerçevelerinin (dataframe) görselleştirilmesini desteklemektedir. Seaborn, Matplotlib'in yeteneklerini genişleten tamamlayıcı bir kütüphane olarak kullanılmaktadır.

4.1.6. Sewar

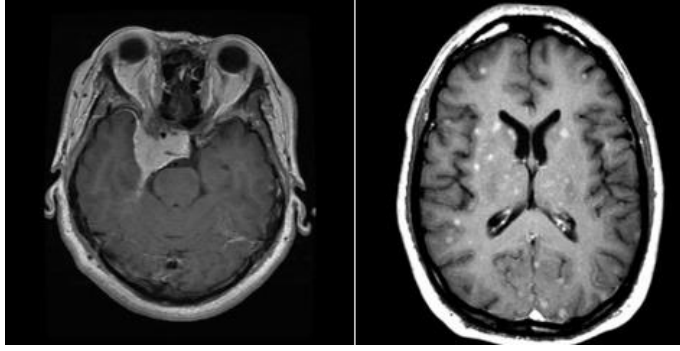
Sewar [90], iki görüntünün benzerlik oranını değerlendirmek için kullanılan bir Python kütüphanesidir. Yapısal benzerlik indeksi (SSIM), tepe sinyal-gürültü oranı (PSNR), kök ortalama kare hatasını (RMSE), piksel tabanlı görsel bilgi doğruluğunu (VIFP) kütüphane içerisinde bulunan metriklerden birkaçıdır. Google Colab üzerinde dahili bulunmamaktadır. Linux komutları ile yüklenerek kullanılmaktadır.

4.2. Veri Setleri

Bu çalışmada beyin tümör MR görüntülerini ve Böbrek BT görüntülerini içeren iki açık veri seti kullanılmaktadır.

4.2.1. Beyin tümör MR görüntü seti

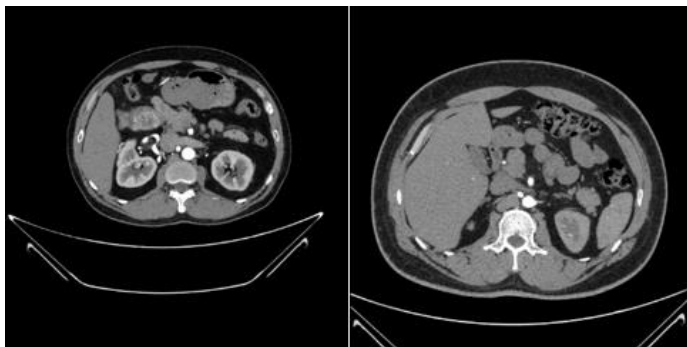
Beyin Tümör MR Veri Seti [91], açık erişimli bir Kaggle veri setidir. Veri seti figshare, SARTAJ veri seti ve Br35H görsellerinin birleşiminden meydana gelmektedir. Veri seti normal beyin görüntülerinin yanı sıra glioma, menenjiyom ve hipofiz tümör görüntüleri içermektedir. Veri setinde dört sınıfa ait toplam 7023 görüntü bulunmaktadır. Bu çalışmada, Beyin Tümör MR veri seti tümör ve normal olarak iki sınıflı bir problem olarak ele alınmıştır. Şekil 4.2' de örnek Beyin Tümör MR görüntüleri yer almaktadır.



Şekil 4.2. Beyin Tümör MR görüntü örnekleri (soldaki tümör, sağdaki normal)

4.2.2. Böbrek BT görüntü seti

Böbrek BT Veri Seti [92, 93], Bangladeş'in Dhaka kentinde bulunan çeşitli hastanelerden DICOM görüntüleri toplanarak oluşturulmuştur. DICOM görüntülerinden hastalara ait kişisel bilgiler ve meta verileri çıkarılmıştır. Ardından kayıpsız şekilde jpg formatına dönüştürülmüştür. Veri setinde yer alan tomografi görüntüleri koronal ve aksiyel kesimler, karın ve ürogram protokolü ile kontrastlı-kontrastsız çalışmalardan elde edilen görüntülerdir. Veri seti normal, kist, taş ve tümör olarak üzere dört sınıfa ait toplam 12 466 görüntü içermektedir. Bu çalışmada, Böbrek BT veri seti tümör ve normal olarak iki sınıflı bir problem olarak ele alınmıştır. Kist ve taş görüntüleri çalışmanın kapsamında değildir. Kullanılan veri alt setinde 7360 görüntü bulunmaktadır. Şekil 4.3' te örnek Böbrek BT görüntüleri yer almaktadır.



Şekil 4.3. Böbrek BT görüntü örnekleri (soldaki tümör, sağdaki normal)

4.2.3. Verilerin ön işleme

Her iki veri seti de aynı ön işlem aşamalarından geçmektedir. Farklı dosyalarda yer alan görüntüler bir ön işlem hattından geçerek tek bir dosyada toplanmaktadır. Ön işlem

sürecinde her bir görüntü RGB renk modunda 224x224 boyutu ile yüklenmektedir. Yüklenen görüntüler numerik olarak etiketlenmiş ve normalize edilmektedir. Görüntüler ve etiketleri birer numpy dizisine kaydedilmektedir. Görüntü ve etiketleri içeren numpy dizilerine bir karıştırma işlemi (shuffle) uygulanmaktadır. Bu sırada deneyin tekrarlanabilirliği sağlamak amacıyla rastgelelik parametresi (random_state) 42 olarak belirlenmektedir.

Veri setlerinde yer alan görüntü örneklerinin sınıf bazında dağılımı Çizelge 4.1' de görülmektedir. Bu verilerin %80'i eğitim aşamasında, %10'u doğrulama aşamasında ve geriye kalan %10'u test aşamasında kullanılmaktadır. Bölümlemede orantılı bir dağılım sağlamak amacıyla katmanlaştırma (stratify) dikkate alınmaktadır. Ayrıca deneyin tekrarlanabilirliği sağlamak amacıyla rastgelelik parametresi 42 olarak kullanılmaktadır.

Çizelge 4.1. Kullanılan görüntü setlerinin ikili sınıflandırmaya göre örnek dağılımları

Veri Seti	Sınıf	Örnek Sayısı	Toplam Örnek
Beyin Tümör MR veri seti	Normal (etiket:0)	2000	7023
	Tümör (etiket:1)	5023	
Böbrek BT veri seti	Normal (etiket:0)	5077	7360
	Tümör (etiket:1)	2283	

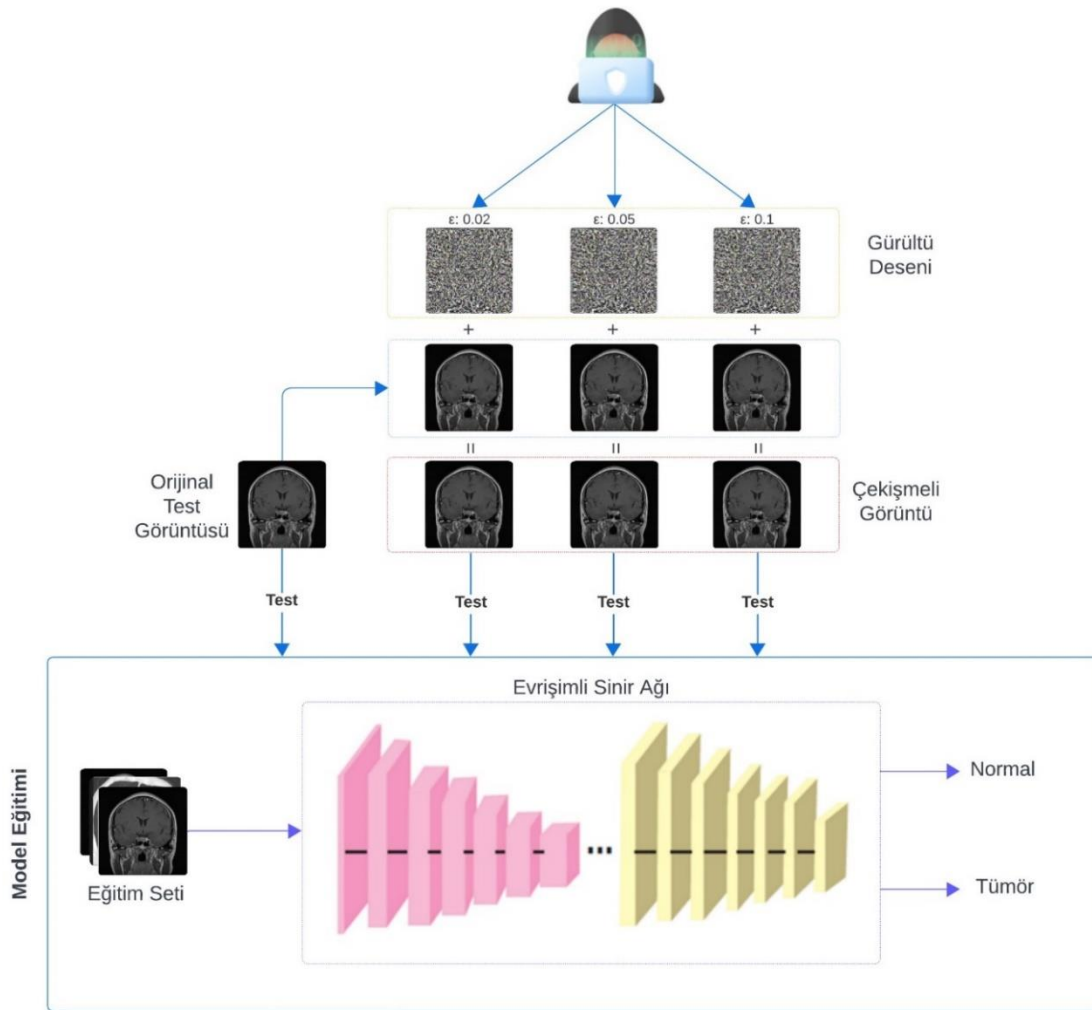
5. SINIFLANDIRMA MODELLERİNİN EĞİTİMİ

5.1. Deney Ortamı

Çalışmada tüm deneyler Google Colab ortamında gerçekleştirilmektedir. Deneyler Nvidia A100-sxm4 40GB vram'li grafik işlem birimi, 83,5GB sistem belleği ve 10 çekirdek sanal işlemcisi olan ortam üzerinde gerçekleştirilmektedir.

5.2. Deney Tasarımı

Tez çalışmasının daha anlaşılır olması için takip edilen adımlar açıklanmaktadır. Deneylerin uygulanması için tasarlanan aşamalar Şekil 5.1' de görülmektedir.

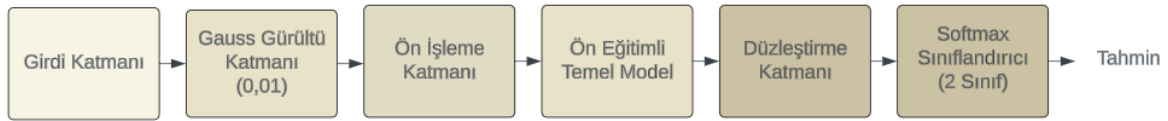


Şekil 5.1. Çalışma boyunca takip edilen deney prosesi

İlk olarak ön eğitilmiş CNN modellerinin eğitimi gerçekleştirilmektedir. Eğitim aşamasında eğitim veri setleri ile modelin eğitimi yapılırken doğrulama seti ile model performansı optimize edilmektedir. Modellerin eğitimi ilerleyen başlıklarda daha detayı ele alınmaktadır. İkinci aşama test aşamasıdır. Eğitilmiş modeller temiz test görüntüleri kullanılarak test edilmektedir. Bu aşama, modellerin temel başarısını değerlendirmek için önemlidir. Üçüncü aşamada FGSM ve PGD saldırıları ile üç farklı gürültü deseni oluşturulmaktadır. Oluşturulan gürültü desenlerinin temiz görüntülere eklenmesi ile çekişmeli örnekler üretilmektedir. Deneyin son aşamasında üretilen çekişmeli örnekler ön eğitilmiş modeller ile analiz edilmektedir. Elde edilen sonuçlar değerlendirilmektedir.

5.3. Hiper Parametre Değerlerinin Belirlenmesi

Görüntü işleme modelleri belirli boyutlarda girdileri kabul etmektedir. Şekil 5.2’ de verilen model tasarımı görüldüğü gibi, görüntü girdileri alındıktan sonra birtakım işlemlerden geçirilerek model eğitimi gerçekleştirilmektedir.



Şekil 5.2. Çalışmada kullanılan genel model tasarımı

Ön işleme katmanından önce bir Gauss gürültü katmanı yer almaktadır. Gauss gürültü katmanının gürültü miktarını belirten parametresi 0,01 olarak belirlenmiştir. Gauss gürültüsü, belirli ortalama ve varyansa sahip normal dağılımın örneklenmesiyle meydana gelmektedir. Modelin eğitim aşamasında görüntülere rastgele ve kontrollü bir gürültü eklenmesi, modellerin genelleştirme kabiliyetinin artmasını ve gerçek dünya problemine adaptasyonunu sağlamaktadır. Gauss gürültü katmanı, özellikle görüntü ve sinyal işleme çalışmalarında kullanılan sinir ağı modellerinin düzenlenmesi için yararlanılan bir yöntemdir. Gürültü katmanı, derin öğrenme katmanlarının önüne yerleştirilebileceği gibi gizli katmanlar arasında da kullanılabilir. Bu çalışmada Gauss katmanının hemen ardında ön işleme katmanı yer almaktadır. Modeller genellikle 244x244 boyutlarında görüntülerle çalışmasına karşın modelin büyüklüğünü ve tasarıma bağlı olarak değişiklik gösterebilmektedir. Örneğin, InceptionV3 modeli 299x299 boyutundaki görüntüler ile

çalışırken, NASNet ise 331x331 boyutlarında görüntülerle çalışmaktadır. Görüntülerin modellere uyumlanmasını sağlamak için ön işleme katmanı kullanılmaktadır.

Gürültü katmanının ardından ön eğitilmiş temel model yer almaktadır. Verilerin analizi için seçilen Keras transfer öğrenme modeli temel model olarak adlandırılmaktadır. Hiper parametre ayarlaması için rastgele bir temel model belirlenmelidir. Çalışmada hiper parametreleri belirlemek için rastgele seçilen model Xception'dır. Ancak hiper parametre seçiminin az sayıda deneme ile gerçekleştirilmesi ve seçilen modelle aşırı uyumdan kaçınılması esastır. Bunun nedeni, belirlenen hiper parametrelerin 20 farklı model üzerinde kullanılacak olmasıdır. Xception ile aşırı uyumlandırılmış parametreler diğer modellerle uyumsuz olabilir ve değerlendirme yanlışlığına neden olabilir. Ön eğitilmiş modelin ardından düzleştirme katmanı (flatten layer) ve 2 sınıflı bir softmax sınıflandırıcısı yer almaktadır. Modelde nöron aktivasyon fonksiyonu olarak ReLU, optimizasyon fonksiyonu olarak Adam ve kayıp fonksiyonu olarak seyrek kategorik çapraz entropi (sparse categorical cross entropy) kullanılmaktadır. Transfer öğrenme aşamasında öğrenme oranı 0,0001 olarak belirlenmiştir. Çalışmanın transfer öğrenme aşamasında örneklerin yığın büyüklüğü 64 olarak belirlenmektedir. Eğitim süresince doğrulama setinin performansı baz alınarak modelin erken durdurma değeri 20 epoch olarak belirlenmektedir. Modellerin eğitiminde esneklik sağlamak amacıyla toplam epoch değeri 999 olarak belirlenmektedir. Bu belirlenen katmanlar ve parametreleri diğer tüm modellere de aynı şekilde uygulanmaktadır.

5.4. İnce Ayarlı Transfer Öğrenme

Transfer öğrenme, modelin bir problem üzerinde elde ettiği kazanımların benzer başka bir problemde kullanılması fikrine dayanmaktadır. Katmanlar eğitime kapatılarak, daha önce ImageNet görüntülerinin sınıflandırılmasında elde edilen model ağırlıkları kullanılmaktadır. Modele ilave olarak eklenen katmanlar eğitilebilir. Transfer öğrenme aşamasında eğitimi tamamlanan ya da durdurulan modeller ince ayar aşamasına geçmektedir. İnce ayar aşamasında model katmanlarını eğitime açılarak doğruluk değerinin artırılması hedeflenmektedir. Çalışmada modellerin tüm katmanları eğitime açılmıştır. Ek olarak, 2 boyutlu (2B) evrişim katmanlarına Elastic Net düzenleme yöntemi uygulanmaktadır. Elastic Net, Lasso (L1) ve Ridge (L2) düzenleme yöntemlerinin birleşiminden meydana gelen toplam maliyet fonksiyonu içermektedir. Lasso,

seyreklik sağlayarak kat sayıları sıfıra doğru yakınlaştırmaktadır. Bu sayede gereksiz özelliklerin etkisi azalmakta ve öznitelik seçimi yapılmaktadır. Ridge ise kat sayıları küçülterek ağırlıklar arasında dengeli bir dağılım olmasını sağlamaktadır. Bu sayede büyük katsayıların etkisi azalmakta ve modelin genelleme kabiliyeti artmaktadır. Elastic Net, her iki yöntemin avantajlarını bir arada kullanır. Gürültülü veri setleri üzerinde gerçekleştirilen regresyon çalışmalarında yaygın kullanılan bir yöntemdir. Bu çalışmada Elastic Net ile ağırlık düzenlemesinde L1 katsayısı $1e-7$ ve L2 katsayısı $1e-6$ olarak belirlenmektedir. İnce ayar aşamasında öğrenme oranı 0,00001 olarak belirlenmektedir. Örnek yığını boyutu, epoch sayısı, erken durdurma, optimizasyon fonksiyonu, kayıp fonksiyonu ve aktivasyon fonksiyonu parametreleri transfer öğrenme aşaması ile aynıdır.

5.5. Modellerin Eğitimi

Modellerin eğitim süreçlerinin daha iyi anlaşılması için her iki veri seti ile gerçekleştirilen eğitim hakkındaki temel bilgiler derlenmektedir. Çizelge 5.1 ve Çizelge 5.2 modellerin parametre sayıları, katman sayısı, transfer öğrenmede geçen epoch sayısı, ince ayar nihayetinde ulaşılan son epoch sayısı, transfer öğrenme süresi, ince ayar süresi ve toplam geçen eğitim süresi gibi çeşitli bilgileri barındırmaktadır. İnce ayar son epoch değeri eğitimin tamamlandığında görülen son epochu temsil etmektedir. İnce ayar aşamasında gerçekleştirilen her bir epochun transfer öğrenme sürecinin üstüne ilave edildiği unutulmamalıdır. Örneğin, Xception modelinde transfer öğrenme aşamasında 90 epoch gerçekleştirilmiştir. İnce ayar 91. epoch'tan başlamış ve 133. epoch'ta tamamlanmıştır.

Çizelge 5.1' de Beyin Tümör MR verileri ile eğitilen 20 modelin eğitim bilgileri görülmektedir. Ayrıca modellerin transfer öğrenme ve ince ayar süreçlerinde elde ettikleri eğitim ve doğrulama performans grafikleri EK-1' de yer almaktadır.

Çizelge 5.1. Beyin Tümör MR görselleri ile eğitilen modellerin eğitimlerinin sayısal Verileri

Modeller	Eğitim Bilgileri							
	Temel Model Parametre Sayısı	Temel Model Katman Sayısı	Transfer Öğrenme Son Epoch	Transfer Öğrenme Süresi (sn)	İnce Ayar Son Epoch	İnce Ayar Süresi (sn)	Toplam Eğitim Süresi (sn)	Toplam Parametre Sayısı
Xception	20861480	132	90	627,06	133	673,52	1300,58	21062186
VGG16	14714688	19	155	747,27	214	639,80	1387,10	14764866
VGG19	20024384	22	90	493,17	151	767,59	1260,76	20074562
DenseNet121	7037504	427	88	557,00	131	942,92	1499,91	7137848
DenseNet169	12642880	595	50	395,11	82	937,69	1332,80	12805954
DenseNet201	18321984	707	49	463,75	79	1059,05	1522,80	18510146
EfficientNetV2B0	5919312	270	39	166,58	130	1189,70	1356,29	6044754
EfficientNetV2B1	6931124	334	55	254,84	148	1463,30	1718,14	7056566
EfficientNetV2B2	8769374	349	142	662,44	210	1139,27	1801,71	8907360
EfficientNetV2B3	12930622	409	88	503,57	136	953,27	1456,84	13081152
EfficientNetV2S	20331360	513	51	372,28	78	727,44	1099,72	20456802
InceptionResNetV2	54336736	780	75	675,46	114	1179,67	1855,13	54413438
InceptionV3	21802784	311	43	194,04	74	422,50	616,54	21905186
MobileNetV1	3228864	86	73	213,88	147	462,85	676,72	3329218
MobileNetV2	2257984	154	71	241,24	158	804,83	1046,07	2383426
MobileNetV3Small	939120	229	57	162,97	112	429,36	592,34	995570
NASNetMobile	4269716	769	94	643,72	143	1364,01	2007,73	4373206
ResNet50V2	23564800	190	41	233,67	73	402,04	635,71	23765506
ResNet101V2	42827266	377	32	280,93	64	667,94	948,86	42626560
ResNet152V2	58331648	564	31	365,03	60	850,46	1215,49	58532354

Çizelge 5.2’ de Böbrek BT verileri ile eğitilen 20 modelin eğitim bilgileri görülmektedir. Ayrıca modellerin transfer öğrenme ve ince ayar süreçlerinde elde ettikleri eğitim ve doğrulama performans grafikleri EK-2’ de yer almaktadır.

Çizelge 5.2. Böbrek BT görselleri ile eğitilen modellerin eğitimlerinin sayısal verileri

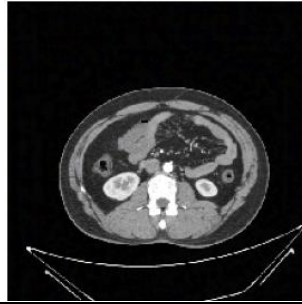
Modeller	Eğitim Bilgileri							
	Temel Model Parametre Sayısı	Temel Model Katman Sayısı	Transfer Öğrenme Son Epoch	Transfer Öğrenme Süresi (sn)	İnce Ayar Son Epoch	İnce Ayar Süresi (sn)	Toplam Eğitim Süresi (sn)	Toplam Parametre Sayısı
Xception	20861480	132	150	1059,61	173	390,32	1449,93	21062186
VGG16	14714688	19	69	344,08	94	294,40	638,48	14764866
VGG19	20024384	22	111	619,04	136	335,43	954,48	20074562
DenseNet121	7037504	427	51	323,40	74	511,34	834,74	7137848
DenseNet169	12642880	595	32	253,28	54	645,27	898,54	12805954
DenseNet201	18321984	707	31	306,34	52	763,06	1069,39	18510146
EfficientNetV2B0	5919312	270	259	1009,36	333	1015,08	2024,44	6044754
EfficientNetV2B1	6931124	334	189	824,14	242	839,13	1663,27	7056566
EfficientNetV2B2	8769374	349	94	440,42	127	574,92	1015,34	8907360
EfficientNetV2B3	12930622	409	164	904,26	196	647,32	1551,58	13081152
EfficientNetV2S	20331360	513	37	271,35	58	565,57	836,92	20456802
InceptionResNetV2	54336736	780	73	650,84	94	663,07	1313,91	54413438
InceptionV3	21802784	311	27	120,98	49	305,57	426,55	21905186
MobileNetV1	3228864	86	105	296,96	134	195,65	492,61	3329218
MobileNetV2	2257984	154	89	285,39	129	351,66	637,05	2383426
MobileNetV3Small	939120	229	114	278,64	146	242,23	520,87	995570
NASNetMobile	4269716	769	89	565,29	112	687,52	1252,81	4373206
ResNet50V2	23564800	190	27	142,73	48	266,06	408,79	23765506
ResNet101V2	42827266	377	23	192,06	44	448,68	640,73	42626560
ResNet152V2	58331648	564	24	280,02	45	646,70	926,72	58532354

5.6. Çekişmeli Test Örneklerinin Üretimi

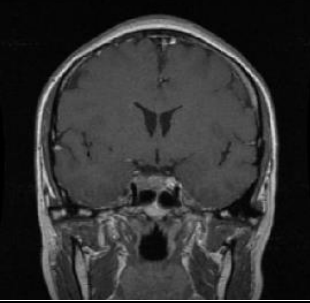
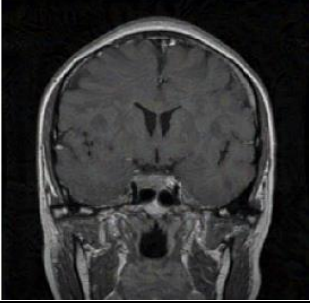
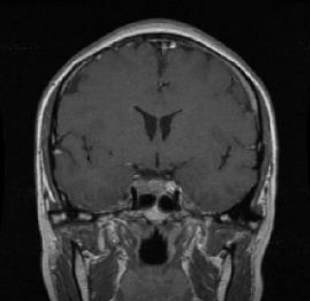
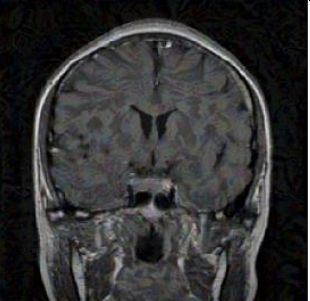
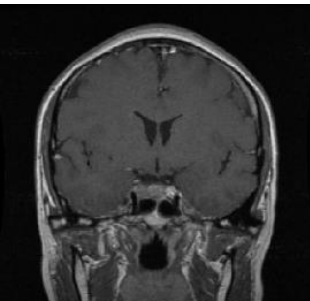
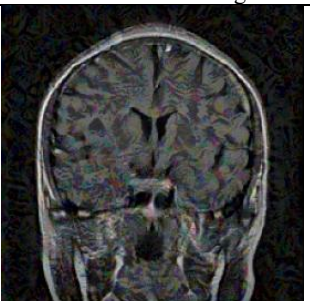
Çekişmeli örneklerin üretiminde FGSM ve PGD saldırıları kullanılmaktadır. FGSM ve PGD ile oluşturulan gürültü desenlerinin orijinal test örneğine eklenmesiyle çekişmeli örnek ortaya çıkmaktadır. Tek adımlı FGSM saldırısı ile gürültü deseni oluşturmak için doğrudan pertürbasyon düzeyi kullanılmaktadır. Çok adımlı bir yöntem olan PGD saldırısında ise pertürbasyon düzeyinin yanı sıra pertürbasyon adım büyüklüğü ve iterasyon sayısına ihtiyaç vardır. Çalışmada her iki saldırı için belirlen pertürbasyon düzeyleri 0,02; 0,05 ve 0,10'dur. PGD saldırısı için iterasyon sayısı 5 olarak belirlenmekte ve deney boyunca sabit tutulmaktadır. Bir diğer PGD parametresi olan pertürbasyon adım büyüklüğünün pertürbasyon düzeyinden küçük olması esastır. Bu nedenle pertürbasyon adım büyüklüğünün pertürbasyon düzeyinin yarısı olarak belirlenmektedir. Yani, 0,02 pertürbasyon ile yapılan deneylerde kullanılan pertürbasyon adım büyüklüğü 0,01'dir.

Pertürbasyon düzeyi 0,05 olan deneylerde kullanılan pertürbasyon adım büyüklüğü 0,025 ve 0,10 olan deneylerde kullanılan pertürbasyon adım büyüklüğü 0,05'tir. Çizelge 5.3' te bir Böbrek BT örneği üzerine FGSM uygulanması ile elde edilen üç farklı gürültü deseni ve üretilen çekişmeli örnek gösterilmektedir. Çizelge 5.4' te ise bir Beyin Tümör MR örneği üzerine PGD uygulanması ile elde edilen üç farklı gürültü deseni ve üretilen çekişmeli örnek gösterilmektedir.

Çizelge 5.3. Bir Böbrek BT görüntüsü üzerinde FGSM saldırısı uygulanması ile meydana gelen gürültü ve çekişmeli örnekler

Orijinal Örnek	Gürültü Deseni	Çekişmeli Örnek
Gerçek: Tümör Değil Tahmin: Tümör Değil	$\epsilon: 0,02$	Gerçek: Tümör Değil Tahmin: Tümör
		
Gerçek: Tümör Değil Tahmin: Tümör Değil	$\epsilon: 0,05$	Gerçek: Tümör Değil Tahmin: Tümör
		
Gerçek: Tümör Değil Tahmin: Tümör Değil	$\epsilon: 0,10$	Gerçek: Tümör Değil Tahmin: Tümör
		

Çizelge 5.4. Bir Beyin Tümör MR görüntüsü üzerinde PGD saldırısı uygulanması ile meydana gelen gürültü ve çekişmeli örnekler

Orijinal Örnek	Gürültü Deseni	Çekişmeli Örnek
Gerçek: Tümör Tahmin: Tümör	$\epsilon: 0,02$	Gerçek: Tümör Tahmin: Tümör Değil
		
Gerçek: Tümör Tahmin: Tümör	$\epsilon: 0,05$	Gerçek: Tümör Tahmin: Tümör Değil
		
Gerçek: Tümör Tahmin: Tümör	$\epsilon: 0,10$	Gerçek: Tümör Tahmin: Tümör Değil
		

Çekişmeli örneklerin üretiminde bilinmesi gereken bir detay bulunmaktadır. FGSM saldırısı ile tek seferde tüm test setinden çekişmeli örnek üretilmektedir. Ancak, PGD ile çekişmeli örnek üretmek için test seti yığınlara ayrılmaktadır. Beyin Tümör MR test setinin her bir yığında 13 örnek alınarak çekişmeli örnek üretilmektedir. Böbrek BT test setinin her bir yığında ise 23 örnek alınarak çekişmeli örnek üretilmektedir. Bu durumun nedeni PGD saldırısının çok adımlı olmasıdır. Kullanılan donanım kaynaklarının sınırları tüm test setine tek seferde PGD saldırısı uygulamak için yeterli değildir. Ancak bu durumun sonuçlara bir etkisi bulunmamaktadır. Deneyin son aşamasında üretilen çekişmeli örnekler ön eğitilmiş model ile analiz edilmektedir. Elde edilen sonuçlar değerlendirilmektedir.

6. DENEYSEL SONUÇLAR VE BULGULAR

6.1. Brain Tümör MR Görüntü Seti Deney Sonuçları

Bu bölümde, Beyin Tümör MR görüntü seti üzerinde gerçekleştirilen deney bulguları değerlendirilmektedir. Bulgular her bir derin öğrenme modeli için temiz test görüntüleri ile yapılan analiz, FGSM saldırısına uğrayan görüntülerin analizi ve PGD saldırısına uğrayan görüntülerin analizi olmak üzere üç kapsamda ele alınmaktadır. Çizelge 6.1’de temiz test görüntülerinin model bazında sınıflandırma başarıları yer almaktadır. Çizelge 6.2’ de farklı pertürbasyon parametreleri ile gerçekleştirilen FGSM saldırısı sonucunda modellerin sınıflandırma başarıları görülmektedir. Çizelge 6.3’ te ise farklı parametre senaryoları ile gerçekleştirilen PGD saldırısı sonucunda modellerin sınıflandırma başarıları verilmektedir.

Çizelge 6.1. Beyin Tümör MR test görüntüleri ile modellerin elde ettiği performans değerleri

Modeller	Orijinal - Temiz				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	99,00	99,40	99,20	99,30	98,5
VGG16	99,00	98,81	99,80	99,30	97,0
VGG19	98,86	99,01	99,40	99,21	97,5
DenseNet121	99,00	99,01	99,60	99,30	97,5
DenseNet169	98,72	98,62	99,60	99,11	96,5
DenseNet201	98,86	99,01	99,40	99,20	97,5
EfficientNetV2B0	98,86	99,01	99,40	99,21	97,5
EfficientNetV2B1	99,00	99,21	99,40	99,30	98,0
EfficientNetV2B2	99,00	98,82	99,80	99,31	97,0
EfficientNetV2B3	99,15	99,21	99,60	99,40	98,0
EfficientNetV2S	99,29	99,01	100,0	99,51	97,5
InceptionResNetV2	99,00	99,01	99,60	99,30	97,5
InceptionV3	98,86	98,81	99,60	99,21	97,0
MobileNet	98,15	98,23	99,20	98,71	95,5
MobileNetV2	98,15	99,20	98,21	98,70	98,0
MobileNetV3Small	97,29	99,19	97,01	98,09	98,0
NasNetMobile	99,00	98,63	100,0	99,31	96,5
ResNet50V2	99,72	99,60	100,0	99,80	99,0
ResNet101V2	99,15	99,01	99,80	99,41	97,5
ResNet152V2	99,15	99,01	99,80	99,41	97,5

Çizelge 6.2. Beyin Tümör MR görüntüleri üzerine uygulanan FGSM saldırıları sonucunda modellerin elde ettiği performans değerleri

Modeller	FGSM- 0,02 epsilon					FGSM - 0,05 epsilon					FGSM - 0,1 epsilon				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	93,45	96,34	94,42	95,37	91,0	62,96	79,09	65,64	71,68	56,5	26,06	46,72	24,10	31,80	31,00
VGG16	91,03	93,29	94,22	93,76	83,0	43,31	64,86	45,22	53,29	38,5	7,98	15,05	6,18	8,76	12,5
VGG19	82,76	90,28	85,06	87,59	77,0	28,92	50,58	26,30	34,60	35,5	5,70	7,90	2,99	4,34	12,5
DenseNet121	89,32	92,45	92,63	92,54	81,0	48,43	68,72	51,20	58,68	41,5	9,26	19,73	8,77	12,14	10,5
DenseNet169	94,02	96,00	95,62	95,81	90,0	62,97	78,27	66,73	72,04	53,5	24,64	45,26	25,70	32,78	22,0
DenseNet201	94,02	95,46	96,22	95,83	88,5	58,55	76,05	61,36	67,92	51,5	15,53	29,41	12,95	17,98	22,0
EfficientNetV2B0	76,64	87,72	78,29	82,74	72,5	19,94	35,98	15,34	21,51	31,5	1,57	1,54	0,60	0,86	4,0
EfficientNetV2B1	82,91	89,79	85,86	87,78	75,5	23,50	42,68	20,32	27,53	31,5	4,56	8,0	31,87	4,56	8,0
EfficientNetV2B2	84,33	89,36	88,65	89,00	73,5	25,21	45,46	22,91	30,46	31,0	2,14	3,05	1,20	1,72	4,5
EfficientNetV2B3	86,90	91,67	89,84	90,75	79,5	36,04	58,80	35,26	44,09	38,0	2,99	3,14	1,20	1,73	7,5
EfficientNetV2S	73,93	85,06	77,09	80,88	66,0	24,50	44,26	21,51	28,95	32,0	13,68	20,46	7,17	10,62	3,0
InceptionResNetV2	94,02	95,64	96,02	95,83	89,0	68,23	80,26	73,71	76,84	54,5	46,72	65,69	53,39	58,90	3,0
InceptionV3	93,16	95,04	95,42	95,23	87,5	71,94	81,71	78,29	79,96	56,0	48,58	67,94	53,19	59,67	37,0
MobileNet	80,48	85,71	87,25	86,48	63,5	25,07	45,39	23,51	30,97	29,0	4,84	5,85	2,19	3,19	11,5
MobileNetV2	81,62	89,94	83,67	86,69	76,5	29,92	51,71	30,08	38,04	29,5	6,41	11,82	4,78	6,81	10,5
MobileNetV3Small	51,71	75,39	48,21	58,81	60,5	13,11	18,97	6,57	9,76	29,5	3,13	0,55	0,20	0,29	10,5
NasNetMobile	84,47	85,92	93,63	89,61	61,5	43,59	63,95	48,41	55,10	31,5	8,26	14,85	5,98	8,52	14,0
ResNet50V2	91,31	95,46	92,23	93,82	89,0	65,10	80,96	66,93	73,28	60,5	34,47	57,5	32,07	41,18	40,5
ResNet101V2	90,03	92,86	93,23	93,04	82,0	66,24	78,74	72,31	75,39	51,0	43,88	64,52	47,81	54,92	34,0
ResNet152V2	89,60	91,98	93,63	92,79	79,5	67,09	76,11	78,69	77,38	38,0	49,15	66,0	59,56	62,62	23,0

Çizelge 6.3. Beyin Tümör MR görüntüleri üzerine uygulanan PGD saldırıları sonucunda modellerin elde ettiği performans değerleri

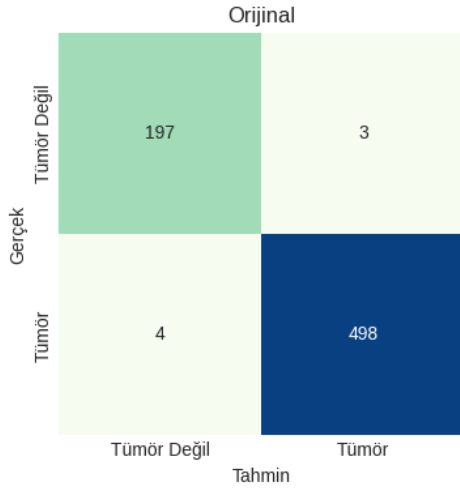
Modeller	PGD- 0,02 epsilon					PGD- 0,05 epsilon					PGD - 0,1 epsilon				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	59,83	78,35	60,56	68,32	58,0	4,42	5,76	2,19	3,18	10,0	1,28	0,0	0,0	-	4,5
VGG16	40,60	63,24	40,44	49,33	41,0	1,85	1,05	0,40	0,58	5,5	0,71	0,0	0,0	-	2,5
VGG19	22,94	41,01	17,73	24,76	36,0	1,28	0,0	0,0	-	4,5	0,14	0,0	0,0	-	0,5
DenseNet121	30,63	52,46	31,87	39,65	27,5	1,85	3,94	1,59	2,27	2,5	0,0	0,0	0,0	-	0,0
DenseNet169	45,73	67,44	46,61	55,12	43,5	2,99	4,57	1,79	2,58	6,0	0,0	0,0	0,0	-	0,0
DenseNet201	45,87	67,33	47,21	55,50	42,5	3,28	4,15	1,59	2,30	7,5	0,14	0,0	0,0	-	0,5
EfficientNetV2B0	47,29	68,64	48,41	56,78	44,5	2,85	3,61	1,39	2,01	6,5	0,0	0,0	0,0	-	0,0
EfficientNetV2B1	48,72	69,29	50,80	58,62	43,5	3,85	6,53	2,59	3,71	7,0	0,0	0,0	0,0	-	0,0
EfficientNetV2B2	56,41	75,13	58,37	65,70	51,5	4,70	7,18	2,79	4,02	9,5	0,0	0,0	0,0	-	0,0
EfficientNetV2B3	61,97	77,52	65,94	71,26	52,0	3,99	5,67	2,19	3,16	8,5	0,0	0,0	0,0	-	0,0
EfficientNetV2S	7,85	9,95	3,59	5,27	18,5	0,14	0,0	0,0	-	0,5	0,0	0,0	0,0	-	0,0
InceptionResNetV2	59,69	75,53	64,64	69,60	47,5	7,41	15,74	6,77	9,47	9,0	0,86	1,50	0,60	0,86	1,5
InceptionV3	55,27	71,76	61,75	66,38	39,0	7,41	14,12	5,98	8,45	11,0	1,71	2,53	1,0	1,43	3,5
MobileNet	19,66	35,51	15,14	21,23	31,0	0,14	0,0	0,0	-	0,5	0,0	0,0	0,0	-	0,0
MobileNetV2	31,34	53,79	28,29	37,08	39,0	1,0	0,0	0,0	-	3,5	0,0	0,0	0,0	-	0,0
MobileNetV3Small	11,82	7,91	2,19	3,43	36,0	0,71	0,0	0,0	-	2,5	0,0	0,0	0,0	-	0,0
NasNetMobile	43,02	63,01	49,20	55,26	27,5	1,28	0,52	0,20	0,29	4,0	0,14	0,0	0,0	-	0,5
ResNet50V2	24,22	43,80	21,12	28,50	32,0	1,0	1,99	0,80	1,14	1,5	0,0	0,0	0,0	-	0,0
ResNet101V2	29,63	51,27	32,07	39,46	23,5	3,42	8,10	3,39	4,76	3,5	0,71	1,49	0,60	0,85	1,0
ResNet152V2	34,76	55,95	41,24	47,48	18,5	3,42	10,0	4,38	6,09	1,0	0,0	0,0	0,0	-	0,0

Çizelge 6.1, Çizelge 6.2 ve Çizelge 6.3’ te görülen ve sınıflandırma başarısını temsil eden performans metrik değerleri bölümün devamında yer alan ilgili derin öğrenme modelinin altında detaylı olarak incelenmektedir.

6.1.1. Xception

Temiz

Xception modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.1’de görülmektedir.

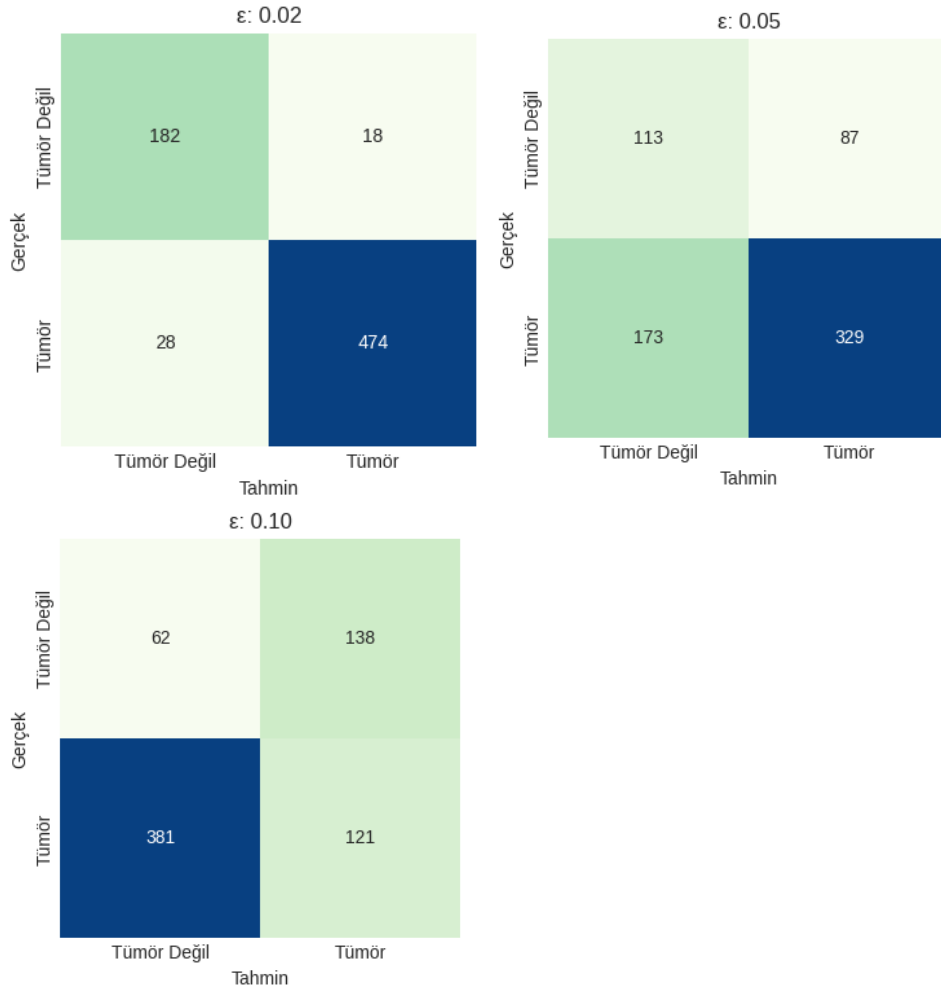


Şekil 6.1. Temiz Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.1’ de görüldüğü gibi temiz test setinin Xception modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 4 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 3 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %99,4 kesinlik, %99,2 geri çağırma, %99,3 F1 skoru ve %98,5 özgüllük elde etmektedir.

FGSM saldırısı

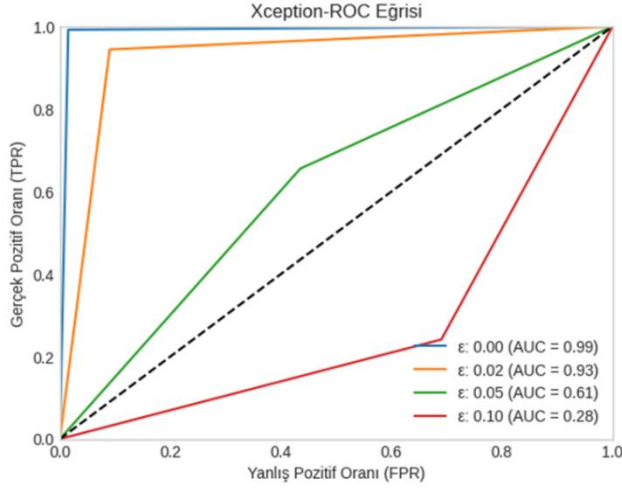
Şekil 6.2’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.2. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.2’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 28 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 18 örnek bulunmaktadır. Çizelge 6.2’ ye göre %93,45 doğruluk, %96,34 kesinlik, %94,42 geri çağırma, %95,37 F1 skoru ve %91 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 173 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 87 örnek bulunmaktadır. Çizelge 6.2’ ye göre %62,96 doğruluk, %79,09 kesinlik, %65,64 geri çağırma, %71,68 F1 skoru ve %56,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 381 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 138 örnek bulunmaktadır. Çizelge 6.2’ ye göre %26,06 doğruluk, %46,72 kesinlik, %24,1 geri

çağırma, %31,8 F1 skoru ve %31 özgüllük elde etmektedir. Şekil 6.3'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

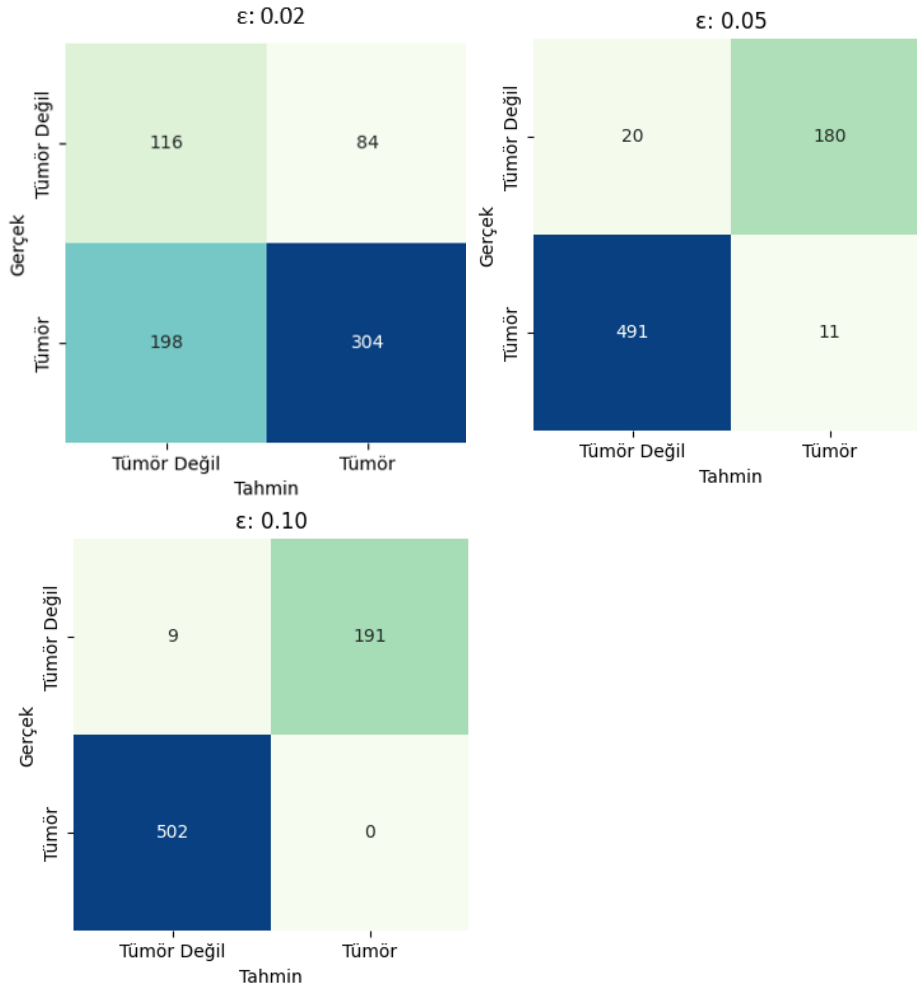


Şekil 6.3. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait ROC eğrileri

Şekil 6.3' e göre Xception modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %28'e kadar düşmektedir.

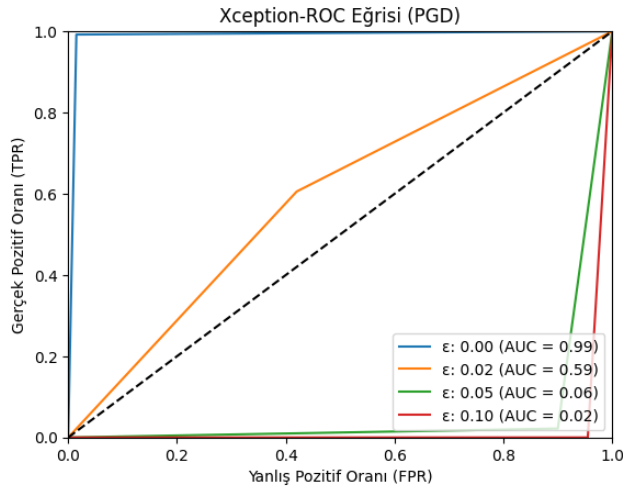
PGD saldırısı

Şekil 6.4' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.4. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.4' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 198 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 84 örnek bulunmaktadır. Çizelge 6.3' e göre %59,83 doğruluk, %78,35 kesinlik, %60,56 geri çağırma, %68,31 F1 skoru ve %58 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 491 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 180 örnek bulunmaktadır. Çizelge 6.3' e göre %4,42 doğruluk, %5,76 kesinlik, %2,19 geri çağırma, %3,18 F1 skoru ve %10 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 191 örnek bulunmaktadır. Çizelge 6.3' e göre %1,28 doğruluk ve %4,5 özgüllük elde etmektedir. Şekil 6.5'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



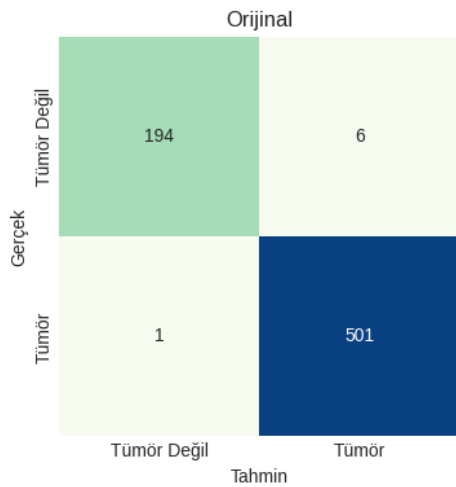
Şekil 6.5. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde Xception modeline ait ROC eğrileri

Şekil 6.5' e göre Xception modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %2'ye kadar düşmektedir.

6.1.2. VGG16

Temiz

VGG16 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.6'da görülmektedir.

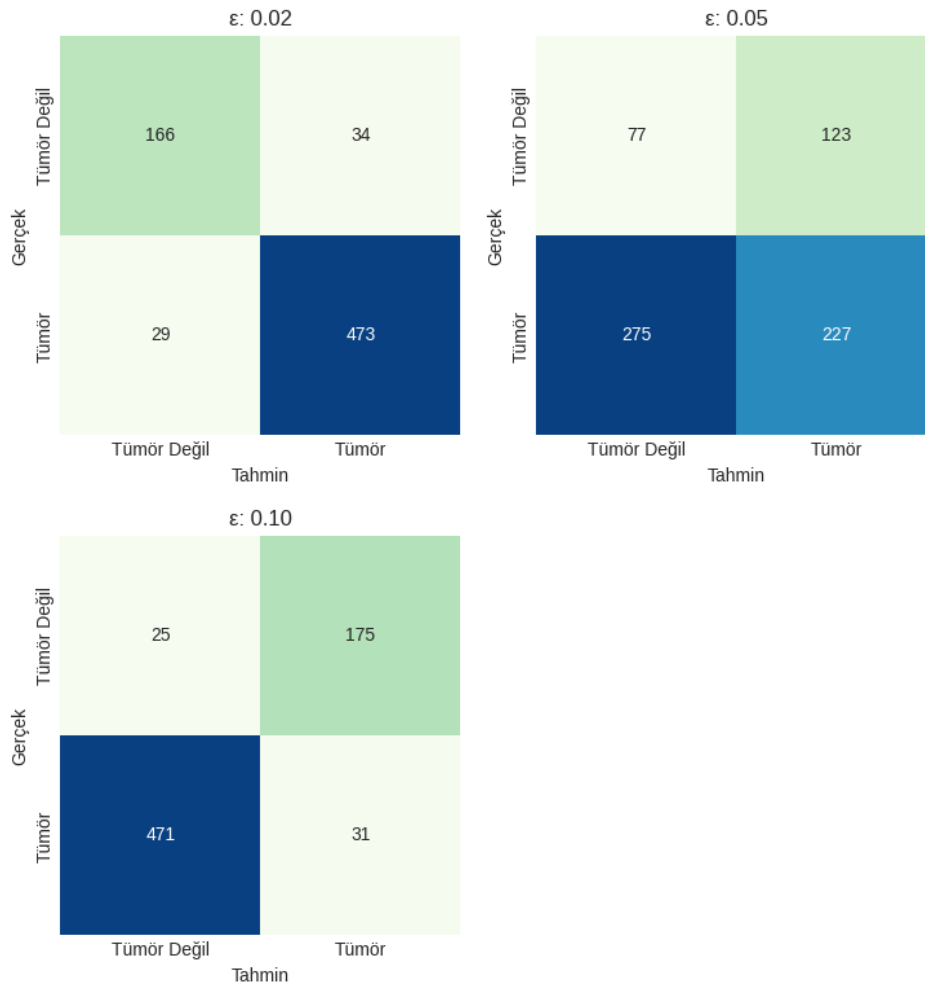


Şekil 6.6. Temiz Beyin Tümör MR görüntüleri üzerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.6’ da görüldüğü gibi temiz test setinin VGG16 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 1 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 6 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %98,81kesinlik, %99,8 geri çağırma, %99,3 F1 skoru ve %97 özgüllük elde etmektedir.

FGSM saldırısı

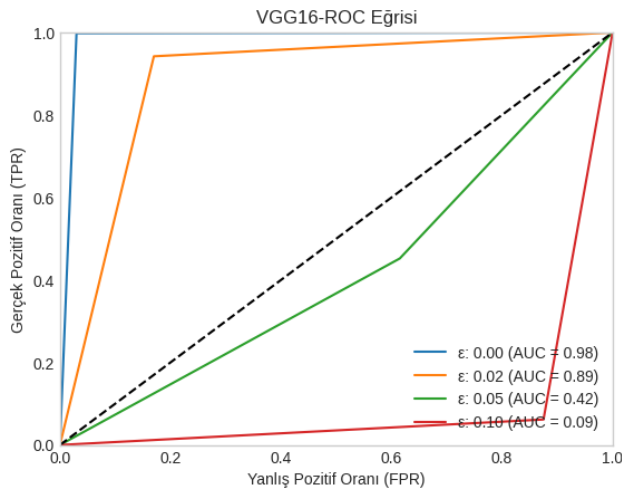
Şekil 6.7’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.7. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.7’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 29

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 34 örnek bulunmaktadır. Çizelge 6.2' ye göre %91,03 doğruluk, %93,29 kesinlik, %94,22 geri çağırma, %93,76 F1 skoru ve %83 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 275 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 123 örnek bulunmaktadır. Çizelge 6.2' ye göre %43,31 doğruluk, %64,86 kesinlik, %45,22 geri çağırma, %53,29 F1 skoru ve %38,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 471 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 175 örnek bulunmaktadır. Çizelge 6.2' ye göre %7,98 doğruluk, %15,05 kesinlik, %6,18 geri çağırma, %8,76 F1 skoru ve %12,5 özgüllük elde etmektedir. Şekil 6.8'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

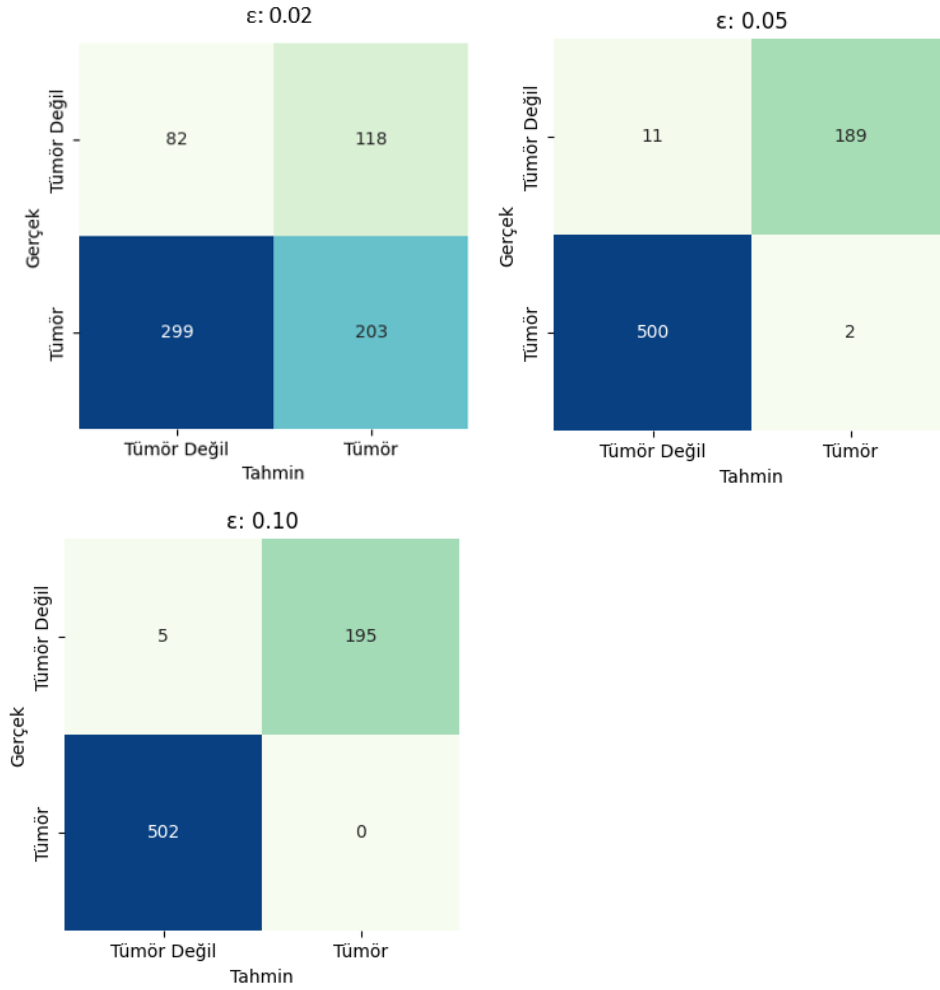


Şekil 6.8. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait ROC eğrileri

Şekil 6.8' e göre VGG16 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %9'a kadar düşmektedir.

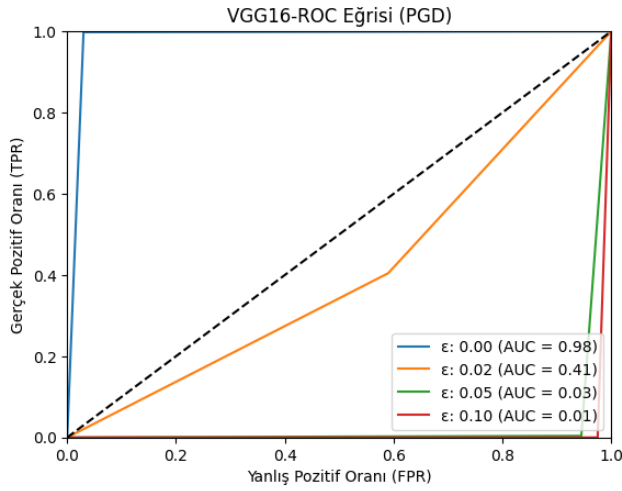
PGD saldırısı

Şekil 6.9' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.9. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.9’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 299 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 118 örnek bulunmaktadır. Çizelge 6.3’ e göre %40,6 doğruluk, %63,24 kesinlik, %40,44 geri çağırma, %49,33 F1 skoru ve %41 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 500 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 189 örnek bulunmaktadır. Çizelge 6.3’ e göre %1,85 doğruluk, %1,05 kesinlik, %0,4 geri çağırma, %0,58 F1 skoru ve %5,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 195 örnek bulunmaktadır. Çizelge 6.3’ e göre %0,71 doğruluk ve %2,5 özgüllük elde etmektedir. Şekil 6.10’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



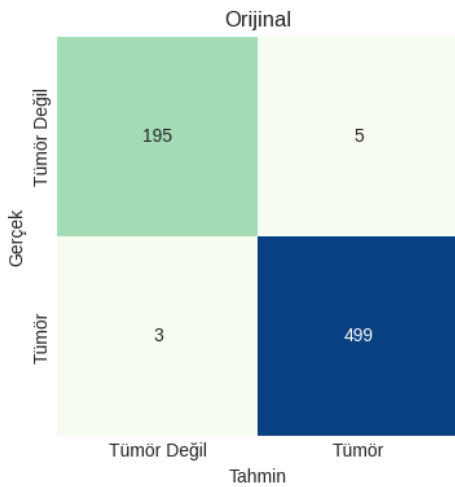
Şekil 6.10. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG16 modeline ait ROC eğrileri

Şekil 6.10' a göre VGG16 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %1'e kadar düşmektedir.

6.1.3. VGG19

Temiz

VGG19 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.11'de görülmektedir.

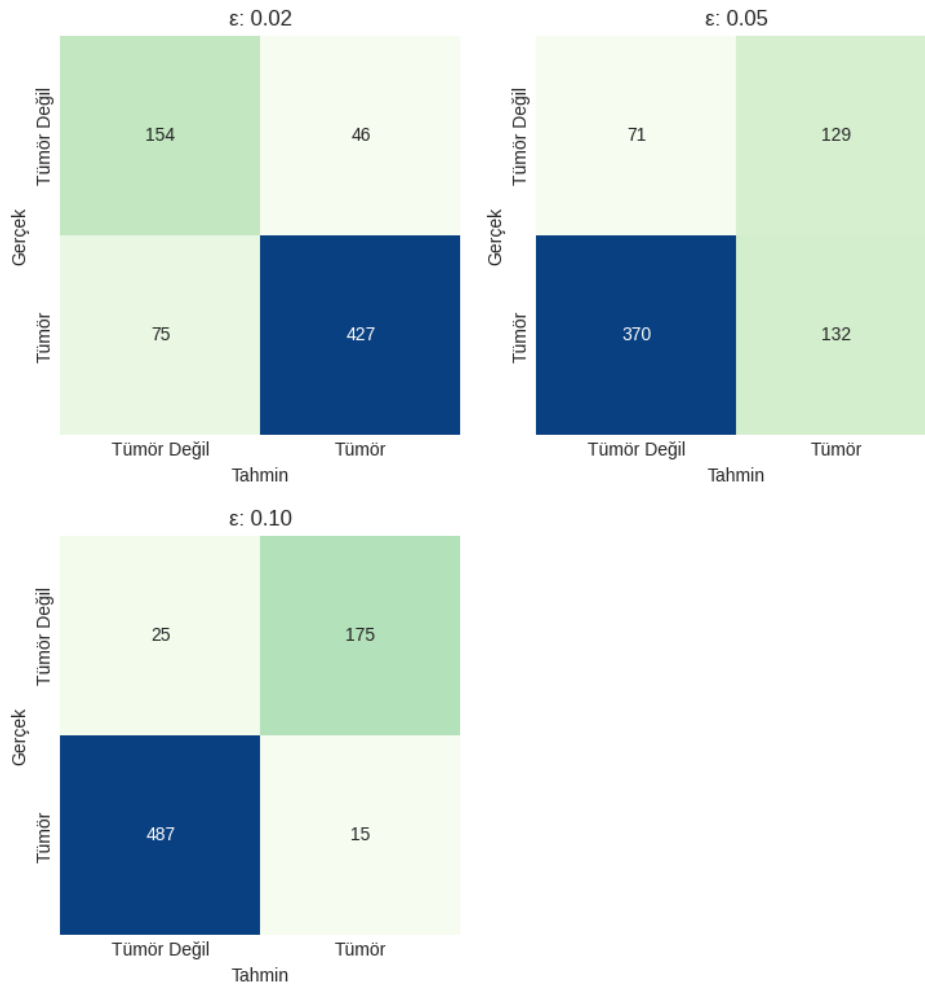


Şekil 6.11. Temiz Beyin Tümör MR görüntüleri üzerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.11’ de görüldüğü gibi temiz test setinin VGG19 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 3 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,86 doğruluk, %99,01 kesinlik, %99,4 geri çağırma, %99,21 F1 skoru ve %97,5 özgüllük elde etmektedir

FGSM saldırısı

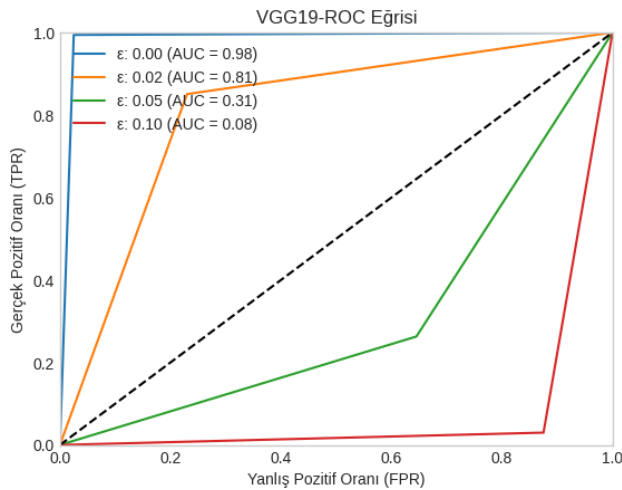
Şekil 6.12’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.12. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.12’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 75

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 46 örnek bulunmaktadır. Çizelge 6.2' ye göre %82,76 doğruluk, %90,28 kesinlik, %85,06 geri çağırma, %87,59 F1 skoru ve %77 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 370 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 129 örnek bulunmaktadır. Çizelge 6.2' ye göre %28,92 doğruluk, %50,58 kesinlik, %26,3 geri çağırma, %34,6 F1 skoru ve %35,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 487 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 175 örnek bulunmaktadır. Çizelge 6.2' ye göre %5,7 doğruluk, %7,9 kesinlik, %2,99 geri çağırma, %4,34 F1 skoru ve %12,5 özgüllük elde etmektedir. Şekil 6.13'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

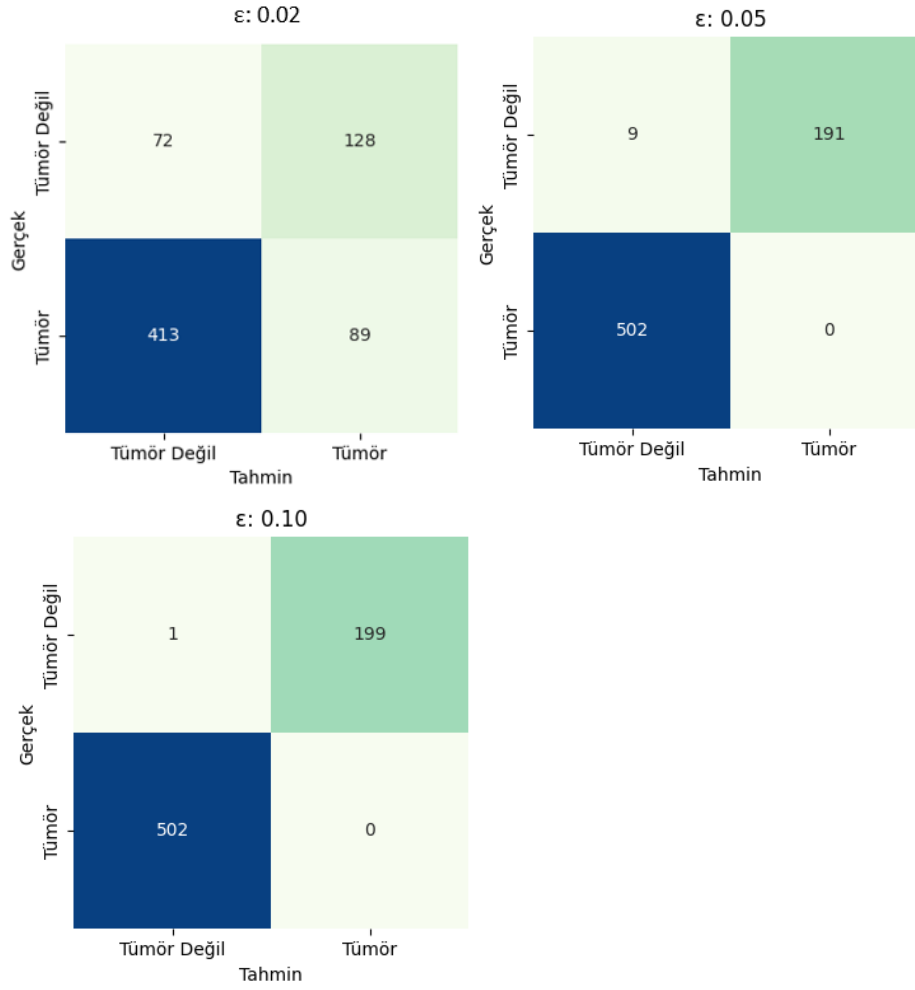


Şekil 6.13. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait ROC eğrileri

Şekil 6.13' e göre VGG19 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' dan %8'e kadar düşmektedir.

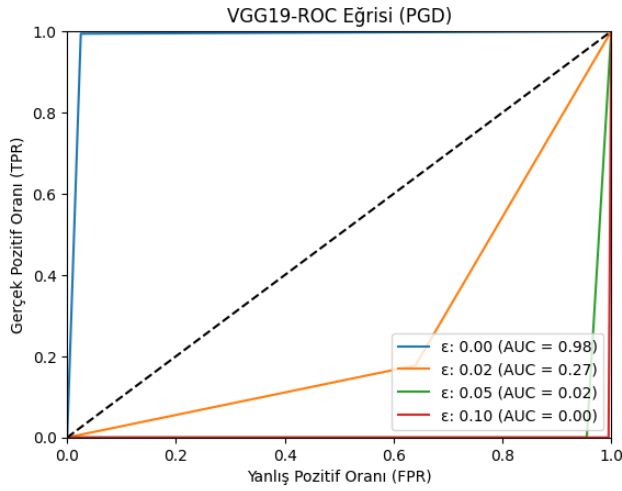
PGD saldırısı

Şekil 6.14' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.14. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.14’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 413 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 128 örnek bulunmaktadır. Çizelge 6.3’ e göre %22,94 doğruluk, %41,01 kesinlik, %17,73 geri çağırma, %24,76 F1 skoru ve %36 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 173 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 87 örnek bulunmaktadır. Çizelge 6.3’ e göre %1,28 doğruluk ve %4,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 199 örnek bulunmaktadır. Çizelge 6.3’ e göre %0,14 doğruluk ve %0,5 özgüllük elde etmektedir. Şekil 6.15’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



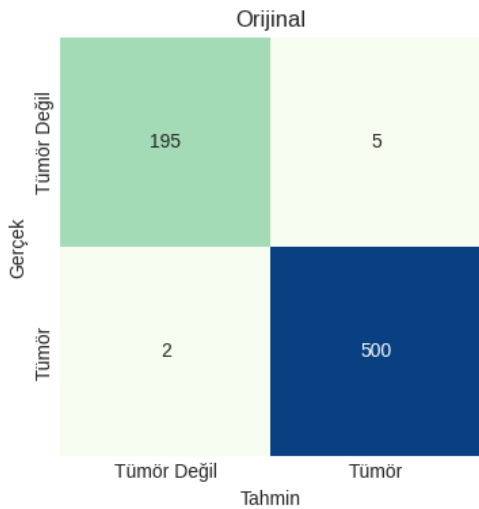
Şekil 6.15. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde VGG19 modeline ait ROC eğrileri

Şekil 6.15' e göre VGG19 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.4. DenseNet121

Temiz

DenseNet121 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.16'da görülmektedir.

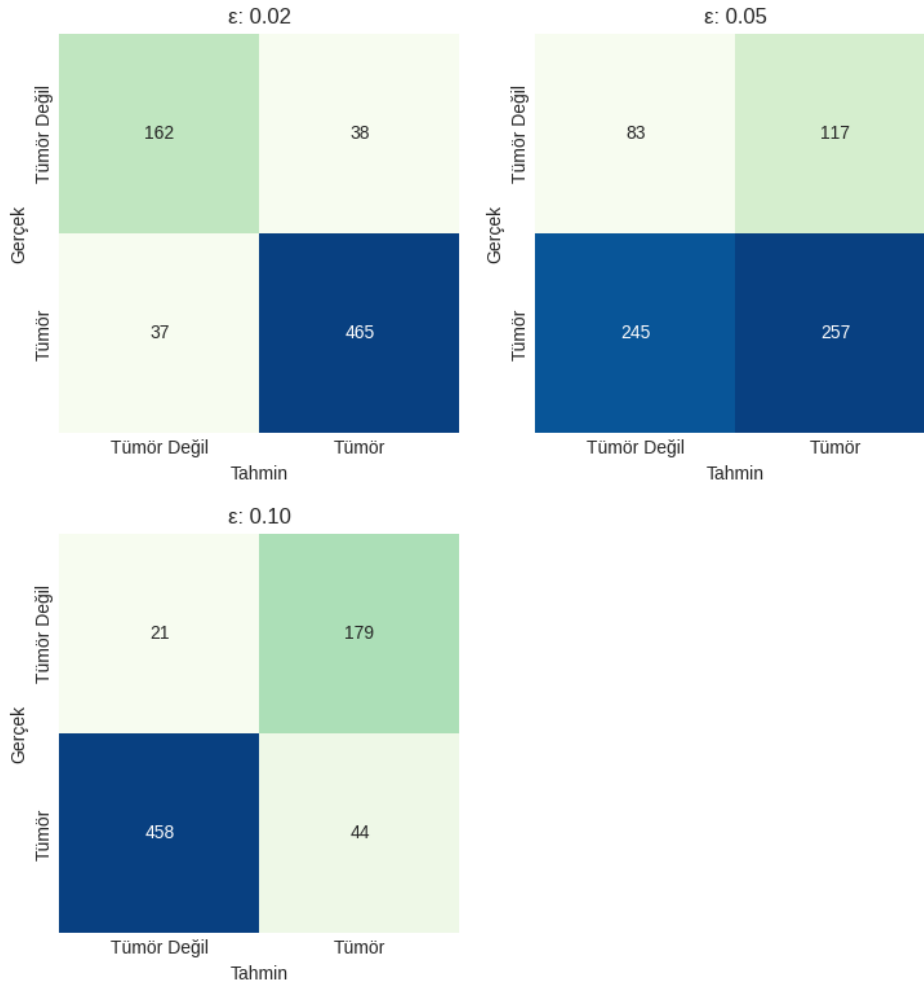


Şekil 6.16. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.16' da görüldüğü gibi temiz test setinin DenseNet121 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 2 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1' e göre %99 doğruluk, %99,01 kesinlik, %99,6 geri çağırma, %99,3 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

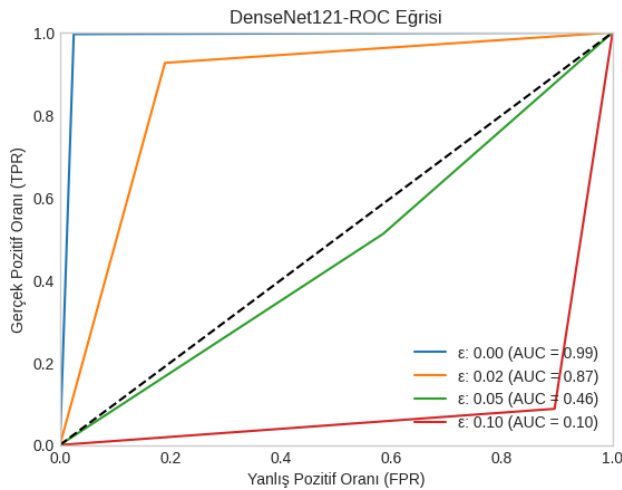
Şekil 6.17' de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.17. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.17' de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 37

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 38 örnek bulunmaktadır. Çizelge 6.2' ye göre %89,32 doğruluk, %92,45 kesinlik, %92,63 geri çağırma, %92,54 F1 skoru ve %81 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 245 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 117 örnek bulunmaktadır. Çizelge 6.2' ye göre %48,43 doğruluk, %68,72 kesinlik, %51,2 geri çağırma, %58,68 F1 skoru ve %41,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 458 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 179 örnek bulunmaktadır. Çizelge 6.2' ye göre %9,26 doğruluk, %19,73 kesinlik, %8,77 geri çağırma, %12,14 F1 skoru ve %10,5 özgüllük elde etmektedir. Şekil 6.18'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

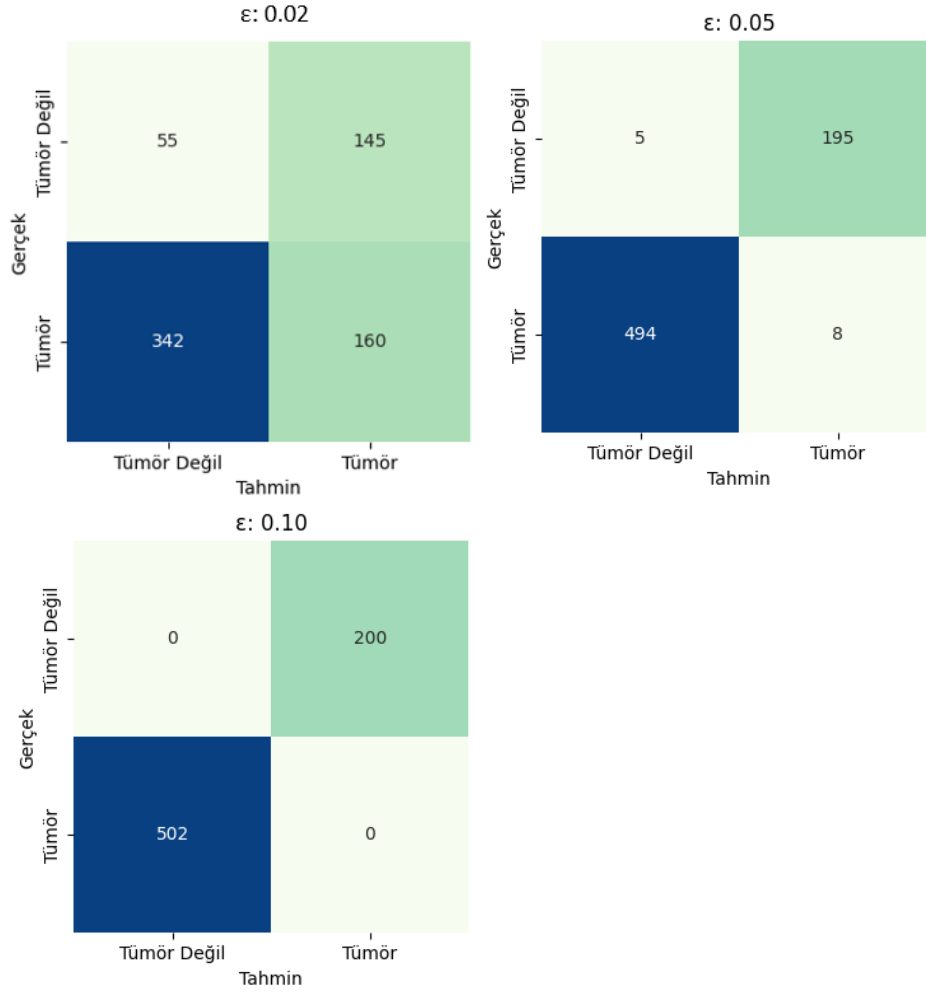


Şekil 6.18. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait ROC eğrileri

Şekil 6.18' e göre DenseNet121 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %10'a kadar düşmektedir.

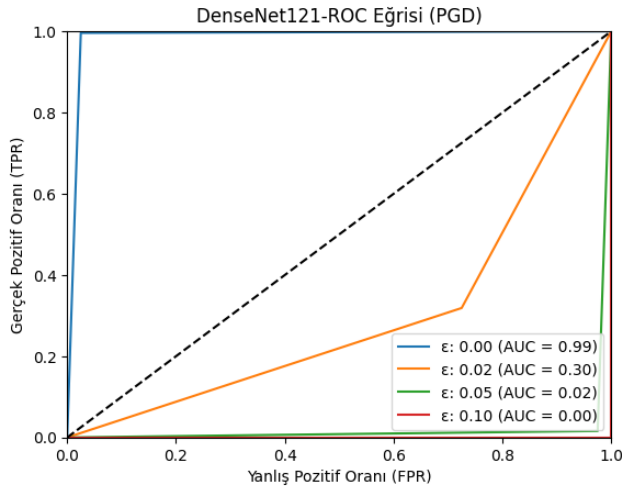
PGD saldırısı

Şekil 6.19' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.19. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.19’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 342 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 145 örnek bulunmaktadır. Çizelge 6.3’ e göre %30,63 doğruluk, %52,46 kesinlik, %31,87 geri çağırma, %39,65 F1 skoru ve %27,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 494 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 195 örnek bulunmaktadır. Çizelge 6.3’ e göre %1,85 doğruluk, %3,94 kesinlik, %1,59 geri çağırma, %2,27 F1 skoru ve %2,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre %0 doğruluk elde etmektedir. Şekil 6.20’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



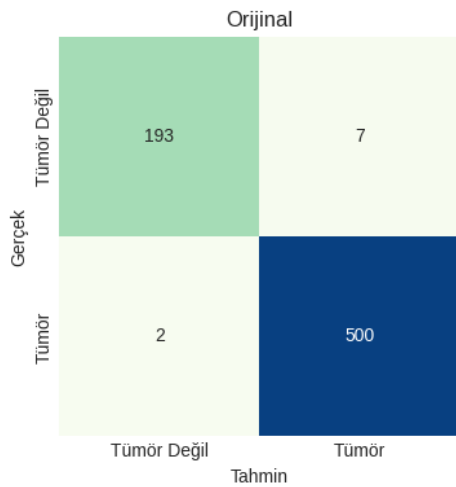
Şekil 6.20. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet121 modeline ait ROC eğrileri

Şekil 6.20' ye göre DenseNet121 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %0'a kadar düşmektedir.

6.1.5. DenseNet169

Temiz

DenseNet169 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.21'de görülmektedir.

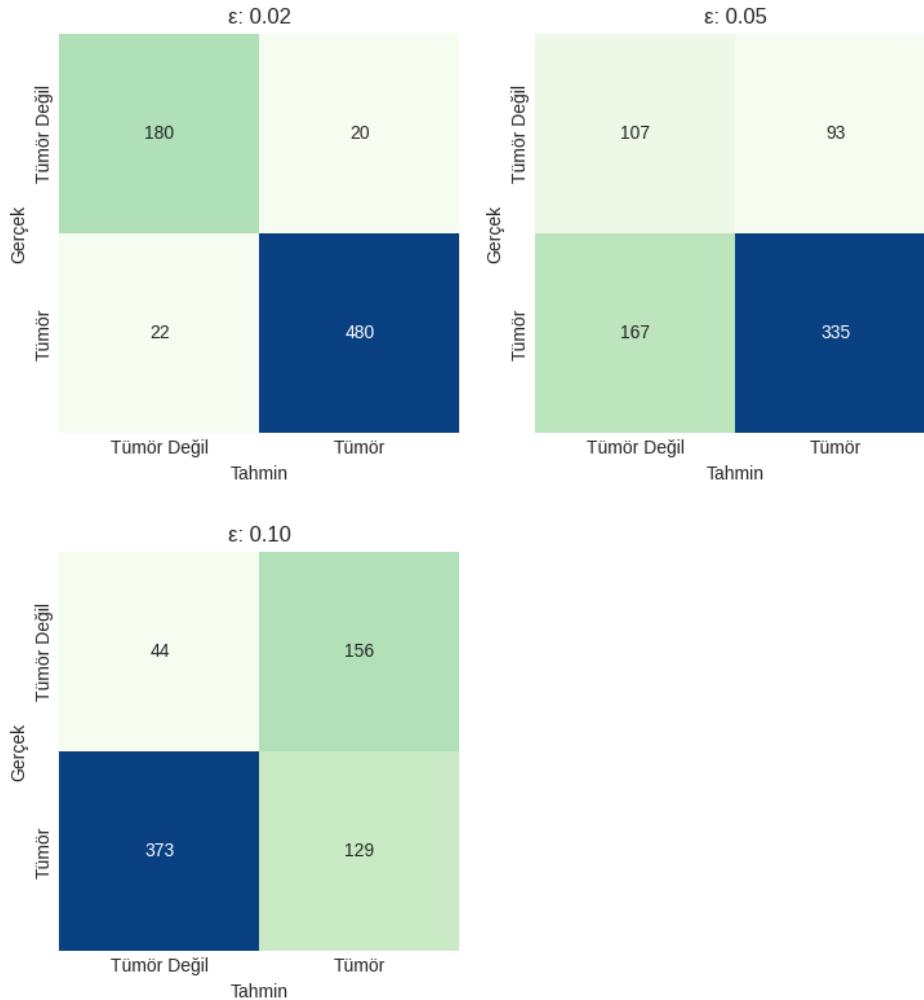


Şekil 6.21. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.21’ de görüldüğü gibi temiz test setinin DenseNet169 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 2 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 7 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,72 doğruluk, %98,62 kesinlik, %99,6 geri çağırma, %99,11 F1 skoru ve %96,5 özgüllük elde etmektedir.

FGSM saldırısı

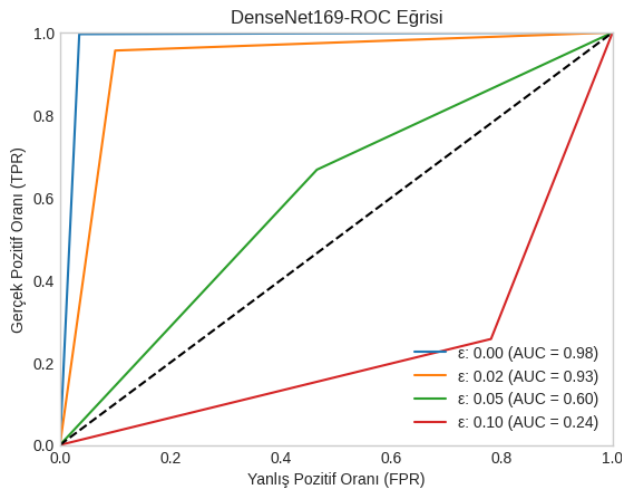
Şekil 6.22’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.22. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.22’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 22

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 20 örnek bulunmaktadır. Çizelge 6.2' ye göre %94,02 doğruluk, %96 kesinlik, %95,62 geri çağırma, %95,81 F1 skoru ve %90 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 167 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 93 örnek bulunmaktadır. Çizelge 6.2' ye göre %62,97 doğruluk, %78,27 kesinlik, %66,73 geri çağırma, %72,04 F1 skoru ve %53,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 373 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 156 örnek bulunmaktadır. Çizelge 6.2' ye göre %24,64 doğruluk, %45,26 kesinlik, %25,7 geri çağırma, %32,78 F1 skoru ve %22 özgüllük elde etmektedir. Şekil 6.23'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

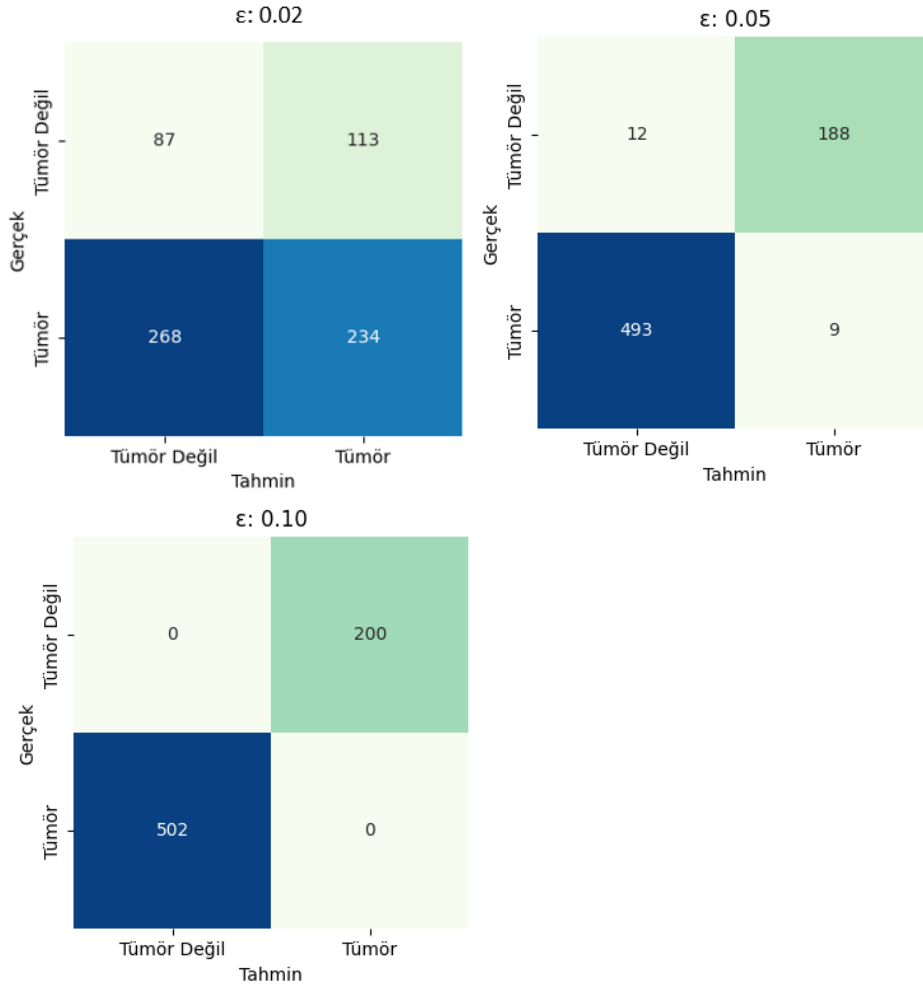


Şekil 6.23. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait ROC eğrileri

Şekil 6.23' e göre DenseNet169 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %24'e kadar düşmektedir.

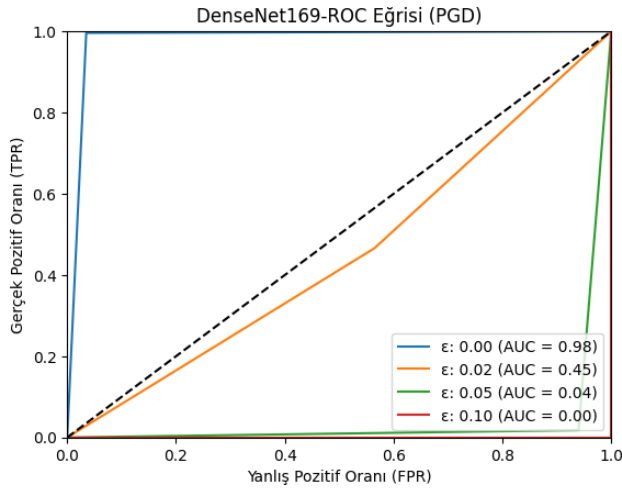
PGD saldırısı

Şekil 6.24' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.24. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.24' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 268 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 113 örnek bulunmaktadır. Çizelge 6.3' e göre %45,73 doğruluk, %67,44 kesinlik, %46,61 geri çağırma, %55,12 F1 skoru ve %43,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 493 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 188 örnek bulunmaktadır. Çizelge 6.3' e göre %2,99 doğruluk, %4,57 kesinlik, %1,79 geri çağırma, %2,58 F1 skoru ve %6 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3' e göre %0 doğruluk elde etmektedir. Şekil 6.25'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



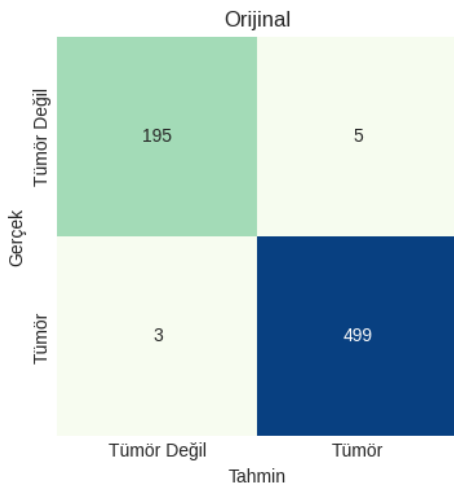
Şekil 6.25. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet169 modeline ait ROC eğrileri

Şekil 6.25' e göre DenseNet169 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.6. DenseNet201

Temiz

DenseNet201 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.26'da görülmektedir.

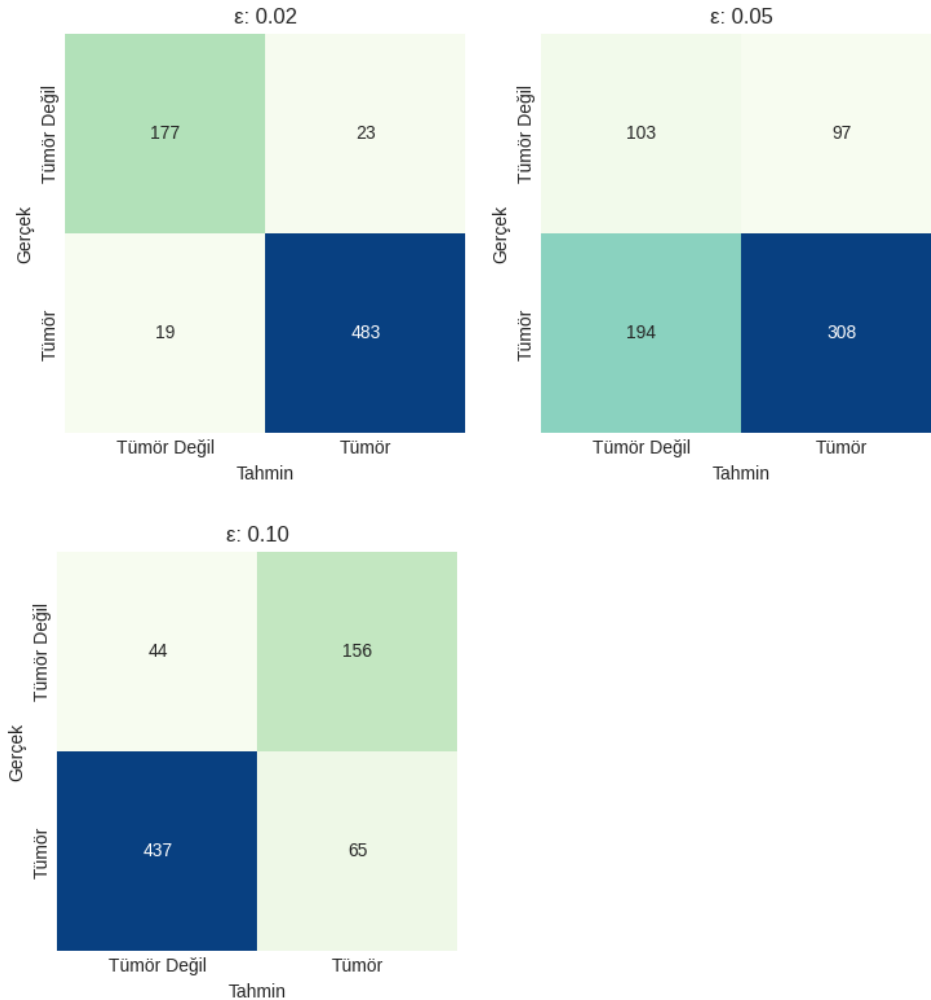


Şekil 6.26. Temiz Beyin Tümör MR görüntüleri üzerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.26' da görüldüğü gibi temiz test setinin DenseNet201 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 3 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1' e göre %98,86 doğruluk, %99,01 kesinlik, %99,4 geri çağırma, %99,2 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

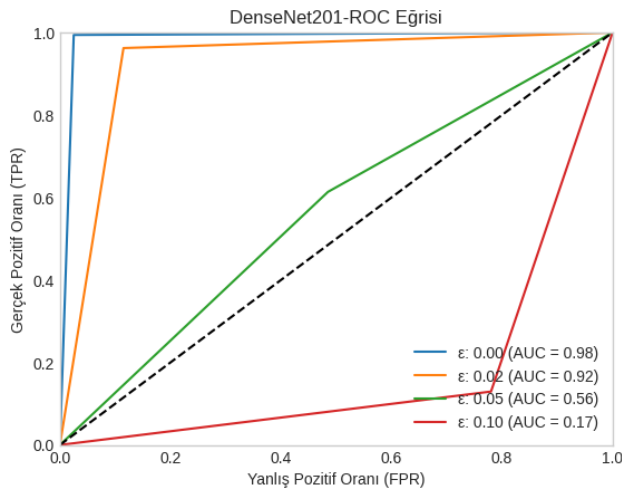
Şekil 6.27' de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.27. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.27' de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 19

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 23 örnek bulunmaktadır. Çizelge 6.2' ye göre %94,02 doğruluk, %95,46 kesinlik, %96,22 geri çağırma, %95,83 F1 skoru ve %88,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 194 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 97 örnek bulunmaktadır. Çizelge 6.2' ye göre %58,55 doğruluk, %76,05 kesinlik, %61,36 geri çağırma, %67,92 F1 skoru ve %51,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 437 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 156 örnek bulunmaktadır. Çizelge 6.2' ye göre %15,53 doğruluk, %29,41 kesinlik, %12,95 geri çağırma, %17,98 F1 skoru ve %22 özgüllük elde etmektedir. Şekil 6.28'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

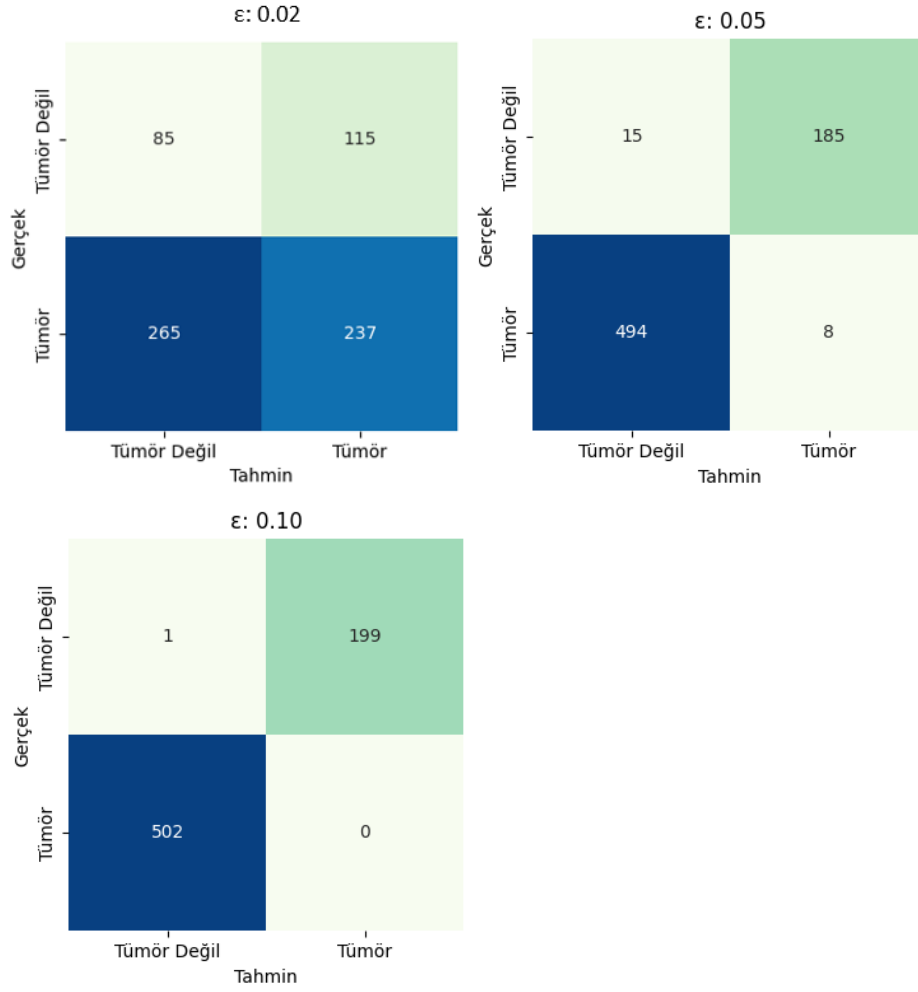


Şekil 6.28. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait ROC eğrileri

Şekil 6.28' e göre DenseNet201 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %17'ye kadar düşmektedir.

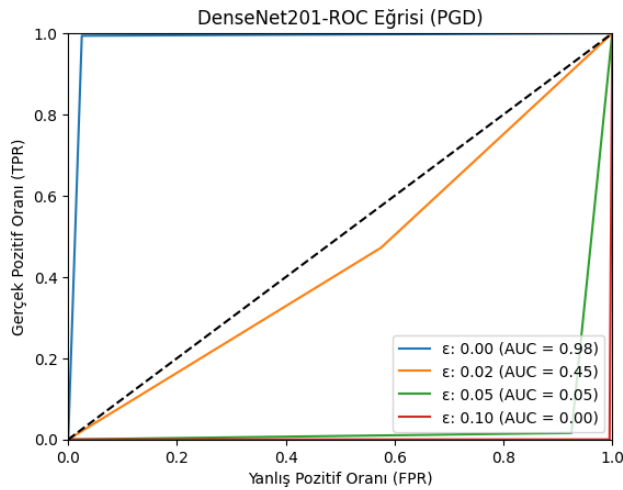
PGD saldırısı

Şekil 6.29' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.29. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.29’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 265 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 115 örnek bulunmaktadır. Çizelge 6.3’ e göre %45,87 doğruluk, %67,33 kesinlik, %47,21 geri çağırma, %55,5 F1 skoru ve %42,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 494 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 185 örnek bulunmaktadır. Çizelge 6.3’ e göre %3,28 doğruluk, %4,15kesinlik, %1,59 geri çağırma, %2,3 F1 skoru ve %7,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 199 örnek bulunmaktadır. Çizelge 6.3’ e göre %0,14 doğruluk ve %0,5 özgüllük elde etmektedir. Şekil 6.30’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



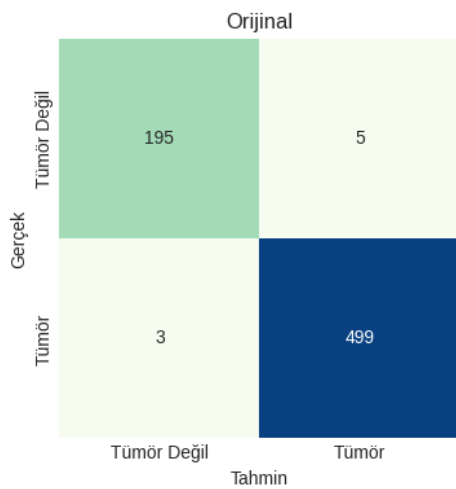
Şekil 6.30. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde DenseNet201 modeline ait ROC eğrileri

Şekil 6.30' a göre DenseNet201 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.7. EfficientNetV2B0

Temiz

EfficientNetV2B0 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.31'de görülmektedir.

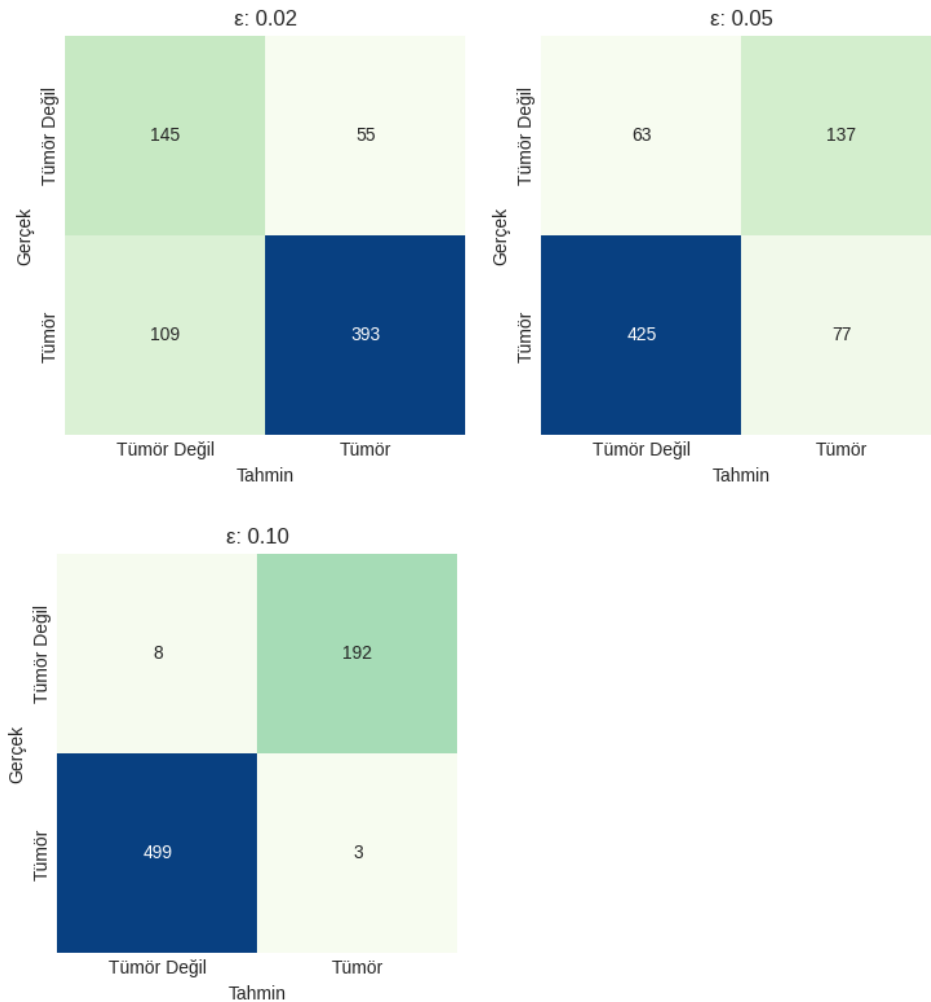


Şekil 6.31. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi

Şekil 6.31’ de görüldüğü gibi temiz test setinin EfficientNetV2B0 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 3 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,86 doğruluk, %99,01 kesinlik, %99,4 geri çağırma, %99,21 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

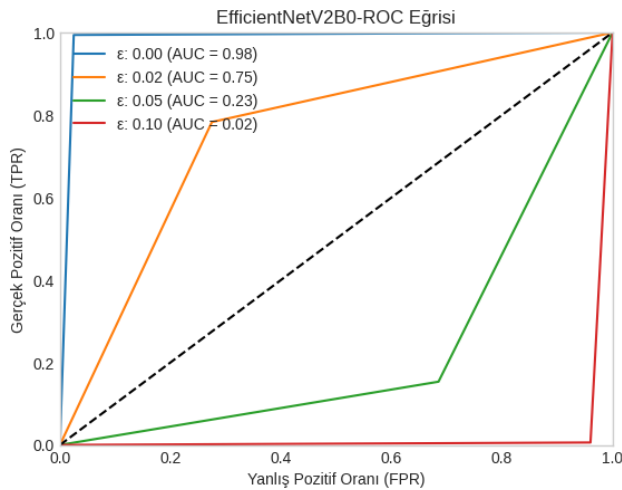
Şekil 6.32’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.32. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi

Şekil 6.32’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 109

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 55 örnek bulunmaktadır. Çizelge 6.2' ye göre %76,46 doğruluk, %87,72 kesinlik, %78,29 geri çağırma, %82,74 F1 skoru ve %72,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 425 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 137 örnek bulunmaktadır. Çizelge 6.2' ye göre %19,94 doğruluk, %35,98 kesinlik, %15,34 geri çağırma, %21,51 F1 skoru ve %31,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 499 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 192 örnek bulunmaktadır. Çizelge 6.2' ye göre %1,57 doğruluk, %1,54 kesinlik, %0,6 geri çağırma, %0,86 F1 skoru ve %4 özgüllük elde etmektedir. Şekil 6.33'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

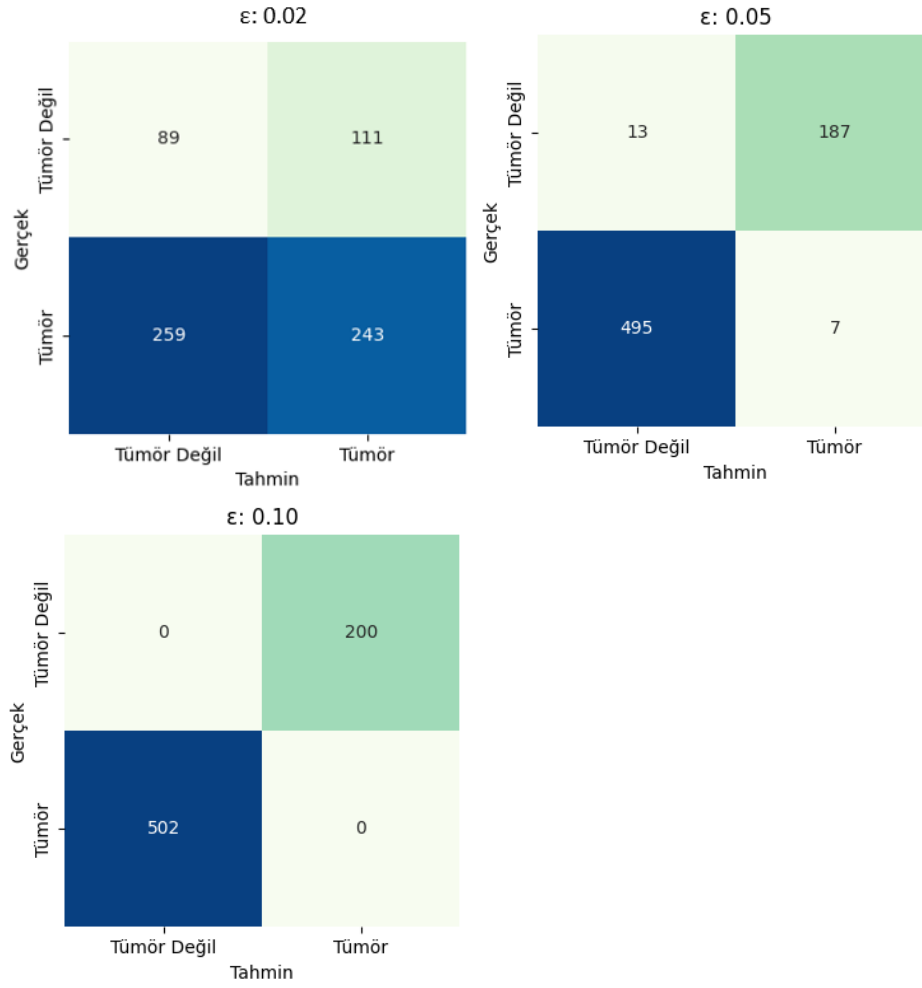


Şekil 6.33. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri

Şekil 6.33' e göre EfficientNetV2B0 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %2'ye kadar düşmektedir.

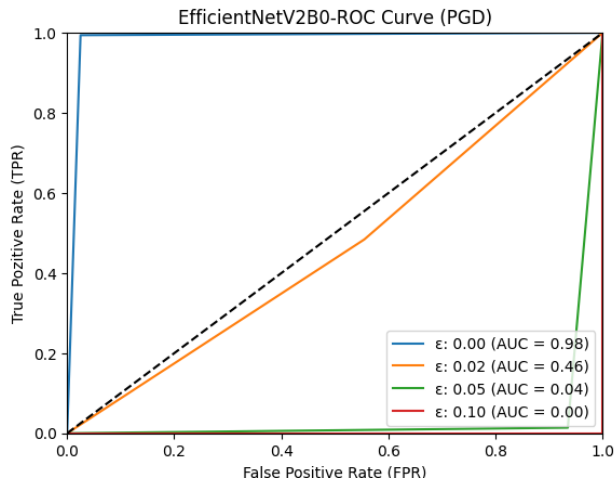
PGD saldırısı

Şekil 6.34' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.34. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi

Şekil 6.34' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 259 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 11örnek bulunmaktadır. Çizelge 6.3' e göre %47,29 doğruluk, %68,64 kesinlik, %48,41 geri çağırma, %56,78 F1 skoru ve %44,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 495 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 187 örnek bulunmaktadır. Çizelge 6.3' e göre %2,85 doğruluk, %3,61 kesinlik, %1,39 geri çağırma, %2,01 F1 skoru ve %6,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3' e göre %0 doğruluk elde etmektedir. Şekil 6.35'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır



Şekil 6.35. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri

Şekil 6.35’ e göre EfficientNetV2B0 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98’ den %0’a kadar düşmektedir.

6.1.8. EfficientNetV2B1

Temiz

EfficientNetV2B1 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.36’da görülmektedir.

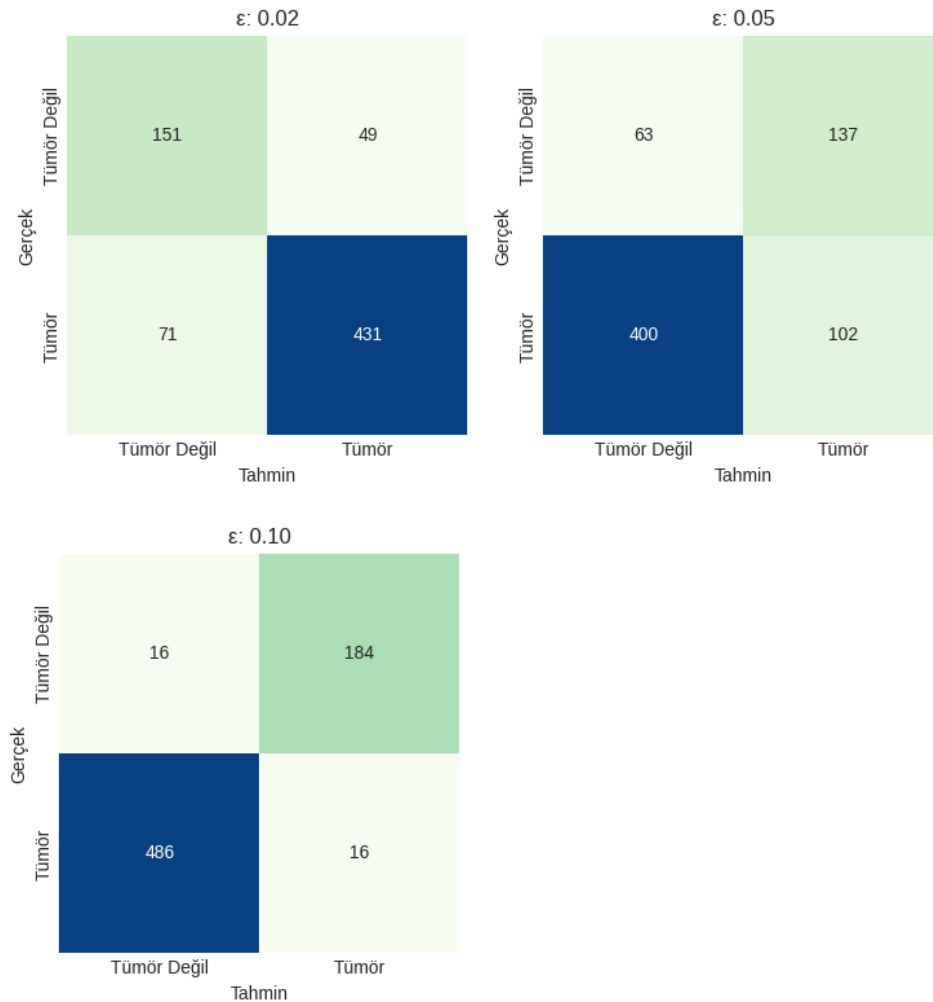
		Orijinal	
Gerçek	Tümör Değil	196	4
	Tümör	3	499
		Tümör Değil	Tümör
		Tahmin	

Şekil 6.36. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.36’ da görüldüğü gibi temiz test setinin EfficientNetV2B1 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 3 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 4 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %99,21 kesinlik, %99,4 geri çağırma, %99,3 F1 skoru ve %98 özgüllük elde etmektedir.

FGSM saldırısı

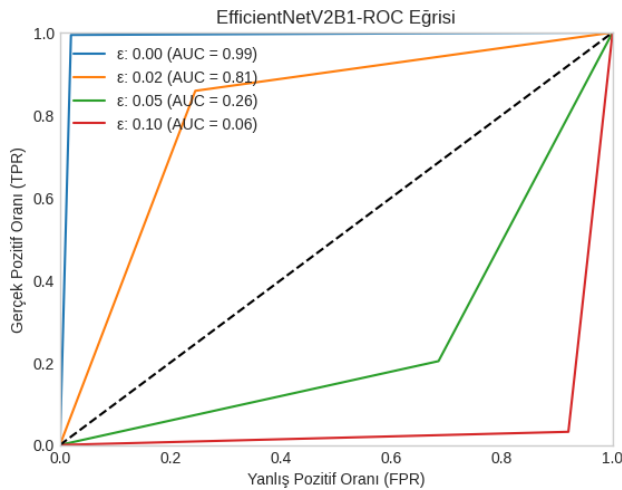
Şekil 6.37’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.37. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.37’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 71

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 49 örnek bulunmaktadır. Çizelge 6.2' ye göre %82,91 doğruluk, %89,79 kesinlik, %85,86 geri çağırma, %87,78 F1 skoru ve %75,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 400 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 137 örnek bulunmaktadır. Çizelge 6.2' ye göre %23,5 doğruluk, %42,68 kesinlik, %20,32 geri çağırma, %27,53 F1 skoru ve %31,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 486 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 184 örnek bulunmaktadır. Çizelge 6.2' ye göre %4,56 doğruluk, %8 kesinlik, %31,87 geri çağırma, %4,56 F1 skoru ve %8 özgüllük elde etmektedir. Şekil 6.38'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

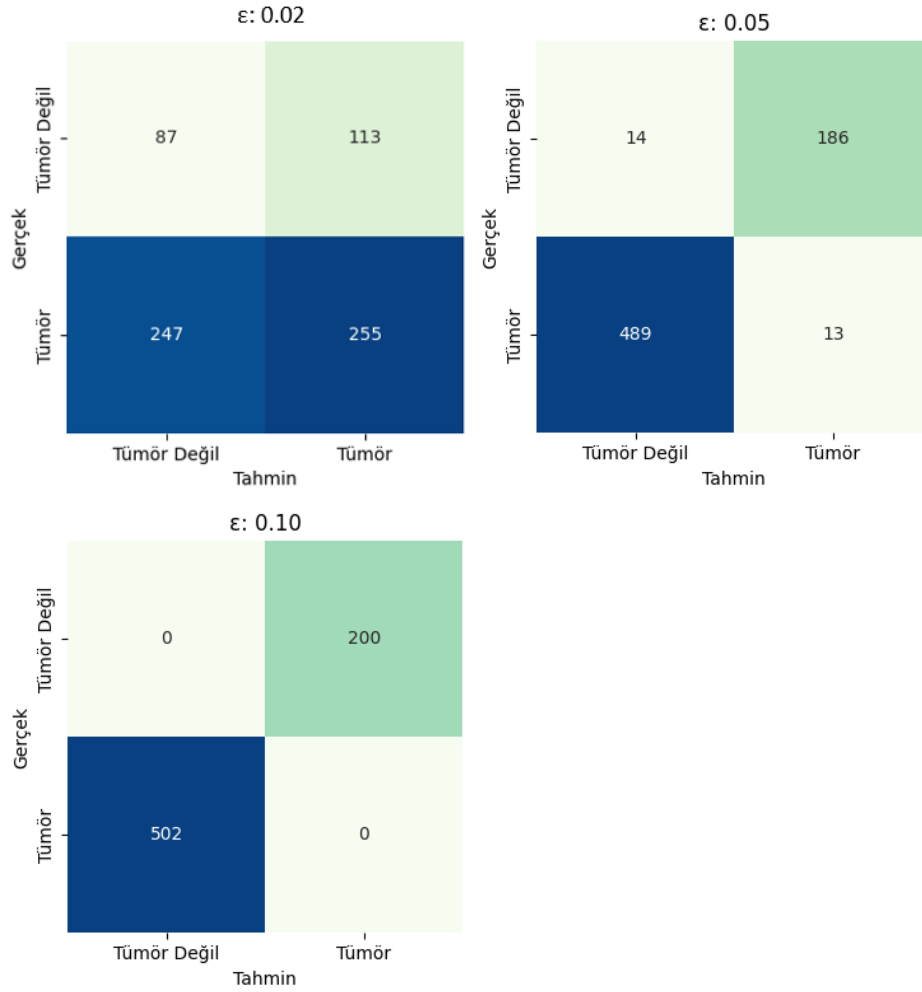


Şekil 6.38. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri

Şekil 6.38' e göre EfficientNetV2B1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %6'ya kadar düşmektedir.

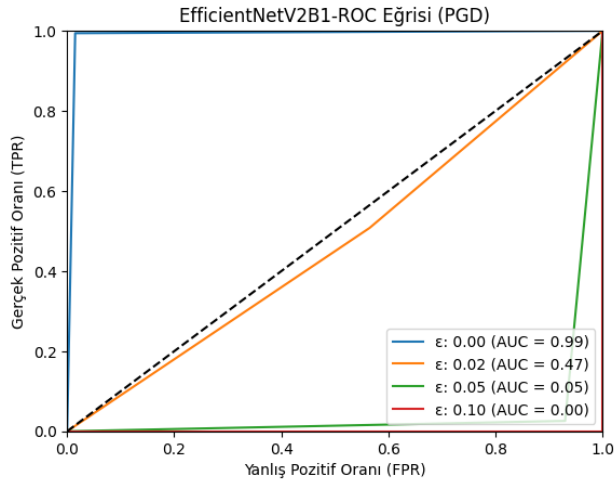
PGD saldırısı

Şekil 6.39' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.39. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.39’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 247 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 113 örnek bulunmaktadır. Çizelge 6.3’ e göre %48,72 doğruluk, %69,29 kesinlik, %50,8 geri çağırma, %58,62 F1 skoru ve %43,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 489 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 186 örnek bulunmaktadır. Çizelge 6.3’ e göre %3,85 doğruluk, %6,53 kesinlik, %2,59 geri çağırma, %3,71 F1 skoru ve %7 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre %0 doğruluk elde etmektedir. Şekil 6.40’ ta pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



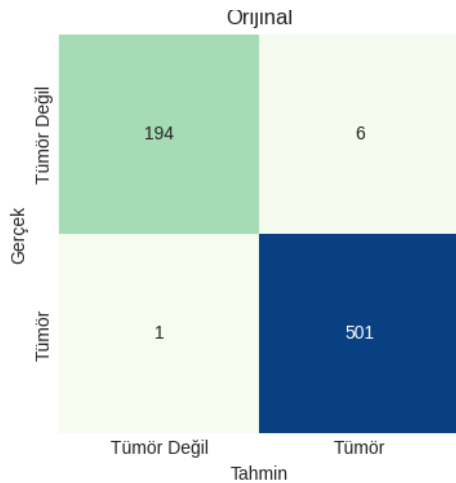
Şekil 6.40. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri

Şekil 6.40' a göre EfficientNetV2B1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %0'a kadar düşmektedir.

6.1.9. EfficientNetV2B2

Temiz

EfficientNetV2B2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.41'de görülmektedir.

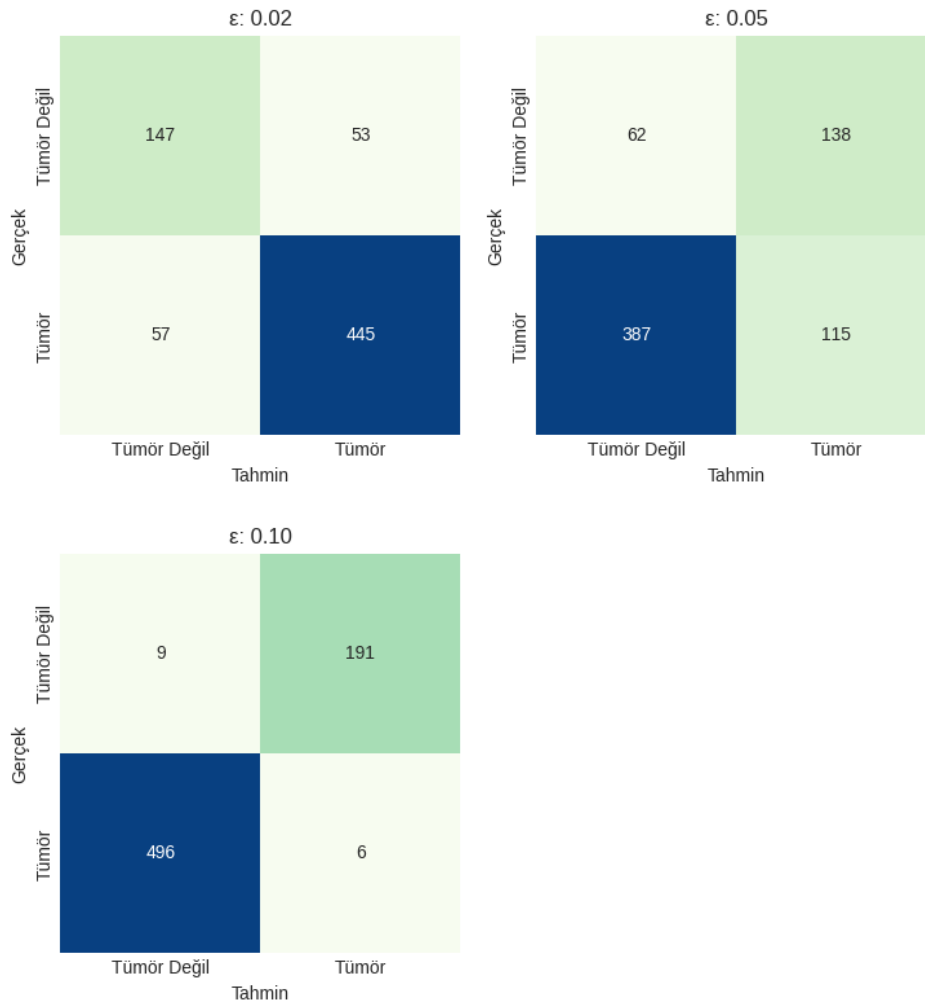


Şekil 6.41. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.41’ de görüldüğü gibi temiz test setinin EfficientNetV2B2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 1 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 6 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %98,82 kesinlik, %99,8 geri çağırma, %99,31 F1 skoru ve %97 özgüllük elde etmektedir.

FGSM saldırısı

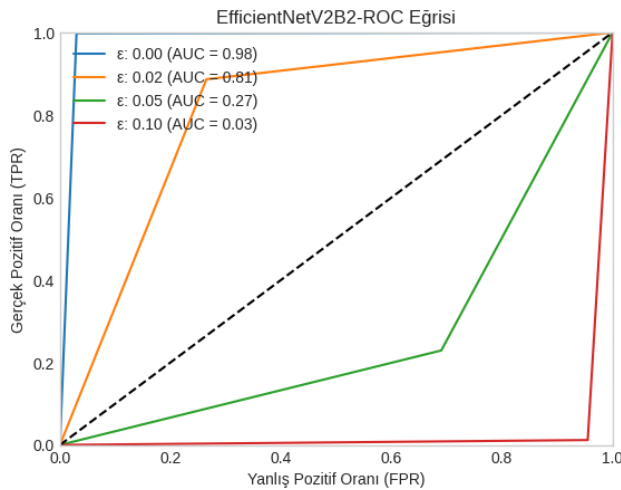
Şekil 6.42’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.42. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.42’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 57

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 53 örnek bulunmaktadır. Çizelge 6.2' ye göre %84,33 doğruluk, %89,36 kesinlik, %88,65 geri çağırma, %89 F1 skoru ve %73,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 387 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 138 örnek bulunmaktadır. Çizelge 6.2' ye göre %25,21 doğruluk, %45,46 kesinlik, %22,91 geri çağırma, %30,46 F1 skoru ve %31 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 496 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 191 örnek bulunmaktadır. Çizelge 6.2' ye göre %2,14 doğruluk, %3,05 kesinlik, %1,2 geri çağırma, %1,72 F1 skoru ve %4,5 özgüllük elde etmektedir. Şekil 6.43'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

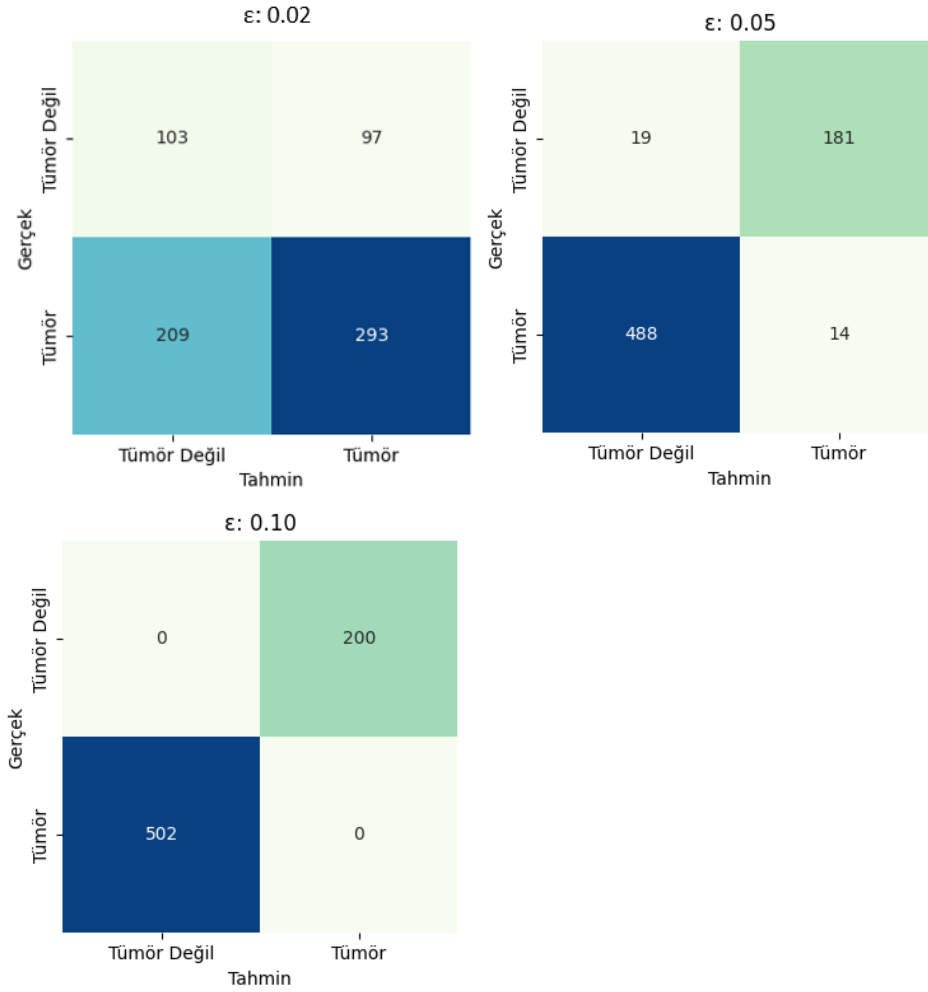


Şekil 6.43. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri

Şekil 6.43' e göre EfficientNetV2B2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %3'e kadar düşmektedir.

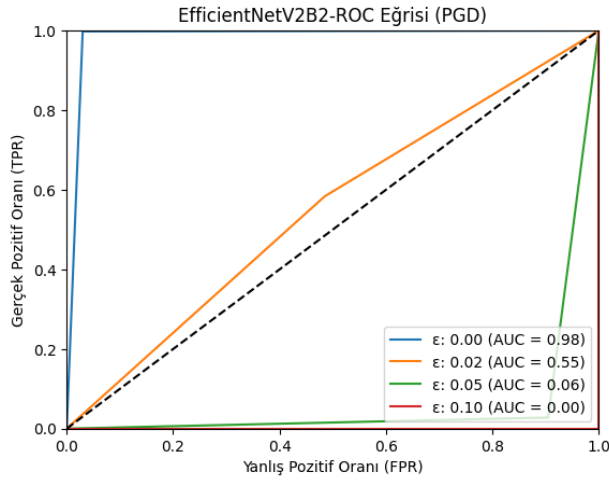
PGD saldırısı

Şekil 6.44' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.44. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.44' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 209 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 97 örnek bulunmaktadır. Çizelge 6.3' e göre %56,41 doğruluk, %75,13 kesinlik, %58,37 geri çağırma, %65,7 F1 skoru ve %51,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 488 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 181 örnek bulunmaktadır. Çizelge 6.3' e göre %4,7 doğruluk, %7,18 kesinlik, %2,79 geri çağırma, %4,02 F1 skoru ve %9,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3' e göre %0 doğruluk elde etmektedir. Şekil 6.45'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



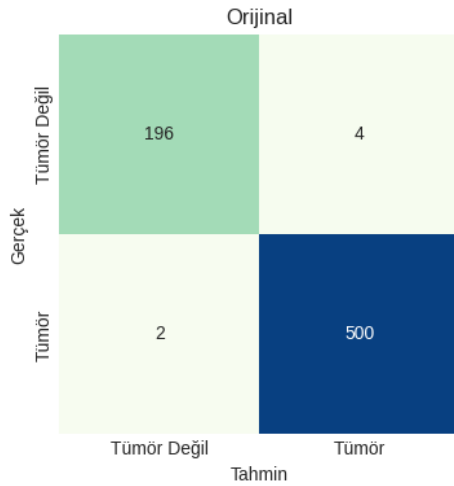
Şekil 6.45. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri

Şekil 6.45' e göre EfficientNetV2B2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.10. EfficientNetV2B3

Temiz

EfficientNetV2B3 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.46'da görülmektedir.

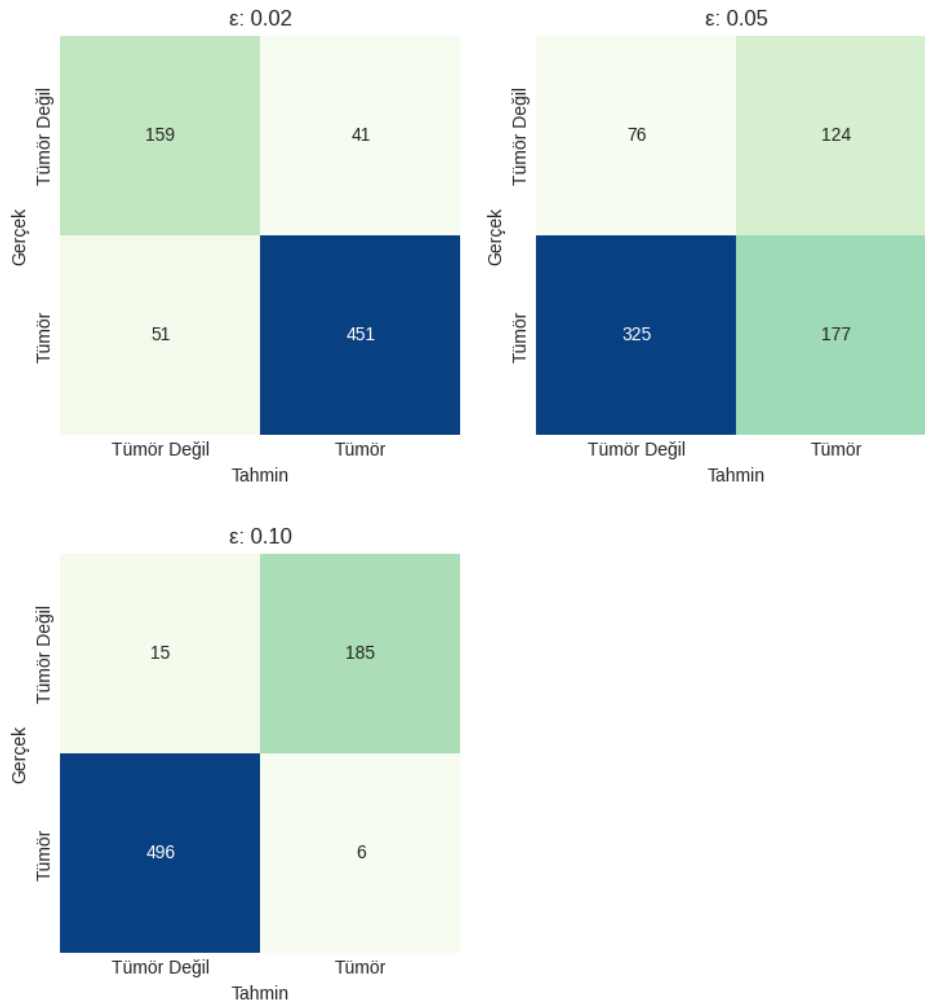


Şekil 6.46. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.46’ da görüldüğü gibi temiz test setinin EfficientNetV2B3 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 2 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 4 örnek bulunmaktadır. Çizelge 6.1’ e göre %99,15 doğruluk, %99,21 kesinlik, %99,6 geri çağırma, %99,4 F1 skoru ve %98 özgüllük elde etmektedir.

FGSM saldırısı

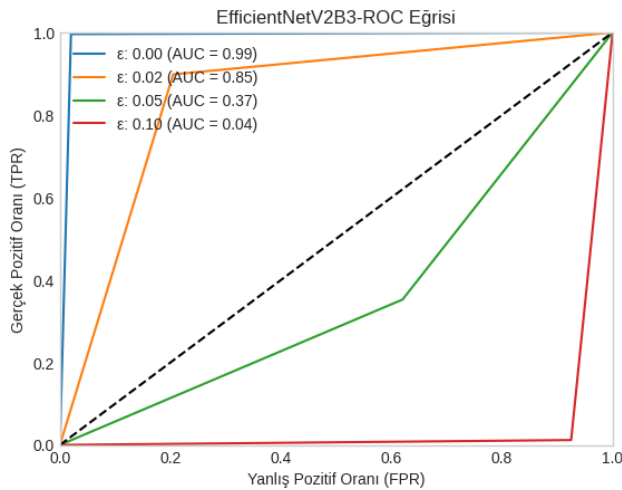
Şekil 6.47’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.47. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.47’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 51

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 41 örnek bulunmaktadır. Çizelge 6.2' ye göre %86,9 doğruluk, %91,67 kesinlik, %89,84 geri çağırma, %90,75 F1 skoru ve %79,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 325 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 124 örnek bulunmaktadır. Çizelge 6.2' ye göre %36,04 doğruluk, %58,8 kesinlik, %35,26 geri çağırma, %44,09 F1 skoru ve %38 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 496 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 185 örnek bulunmaktadır. Çizelge 6.2' ye göre %2,99 doğruluk, %3,14 kesinlik, %1,2 geri çağırma, %1,73 F1 skoru ve %7,5 özgüllük elde etmektedir. Şekil 6.48'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

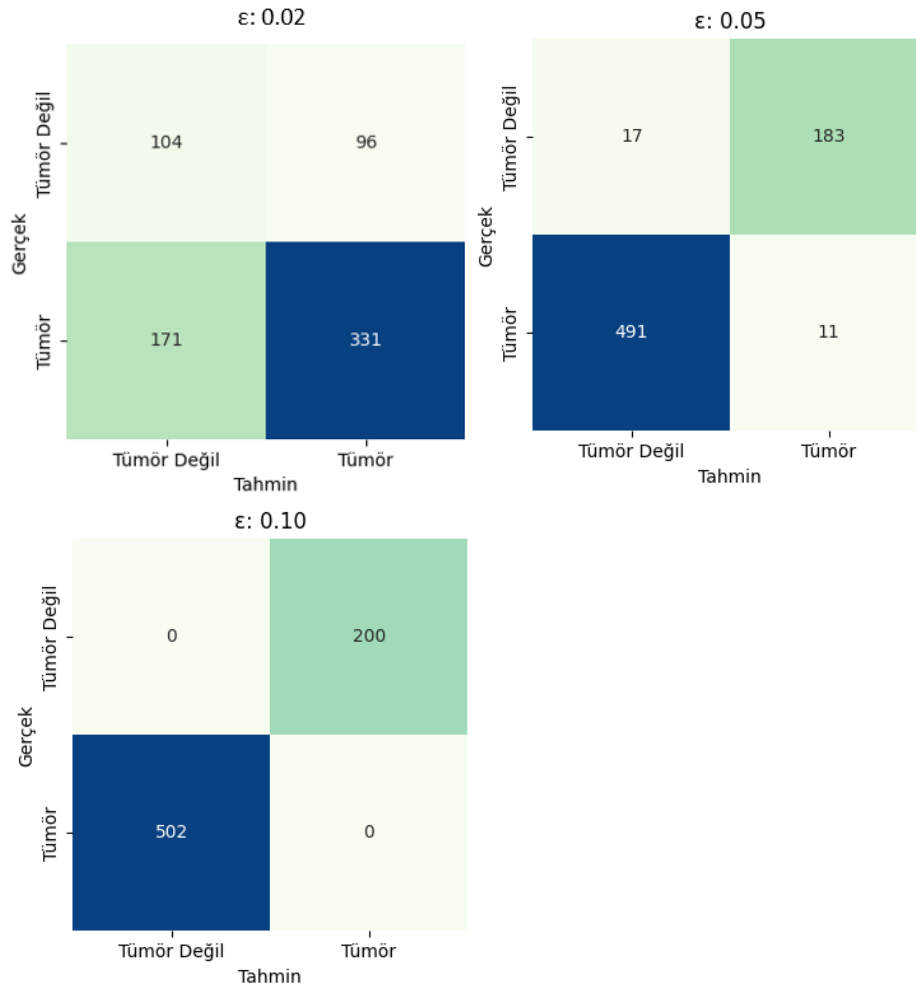


Şekil 6.48. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri

Şekil 6.48' e göre EfficientNetV2B3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %4'e kadar düşmektedir.

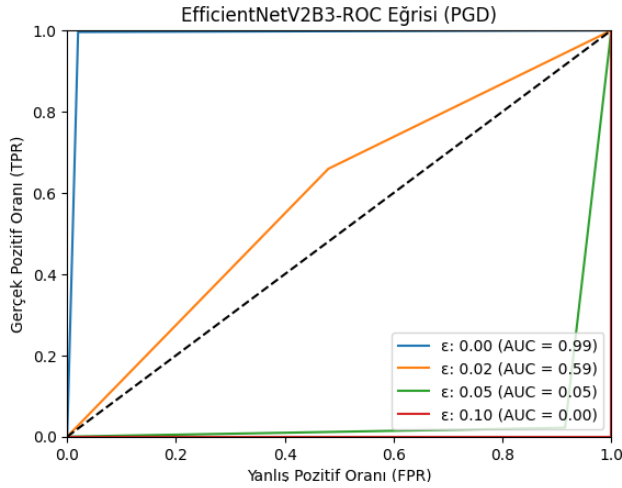
PGD saldırısı

Şekil 6.49' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.49. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.49’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 171 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 96 örnek bulunmaktadır. Çizelge 6.3’ e göre %61,97 doğruluk, %77,52 kesinlik, %65,94 geri çağırma, %71,26 F1 skoru ve %52 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 491 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 183 örnek bulunmaktadır. Çizelge 6.3’ e göre %3,99 doğruluk, %5,67 kesinlik, %2,19 geri çağırma, %3,16 F1 skoru ve %8,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.50’ ye göre %0 doğruluk elde etmektedir. Şekil 6.50’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



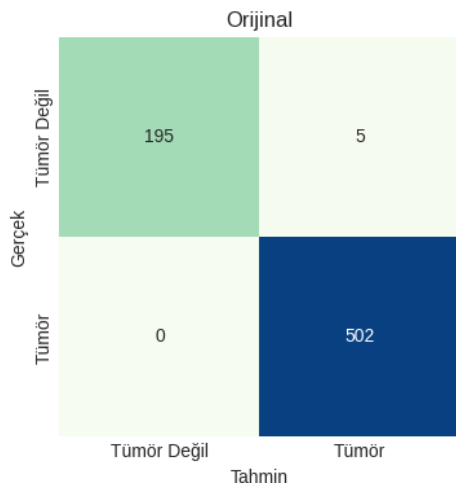
Şekil 6.50. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri

Şekil 6.50' ye göre EfficientNetV2B3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %0'a kadar düşmektedir.

6.1.11. EfficientNetV2S

Temiz

EfficientNetV2S modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.51'de görülmektedir.

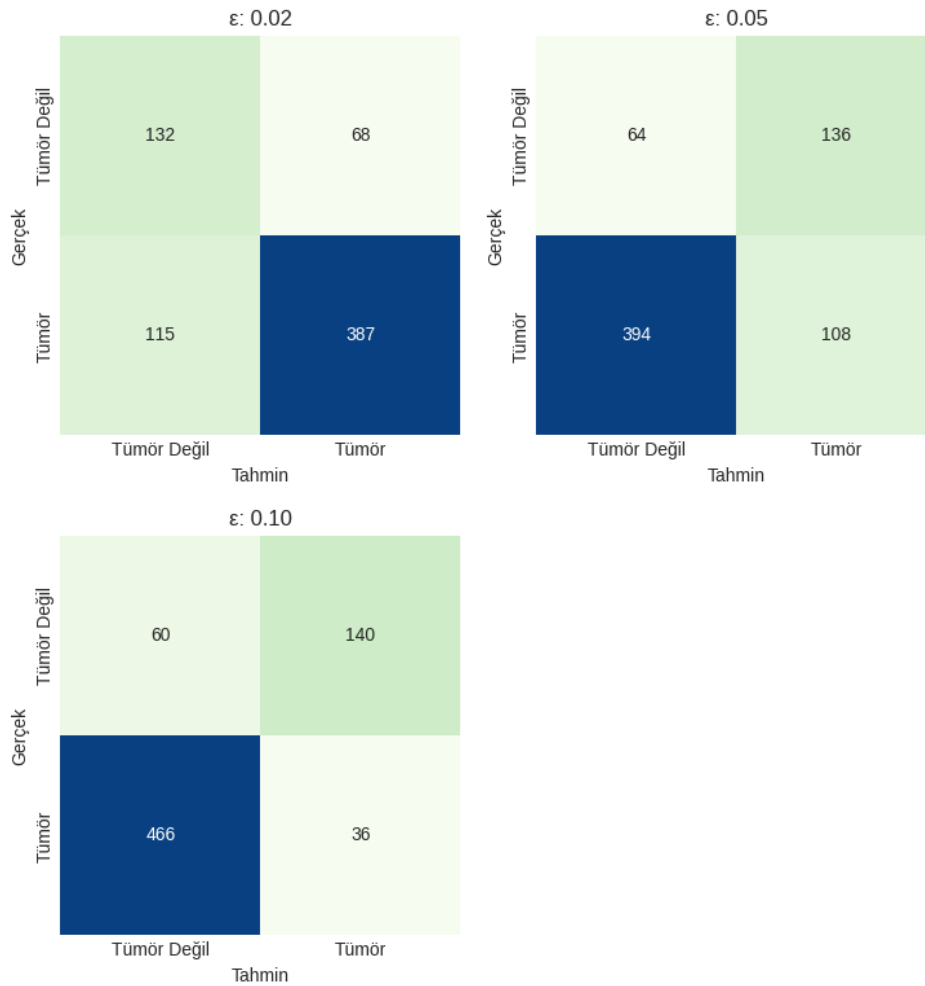


Şekil 6.51. Temiz Beyin Tümör MR görüntüleri üzerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.51’ de görüldüğü gibi temiz test setinin EfficientNetV2S modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) örnek bulunmamaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %99,29 doğruluk, %99,01 kesinlik, %100 geri çağırma, %99,51 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

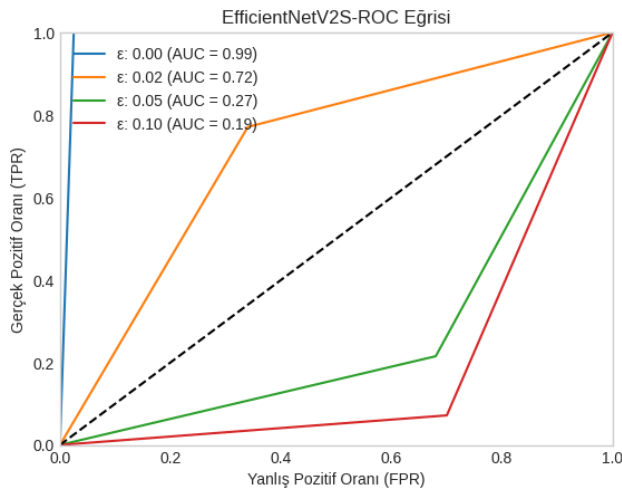
Şekil 6.52’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.52. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.52’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 115

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 68 örnek bulunmaktadır. Çizelge 6.2' ye göre %73,93 doğruluk, %85,06 kesinlik, %77,09 geri çağırma, %80,88 F1 skoru ve %66 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 394 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 136 örnek bulunmaktadır. Çizelge 6.2' ye göre %24,5 doğruluk, %44,26 kesinlik, %21,51 geri çağırma, %28,95 F1 skoru ve %32 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 466 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 140 örnek bulunmaktadır. Çizelge 6.2' ye göre %13,68 doğruluk, %20,46 kesinlik, %7,17 geri çağırma, %10,62 F1 skoru ve %3 özgüllük elde etmektedir. Şekil 6.53'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

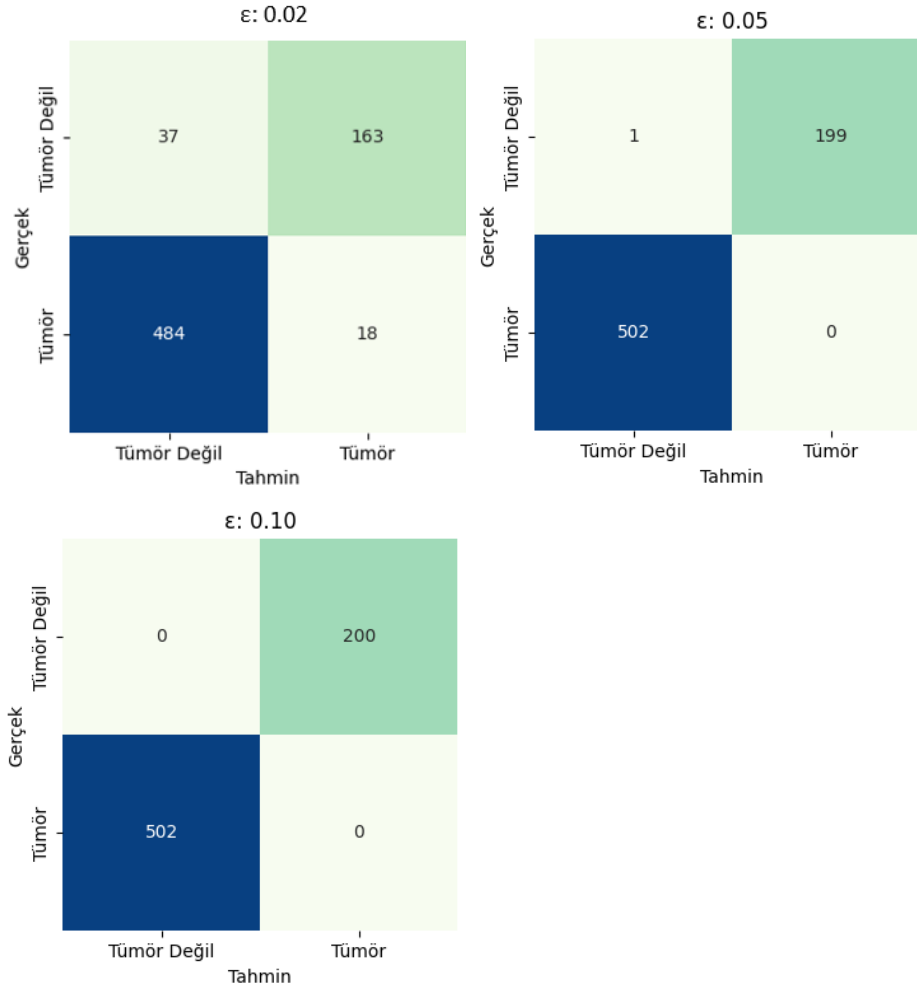


Şekil 6.53. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait ROC eğrileri

Şekil 6.53' e göre EfficientNetV2S modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %19'a kadar düşmektedir.

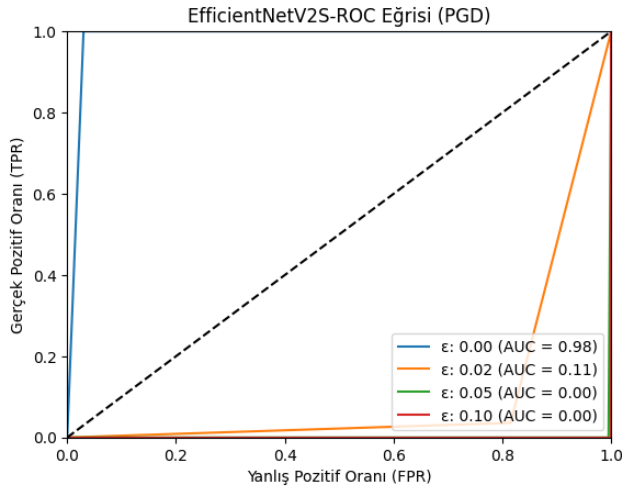
PGD saldırısı

Şekil 6.54' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.54. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.54' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 484 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 163 örnek bulunmaktadır. Çizelge 6.3' e göre %7,85 doğruluk, %9,95 kesinlik, %3,59 geri çağırma, %5,27 F1 skoru ve %18,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 ve 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3' e göre 0,05 pertürbasyon için %0,14 doğruluk ve 0,10 pertürbasyon için %0 doğruluk elde etmektedir. Şekil 6.55'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



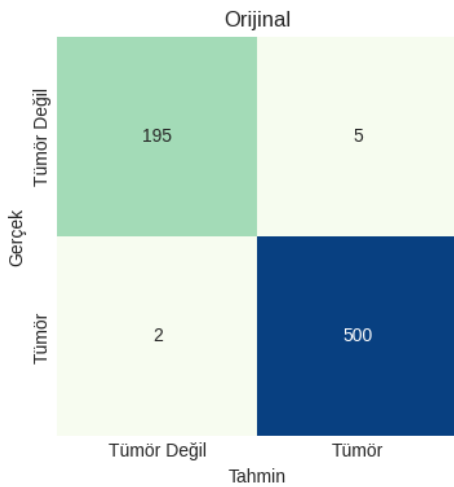
Şekil 6.55. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde EfficientNetV2S modeline ait ROC eğrileri

Şekil 6.55' e göre EfficientNetV2S modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.12. InceptionResNetV2

Temiz

InceptionResNetV2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.56'da görülmektedir.

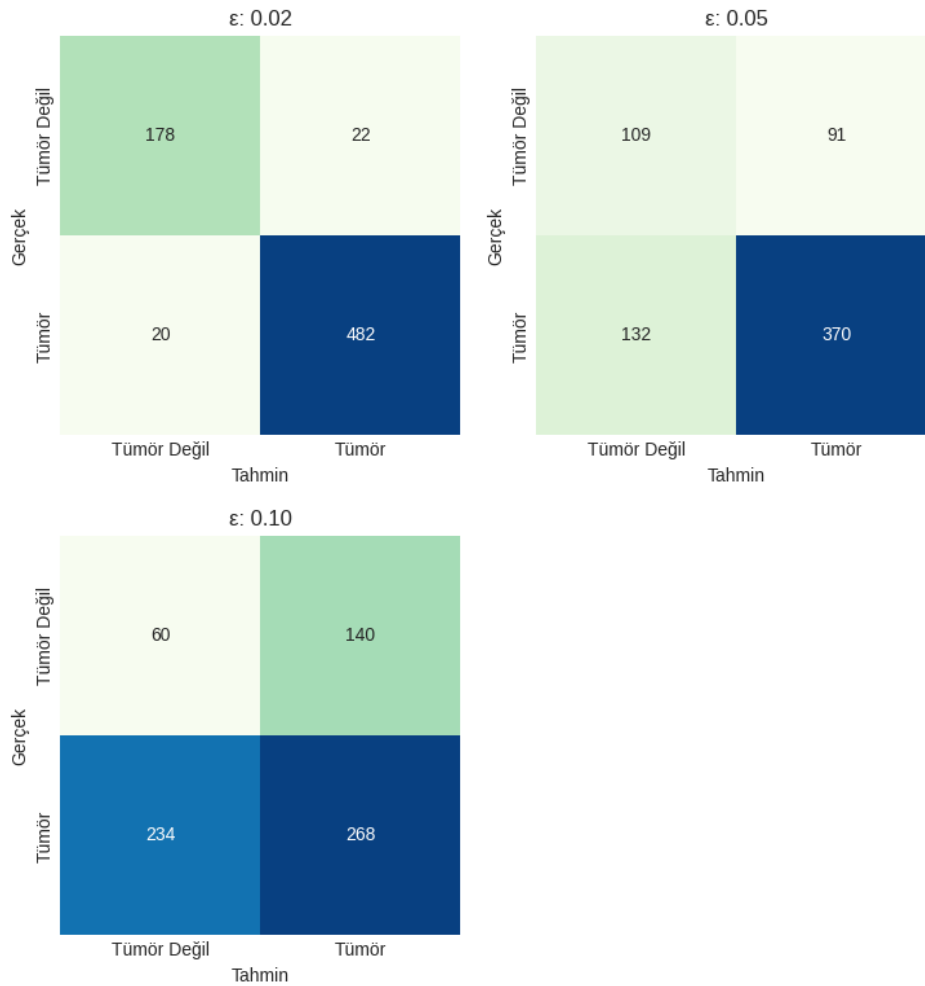


Şekil 6.56. Temiz Beyin Tümör MR görüntüleri üzerinde InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.56’ da görüldüğü gibi temiz test setinin InceptionResNetV2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 2 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %99,01 kesinlik, %99,6 geri çağırma, %99,4 F1 skoru ve %98 özgüllük elde etmektedir.

FGSM saldırısı

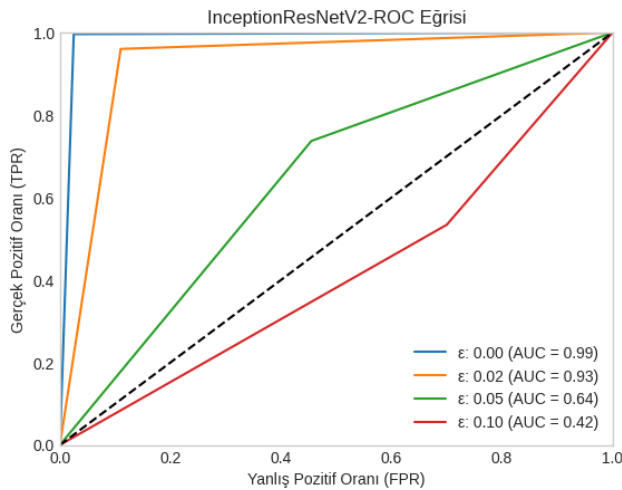
Şekil 6.57’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.57. FGSM saldırısına uğrayan Beyin Tümör MR görüntüleride InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.57’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 20

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 22 örnek bulunmaktadır. Çizelge 6.2' ye göre %94,02 doğruluk, %95,64 kesinlik, %96,02 geri çağırma, %95,83 F1 skoru ve %89 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 132 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 91 örnek bulunmaktadır. Çizelge 6.2' ye göre %68,23 doğruluk, %80,26 kesinlik, %73,71 geri çağırma, %76,84 F1 skoru ve %54,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 234 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 140 örnek bulunmaktadır. Çizelge 6.2' ye göre %46,72 doğruluk, %65,69 kesinlik, %53,39 geri çağırma, %58,9 F1 skoru ve %3 özgüllük elde etmektedir. Şekil 6.58'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

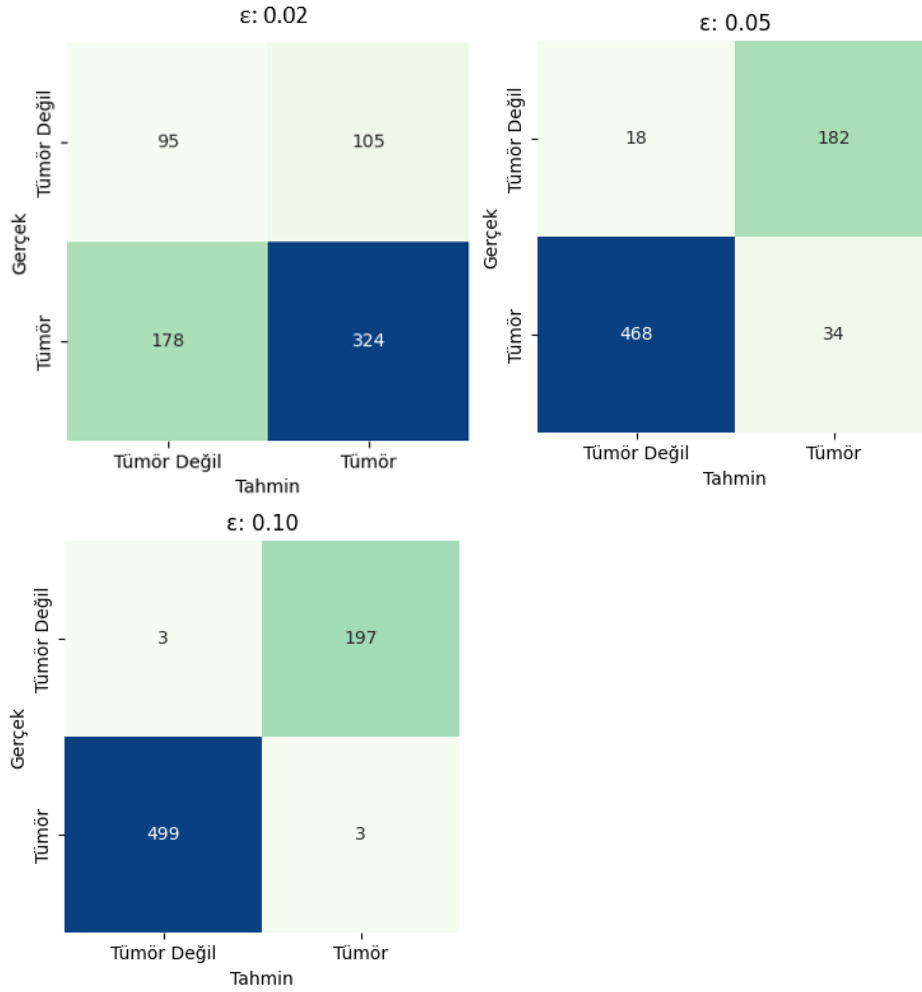


Şekil 6.58. FGSM saldırısına uğrayan Beyin Tümör MR görüntüleride InceptionResNetV2 modeline ait ROC eğrileri

Şekil 6.58' e göre InceptionResNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %42'ye kadar düşmektedir.

PGD saldırısı

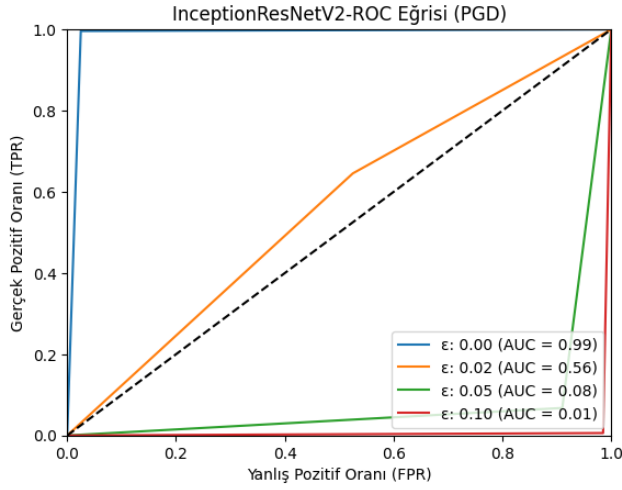
Şekil 6.59' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.59. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.59’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 178 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 105 örnek bulunmaktadır. Çizelge 6.3’ e göre %59,69 doğruluk, %75,53 kesinlik, %64,64 geri çağırma, %69,6 F1 skoru ve %47,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 468 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 182 örnek bulunmaktadır. Çizelge 6.3’ e göre %7,41 doğruluk, %15,74 kesinlik, %6,77 geri çağırma, %9,47 F1 skoru ve %9 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 499 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 197 örnek bulunmaktadır. Çizelge 6.3’ e göre %0,86 doğruluk, %1,5 kesinlik, %0,6 geri çağırma, %0,86 F1 skoru ve %1,5

özgüllük elde etmektedir. Şekil 6.60'ta pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.60. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri

Şekil 6.60' a göre InceptionResNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %1'e kadar düşmektedir.

6.1.13. InceptionV3

Temiz

InceptionV3 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.61'de görülmektedir.

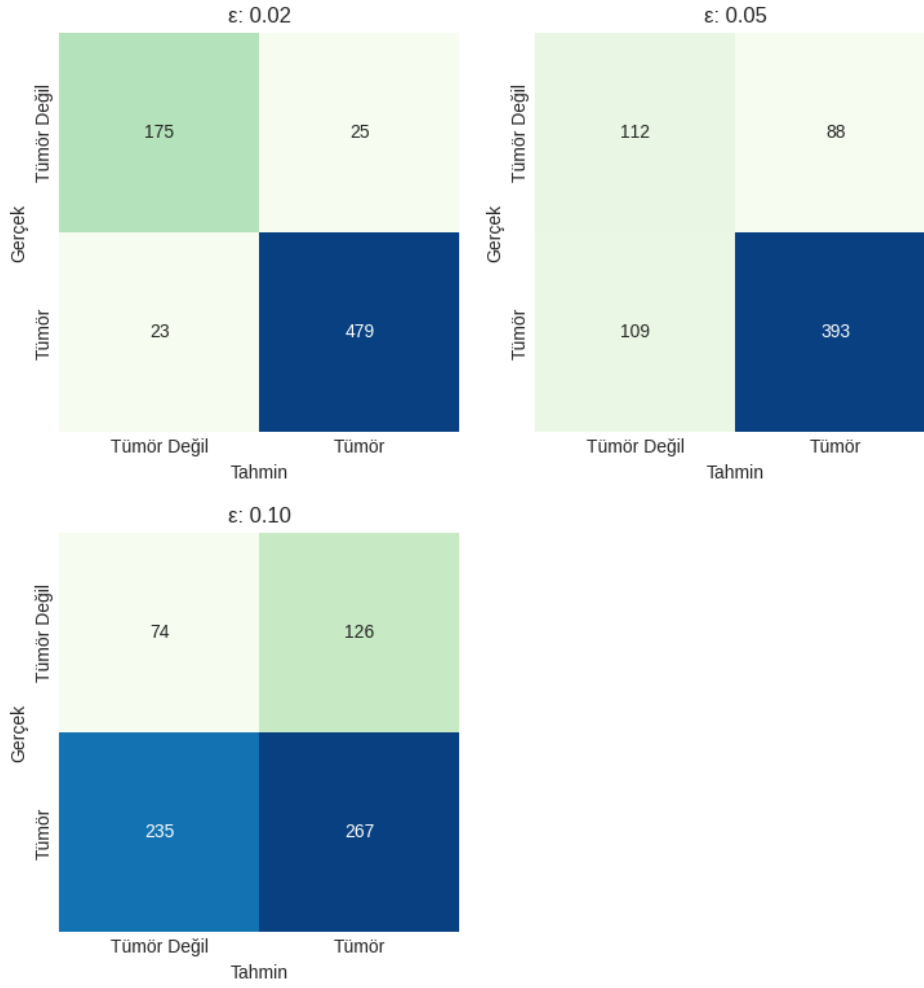
		Orjinal	
Gerçek	Tümör Değil	194	6
	Tümör	2	500
		Tümör Değil	Tümör
		Tahmin	

Şekil 6.61. Temiz Beyin Tümör MR görüntüleri üzerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.61’ de görüldüğü gibi temiz test setinin InceptionV3 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 2 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 6 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,86 doğruluk, %98,81 kesinlik, %99,6 geri çağırma, %99,21 F1 skoru ve %97 özgüllük elde etmektedir.

FGSM saldırısı

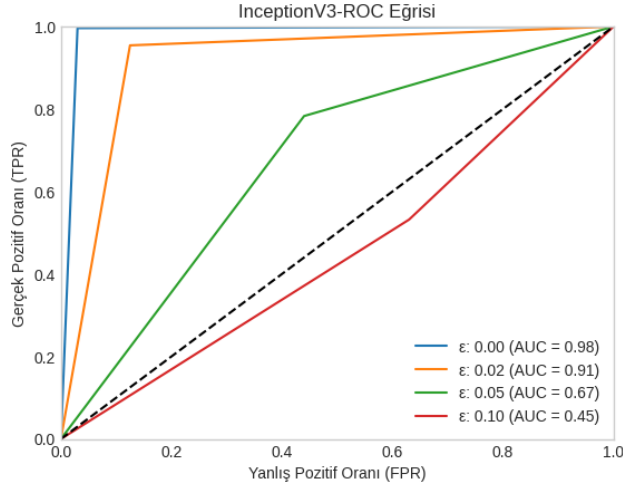
Şekil 6.62’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.62. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.62’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 23 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 25 örnek bulunmaktadır. Çizelge 6.2’ ye göre %93,16 doğruluk, %95,04 kesinlik, %95,42 geri çağırma, %95,23 F1 skoru ve %87,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 109 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 88 örnek bulunmaktadır. Çizelge 6.2’ ye göre %71,94 doğruluk, %81,71 kesinlik, %78,29 geri çağırma, %79,96 F1 skoru ve %56 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 235 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 126 örnek bulunmaktadır. Çizelge 6.2’ ye göre %48,58 doğruluk, %67,94 kesinlik, %53,19 geri

çağırma, %59,67 F1 skoru ve %37 özgüllük elde etmektedir. Şekil 6.63'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

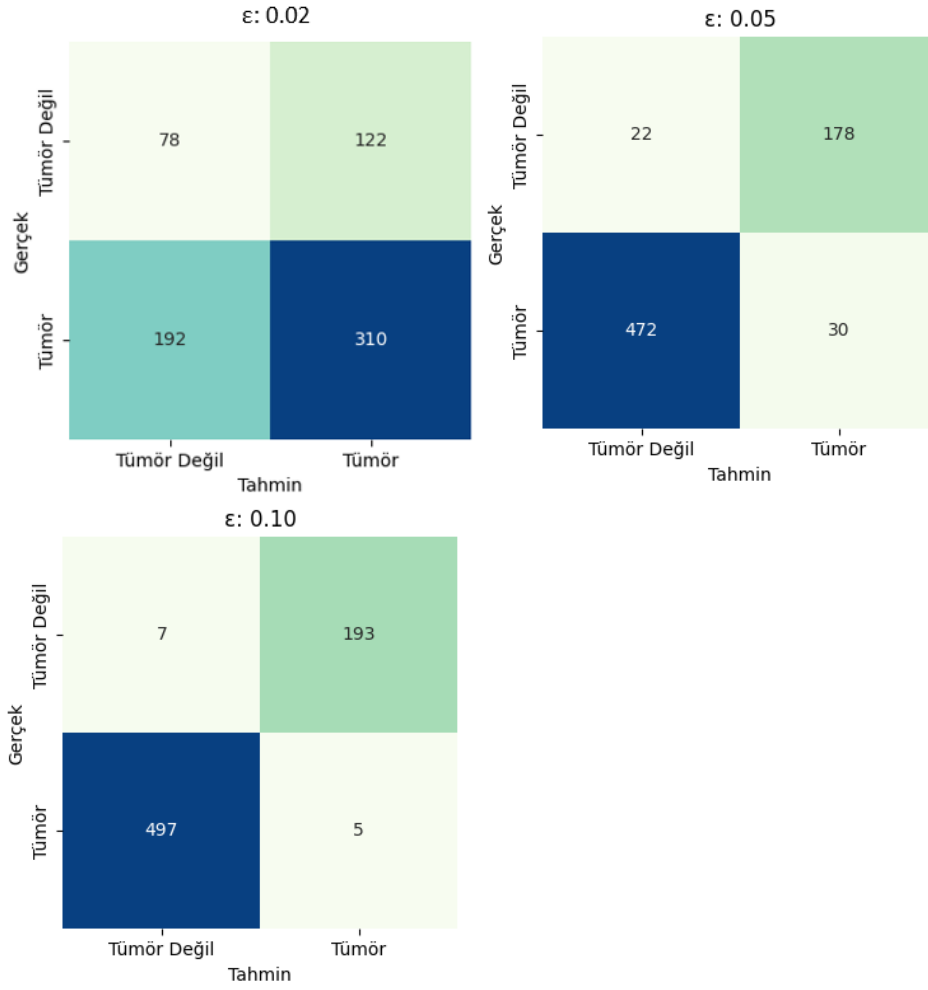


Şekil 6.63. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait ROC eğrileri

Şekil 6.63' e göre InceptionV3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %45'e kadar düşmektedir.

PGD saldırı

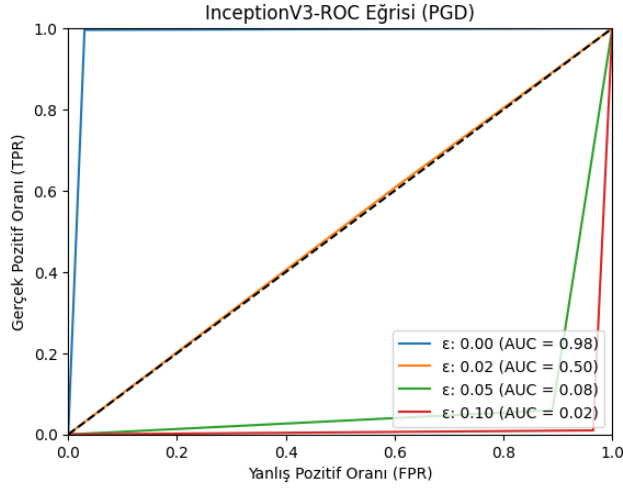
Şekil 6.64' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.64. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.64' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 192 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 122 örnek bulunmaktadır. Çizelge 6.3' e göre %55,27 doğruluk, %71,76 kesinlik, %61,75 geri çağırma, %66,38 F1 skoru ve %39 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 472 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 178 örnek bulunmaktadır. Çizelge 6.3' e göre %7,41 doğruluk, %14,12 kesinlik, %5,98 geri çağırma, %8,45 F1 skoru ve %11 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 497 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 193 örnek bulunmaktadır. Çizelge 6.3' e göre %1,71 doğruluk, %2,53 kesinlik, %1 geri çağırma, %1,43 F1 skoru ve %3,5

özgüllük elde etmektedir. Şekil 6.65'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.65. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde InceptionV3 modeline ait ROC eğrileri

Şekil 6.65' e göre InceptionV3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %2'ye kadar düşmektedir.

6.1.14. MobileNetV1

Temiz

MobileNetV1 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.66'da görülmektedir.

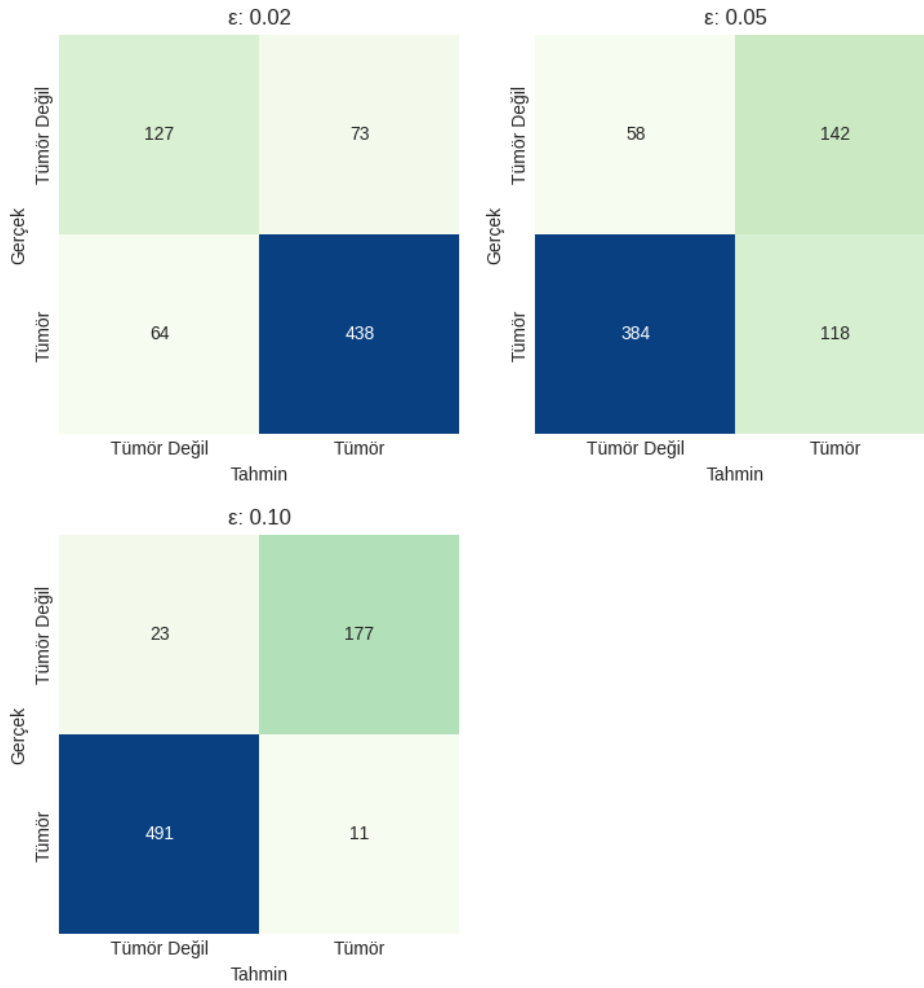
		Orjinal	
Gerçek	Tümör Değil	191	9
	Tümör	4	498
		Tümör Değil	Tümör
		Tahmin	

Şekil 6.66. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.66’ da görüldüğü gibi temiz test setinin MobileNetV1 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 4 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 9 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,15 doğruluk, %98,23 kesinlik, %99,2 geri çağırma, %98,71 F1 skoru ve %95,5 özgüllük elde etmektedir.

FGSM saldırısı

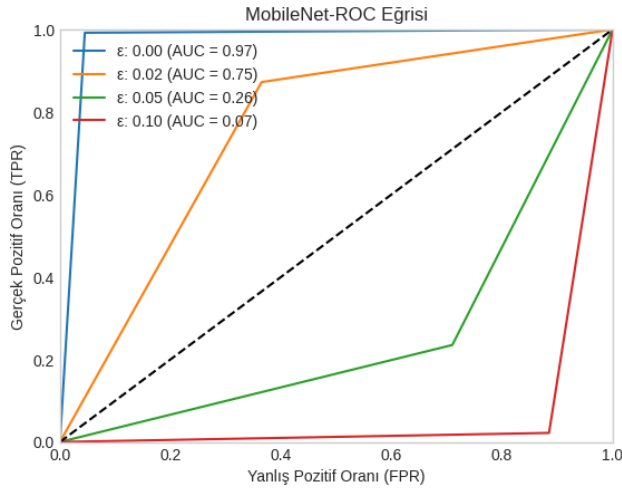
Şekil 6.67’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.67. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.67’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 64 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 73 örnek bulunmaktadır. Çizelge 6.2’ ye göre %80,48 doğruluk, %85,71 kesinlik, %87,25 geri çağırma, %86,48 F1 skoru ve %63,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 384 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 142 örnek bulunmaktadır. Çizelge 6.2’ ye göre %25,07 doğruluk, %45,39 kesinlik, %23,51 geri çağırma, %30,97 F1 skoru ve %29 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 491 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 177 örnek bulunmaktadır. Çizelge 6.2’ ye göre %4,84 doğruluk, %5,85 kesinlik, %2,19 geri çağırma,

%3,19 F1 skoru ve %11,5 özgüllük elde etmektedir. Şekil 6.68’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

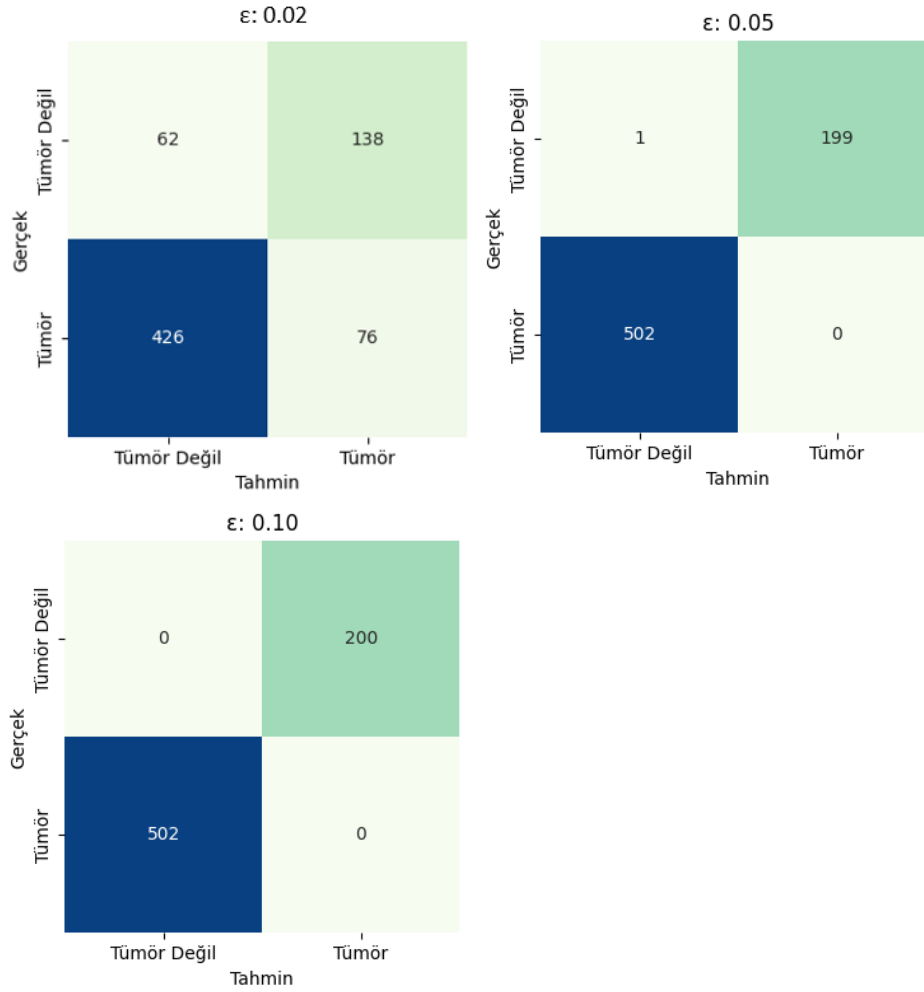


Şekil 6.68. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait ROC eğrileri

Şekil 6.68’ e göre MobileNetV1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %97 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %97’ den %7’ye kadar düşmektedir.

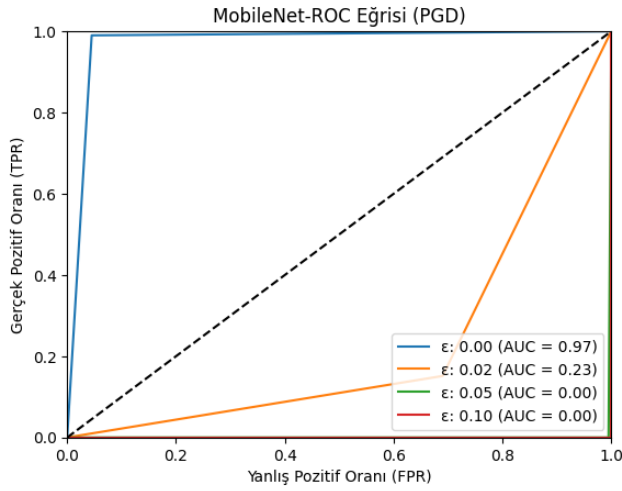
PGD saldırısı

Şekil 6.69’ da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.69. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.69’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 426 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 138 örnek bulunmaktadır. Çizelge 6.3’ e göre %19,66 doğruluk, %35,51 kesinlik, %15,14 geri çağırma, %21,23 F1 skoru ve %31 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 199 örnek bulunmaktadır. Çizelge 6.3’ e göre %0,14 doğruluk elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre %0 doğruluk elde etmektedir. Şekil 6.70’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



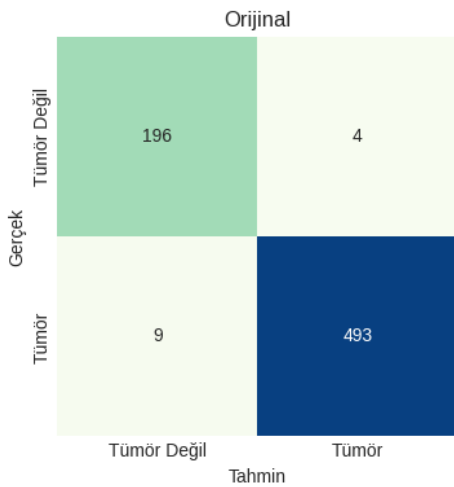
Şekil 6.70. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV1 modeline ait ROC eğrileri

Şekil 6.70' e göre MobileNetV1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %97 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %97' den %0'a kadar düşmektedir.

6.1.15. MobileNetV2

Temiz

MobileNetV2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.71'de görülmektedir.

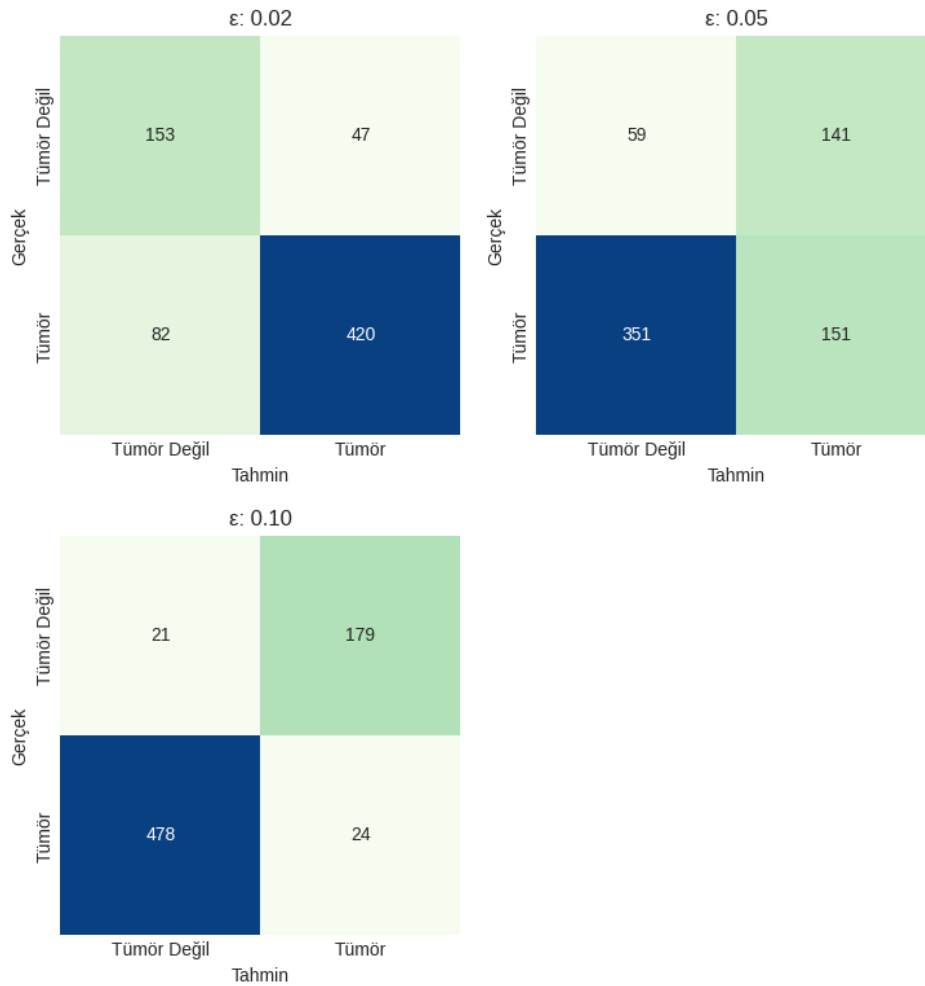


Şekil 6.71. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.71’ de görüldüğü gibi temiz test setinin MobileNetV2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 9 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 4 örnek bulunmaktadır. Çizelge 6.1’ e göre %98,15 doğruluk, %98,2 kesinlik, %99,21 geri çağırma, %98,7 F1 skoru ve %98 özgüllük elde etmektedir.

FGSM saldırısı

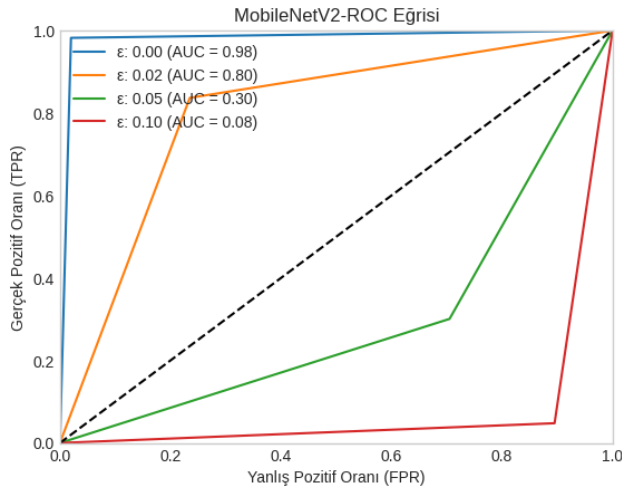
Şekil 6.72’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.72. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.72’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 82

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 47 örnek bulunmaktadır. Çizelge 6.2' ye göre %81,62 doğruluk, %89,94 kesinlik, %83,67 geri çağırma, %86,69 F1 skoru ve %76,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 351 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 141 örnek bulunmaktadır. Çizelge 6.2' ye göre %29,92 doğruluk, %51,71 kesinlik, %30,08 geri çağırma, %38,04 F1 skoru ve %29,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 478 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 179 örnek bulunmaktadır. Çizelge 6.2' ye göre %6,41 doğruluk, %11,82 kesinlik, %4,78 geri çağırma, %6,81 F1 skoru ve %10,5 özgüllük elde etmektedir. Şekil 6.73'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

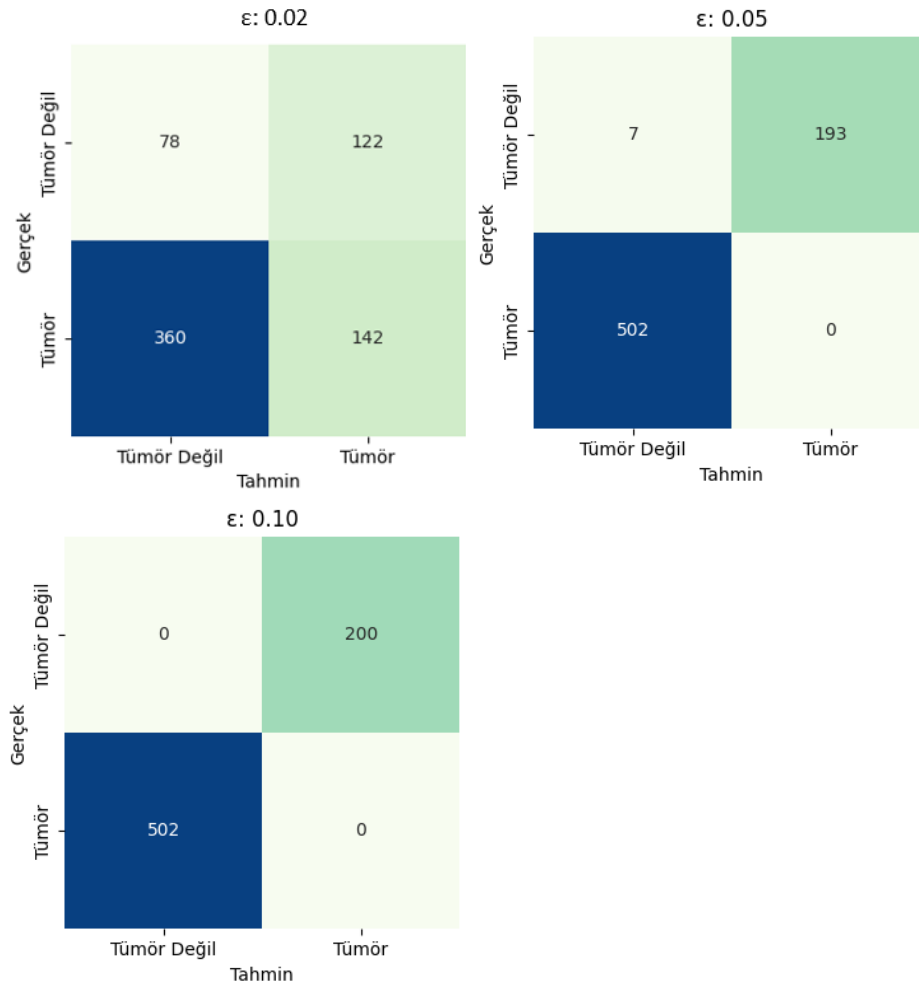


Şekil 6.73. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait ROC eğrileri

Şekil 6.73' e göre MobileNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %8'e kadar düşmektedir.

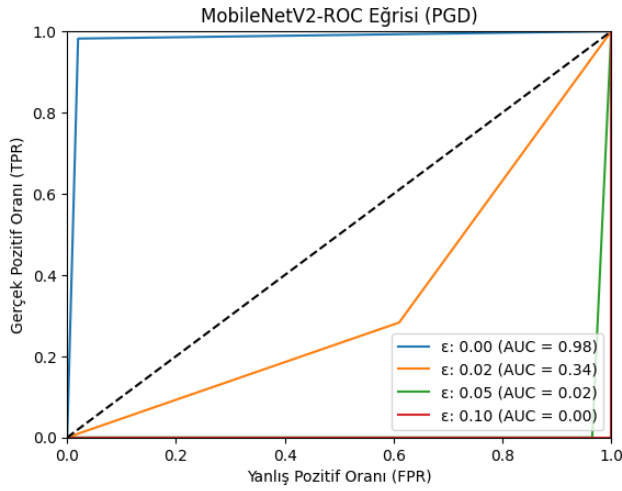
PGD saldırısı

Şekil 6.74' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.74. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.74’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 360 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 122 örnek bulunmaktadır. Çizelge 6.3’ e göre %31,34 doğruluk, %53,79 kesinlik, %28,29 geri çağırma, %37,08 F1 skoru ve %39 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 193 örnek bulunmaktadır. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre 0,05 pertürbasyon için %1 doğruluk ve 0,10 pertürbasyon için %0 doğruluk elde etmektedir. Şekil 6.75’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



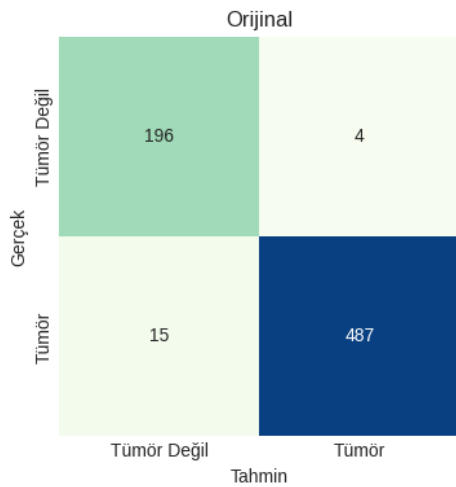
Şekil 6.75. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV2 modeline ait ROC eğrileri

Şekil 6.75' e göre MobileNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.16. MobileNetV3Small

Temiz

MobileNetV3Small modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.76'da görülmektedir.

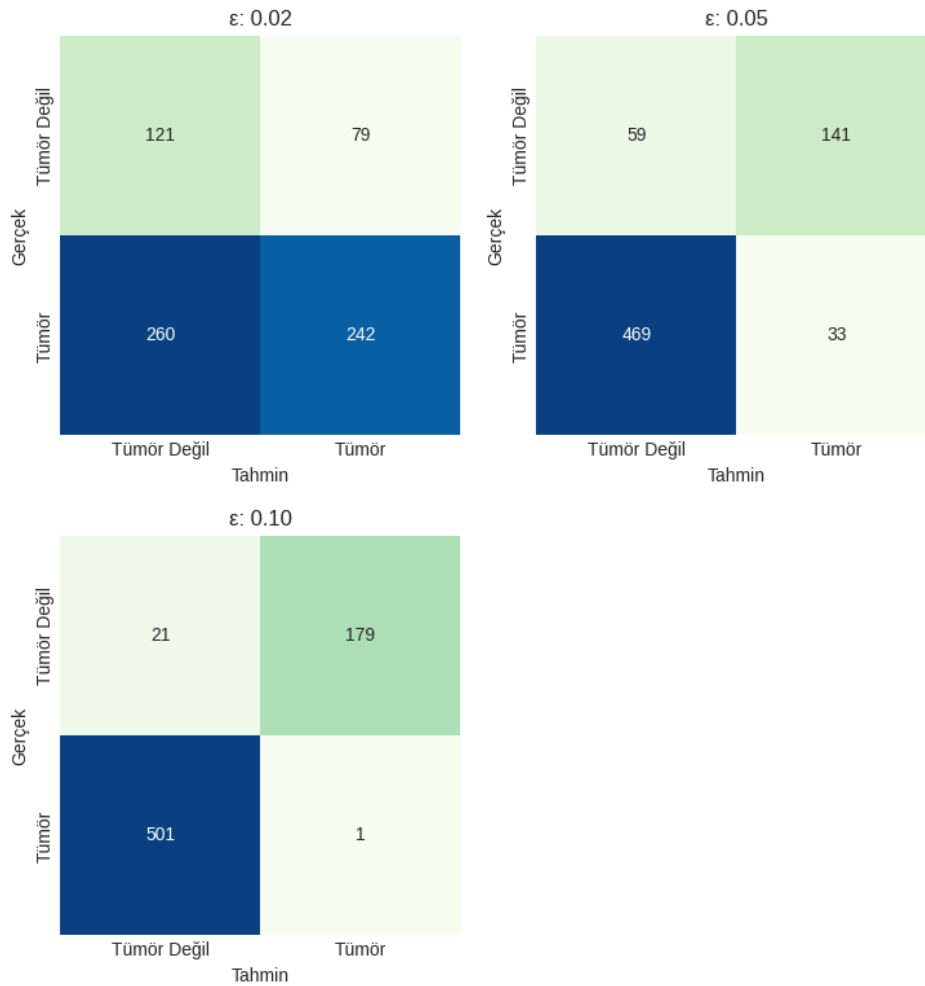


Şekil 6.76. Temiz Beyin Tümör MR görüntüleri üzerinde MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.76' da görüldüğü gibi temiz test setinin MobileNetV3Small modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 15 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 4 örnek bulunmaktadır. Çizelge 6.1' e göre %97,29 doğruluk, %99,19 kesinlik, %97,01 geri çağırma, %98,09 F1 skoru ve %98 özgüllük elde etmektedir.

FGSM saldırısı

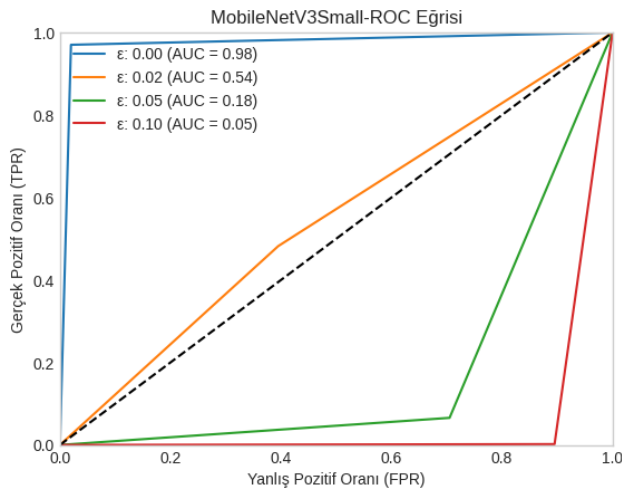
Şekil 6.77' de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.77. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerine MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.77' de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 260

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 79 örnek bulunmaktadır. Çizelge 6.2' ye göre %51,71 doğruluk, %75,39 kesinlik, %48,21 geri çağırma, %58,81 F1 skoru ve %60,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 469 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 141 örnek bulunmaktadır. Çizelge 6.2' ye göre %13,11 doğruluk, %18,97 kesinlik, %6,57 geri çağırma, %9,76 F1 skoru ve %29,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 501 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 179 örnek bulunmaktadır. Çizelge 6.2' ye göre %3,13 doğruluk, %0,55 kesinlik, %0,2 geri çağırma, %0,29 F1 skoru ve %10,5 özgüllük elde etmektedir. Şekil 6.78'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

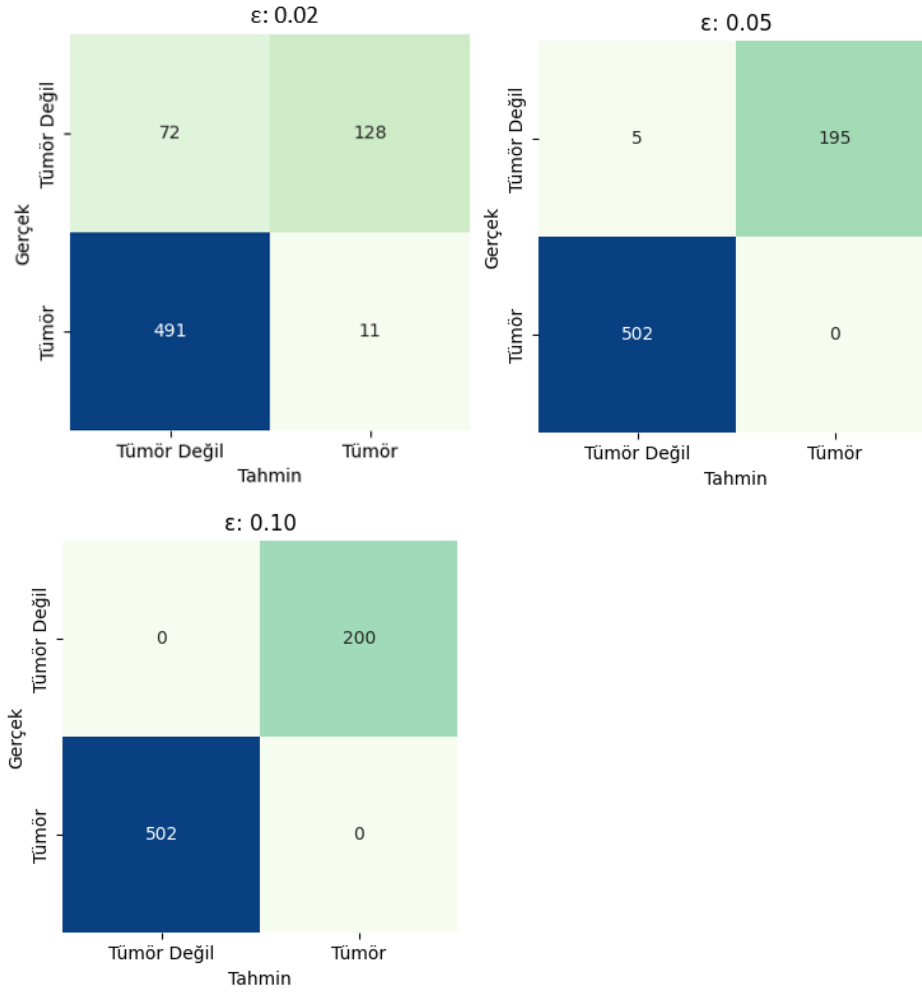


Şekil 6.78. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerine MobileNetV3Small modeline ait ROC eğrileri

Şekil 6.78' e göre MobileNetV3Small modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %5'e kadar düşmektedir.

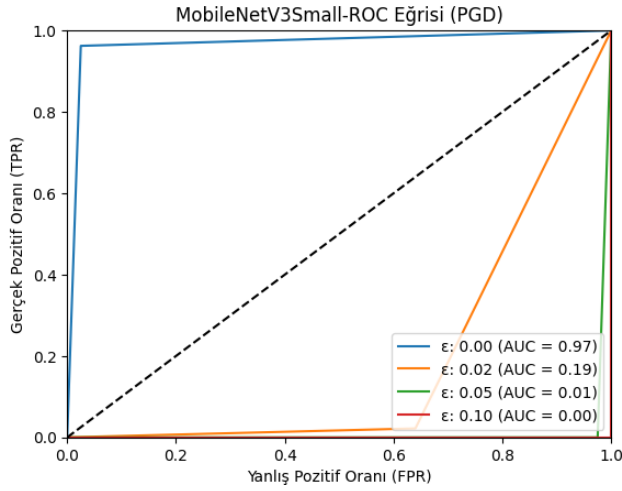
PGD saldırısı

Şekil 6.79' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.79. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.79’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 491 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 128 örnek bulunmaktadır. Çizelge 6.3’ e göre %11,82 doğruluk, %7,91 kesinlik, %2,19 geri çağırma, %3,43 F1 skoru ve %36 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 195 örnek bulunmaktadır. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre 0,05 pertürbasyon için %0,71 doğruluk ve 0,10 pertürbasyon için %0 doğruluk elde etmektedir. Şekil 6.80’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



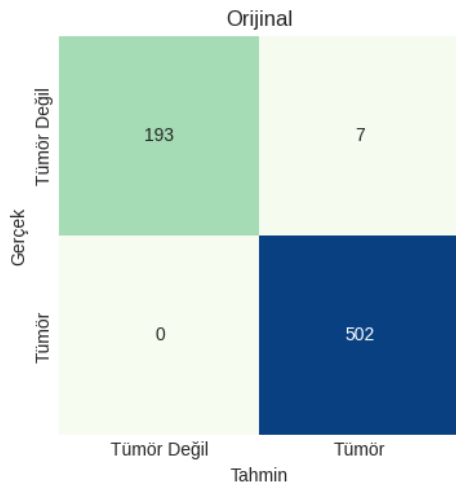
Şekil 6.80. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde MobileNetV3Small modeline ait ROC eğrileri

Şekil 6.80' e göre MobileNetV3Small modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %97 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %97' den %0'a kadar düşmektedir.

6.1.17. NASNetMobile

Temiz

NASNetMobile modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.81'da görülmektedir.

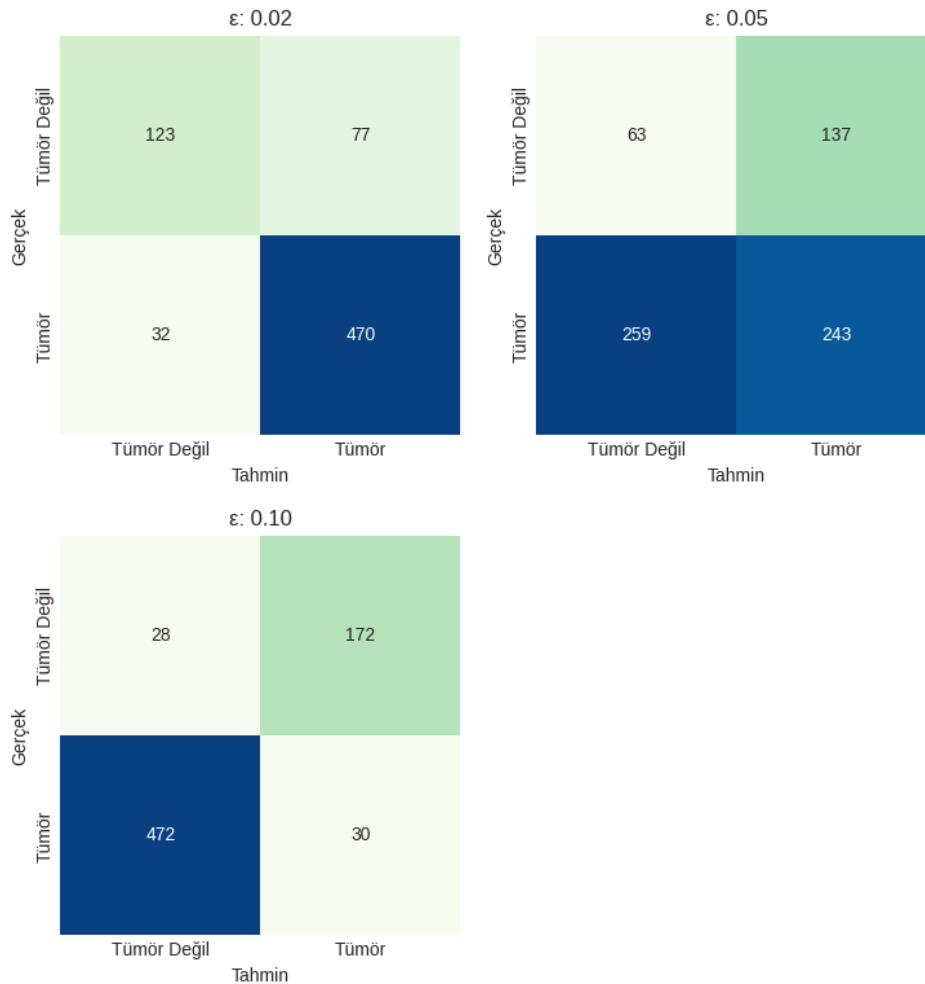


Şekil 6.81. Temiz Beyin Tümör MR görüntüleri üzerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.81’ de görüldüğü gibi temiz test setinin NASNetMobile modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) bulunmamaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 7 örnek bulunmaktadır. Çizelge 6.1’ e göre %99 doğruluk, %98,63 kesinlik, %100 geri çağırma, %99,31 F1 skoru ve %96,5 özgüllük elde etmektedir.

FGSM saldırısı

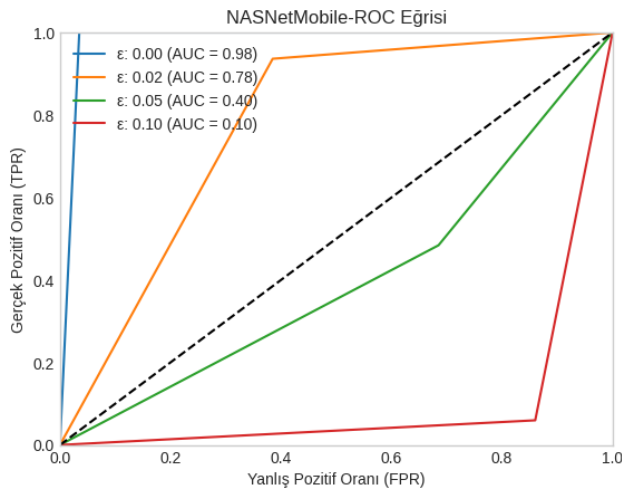
Şekil 6.82’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.82. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.82’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 32

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 77 örnek bulunmaktadır. Çizelge 6.2' ye göre %84,47 doğruluk, %85,92 kesinlik, %93,63 geri çağırma, %89,61 F1 skoru ve %61,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 259 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 137 örnek bulunmaktadır. Çizelge 6.2' ye göre %43,59 doğruluk, %63,95 kesinlik, %48,41 geri çağırma, %55,1 F1 skoru ve %31,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 472 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 172 örnek bulunmaktadır. Çizelge 6.2' ye göre %8,26 doğruluk, %14,85 kesinlik, %5,98 geri çağırma, %8,52 F1 skoru ve %14 özgüllük elde etmektedir. Şekil 6.83'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

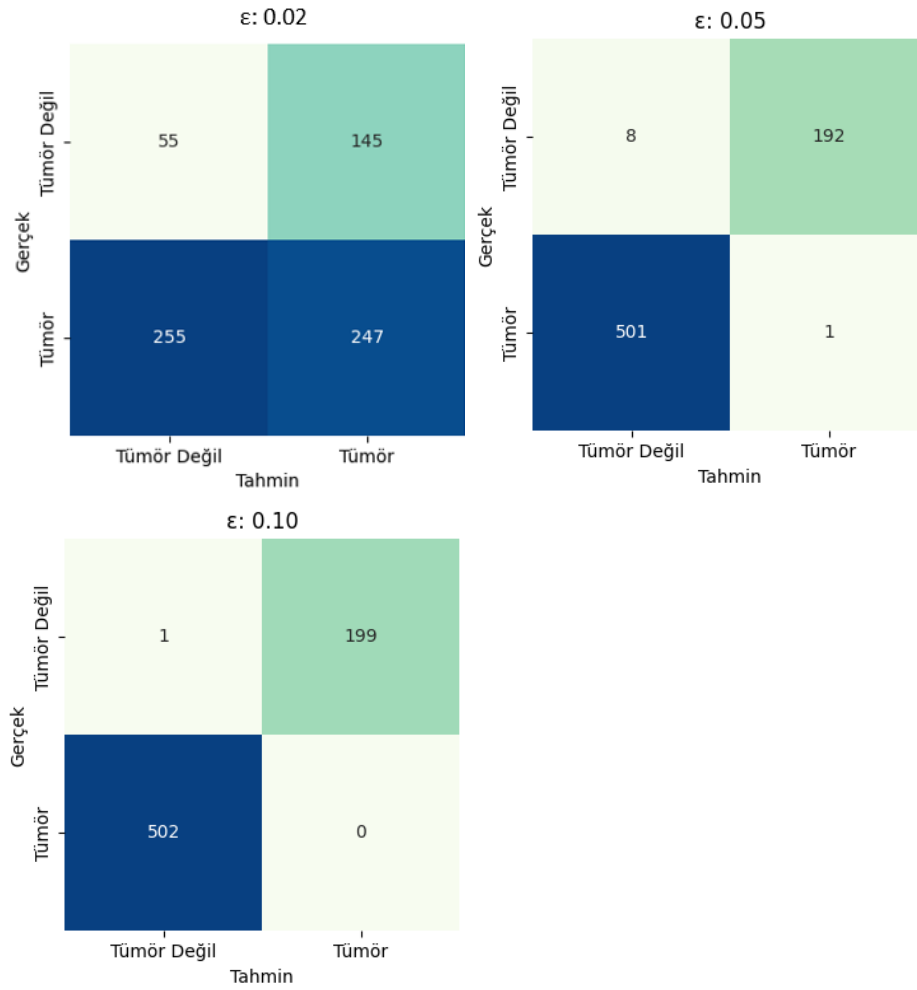


Şekil 6.83. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait ROC eğrileri

Şekil 6.83' e göre NASNetMobile modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %10'a kadar düşmektedir.

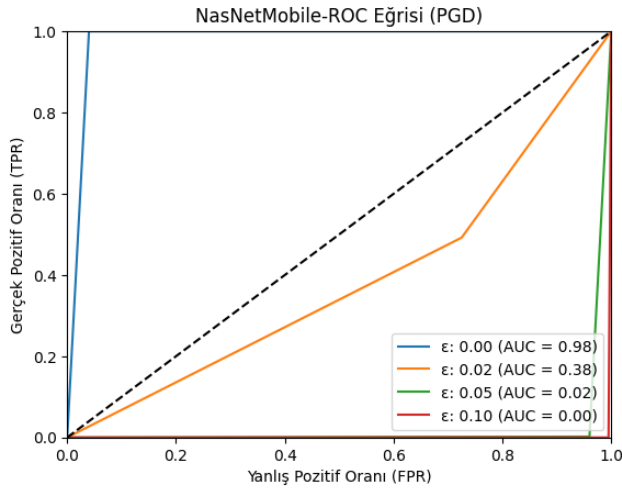
PGD saldırısı

Şekil 6.84' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.84. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.84’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 255 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 145 örnek bulunmaktadır. Çizelge 6.3’ e göre %43,02 doğruluk, %63,01 kesinlik, %49,2 geri çağırma, %55,26 F1 skoru ve %27,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 501 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 192 örnek bulunmaktadır. Çizelge 6.3’ e göre %1,28 doğruluk, %0,52 kesinlik, %0,2 geri çağırma, %0,29 F1 skoru ve %4 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 199 örnek bulunmaktadır. Çizelge 6.85’ e göre %14 doğruluk elde etmektedir. Şekil 6.5’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



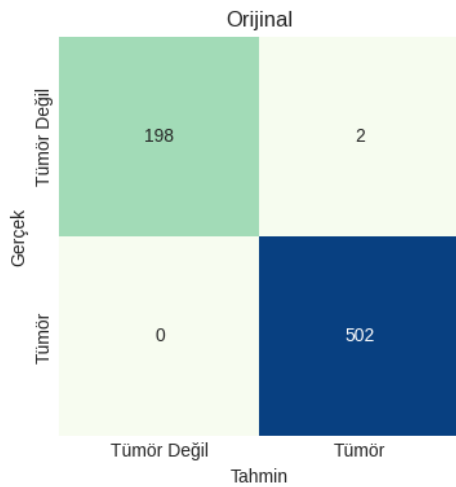
Şekil 6.85. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde NASNetMobile modeline ait ROC eğrileri

Şekil 6.85' e göre NASNetMobile modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %98 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %98' den %0'a kadar düşmektedir.

6.1.18. ResNet50V2

Temiz

ResNet50V2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.86'da görülmektedir.

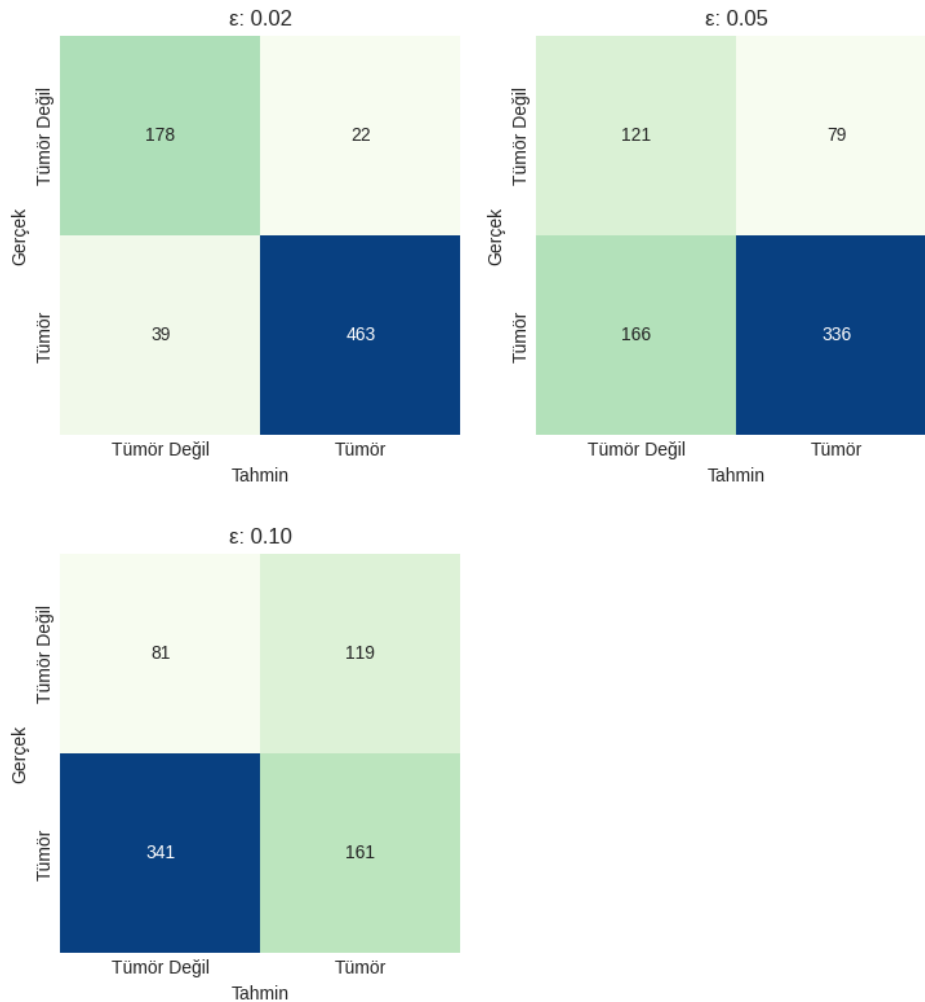


Şekil 6.86. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.86’ da görüldüğü gibi temiz test setinin ResNet50V2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) bulunmamaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 2 örnek bulunmaktadır. Çizelge 6.1’ e göre %99,72 doğruluk, %99,6 kesinlik, %100 geri çağırma, %99,8 F1 skoru ve %99 özgüllük elde etmektedir.

FGSM saldırısı

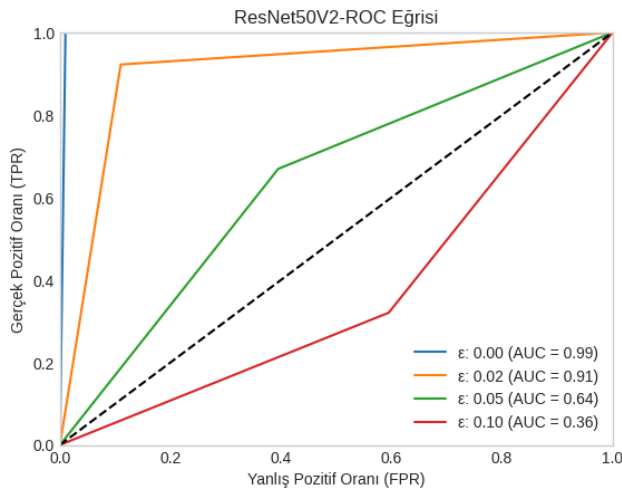
Şekil 6.87’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.87. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.87’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 39

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 22 örnek bulunmaktadır. Çizelge 6.2' ye göre %91,31 doğruluk, %95,46 kesinlik, %92,23 geri çağırma, %93,82 F1 skoru ve %89 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 166 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 79 örnek bulunmaktadır. Çizelge 6.2' ye göre %65,1 doğruluk, %80,96 kesinlik, %66,93 geri çağırma, %73,28 F1 skoru ve %60,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 341 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 119 örnek bulunmaktadır. Çizelge 6.2' ye göre %34,47 doğruluk, %57,5 kesinlik, %32,07 geri çağırma, %41,18 F1 skoru ve %40,5 özgüllük elde etmektedir. Şekil 6.88'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

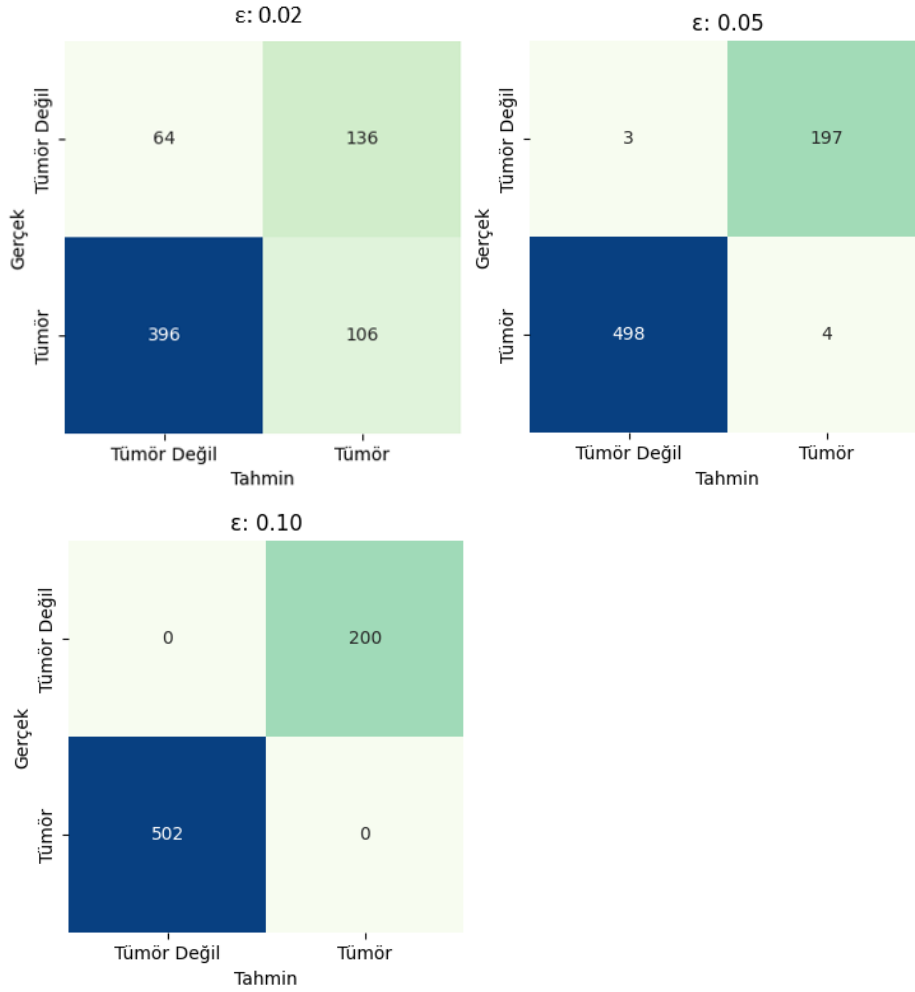


Şekil 6.88. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait ROC eğrileri

Şekil 6.88' e göre ResNet50V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %36'ya kadar düşmektedir.

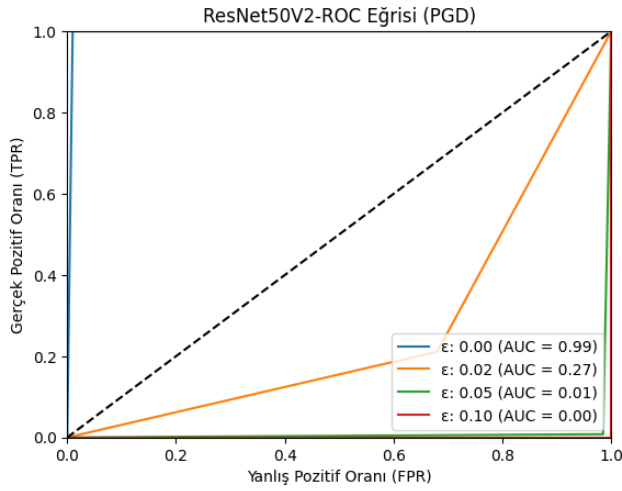
PGD saldırısı

Şekil 6.89' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.89. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.89’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 396 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 136 örnek bulunmaktadır. Çizelge 6.3’ e göre %24,22 doğruluk, %43,8 kesinlik, %21,12 geri çağırma, %28,5 F1 skoru ve %32 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 498 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 197 örnek bulunmaktadır. Çizelge 6.3’ e göre %1 doğruluk, %1,99 kesinlik, %0,8 geri çağırma, %1,14 F1 skoru ve %1,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre %0 doğruluk elde etmektedir. Şekil 6.90’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.90. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet50V2 modeline ait ROC eğrileri

Şekil 6.90' a göre ResNet50V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %0'a kadar düşmektedir.

6.1.19. ResNet101V2

Temiz

ResNet101V2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.91'de görülmektedir.

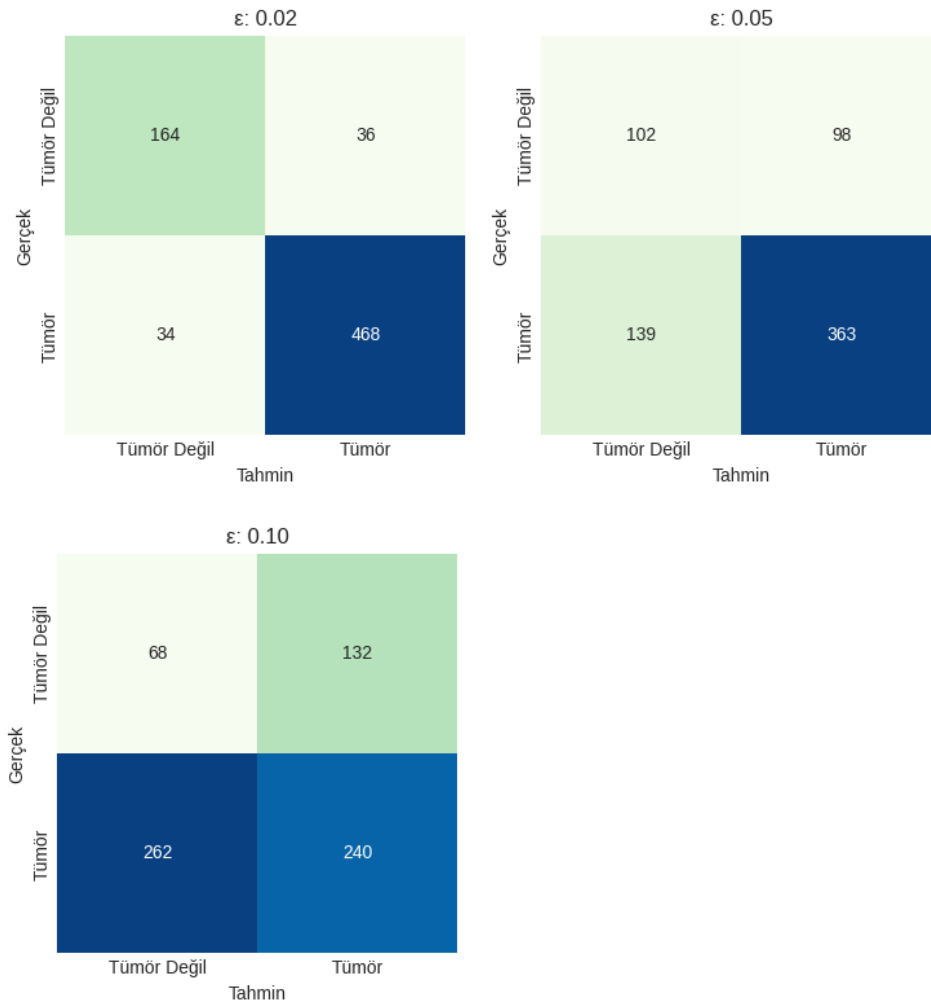
		Orijinal	
Gerçek	Tümör Değil	195	5
	Tümör	1	501
		Tümör Değil	Tümör
		Tahmin	

Şekil 6.91. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.91’ de görüldüğü gibi temiz test setinin ResNet101V2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 1 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %99,15 doğruluk, %99,01 kesinlik, %99,8 geri çağırma, %99,41 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

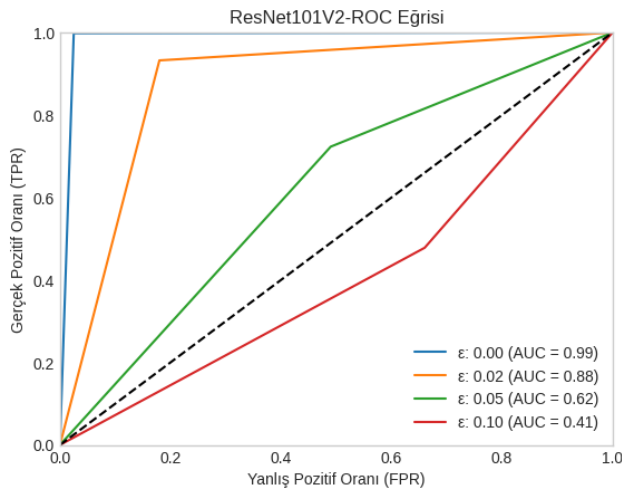
Şekil 6.92’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.92. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.92’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 34

örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 36 örnek bulunmaktadır. Çizelge 6.2' ye göre %90,03 doğruluk, %92,86 kesinlik, %93,23 geri çağırma, %93,04 F1 skoru ve %82 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 139 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 98 örnek bulunmaktadır. Çizelge 6.2' ye göre %66,24 doğruluk, %78,74 kesinlik, %72,31 geri çağırma, %75,39 F1 skoru ve %51 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 262 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 132 örnek bulunmaktadır. Çizelge 6.2' ye göre %43,88 doğruluk, %64,52 kesinlik, %47,81 geri çağırma, %54,92 F1 skoru ve %34 özgüllük elde etmektedir. Şekil 6.93'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

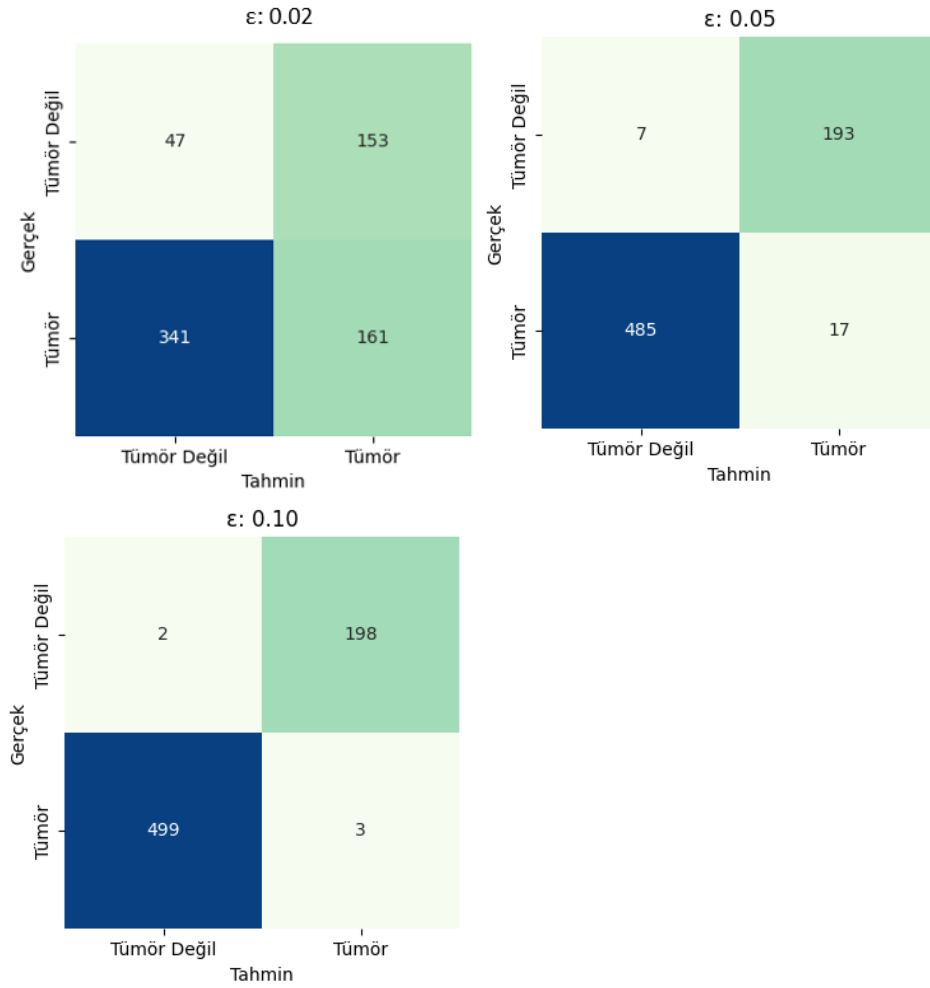


Şekil 6.93. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 modeline ait ROC eğrileri

Şekil 6.93' e göre ResNet101V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %41'e kadar düşmektedir.

PGD saldırısı

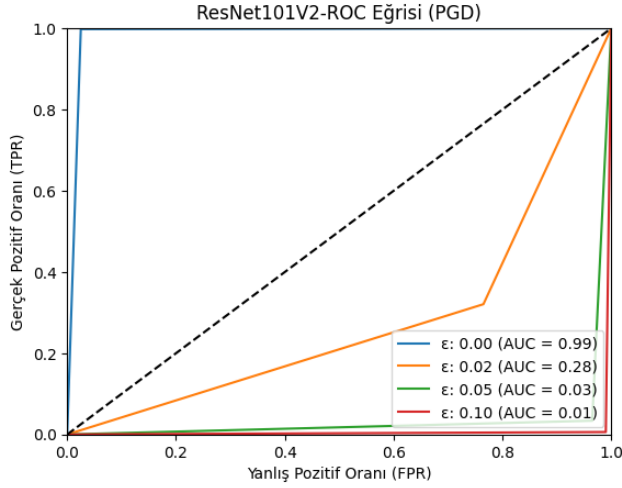
Şekil 6.94' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.94. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.94' te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 341 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 153 örnek bulunmaktadır. Çizelge 6.3' e göre %29,63 doğruluk, %51,27 kesinlik, %32,07 geri çağırma, %39,46 F1 skoru ve %23,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 485 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 193 örnek bulunmaktadır. Çizelge 6.3' e göre %3,42 doğruluk, %8,10 kesinlik, %3,39 geri çağırma, %4,76 F1 skoru ve %3,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 499 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 198 örnek bulunmaktadır. Çizelge 6.3' e göre %0,71 doğruluk, %1,49 kesinlik, %0,6 geri çağırma, %0,85 F1 skoru ve %1

özgüllük elde etmektedir. Şekil 6.95'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.95. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet101V2 modeline ait ROC eğrileri

Şekil 6.95' e göre ResNet101V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %1'e kadar düşmektedir.

6.1.20. ResNet152V2

Temiz

ResNet152V2 modeli ile saldırıya uğramamış orijinal Beyin Tümör MR test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.96'da görülmektedir.

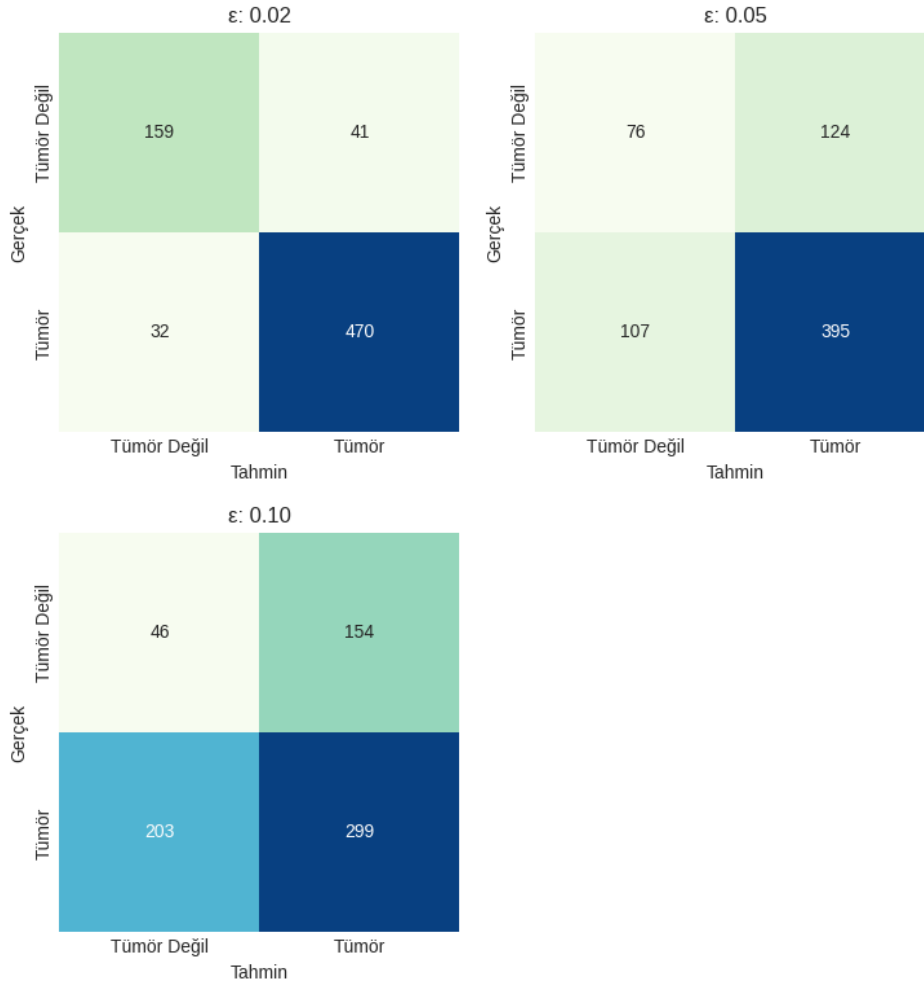
		Orjinal	
Gerçek	Tümör Değil	195	5
	Tümör	1	501
		Tümör Değil	Tümör
		Tahmin	

Şekil 6.96. Temiz Beyin Tümör MR görüntüleri üzerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.96’ da görüldüğü gibi temiz test setinin ResNet152V2 modeli ile analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 1 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 5 örnek bulunmaktadır. Çizelge 6.1’ e göre %99,15 doğruluk, %99,01 kesinlik, %99,8 geri çağırma, %99,41 F1 skoru ve %97,5 özgüllük elde etmektedir.

FGSM saldırısı

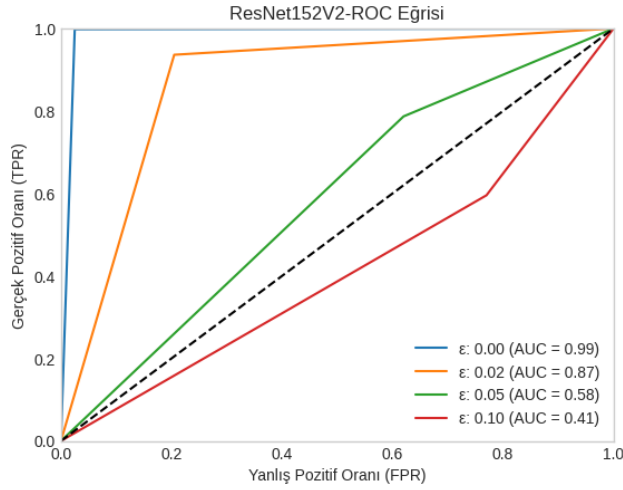
Şekil 6.97’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.97. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.97’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 32 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 41 örnek bulunmaktadır. Çizelge 6.2’ ye göre %89,6 doğruluk, %91,98 kesinlik, %93,63 geri çağırma, %92,79 F1 skoru ve %79,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 107 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 124 örnek bulunmaktadır. Çizelge 6.2’ ye göre %67,09 doğruluk, %76,11 kesinlik, %78,69 geri çağırma, %77,38 F1 skoru ve %38 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 203 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 154 örnek bulunmaktadır. Çizelge 6.2’ ye göre %49,15 doğruluk, %66 kesinlik, %59,56 geri çağırma,

%62,62 F1 skoru ve %23 özgüllük elde etmektedir. Şekil 6.98’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

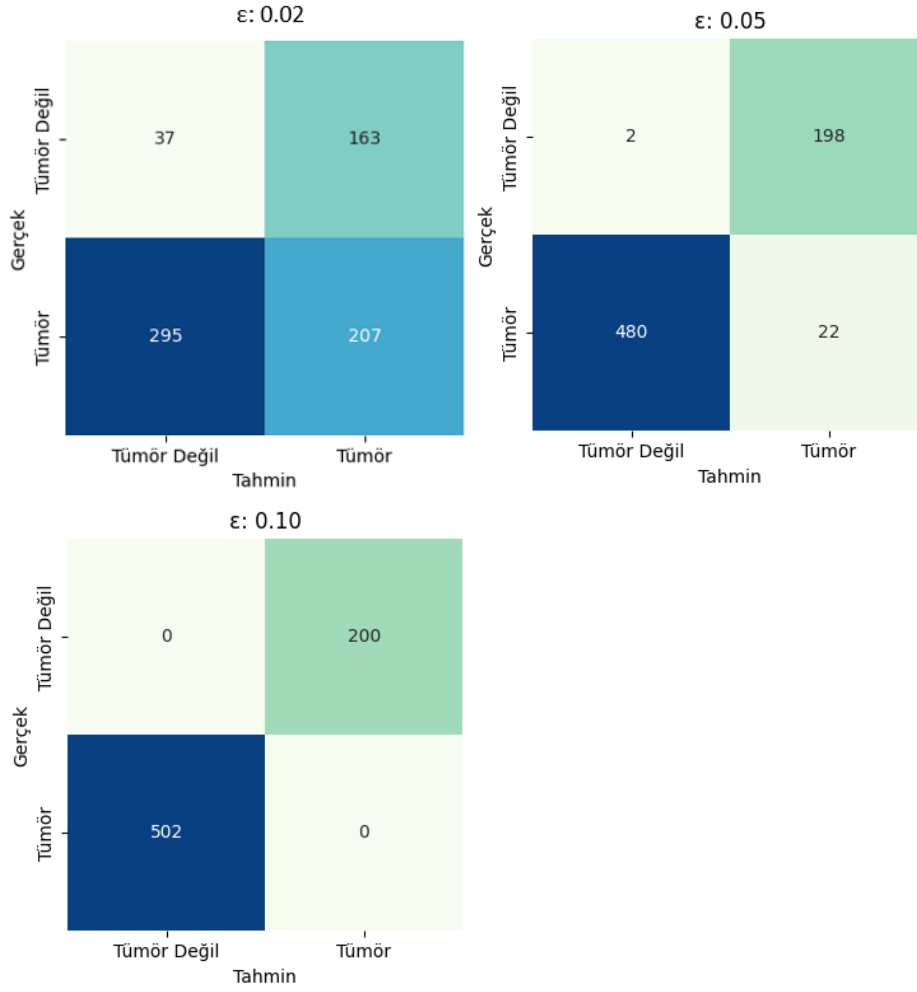


Şekil 6.98. FGSM saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait ROC eğrileri

Şekil 6.98’ e göre ResNet152V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99’ dan %41’e kadar düşmektedir.

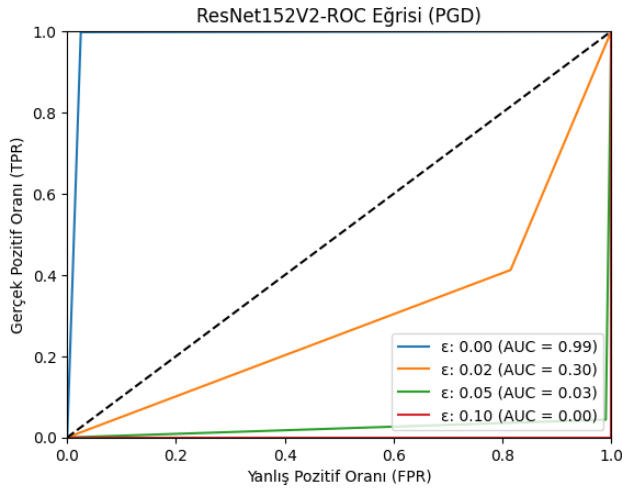
PGD saldırısı

Şekil 6.99’ da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.99. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.99’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 295 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 163 örnek bulunmaktadır. Çizelge 6.3’ e göre %34,76 doğruluk, %55,95 kesinlik, %41,24 geri çağırma, %47,48 F1 skoru ve %18,5 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 480 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 198 örnek bulunmaktadır. Çizelge 6.3’ e göre %3,42 doğruluk, %10 kesinlik, %4,38 geri çağırma, %6,09 F1 skoru ve %1 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 502 örnek bulunmaktadır. Normal MR görüntülerinin tümör olarak sınıflandırıldığı (HP) 200 örnek bulunmaktadır. Çizelge 6.3’ e göre %0 doğruluk elde etmektedir. Şekil 6.100’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.100. PGD saldırısına uğrayan Beyin Tümör MR görüntülerinde ResNet152V2 modeline ait ROC eğrileri

Şekil 6.100' e göre ResNet152V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %99 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %99' dan %0'a kadar düşmektedir.

6.2. Böbrek BT Görüntü Seti Deney Sonuçları

Bu bölümde, Böbrek BT görüntü seti üzerinde gerçekleştirilen deney bulguları değerlendirilmektedir. Bulgular her bir derin öğrenme modeli için temiz test görüntüleri ile yapılan analiz, FGSM saldırısına uğrayan görüntülerin analizi ve PGD saldırısına uğrayan görüntülerin analizi olmak üzere üç kapsamda ele alınmaktadır. Çizelge 6.4'te temiz test görüntülerinin model bazında sınıflandırma başarıları yer almaktadır. Çizelge 6.5' te farklı pertürbasyon parametreleri ile gerçekleştirilen FGSM saldırısı sonucunda modellerin sınıflandırma başarıları görülmektedir. Çizelge 6.6' da ise farklı parametre senaryoları ile gerçekleştirilen PGD saldırısı sonucunda modellerin sınıflandırma başarıları verilmektedir.

Çizelge 6.4. Böbrek BT test görüntüleri ile modellerin elde ettiği performans değerleri

Modeller	Orijinal - Temiz				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	100,0	100,0	100,0	100,0	100,0
VGG16	100,0	100,0	100,0	100,0	100,0
VGG19	100,0	100,0	100,0	100,0	100,0
DenseNet121	100,0	100,0	100,0	100,0	100,0
DenseNet169	100,0	100,0	100,0	100,0	100,0
DenseNet201	100,0	100,0	100,0	100,0	100,0
EfficientNetV2B0	100,0	100,0	100,0	100,0	100,0
EfficientNetV2B1	99,86	100,0	99,56	99,78	100,0
EfficientNetV2B2	99,86	100,0	99,56	99,78	100,0
EfficientNetV2B3	100,0	100,0	100,0	100,0	100,0
EfficientNetV2S	100,0	100,0	100,0	100,0	100,0
InceptionResNetV2	99,86	99,57	100,0	99,78	99,80
InceptionV3	100,0	100,0	100,0	100,0	100,0
MobileNet	100,0	100,0	100,0	100,0	100,0
MobileNetV2	100,0	100,0	100,0	100,0	100,0
MobileNetV3Small	100,0	100,0	100,0	100,0	100,0
NasNetMobile	100,0	100,0	100,0	100,0	100,0
ResNet50V2	100,0	100,0	100,0	100,0	100,0
ResNet101V2	100,0	100,0	100,0	100,0	100,0
ResNet152V2	100,0	100,0	100,0	100,0	100,0

Çizelge 6.5. Böbrek BT görüntüleri üzerine uygulanan FGSM saldırıları sonucunda modellerin elde ettiği performans değerleri

Modeller	FGSM- 0,02 epsilon					FGSM - 0,05 epsilon					FGSM - 0,1 epsilon				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	88,72	82,30	81,22	81,76	92,11	19,02	4,01	6,99	5,10	24,46	1,77	0,99	2,18	1,36	1,58
VGG16	90,08	80,95	89,08	84,82	90,53	19,02	7,03	13,10	9,15	21,70	0,14	0,0	0,0	-	0,20
VGG19	92,94	90,05	86,90	88,45	95,66	19,29	3,09	5,24	3,88	25,64	0,41	0,0	0,0	-	0,59
DenseNet121	81,79	65,47	87,77	75,0	79,09	39,95	24,46	44,54	31,58	37,87	0,54	0,40	0,87	0,54	0,40
DenseNet169	85,87	72,73	87,34	79,37	85,21	33,15	23,0	48,91	31,29	26,04	2,72	1,97	4,37	2,72	1,97
DenseNet201	93,75	87,35	93,45	90,30	93,89	56,93	38,36	63,32	47,78	54,04	3,13	1,60	3,49	2,20	2,96
EfficientNetV2B0	50,0	26,12	33,19	29,23	57,59	4,89	0,21	0,44	0,29	6,90	0,0	0,0	0,0	-	0,0
EfficientNetV2B1	64,40	43,27	46,29	44,73	72,58	4,89	0,21	0,44	0,29	6,90	0,0	0,0	0,0	-	0,0
EfficientNetV2B2	87,09	79,13	79,48	79,30	90,53	17,66	3,23	5,68	4,11	23,08	1,63	0,0	0,0	-	2,37
EfficientNetV2B3	81,11	69,57	69,87	69,72	86,19	21,06	5,10	8,73	6,44	26,63	1,36	0,0	0,0	-	1,97
EfficientNetV2S	62,09	41,78	55,46	47,66	65,09	7,47	1,50	3,06	2,01	9,47	1,90	0,0	0,0	-	2,76
InceptionResNetV2	83,70	68,86	86,90	76,83	82,25	37,64	19,74	32,75	24,63	39,74	15,90	5,28	10,04	6,92	18,54
InceptionV3	94,02	89,03	92,14	90,56	94,87	40,76	20,51	31,44	24,83	44,97	11,14	3,09	6,11	4,11	13,41
MobileNet	78,80	67,14	62,45	64,71	86,19	6,79	0,65	1,31	0,87	9,27	0,14	0,0	0,0	-	0,20
MobileNetV2	86,69	81,34	74,24	77,63	92,31	20,92	5,54	9,61	7,03	26,04	0,54	0,59	1,31	0,81	0,20
MobileNetV3Small	16,03	7,06	13,97	9,38	16,96	0,14	0,20	0,44	0,27	0,0	0,0	0,0	0,0	-	0,0
NasNetMobile	44,16	21,20	29,26	24,59	50,89	0,27	0,20	0,44	0,27	0,20	0,0	0,0	0,0	-	0,0
ResNet50V2	93,61	88,56	91,27	89,89	94,68	26,09	11,86	21,40	15,27	28,21	4,89	2,23	4,80	3,05	4,93
ResNet101V2	94,97	89,67	94,76	92,14	95,07	30,03	15,29	27,51	19,66	31,16	9,24	3,79	7,86	5,11	9,86
ResNet152V2	98,51	98,66	96,51	97,57	99,41	42,53	20,06	28,38	23,51	48,92	17,80	5,87	10,92	7,63	20,91

Çizelge 6.6. Böbrek BT görüntüleri üzerine uygulanan PGD saldırıları sonucunda modellerin elde ettiği performans değerleri

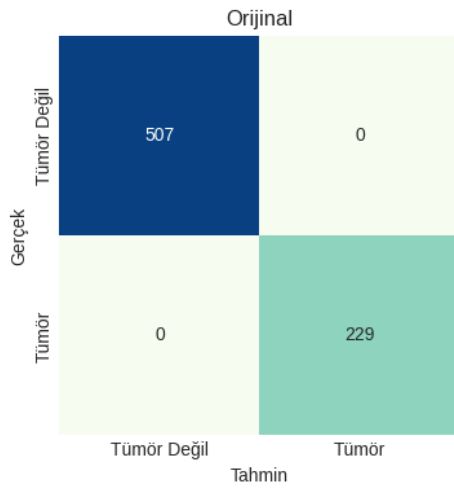
Modeller	PGD- 0,02 epsilon					PGD- 0,05 epsilon					PGD - 0,1 epsilon				
	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük	Doğruluk	Kesinlik	Geri Çağırma	F1 skoru	Özgüllük
Xception	22,55	4,78	7,86	5,94	29,19	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
VGG16	37,64	22,75	41,92	29,49	35,70	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
VGG19	39,54	22,02	37,12	27,64	40,63	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
DenseNet121	15,63	6,83	13,54	9,08	16,57	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
DenseNet169	11,41	7,45	16,15	10,19	9,27	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
DenseNet201	29,62	17,08	32,75	22,46	28,21	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
EfficientNetV2B0	19,43	5,17	9,17	6,61	24,06	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
EfficientNetV2B1	22,01	6,55	11,35	8,31	26,83	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
EfficientNetV2B2	44,29	24,65	38,43	30,03	46,94	0,82	0,0	0,0	-	1,18	0,0	0,0	0,0	-	0,0
EfficientNetV2B3	41,71	18,35	25,33	21,28	49,11	2,72	0,0	0,0	-	3,95	0,0	0,0	0,0	-	0,0
EfficientNetV2S	7,75	2,33	4,80	3,14	9,07	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
InceptionResNetV2	34,92	20,66	38,43	26,87	33,33	1,22	1,18	2,62	1,62	0,59	0,0	0,0	0,0	-	0,0
InceptionV3	25,68	14,51	28,38	19,20	24,46	1,50	0,40	0,87	0,55	1,78	0,0	0,0	0,0	-	0,0
MobileNet	8,42	2,56	5,24	3,44	9,86	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
MobileNetV2	30,44	12,86	21,40	16,07	34,52	0,27	0,39	0,87	0,54	0,0	0,0	0,0	0,0	-	0,0
MobileNetV3Small	0,68	0,59	1,31	0,81	0,40	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
NasNetMobile	2,45	0,61	1,31	0,83	2,96	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
ResNet50V2	6,52	3,82	8,30	5,23	5,72	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
ResNet101V2	8,83	3,57	7,42	4,82	9,47	0,0	0,0	0,0	-	0,0	0,0	0,0	0,0	-	0,0
ResNet152V2	22,28	6,58	11,35	8,33	27,22	0,54	0,0	0,0	-	0,79	0,0	0,0	0,0	-	0,0

Çizelge 6.4, Çizelge 6.5 ve Çizelge 6.6’ da verilen ve sınıflandırma başarısını temsil eden performans metrik değerleri bölümün devamında yer alan ilgili derin öğrenme modelinin altında detaylı olarak incelenmektedir.

6.2.1. Xception

Temiz

Xception modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.101’de görülmektedir.

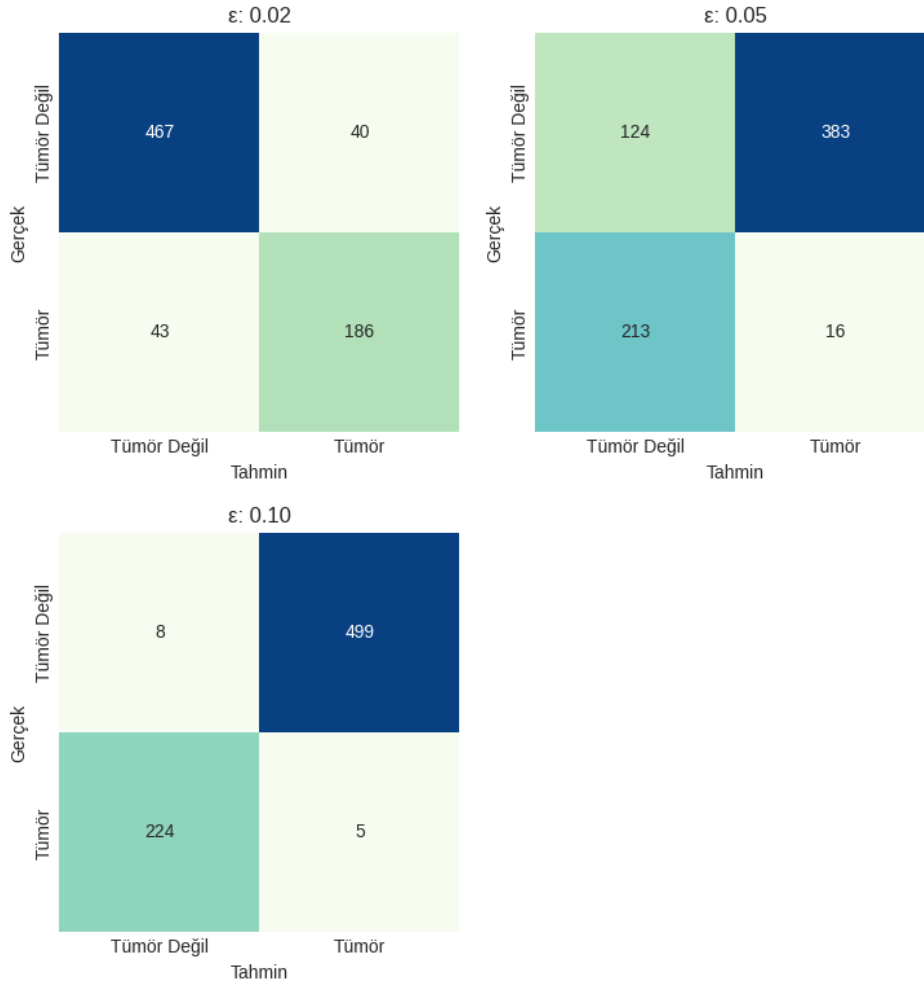


Şekil 6.101. Temiz Böbrek BT görüntüleri üzerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.101’ de görüldüğü gibi temiz test setinin Xception modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

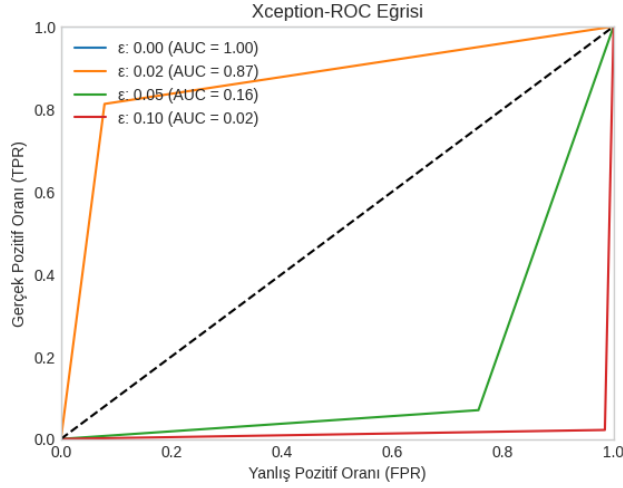
Şekil 6.102’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.102. FGSM saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.102’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 43 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 40 örnek bulunmaktadır. Çizelge 6.5’ e göre %88,72 doğruluk, %82,3 kesinlik, %81,22 geri çağırma, %81,76 F1 skoru ve %92,11 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 213 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 383 örnek bulunmaktadır. Çizelge 6.5’ e göre %19,02 doğruluk, %4,01 kesinlik, %6,99 geri çağırma, %5,1 F1 skoru ve %24,46 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 224 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 499 örnek bulunmaktadır. Çizelge 6.5’ e göre %1,77 doğruluk, %0,99 kesinlik, %2,18 geri çağırma, %1,36 F1 skoru ve

%1,58 özgüllük elde etmektedir. Şekil 6.103'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

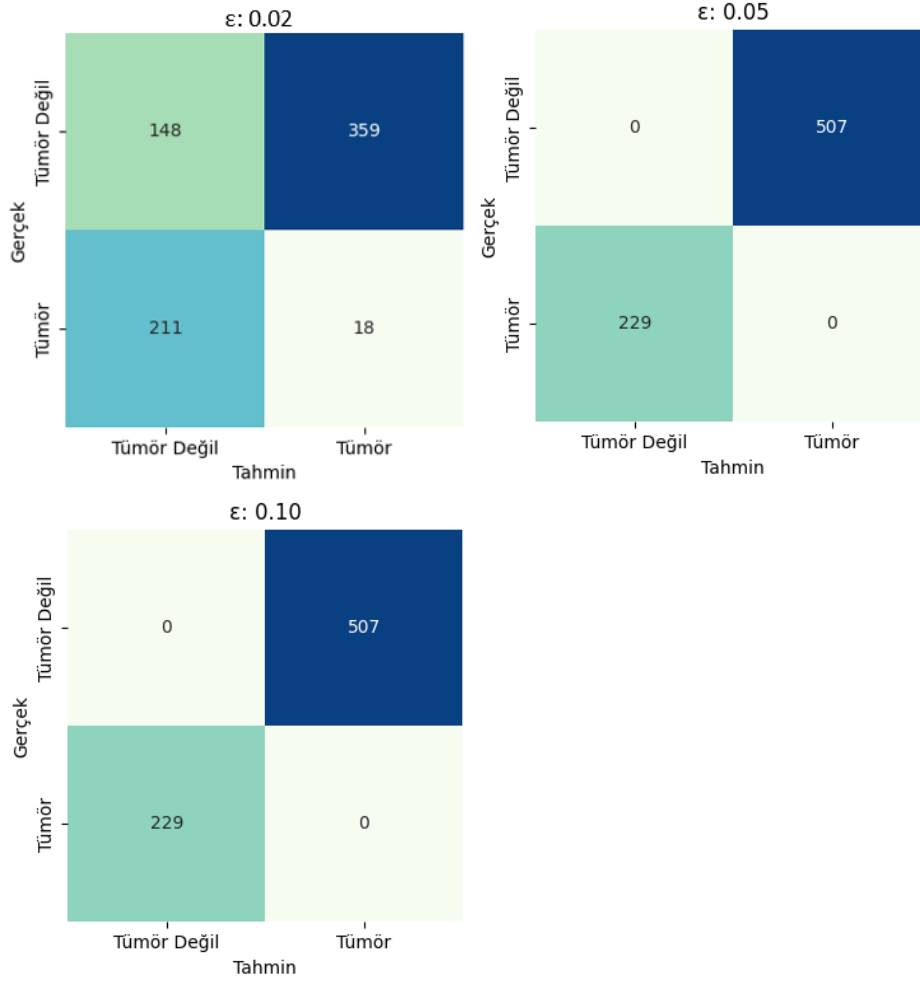


Şekil 6.103. FGSM saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait ROC eğrileri

Şekil 6.103' e göre Xception modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %2'ye kadar düşmektedir.

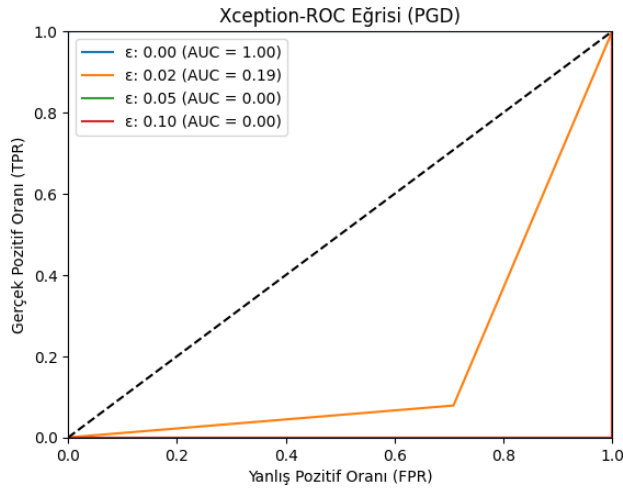
PGD saldırısı

Şekil 6.104' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.104. PGD saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait karmaşıklık matrisi

Şekil 6.104’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 211 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 359 örnek bulunmaktadır. Çizelge 6.6’ ya göre %22,55 doğruluk, %4,78 kesinlik, %7,86 geri çağırma, %5,94 F1 skoru ve %29,19 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.105’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



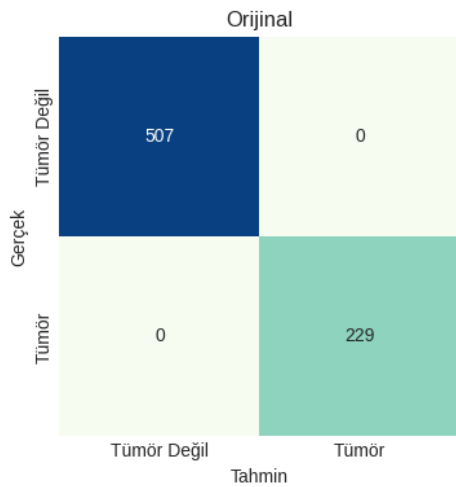
Şekil 6.105. PGD saldırısına uğrayan Böbrek BT görüntülerinde Xception modeline ait ROC eğrileri

Şekil 6.105' e göre Xception modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.2. VGG16

Temiz

Xception modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.106'da görülmektedir.

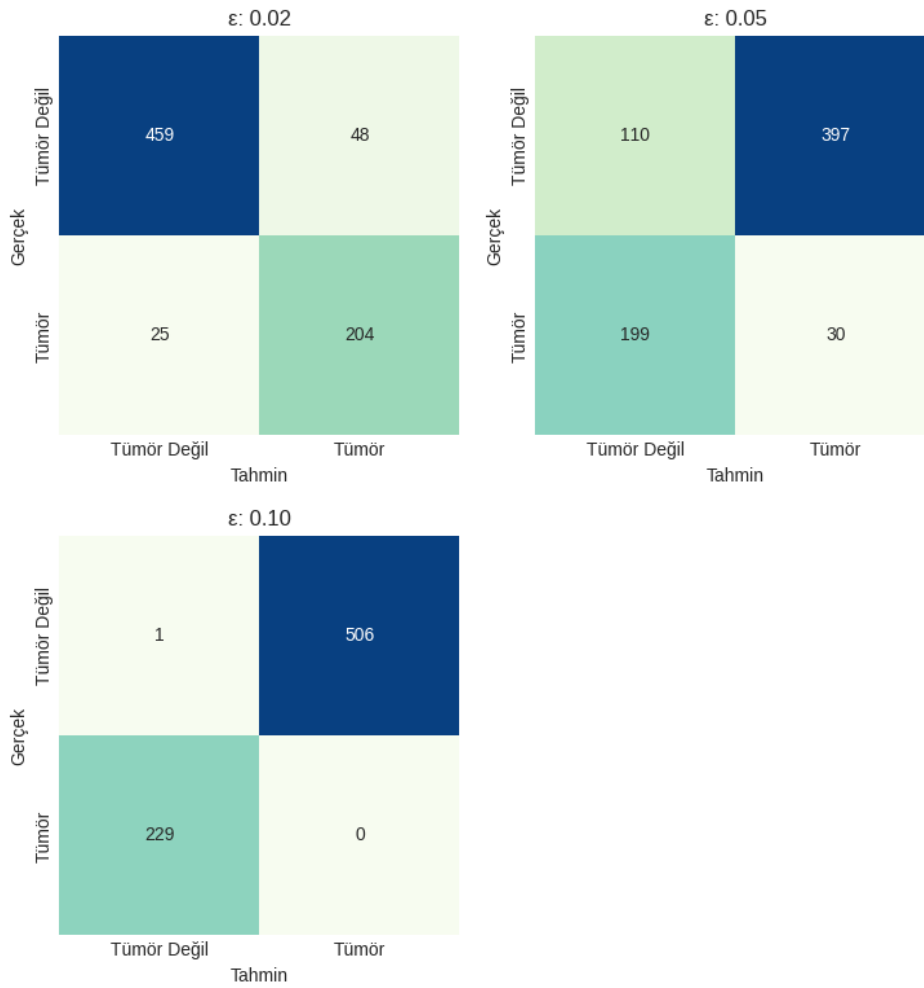


Şekil 6.106. Temiz Böbrek BT görüntüleri üzerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.106’ da görüldüğü gibi temiz test setinin VGG16 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

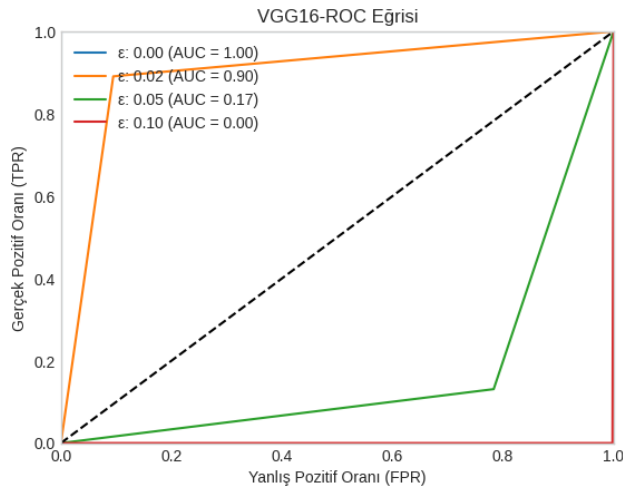
Şekil 6.107’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.107. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.107’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 25 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 48 örnek bulunmaktadır. Çizelge 6.5’ e göre %90,08 doğruluk, %80,95 kesinlik, %89,08 geri

çağırma, %84,82 F1 skoru ve %90,53 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 199 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 397 örnek bulunmaktadır. Çizelge 6.5' e göre %19,02 doğruluk, %7,03 kesinlik, %13,1 geri çağırma, %9,15 F1 skoru ve %21,7 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 506 örnek bulunmaktadır. Çizelge 6.5' e göre %0,14 doğruluk ve %0,2 özgüllük elde etmektedir. Şekil 6.108'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

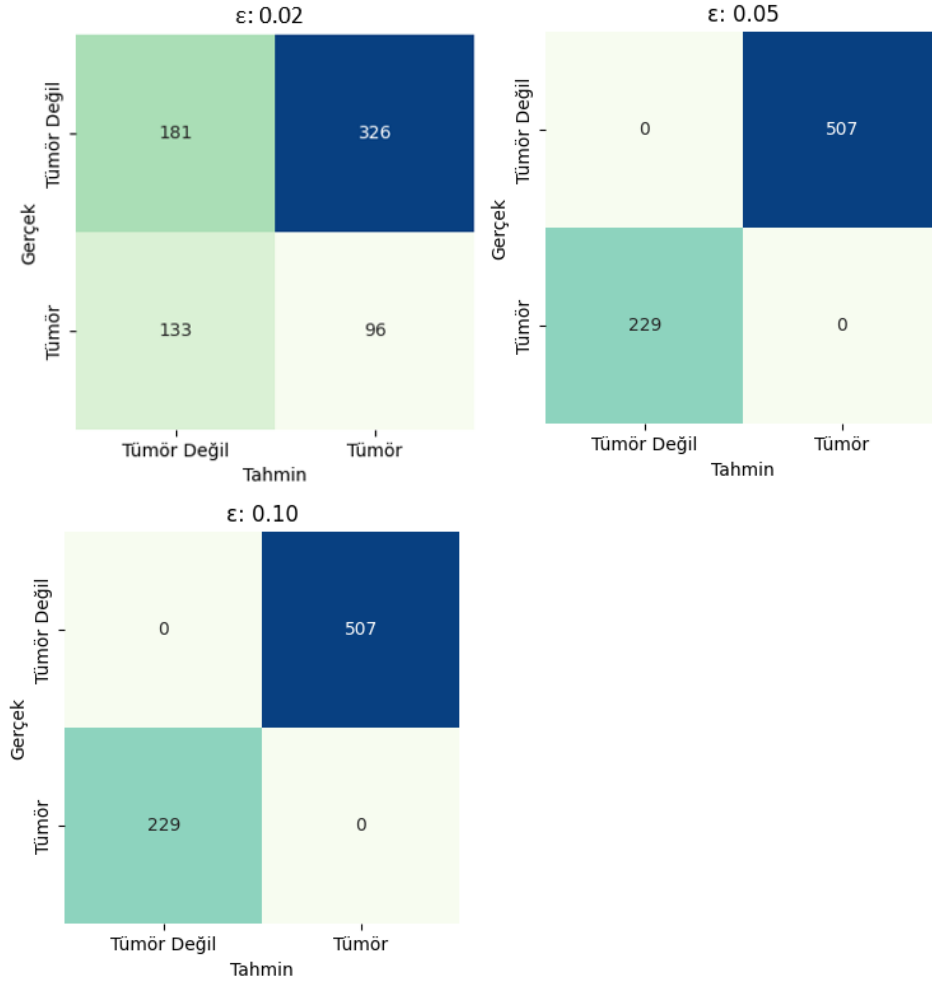


Şekil 6.108. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait ROC eğrileri

Şekil 6.108' e göre VGG16 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

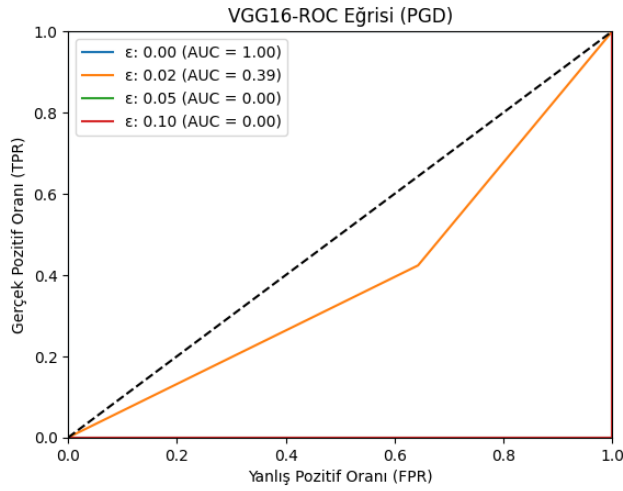
PGD saldırısı

Şekil 6.109' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.109. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.109’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 133 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 326 örnek bulunmaktadır. Çizelge 6.6’ ya göre %37,64 doğruluk, %22,75 kesinlik, %41,92 geri çağırma, %29,49 F1 skoru ve %35,7 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.110’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



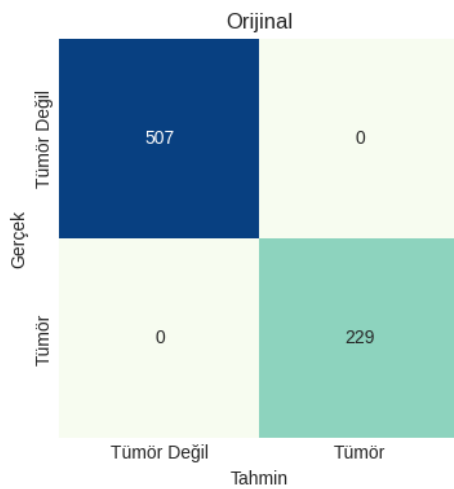
Şekil 6.110. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait ROC eğrileri

Şekil 6.110' a göre VGG16 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.3. VGG19

Temiz

VGG19 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.111'de görülmektedir.

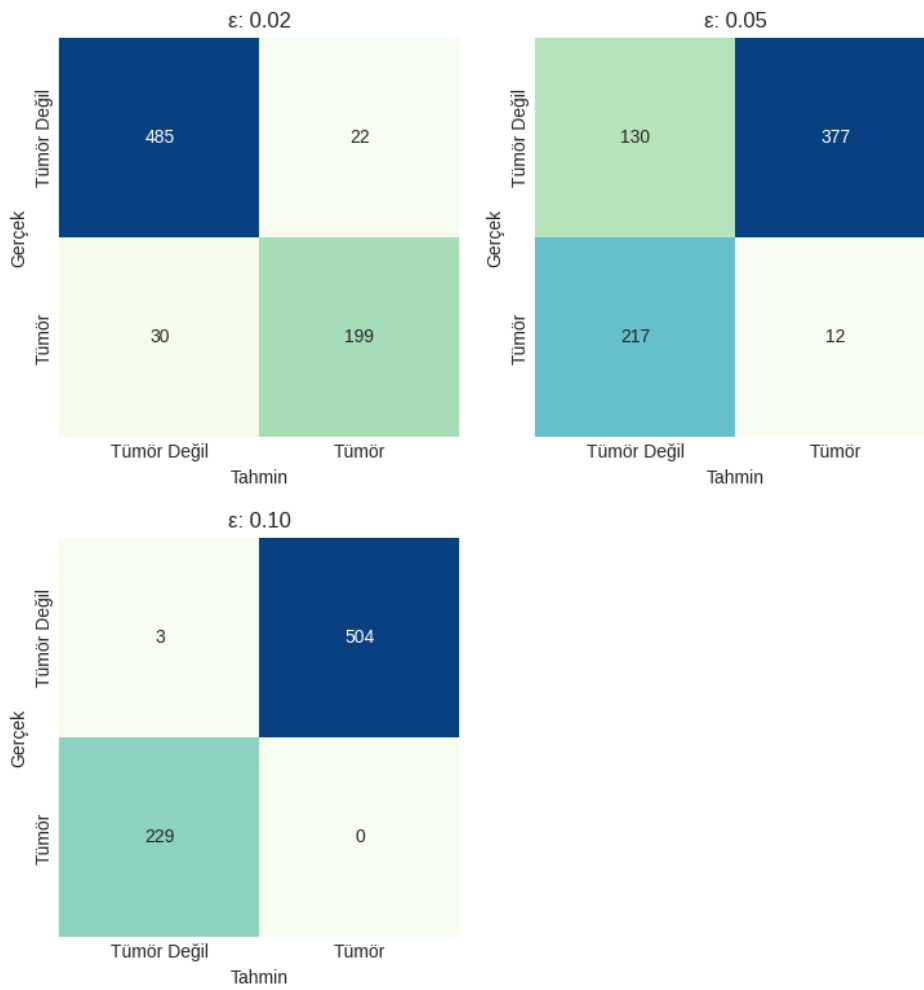


Şekil 6.111. Temiz Böbrek BT görüntüleri üzerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.111’ de görüldüğü gibi temiz test setinin VGG19 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

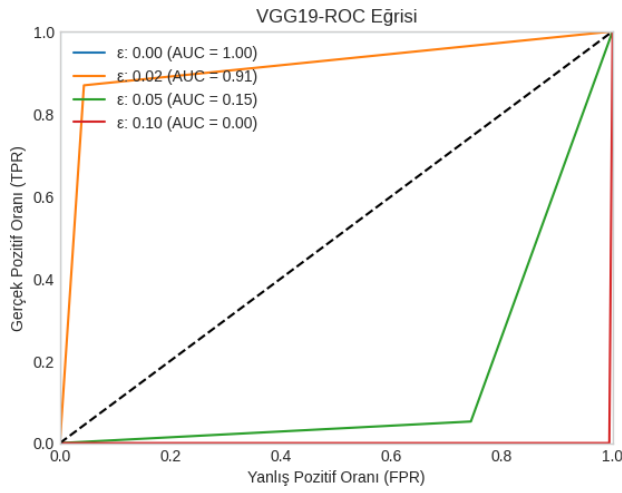
Şekil 6.112’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.112. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.112’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 30 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 22 örnek bulunmaktadır. Çizelge 6.5’ e göre %92,94 doğruluk, %90,05 kesinlik, %86,9 geri

çağırma, %88,45 F1 skoru ve %95,66 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 217 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 377 örnek bulunmaktadır. Çizelge 6.5' e göre %19,29 doğruluk, %3,09 kesinlik, %5,24 geri çağırma, %3,88 F1 skoru ve %25,64 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 504 örnek bulunmaktadır. Çizelge 6.5' e göre %0,41 doğruluk ve %0,59 özgüllük elde etmektedir. Şekil 6.113'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

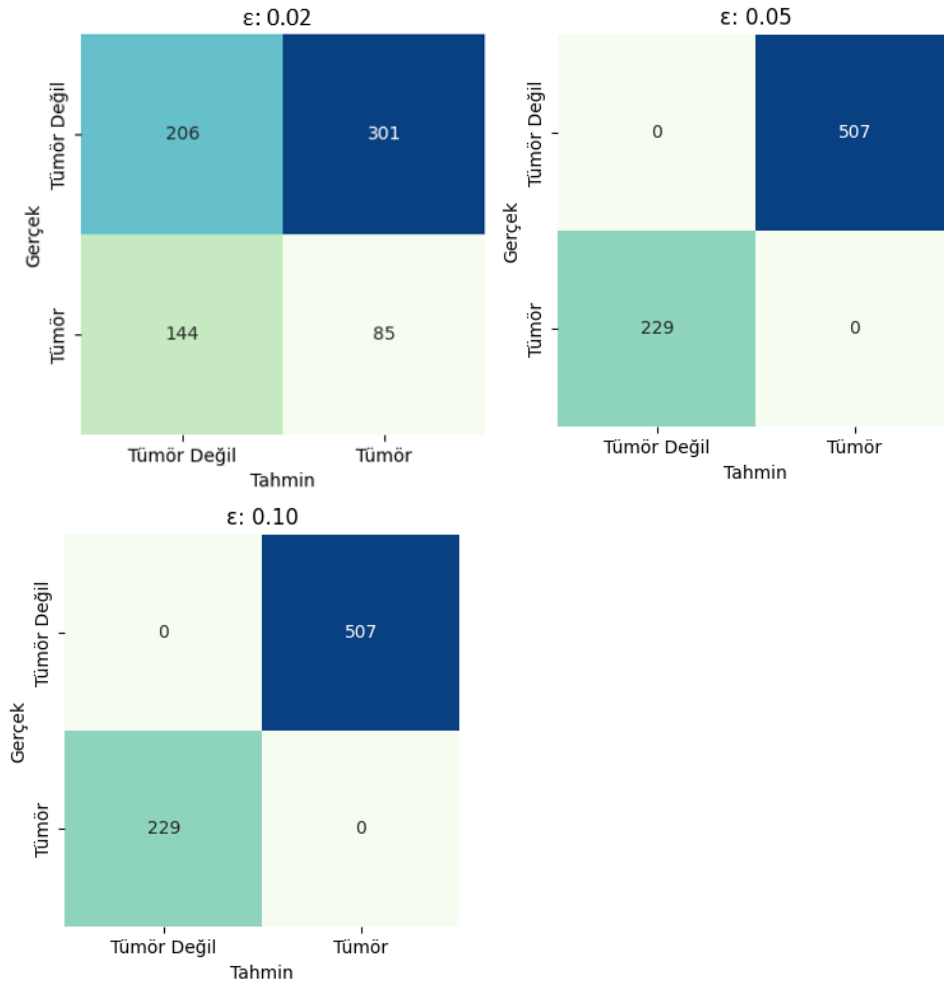


Şekil 6.113. FGSM saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait ROC eğrileri

Şekil 6.113' e göre VGG19 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

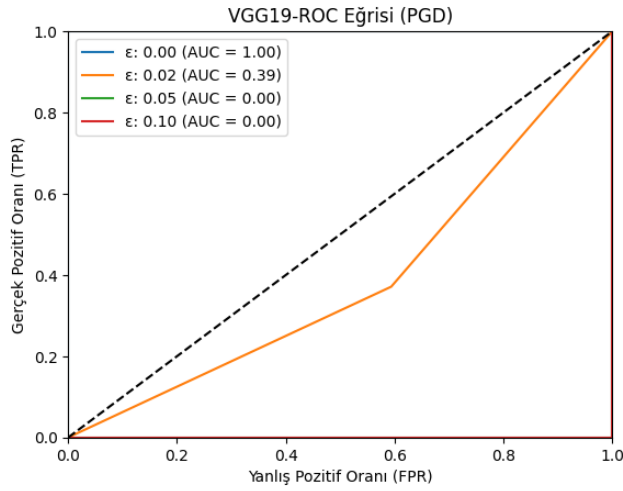
PGD saldırısı

VGG19 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.114'te görülmektedir.



Şekil 6.114. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait karmaşıklık matrisi

Şekil 6.114’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 144 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 301 örnek bulunmaktadır. Çizelge 6.6’ ya göre %39,54 doğruluk, %22,02 kesinlik, %37,12 geri çağırma, %27,64 F1 skoru ve %40,63 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.115’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



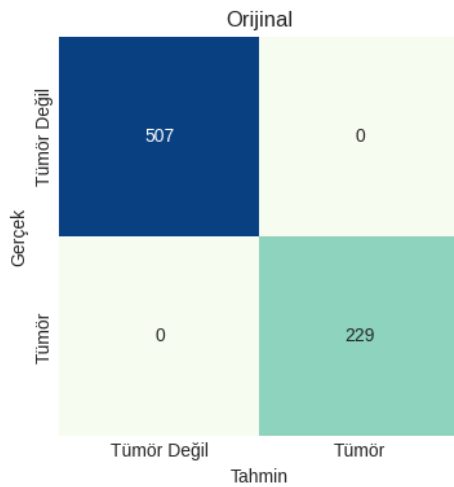
Şekil 6.115. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG19 modeline ait ROC eğrileri

Şekil 6.115' e göre VGG19 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.4. DenseNet121

Temiz

DenseNet121 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.116'da görülmektedir.

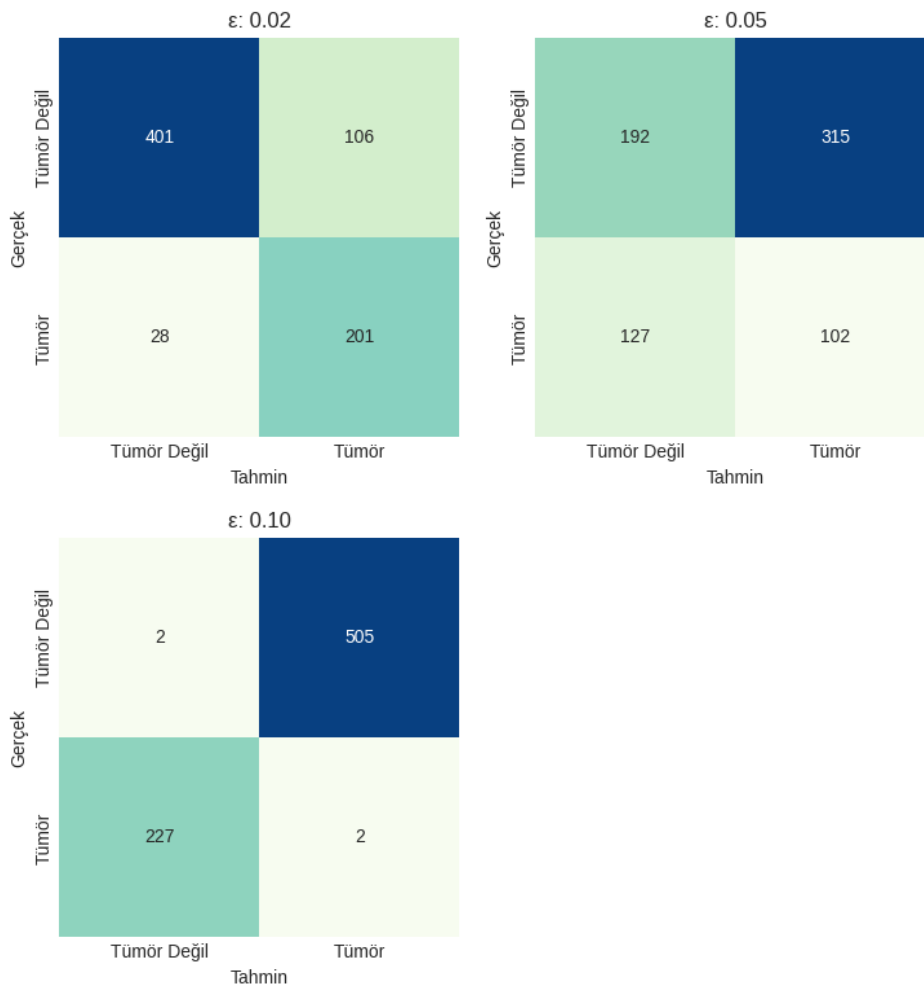


Şekil 6.116. Temiz Böbrek BT görüntüleri üzerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.116’ da görüldüğü gibi temiz test setinin DenseNet121 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

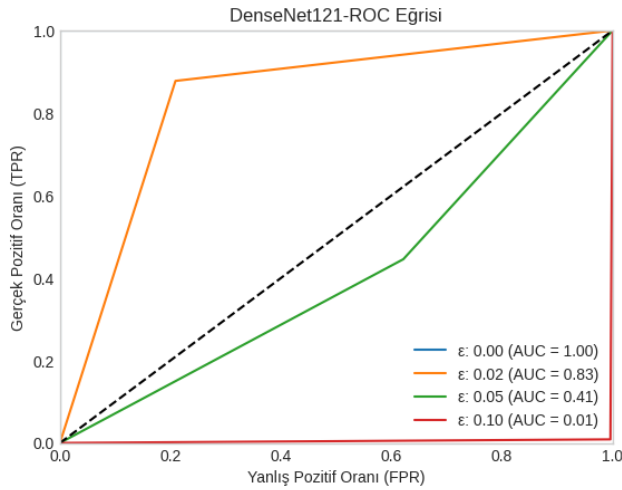
Şekil 6.117’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.117. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.117’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 28 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 106 örnek bulunmaktadır. Çizelge 6.5’ e göre %81,79 doğruluk, %65,47 kesinlik, %87,77 geri

çağırma, %75 F1 skoru ve %79,09 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 127 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 315 örnek bulunmaktadır. Çizelge 6.5' e göre %39,95 doğruluk, %24,46 kesinlik, %44,54 geri çağırma, %31,58 F1 skoru ve %37,87 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 227 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 505 örnek bulunmaktadır. Çizelge 6.5' e göre %0,54 doğruluk, %0,40 kesinlik, %0,87 geri çağırma, %0,54 F1 skoru ve %0,4 özgüllük elde etmektedir. Şekil 6.118'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

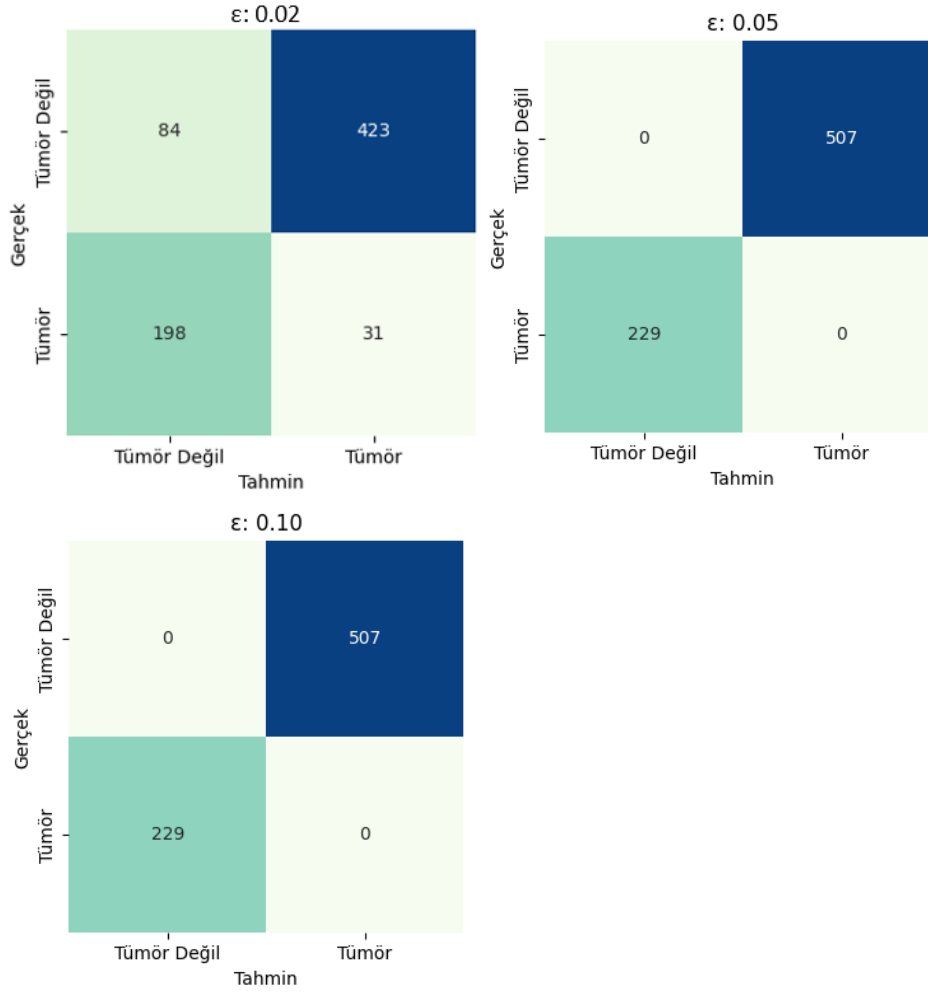


Şekil 6.118. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait ROC eğrileri

Şekil 6.118' e göre DenseNet121 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %1'e kadar düşmektedir.

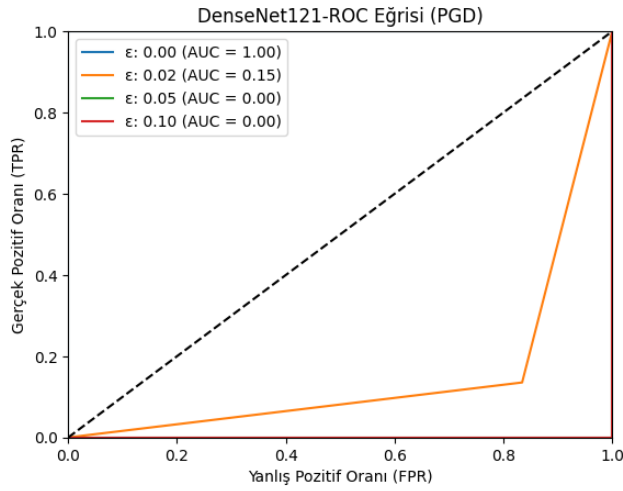
PGD saldırısı

Şekil 6.119' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.119. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait karmaşıklık matrisi

Şekil 6.119’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 198 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 423 örnek bulunmaktadır. Çizelge 6.6’ ya göre %15,63 doğruluk, %6,83 kesinlik, %13,54 geri çağırma, %9,08 F1 skoru ve %16,57 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.120’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



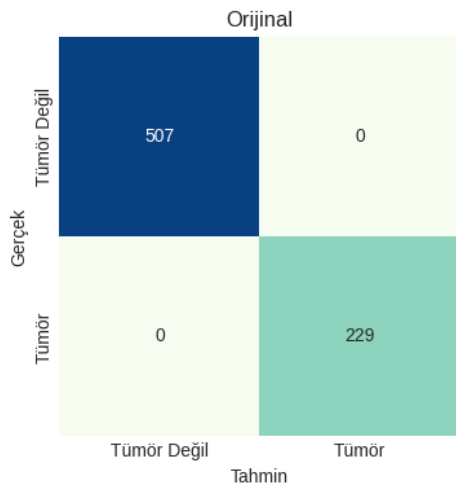
Şekil 6.120. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet121 modeline ait ROC eğrileri

Şekil 6.120' ye göre DenseNet121 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.5. DenseNet169

Temiz

DenseNet169 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.121'de görülmektedir.

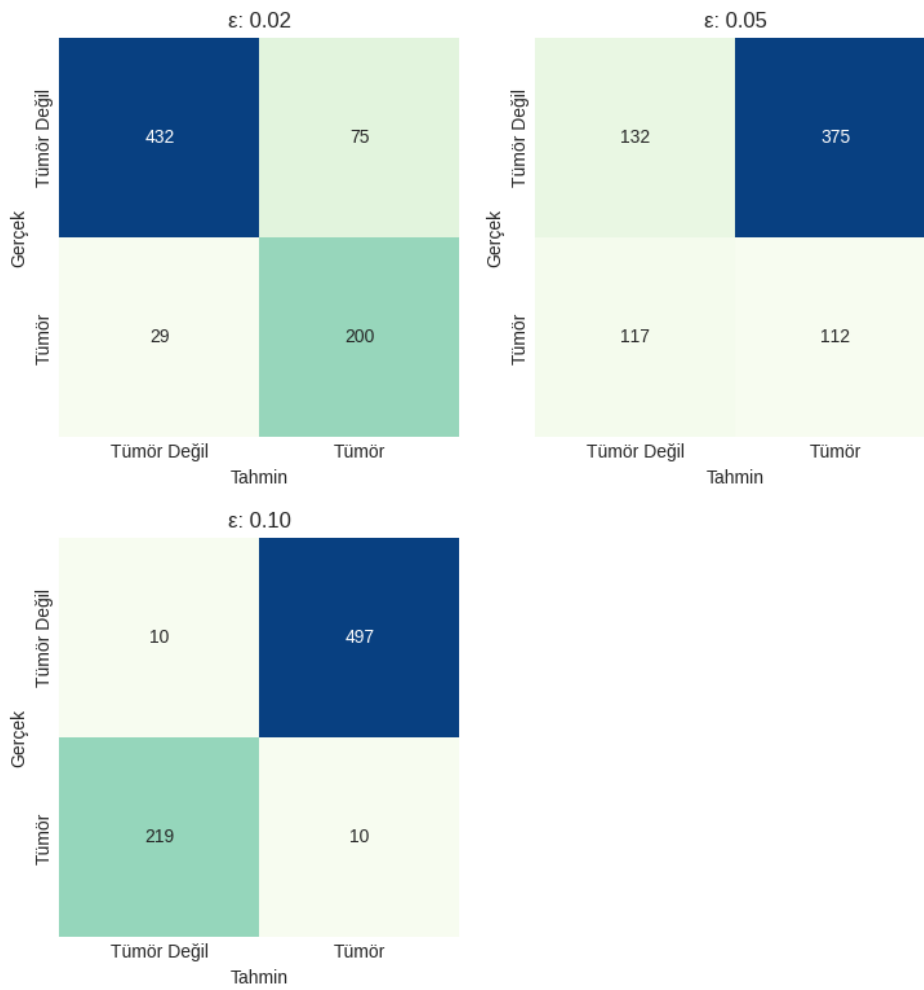


Şekil 6.121. Temiz Böbrek BT görüntüleri üzerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.121’ de görüldüğü gibi temiz test setinin DenseNet169 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

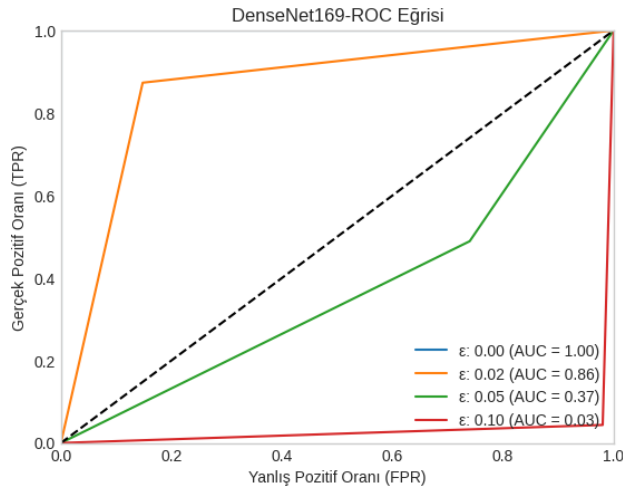
Şekil 6.122’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.122. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.122’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 29 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 75 örnek bulunmaktadır. Çizelge 6.5’ e göre %85,87 doğruluk, %72,73 kesinlik, %87,34 geri

çağırma, %79,37 F1 skoru ve %85,21 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 117 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 375 örnek bulunmaktadır. Çizelge 6.5' e göre %33,15 doğruluk, %23 kesinlik, %48,91 geri çağırma, %31,29 F1 skoru ve %26,04 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 219 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 497 örnek bulunmaktadır. Çizelge 6.5' e göre %2,72 doğruluk, %1,97 kesinlik, %4,37 geri çağırma, %2,72 F1 skoru ve %1,97 özgüllük elde etmektedir. Şekil 6.123'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

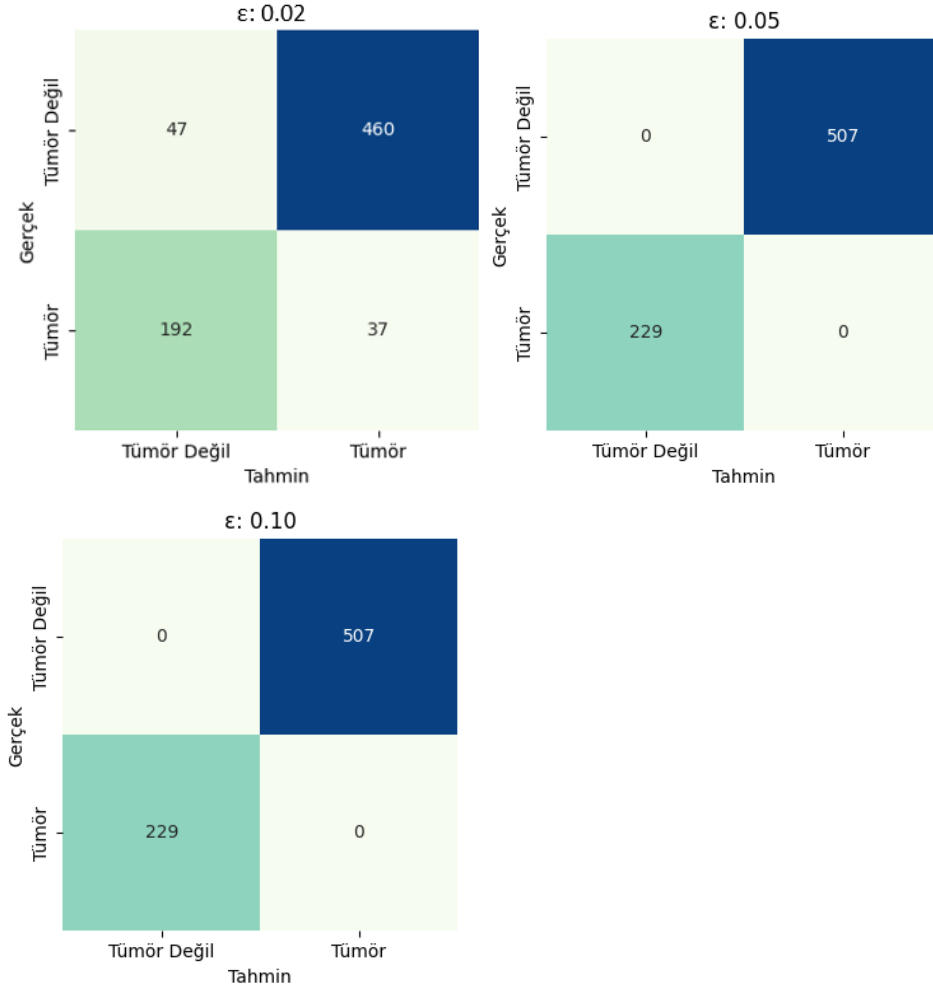


Şekil 6.123. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait ROC eğrileri

Şekil 6.123' e göre DenseNet169 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %3'e kadar düşmektedir.

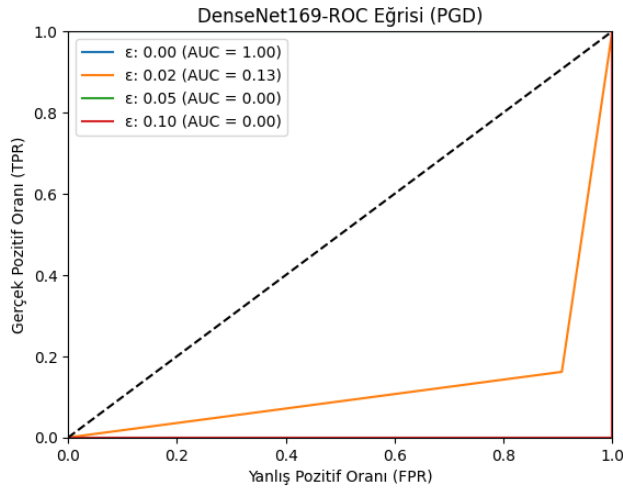
PGD saldırısı

Şekil 6.124' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.124. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait karmaşıklık matrisi

Şekil 6.124’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 192 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 460 örnek bulunmaktadır. Çizelge 6.6’ ya göre %11,41 doğruluk, %7,45 kesinlik, %16,15 geri çağırma, %10,19 F1 skoru ve %9,27 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.125’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



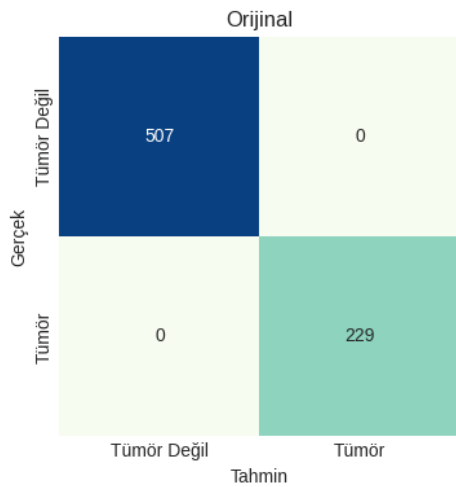
Şekil 6.125. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet169 modeline ait ROC eğrileri

Şekil 6.125' e göre DenseNet169 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.6. DenseNet201

Temiz

DenseNet201 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.126'da görülmektedir.

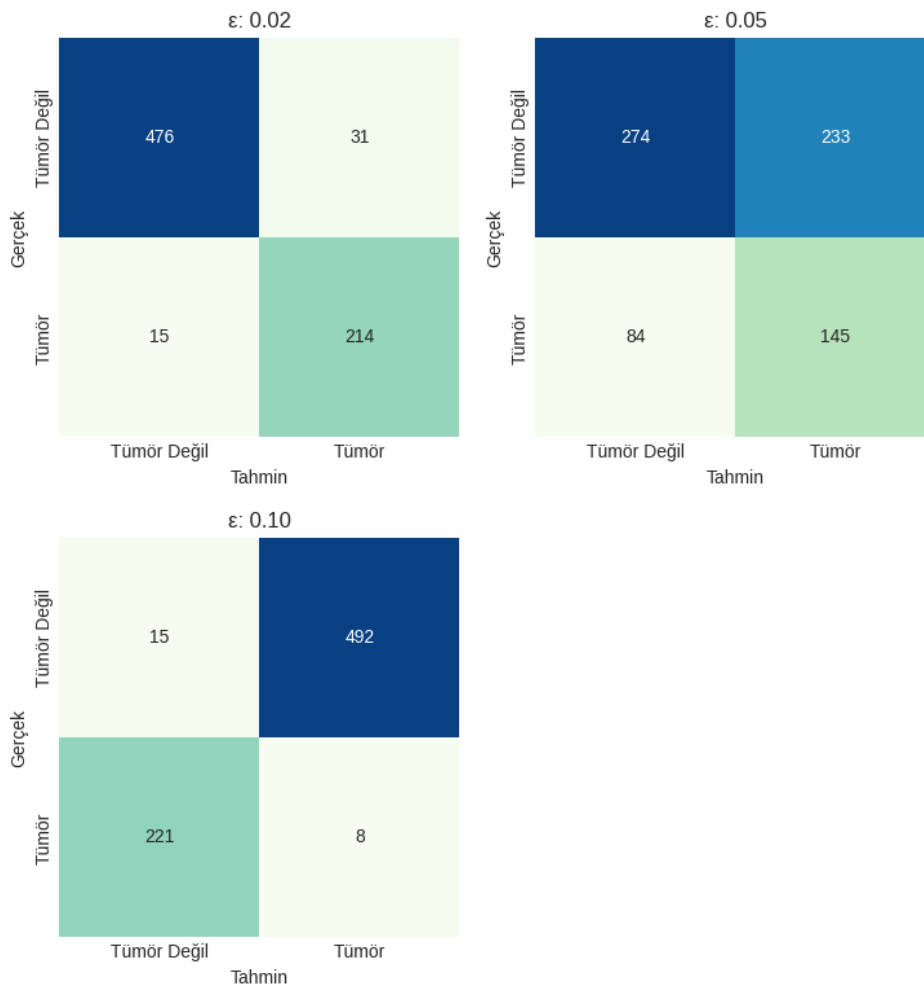


Şekil 6.126. Temiz Böbrek BT görüntüleri üzerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.126’ da görüldüğü gibi temiz test setinin DenseNet201 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

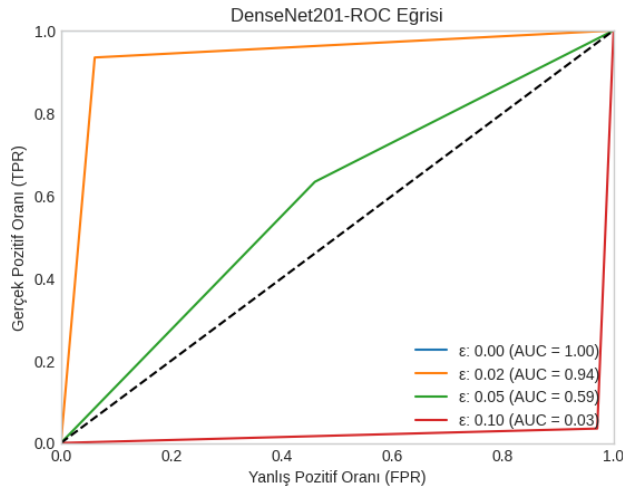
Şekil 6.127’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.127. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.127’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 15 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 31 örnek bulunmaktadır. Çizelge 6.5’ e göre %93,75 doğruluk, %87,35 kesinlik, %93,45 geri

çağırma, %90,3 F1 skoru ve %93,89 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 84 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 233 örnek bulunmaktadır. Çizelge 6.5' e göre %56,93 doğruluk, %38,36 kesinlik, %63,32 geri çağırma, %47,48 F1 skoru ve %54,04 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 221 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 492 örnek bulunmaktadır. Çizelge 6.5' e göre %3,13 doğruluk, %1,6 kesinlik, %3,49 geri çağırma, %2,2 F1 skoru ve %2,96 özgüllük elde etmektedir. Şekil 6.128'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

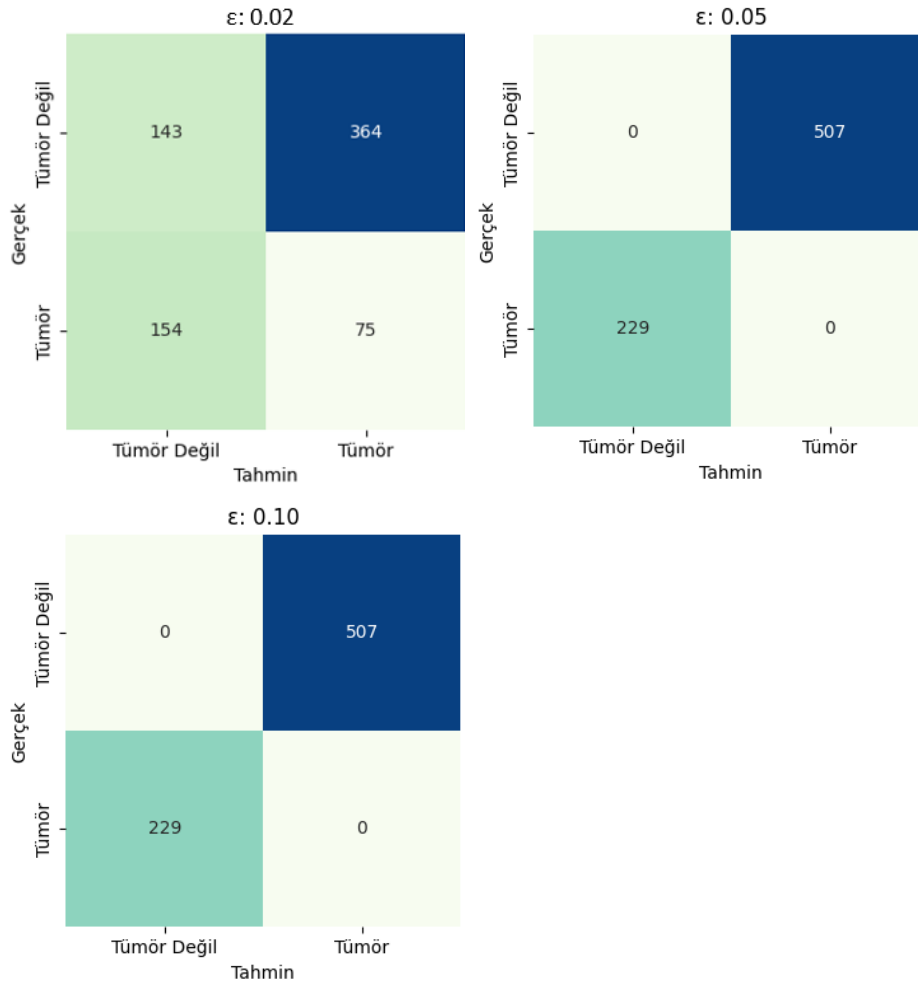


Şekil 6.128. FGSM saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait ROC eğrileri

Şekil 6.128' e göre DenseNet201 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %3'e kadar düşmektedir.

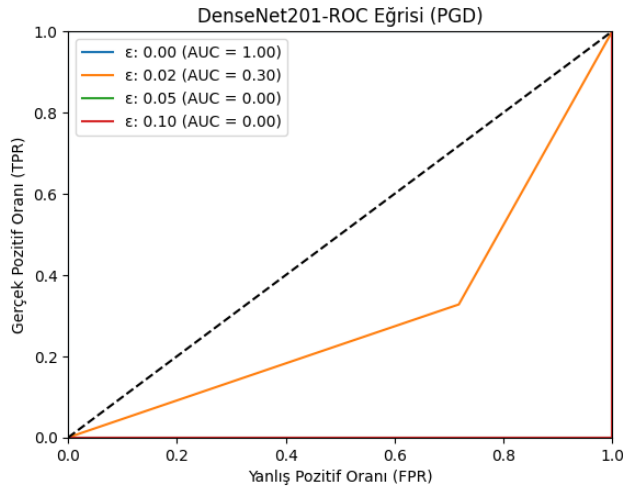
PGD saldırısı

Şekil 6.129' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.129. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait karmaşıklık matrisi

Şekil 6.129’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 154 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 364 örnek bulunmaktadır. Çizelge 6.6’ ya göre %29,62 doğruluk, %17,08 kesinlik, %32,75 geri çağırma, %22,46 F1 skoru ve %28,21 özgülük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.130’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



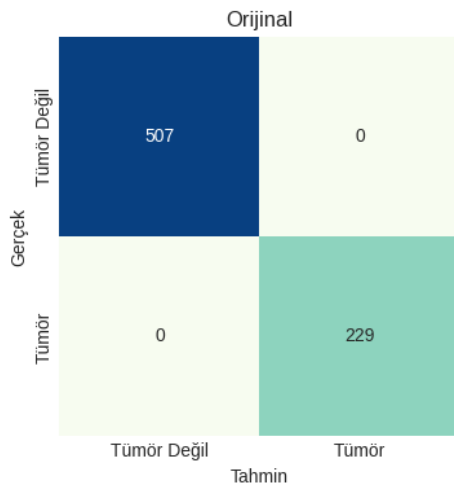
Şekil 6.130. PGD saldırısına uğrayan Böbrek BT görüntülerinde DenseNet201 modeline ait ROC eğrileri

Şekil 6.130' a göre DenseNet201 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.7. EfficientNetV2B0

Temiz

EfficientNetV2B0 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.131'de görülmektedir.

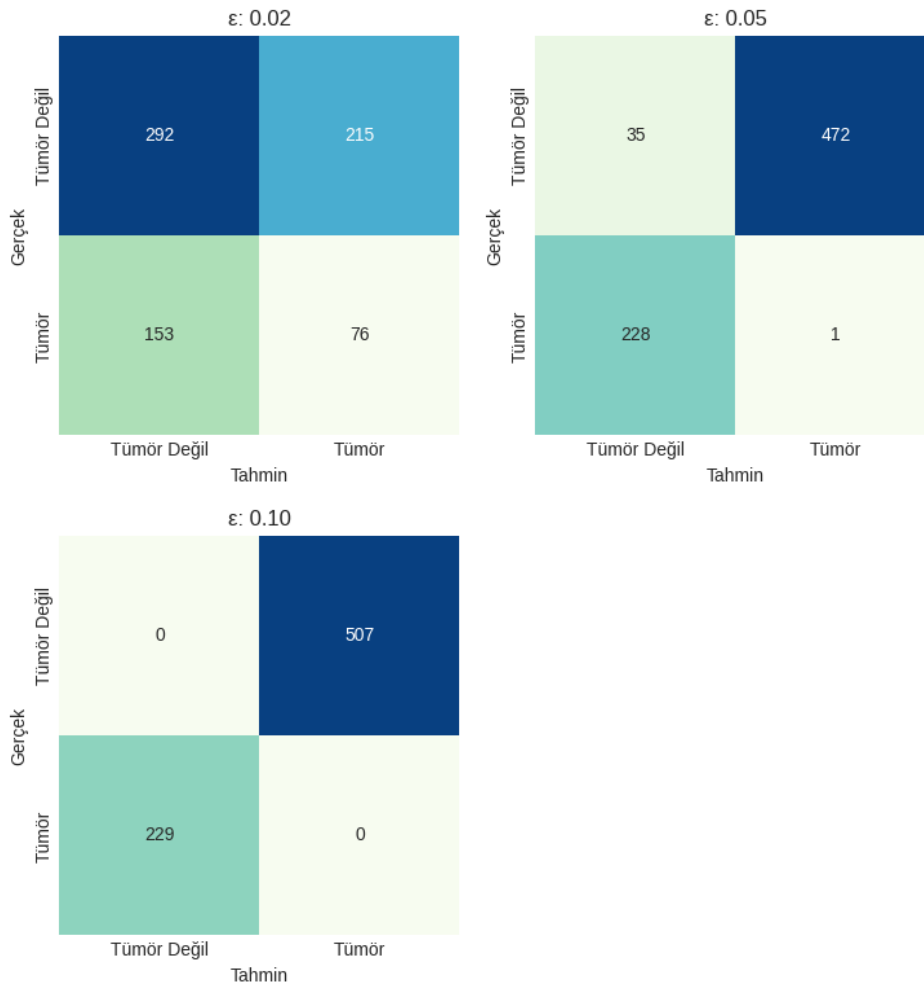


Şekil 6.131. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi

Şekil 6.131’ de görüldüğü gibi temiz test setinin EfficientNetV2B0 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

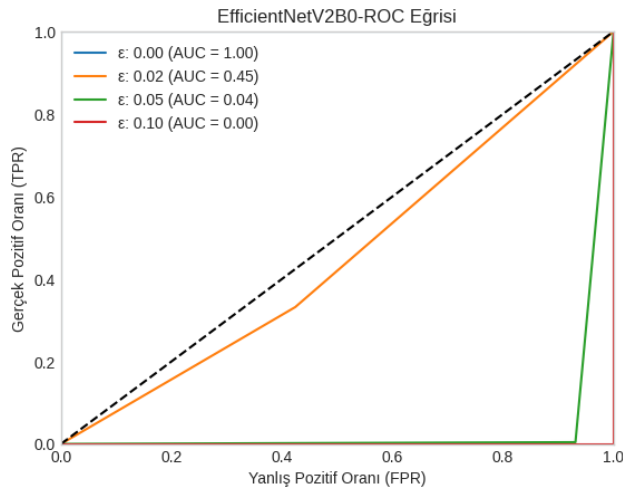
Şekil 6.132’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.132. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait karmaşıklık matrisi

Şekil 6.132’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 153 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 215 örnek bulunmaktadır. Çizelge 6.5’ e göre %50 doğruluk, %26,12 kesinlik, %33,19 geri

çağırma, %29,23 F1 skoru ve %57,59 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 228 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 472 örnek bulunmaktadır. Çizelge 6.5' e göre %4,89 doğruluk, %0,21 kesinlik, %0,44 geri çağırma, %0,29 F1 skoru ve %6,9 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.133'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

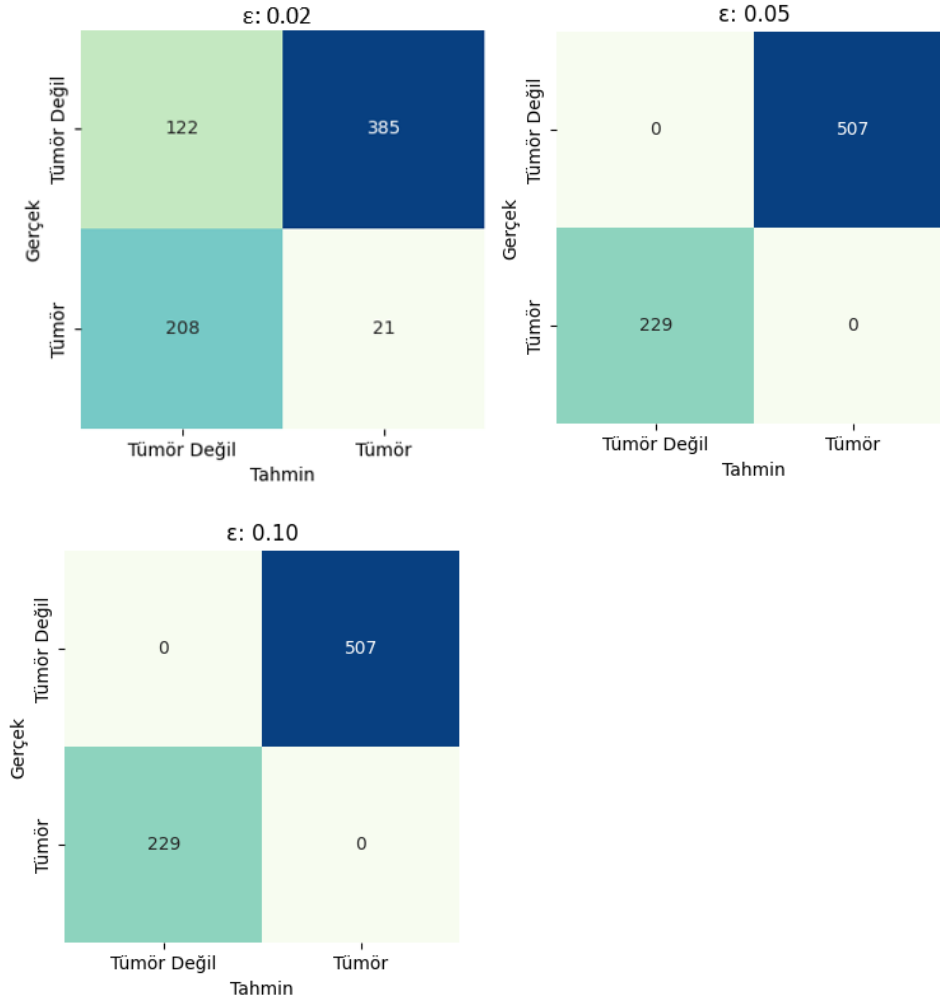


Şekil 6.133. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri

Şekil 6.133' e göre EfficientNetV2B0 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

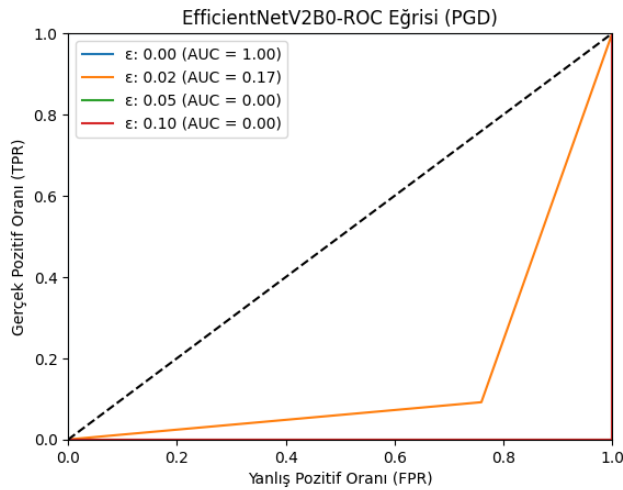
PGD saldırısı

Şekil 6.134' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.134. PGD saldırısına uğrayan Böbrek BT görüntülerinde VGG16 modeline ait karmaşıklık matrisi

Şekil 6.134’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 208 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 385 örnek bulunmaktadır. Çizelge 6.6’ ya göre %19,43 doğruluk, %5,17 kesinlik, %9,17 geri çağırma, %6,61 F1 skoru ve %24,06 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.135’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



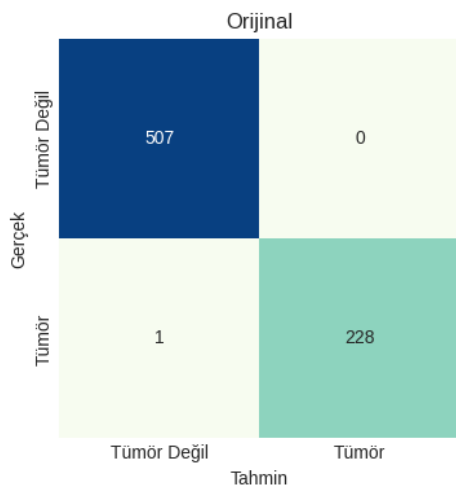
Şekil 6.135. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B0 modeline ait ROC eğrileri

Şekil 6.135' e göre EfficientNetV2B0 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.8. EfficientNetV2B1

Temiz

EfficientNetV2B1 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.136'da görülmektedir.

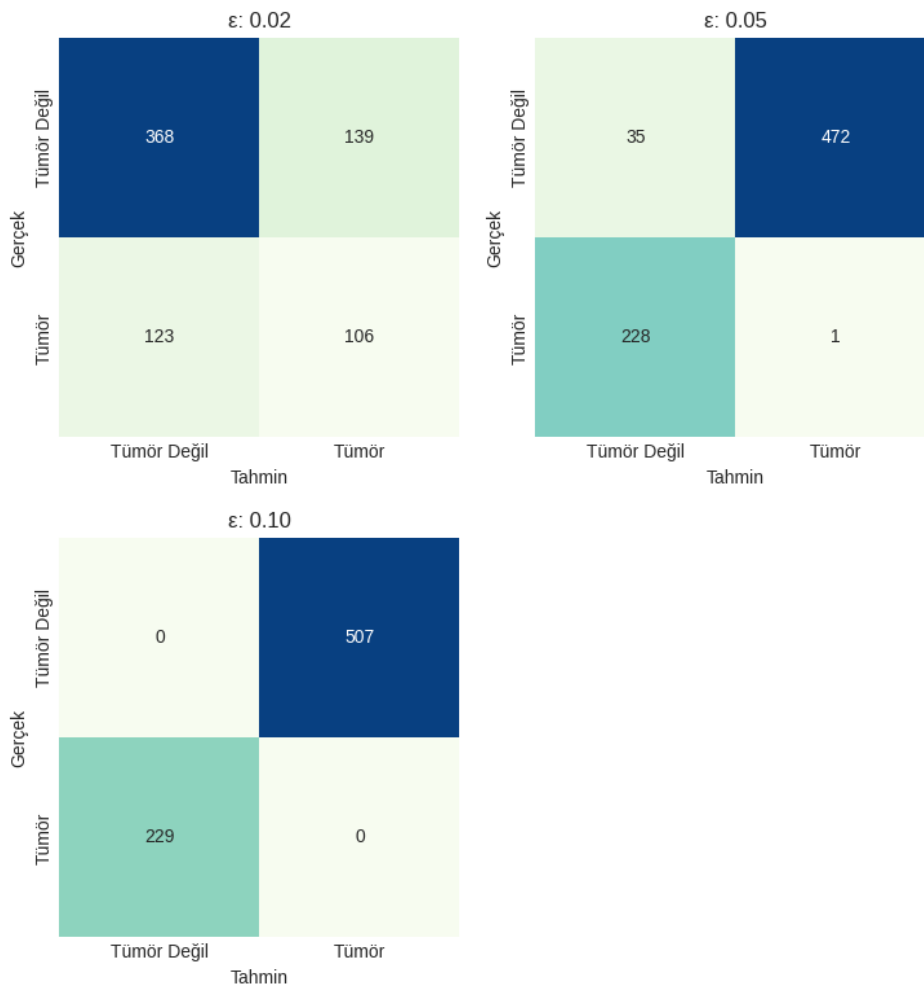


Şekil 6.136. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.136’ da görüldüğü gibi temiz test setinin EfficientNetV2B1 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre %99,86 doğruluk, %100 kesinlik, %99,56 geri çağırma, %99,78 F1 skoru ve %100 özgüllük elde etmektedir.

FGSM saldırısı

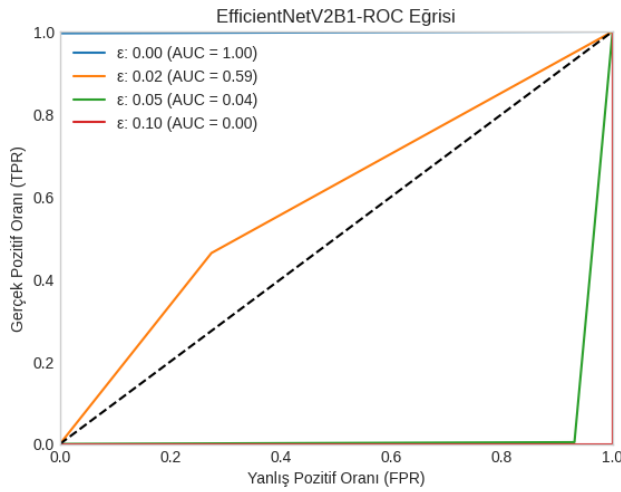
Şekil 6.137’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.137. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.137’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 123 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 139 örnek bulunmaktadır. Çizelge 6.5’ e göre %64,4 doğruluk, %43,27 kesinlik, %46,29

geri çağırma, %44,73 F1 skoru ve %72,58 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 228 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 472 örnek bulunmaktadır. Çizelge 6.5' e göre %4,89 doğruluk, %0,21 kesinlik, %0,44 geri çağırma, %0,29 F1 skoru ve %6,9 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.138'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

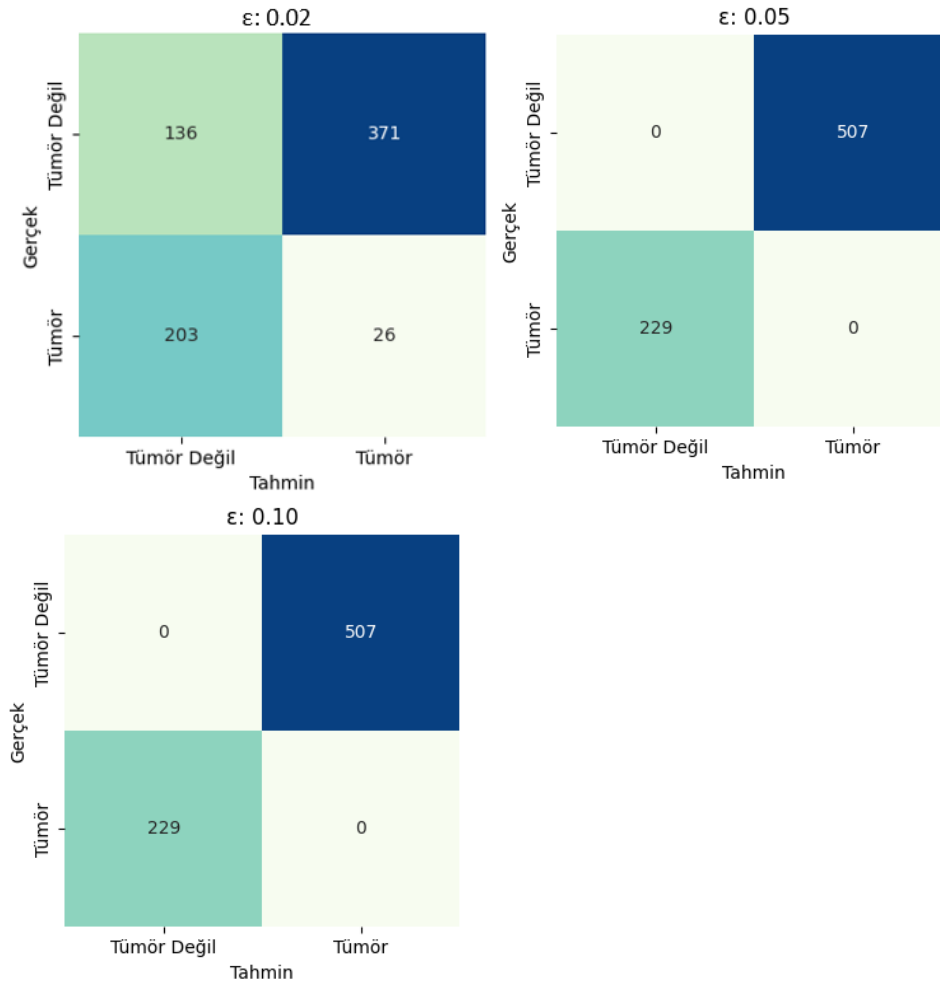


Şekil 6.138. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri

Şekil 6.138' e göre EfficientNetV2B1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

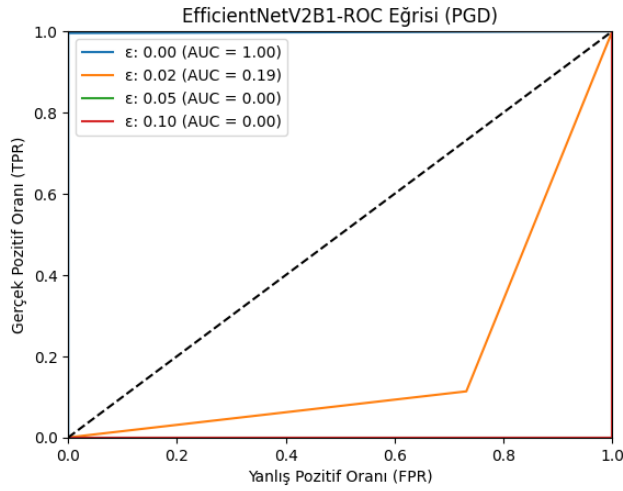
PGD saldırısı

Şekil 6.139' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.139. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait karmaşıklık matrisi

Şekil 6.139’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 203 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 371 örnek bulunmaktadır. Çizelge 6.6’ ya göre %22,01 doğruluk, %6,55 kesinlik, %11,35 geri çağırma, %8,31 F1 skoru ve %26,83 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.140’ta pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



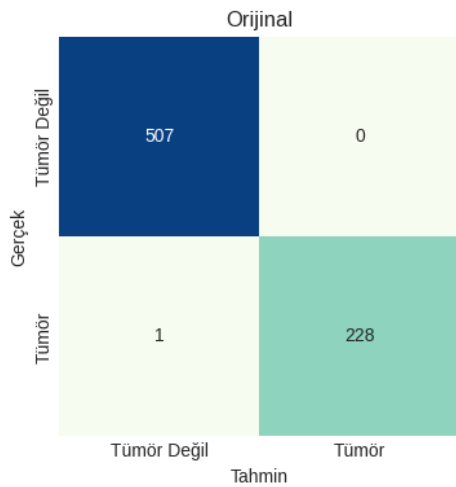
Şekil 6.140. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B1 modeline ait ROC eğrileri

Şekil 6.140' a göre EfficientNetV2B1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.9. EfficientNetV2B2

Temiz

EfficientNetV2B2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.141'de görülmektedir.

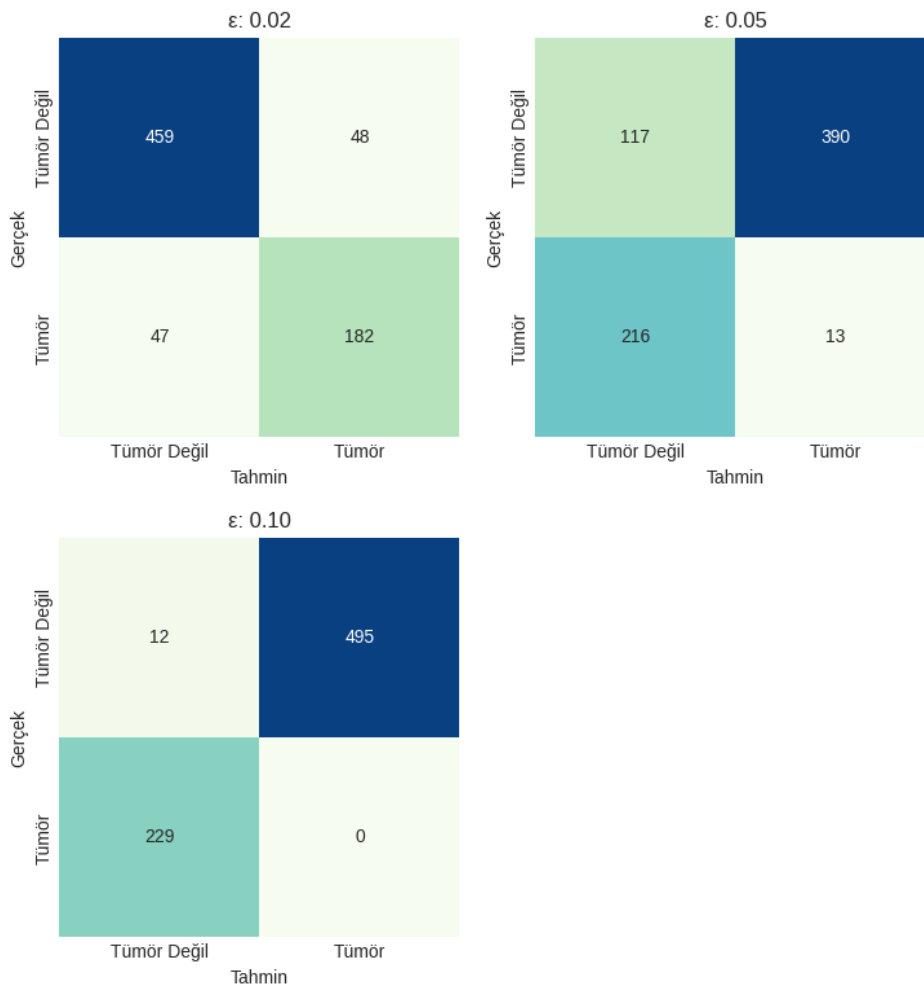


Şekil 6.141. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.141’ de görüldüğü gibi temiz test setinin EfficientNetV2B2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre %99,86 doğruluk, %100 kesinlik, %99,56 geri çağırma, %99,78 F1 skoru ve %100 özgüllük elde etmektedir.

FGSM saldırısı

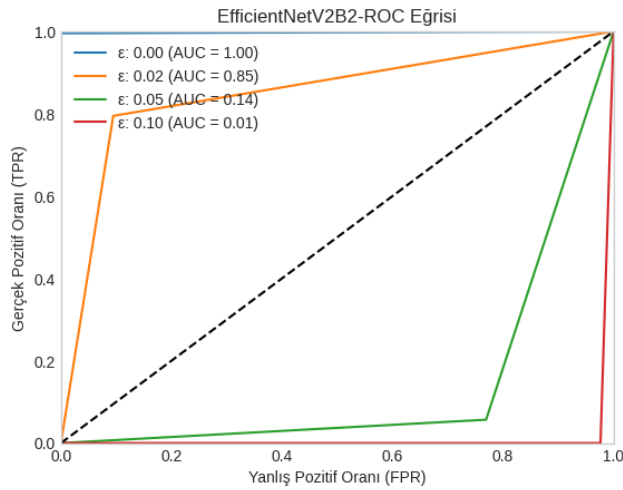
Şekil 6.142’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.142. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.142’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 47 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 48 örnek bulunmaktadır. Çizelge 6.5’ e göre %87,09 doğruluk, %79,13 kesinlik, %79,48 geri

çağırma, %79,3 F1 skoru ve %90,53 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 216 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 390 örnek bulunmaktadır. Çizelge 6.5' e göre %17,66 doğruluk, %3,23 kesinlik, %5,68 geri çağırma, %4,11 F1 skoru ve %23,08 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 495 örnek bulunmaktadır. Çizelge 6.5' e göre %1,63 doğruluk ve %2,37 özgüllük elde etmektedir. Şekil 6.143'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

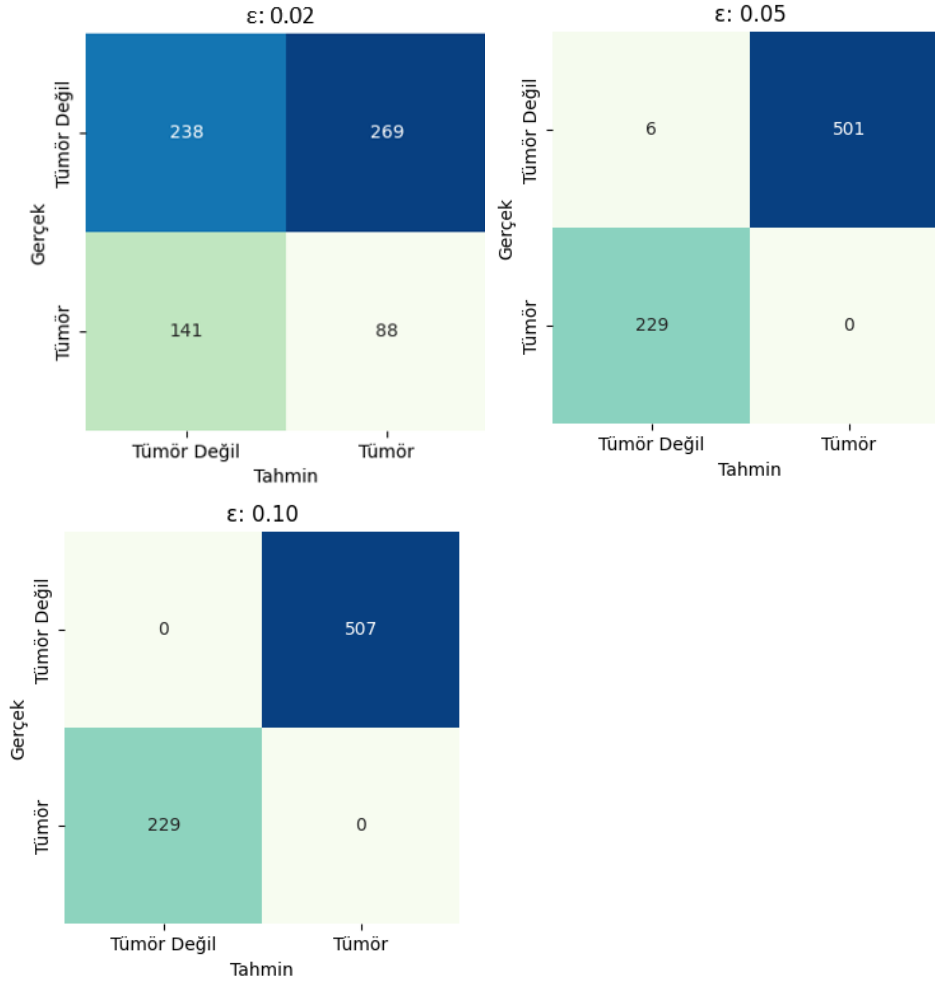


Şekil 6.143. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri

Şekil 6.143' e göre EfficientNetV2B2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %1'e kadar düşmektedir.

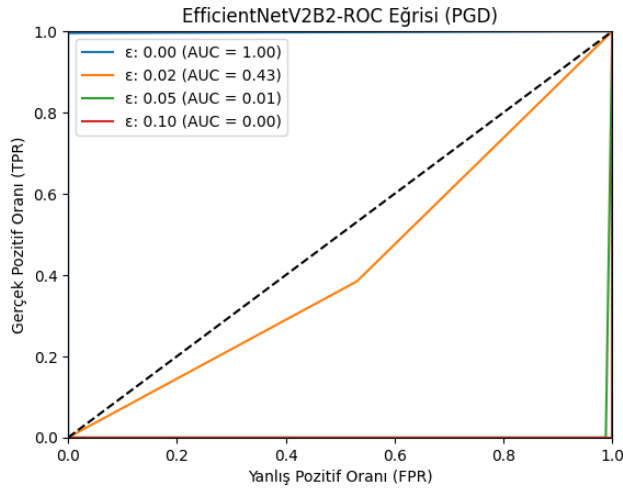
PGD saldırısı

Şekil 6.144' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.144. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait karmaşıklık matrisi

Şekil 6.144’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 141 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 269 örnek bulunmaktadır. Çizelge 6.6’ ya göre %44,29 doğruluk, %24,65 kesinlik, %38,43 geri çağırma, %30,03 F1 skoru ve %46,94 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 501 örnek bulunmaktadır. Çizelge 6.6’ ya göre %0,82 doğruluk ve %1,18 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.145’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



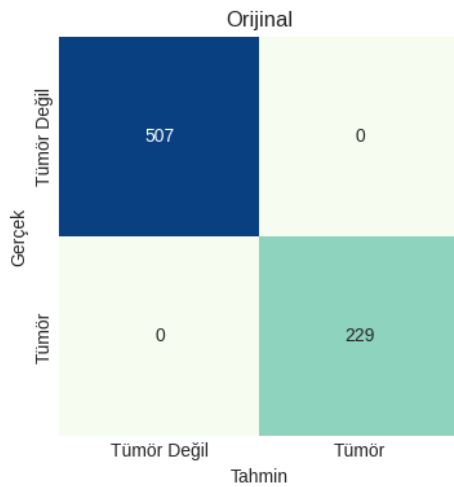
Şekil 6.145. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B2 modeline ait ROC eğrileri

Şekil 6.145' e göre EfficientNetV2B2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.10. EfficientNetV2B3

Temiz

EfficientNetV2B3 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.146'da görülmektedir.

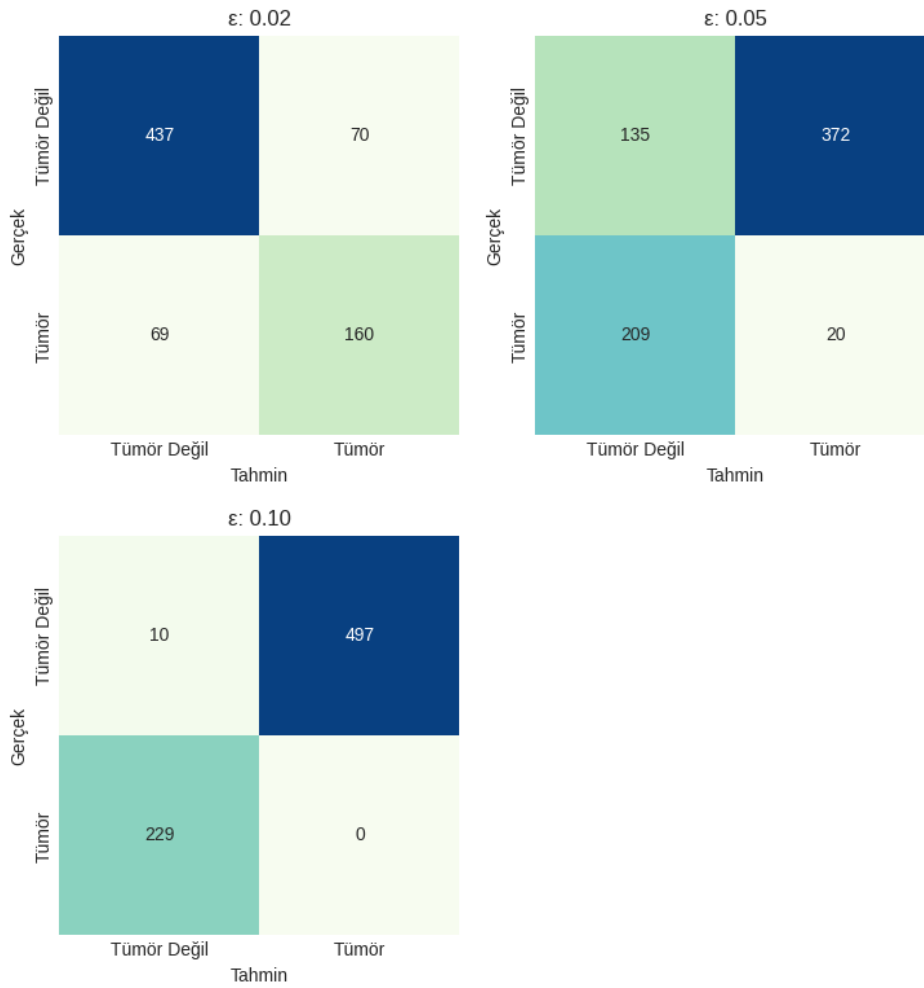


Şekil 6.146. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.146’ da görüldüğü gibi temiz test setinin EfficientNetV2B3 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgülük değerleri %100’dür.

FGSM saldırısı

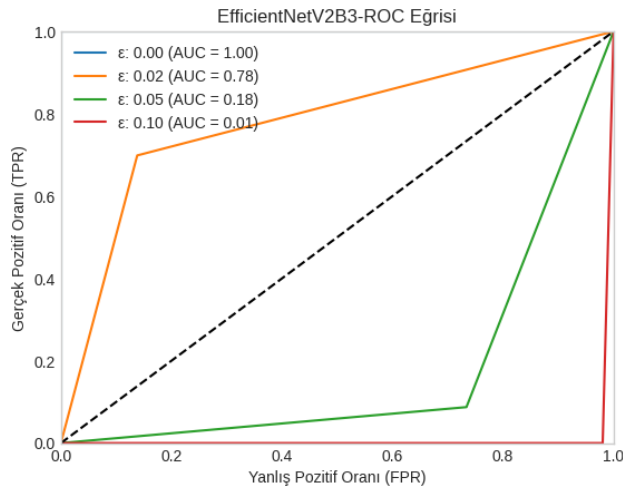
Şekil 6.147’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.147. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.147’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 69 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 70 örnek bulunmaktadır. Çizelge 6.5’ e göre %81,11 doğruluk, %69,57 kesinlik, %69,87 geri

çağırma, %69,72 F1 skoru ve %86,19 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 209 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 372 örnek bulunmaktadır. Çizelge 6.5' e göre %21,06 doğruluk, %5,1 kesinlik, %8,73 geri çağırma, %6,44 F1 skoru ve %26,63 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 497 örnek bulunmaktadır. Çizelge 6.5' e göre %1,36 doğruluk ve %1,97 özgüllük elde etmektedir. Şekil 6.148'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

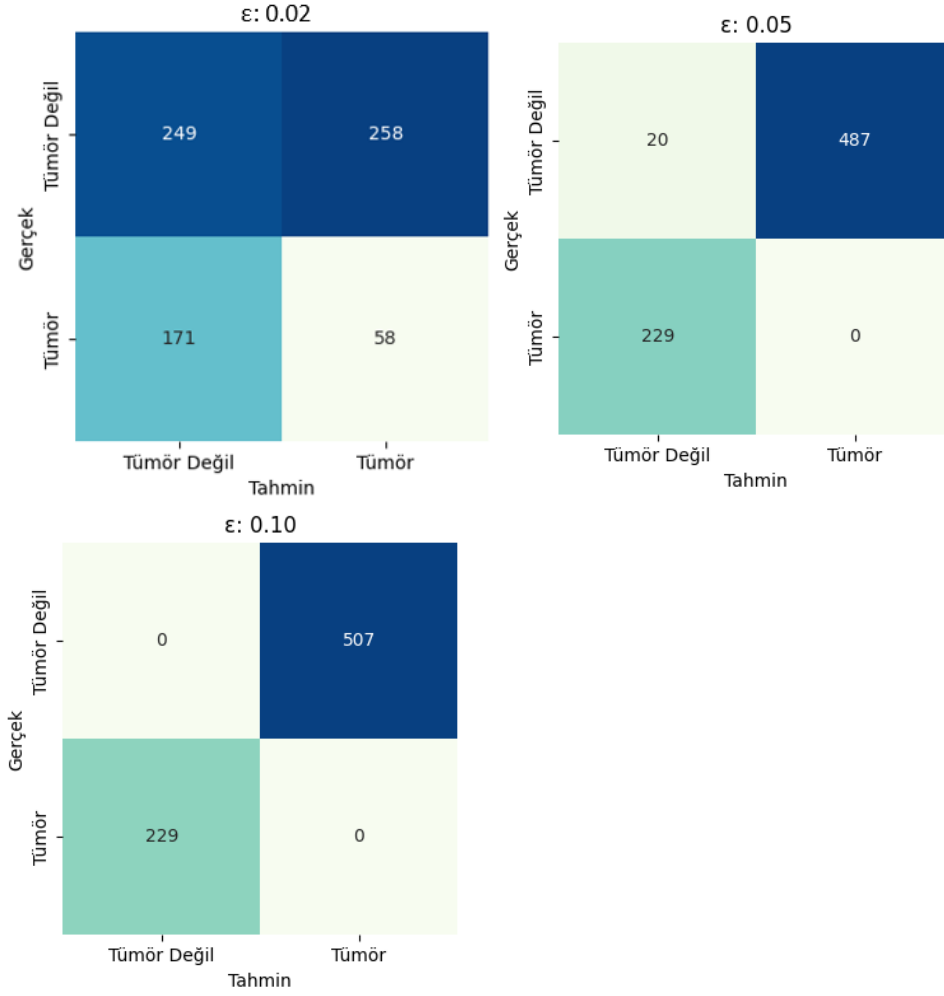


Şekil 6.148. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri

Şekil 6.148' e göre EfficientNetV2B3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %1'e kadar düşmektedir.

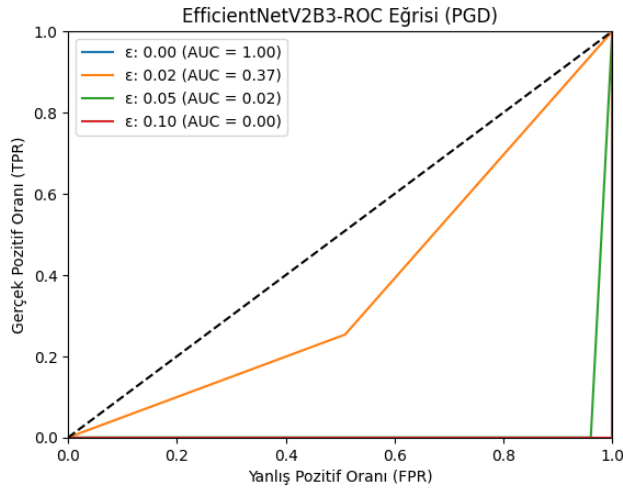
PGD saldırısı

Şekil 6.149' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.149. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait karmaşıklık matrisi

Şekil 6.149’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 171 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 258 örnek bulunmaktadır. Çizelge 6.6’ ya göre %41,71 doğruluk, %18,35 kesinlik, %25,33 geri çağırma, %21,28 F1 skoru ve %49,11 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 487 örnek bulunmaktadır. Çizelge 6.6’ ya göre %2,72 doğruluk ve %3,95 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.150’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



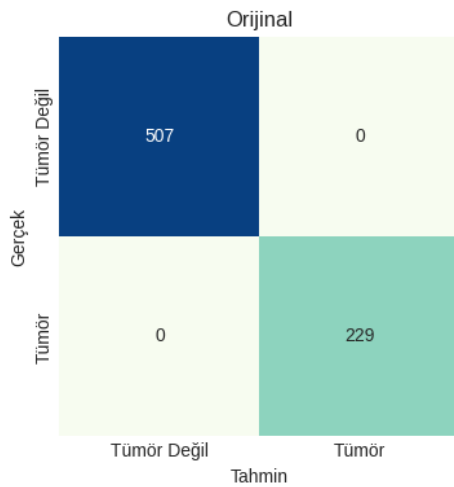
Şekil 6.150. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2B3 modeline ait ROC eğrileri

Şekil 6.150' ye göre EfficientNetV2B3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.11. EfficientNetV2S

Temiz

EfficientNetV2S modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.151'de görülmektedir.

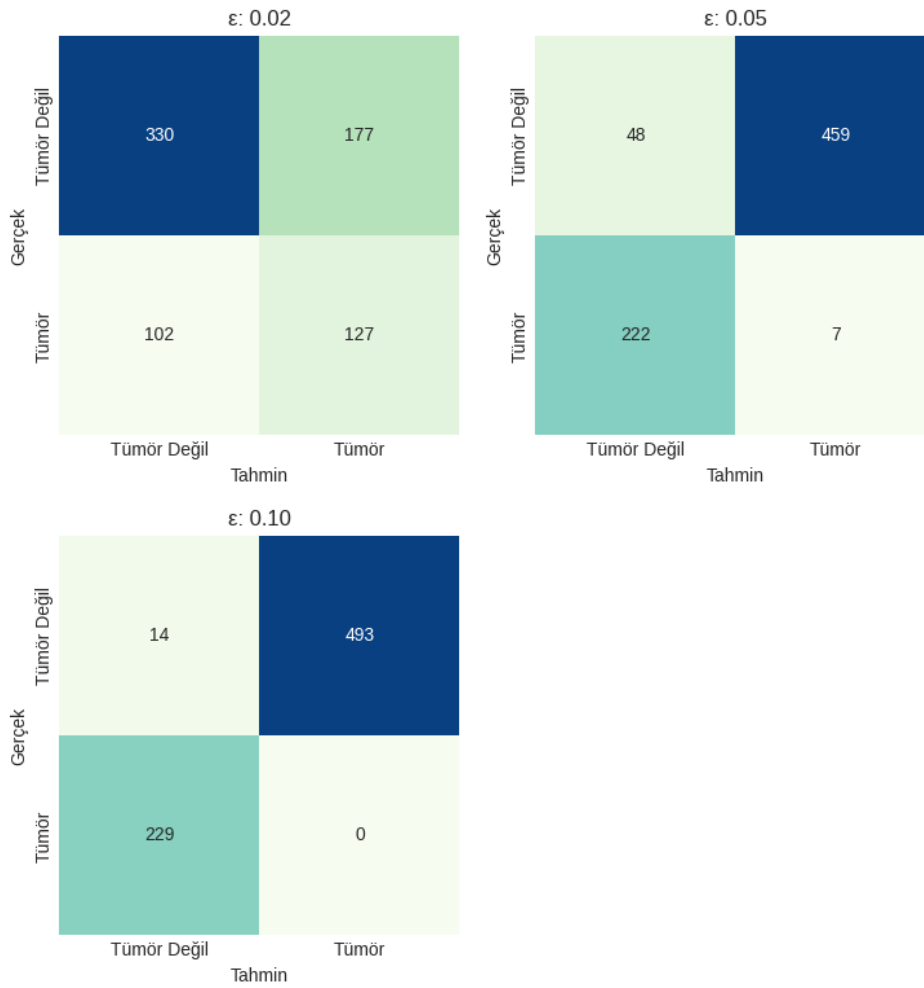


Şekil 6.151. Temiz Böbrek BT görüntüleri üzerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.151’ de görüldüğü gibi temiz test setinin EfficientNetV2S modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

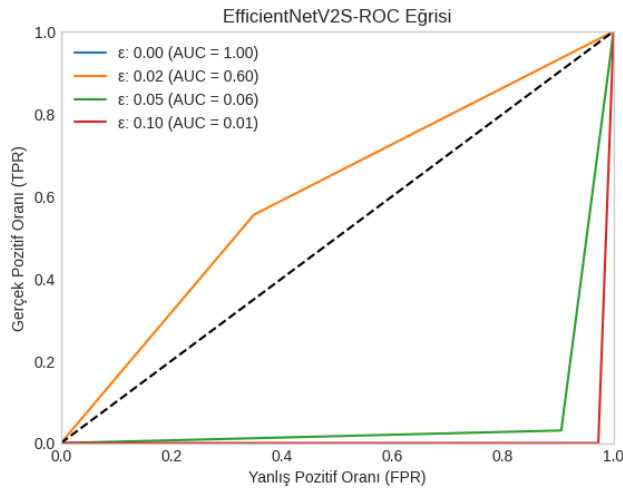
Şekil 6.152’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.152. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.152’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 102 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 177 örnek bulunmaktadır. Çizelge 6.5’ e göre %62,09 doğruluk, %41,78 kesinlik, %55,46

geri çağırma, %47,66 F1 skoru ve %65,09 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 222 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 459 örnek bulunmaktadır. Çizelge 6.5' e göre %7,47 doğruluk, %1,5 kesinlik, %3,06 geri çağırma, %2,01 F1 skoru ve %9,47 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 493 örnek bulunmaktadır. Çizelge 6.5' e göre %1,9 doğruluk ve %2,76 özgüllük elde etmektedir. Şekil 6.153'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

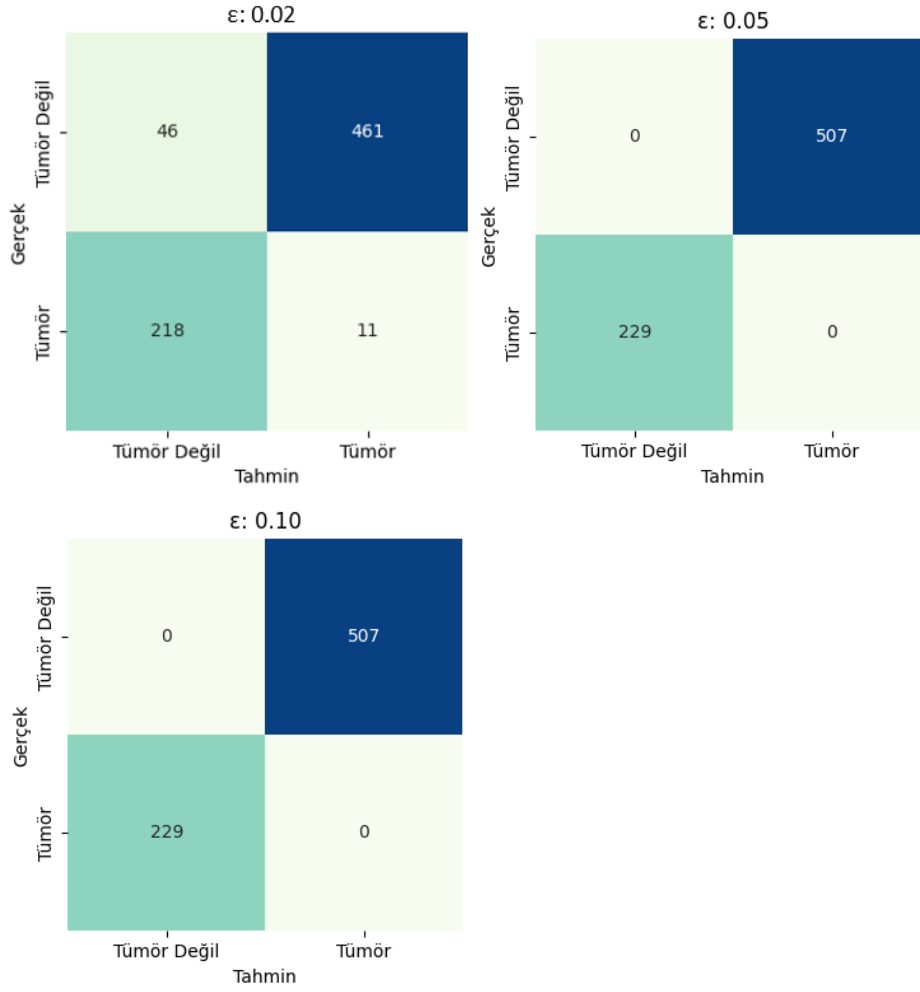


Şekil 6.153. FGSM saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait ROC eğrileri

Şekil 6.153' e göre EfficientNetV2S modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %1'e kadar düşmektedir.

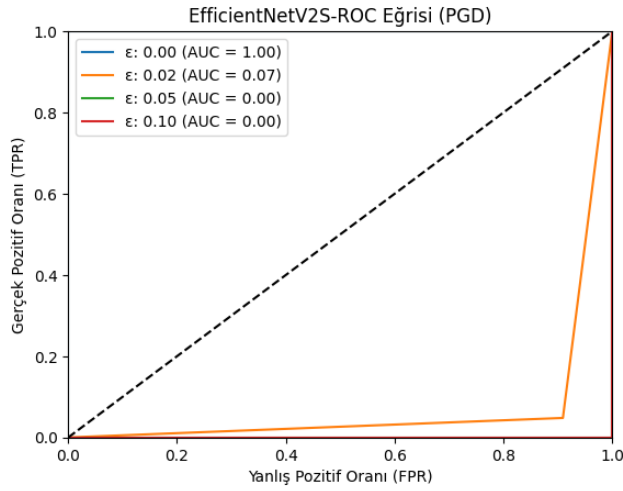
PGD saldırısı

Şekil 6.154' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.154. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait karmaşıklık matrisi

Şekil 6.154’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 218 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 461 örnek bulunmaktadır. Çizelge 6.6’ ya göre %7,75 doğruluk, %2,33 kesinlik, %4,8 geri çağırma, %3,14 F1 skoru ve %9,07 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.155’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



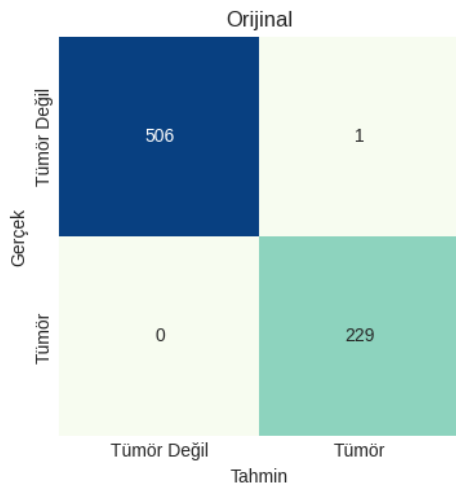
Şekil 6.155. PGD saldırısına uğrayan Böbrek BT görüntülerinde EfficientNetV2S modeline ait ROC eğrileri

Şekil 6.155' e göre EfficientNetV2S modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.12. InceptionResNetV2

Temiz

InceptionResNetV2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.156'da görülmektedir.

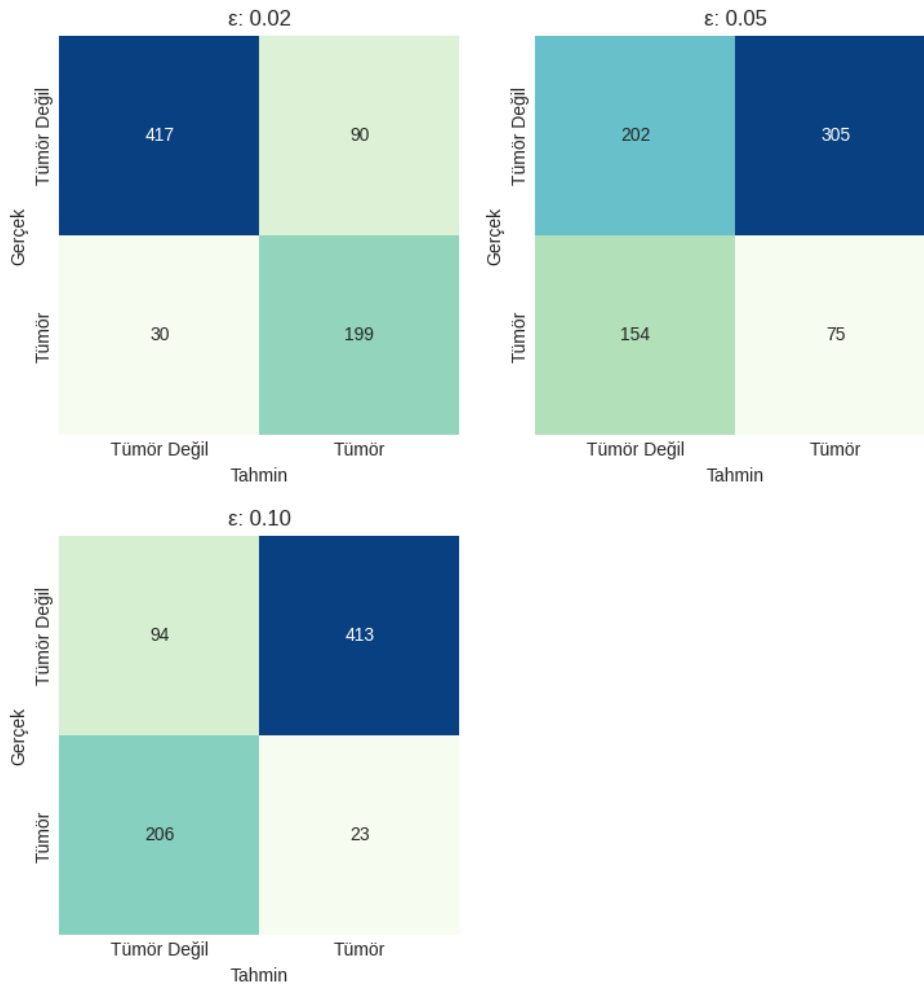


Şekil 6.156. Temiz Böbrek BT görüntüleri üzerinde InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.156’ da görüldüğü gibi temiz test setinin InceptionResNetV2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre %99,86 doğruluk, %99,57 kesinlik, %100 geri çağırma, %99,78 F1 skoru ve %99,8 özgüllük elde etmektedir.

FGSM saldırısı

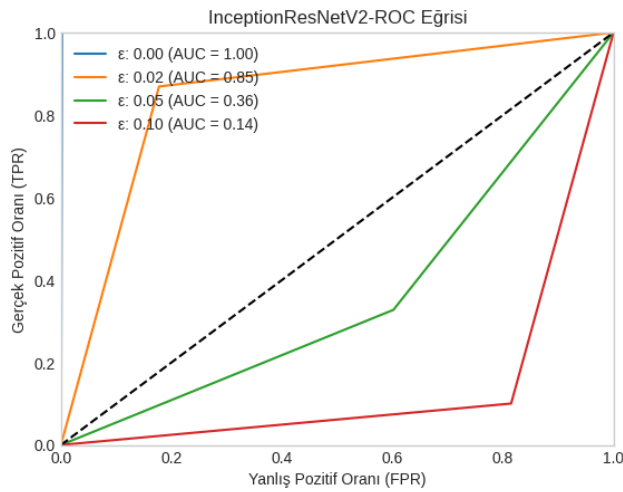
Şekil 6.157’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.157. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.157’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 30 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 90 örnek bulunmaktadır. Çizelge 6.5’ e göre %83,7 doğruluk, %68,86 kesinlik, %86,9 geri

çağırma, %76,83 F1 skoru ve %82,25 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 154 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 305 örnek bulunmaktadır. Çizelge 6.5' e göre %37,64 doğruluk, %19,74 kesinlik, %32,75 geri çağırma, %24,63 F1 skoru ve %39,74 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 206 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 413 örnek bulunmaktadır. Çizelge 6.5' e göre %15,9 doğruluk, %5,28 kesinlik, %10,04 geri çağırma, %6,92 F1 skoru ve %18,54 özgüllük elde etmektedir. Şekil 6.158'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

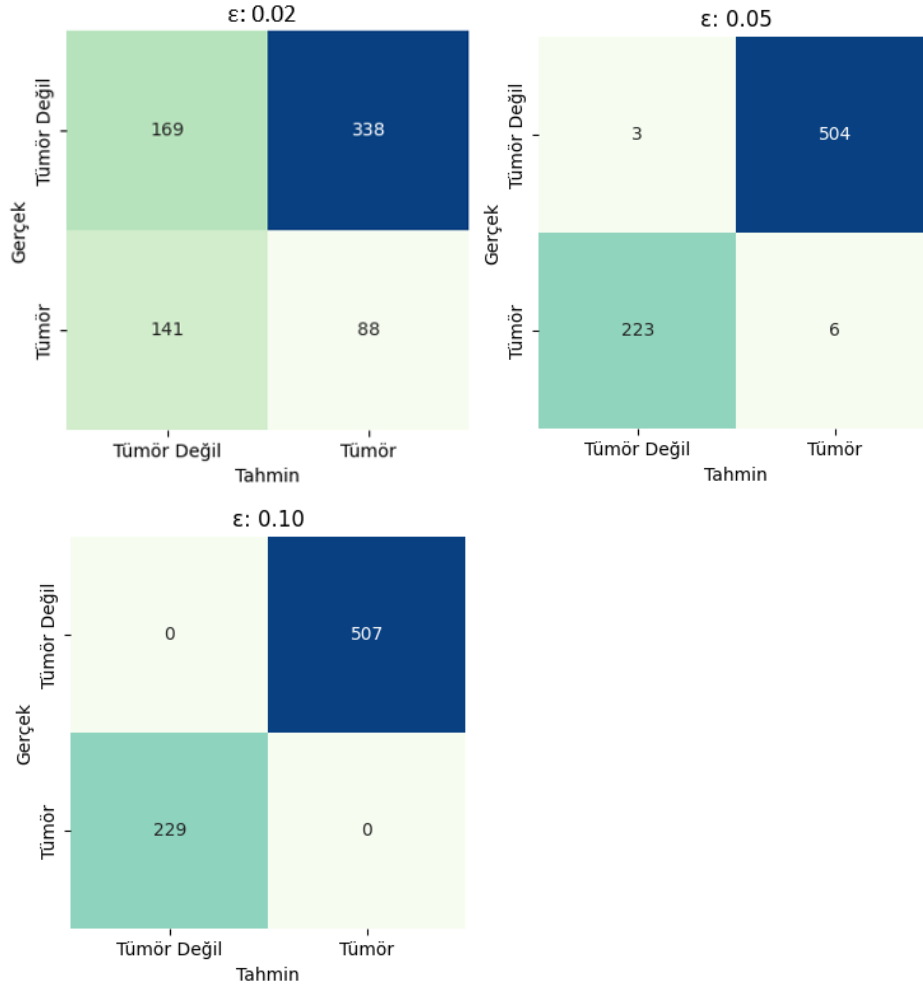


Şekil 6.158. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri

Şekil 6.158' e göre InceptionResNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %14'e kadar düşmektedir.

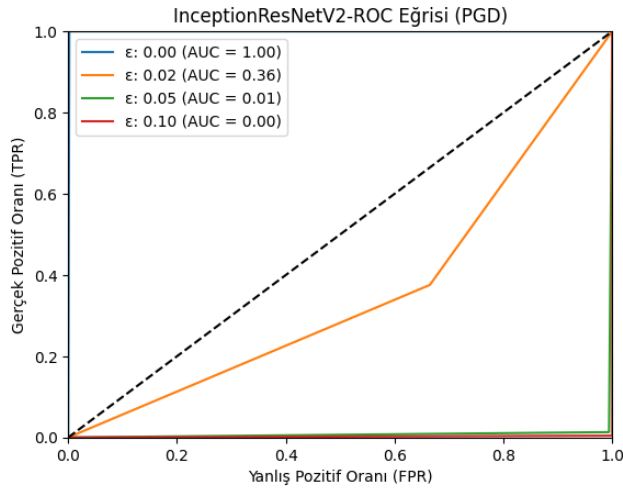
PGD saldırısı

Şekil 6.159' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.159. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait karmaşıklık matrisi

Şekil 6.159’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 141 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 338 örnek bulunmaktadır. Çizelge 6.6’ ya göre %34,92 doğruluk, %20,66 kesinlik, %38,43 geri çağırma, %26,87 F1 skoru ve %33,33 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 223 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 504 örnek bulunmaktadır. Çizelge 6.6’ ya göre %1,22 doğruluk, %1,18 kesinlik, %2,62 geri çağırma, %1,62 F1 skoru ve %0,59 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.160’da pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



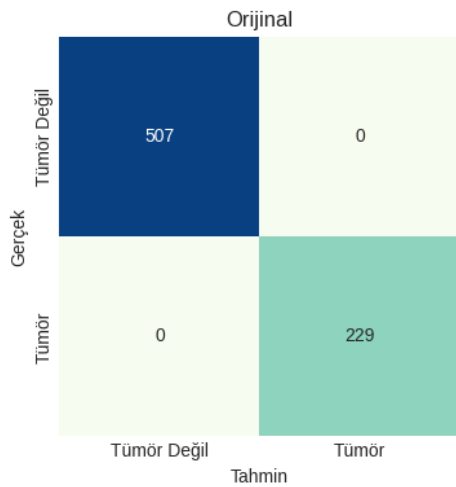
Şekil 6.160. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionResNetV2 modeline ait ROC eğrileri

Şekil 6.160' a göre InceptionResNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.13. InceptionV3

Temiz

InceptionV3 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.161'de görülmektedir.

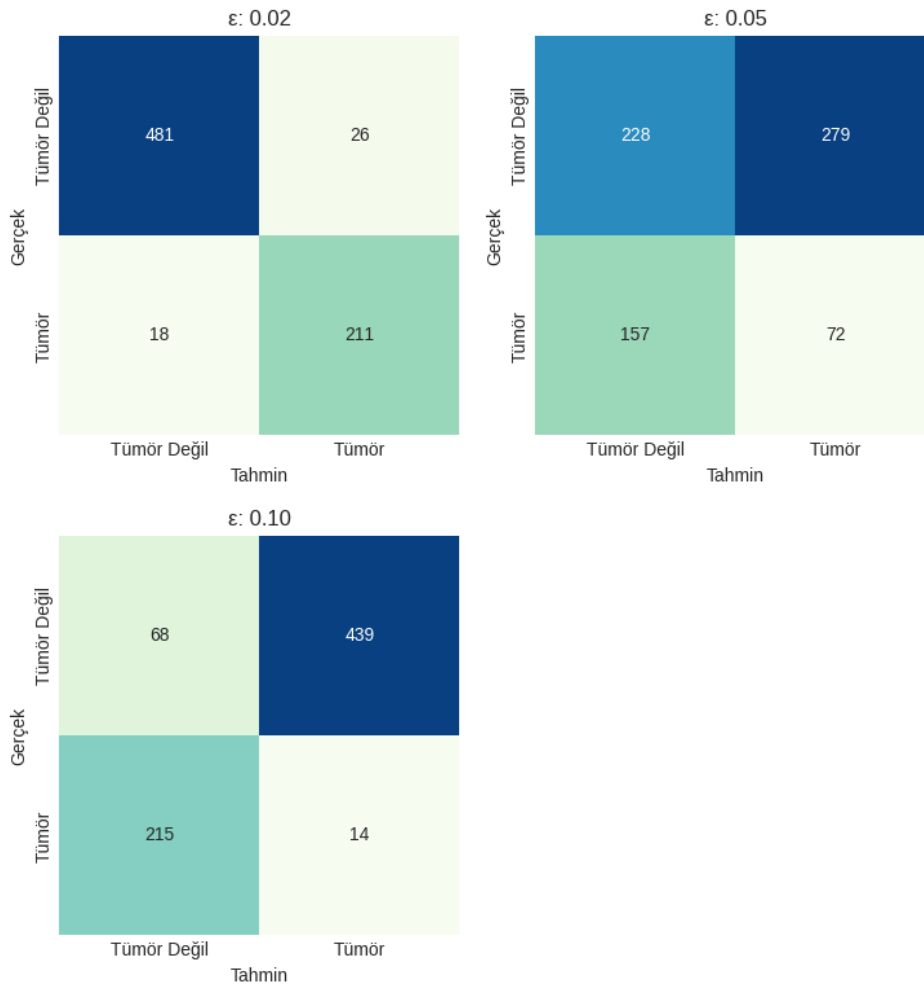


Şekil 6.161. Temiz Böbrek BT görüntüleri üzerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.161’ de görüldüğü gibi temiz test setinin InceptionV3 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

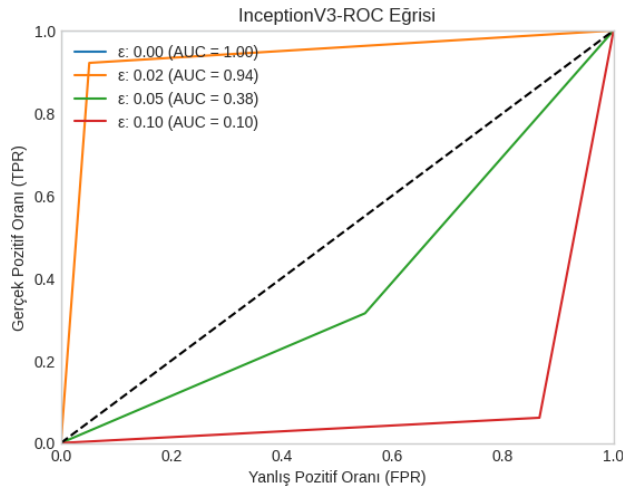
Şekil 6.162’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.162. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.162’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 18 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 26 örnek bulunmaktadır. Çizelge 6.5’ e göre %94,02 doğruluk, %89,03 kesinlik, %92,14 geri

çağırma, %90,56 F1 skoru ve %94,87 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 157 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 279 örnek bulunmaktadır. Çizelge 6.5' e göre %40,76 doğruluk, %20,51 kesinlik, %31,44 geri çağırma, %24,83 F1 skoru ve %44,97 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 215 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 439 örnek bulunmaktadır. Çizelge 6.5' e göre %11,14 doğruluk, %3,09 kesinlik, %6,11 geri çağırma, %4,11 F1 skoru ve %13,41 özgüllük elde etmektedir. Şekil 6.163'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

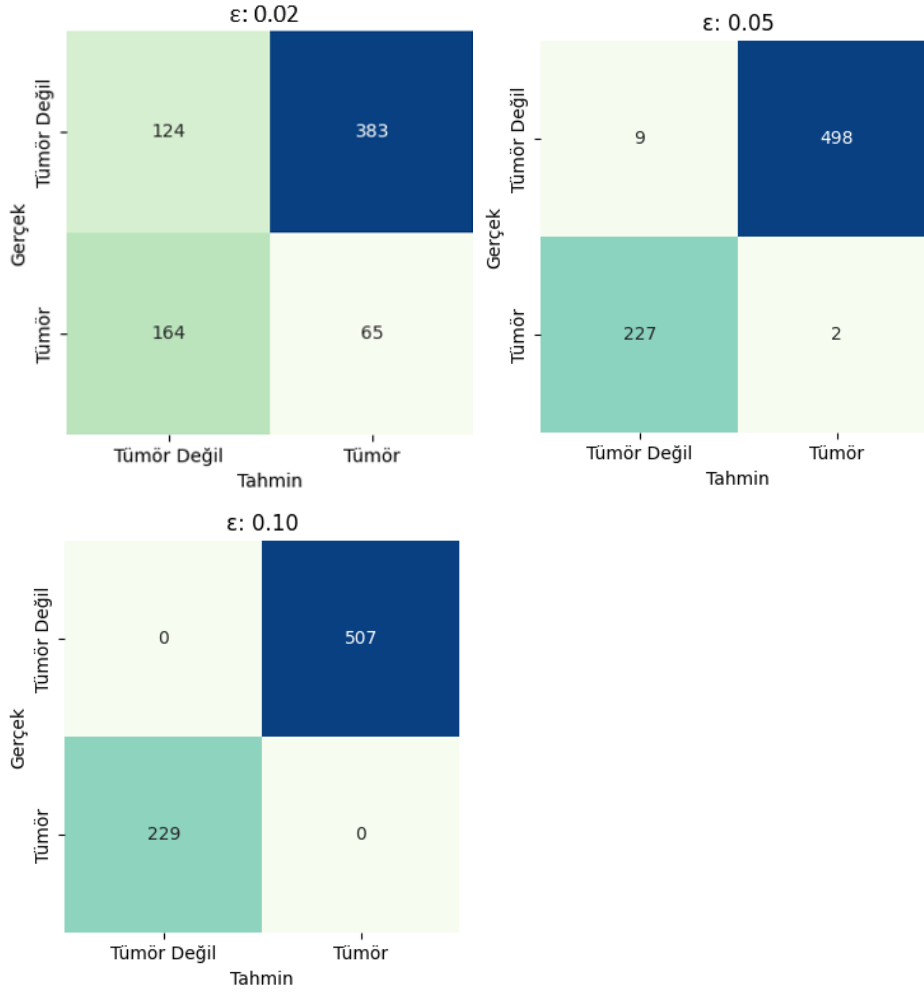


Şekil 6.163. FGSM saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait ROC eğrileri

Şekil 6.163' e göre InceptionV3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %10'a kadar düşmektedir.

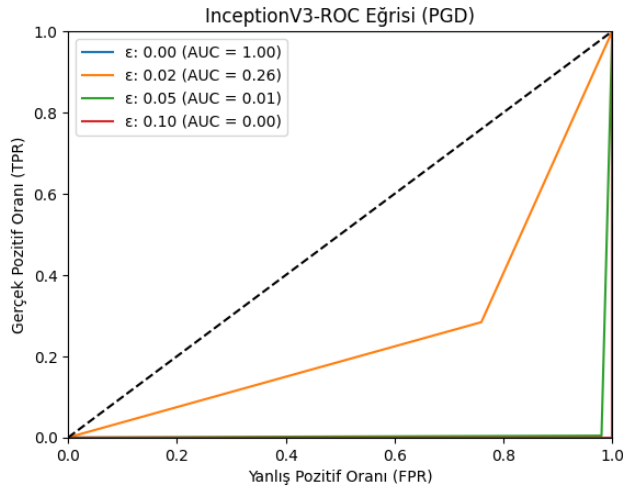
PGD saldırısı

Şekil 6.164' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.164. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait karmaşıklık matrisi

Şekil 6.164’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 164 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 383 örnek bulunmaktadır. Çizelge 6.6’ ya göre %25,68 doğruluk, %14,51 kesinlik, %28,38 geri çağırma, %19,2 F1 skoru ve %24,46 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 227 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 498 örnek bulunmaktadır. Çizelge 6.6’ ya göre %1,5 doğruluk, %0,4 kesinlik, %0,87 geri çağırma, %0,55 F1 skoru ve %1,78 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.165’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



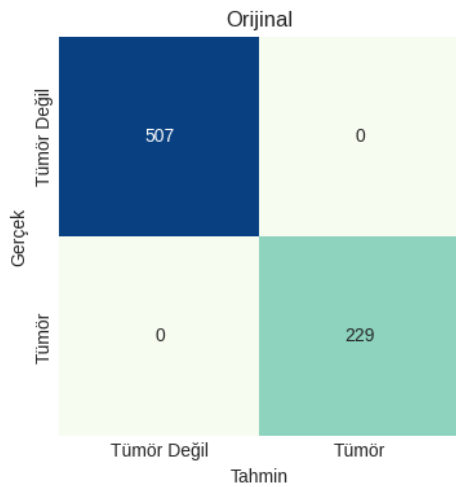
Şekil 6.165. PGD saldırısına uğrayan Böbrek BT görüntülerinde InceptionV3 modeline ait ROC eğrileri

Şekil 6.165' e göre InceptionV3 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.14. MobileNetV1

Temiz

MobileNetV1 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.166'da görülmektedir.

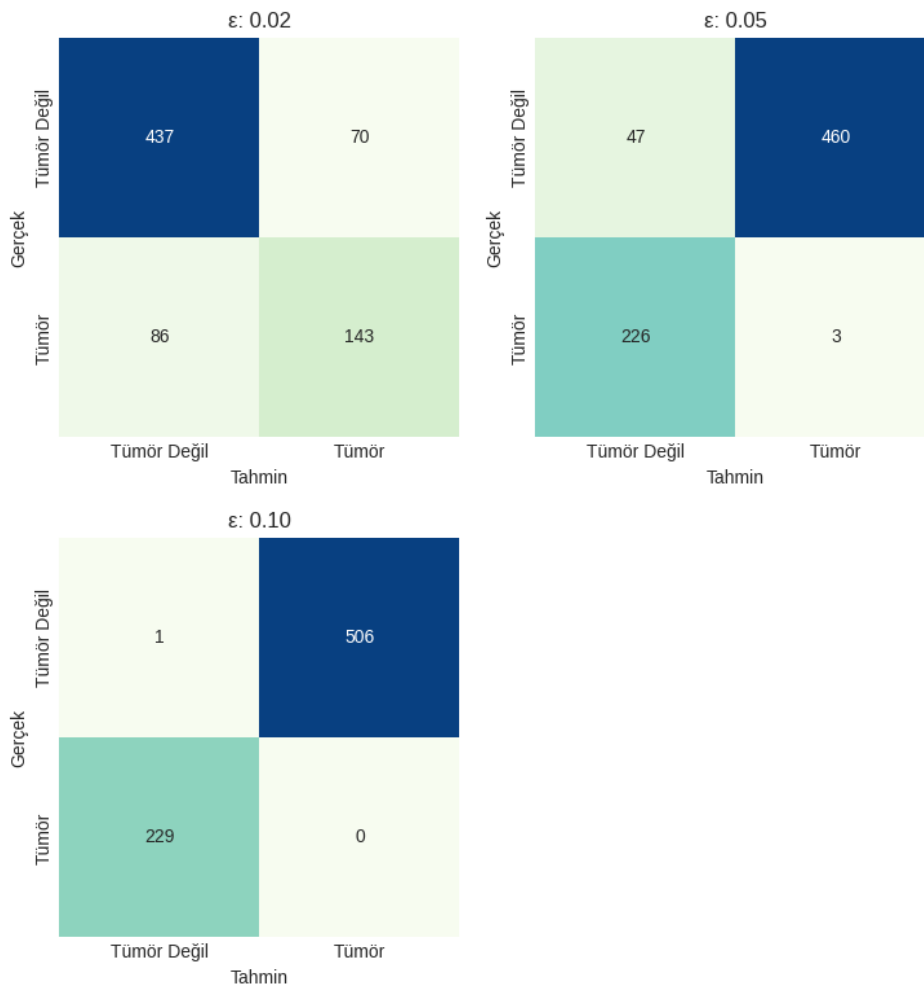


Şekil 6.166. Temiz Böbrek BT görüntüleri üzerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.166’ da görüldüğü gibi temiz test setinin MobileNetV1 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

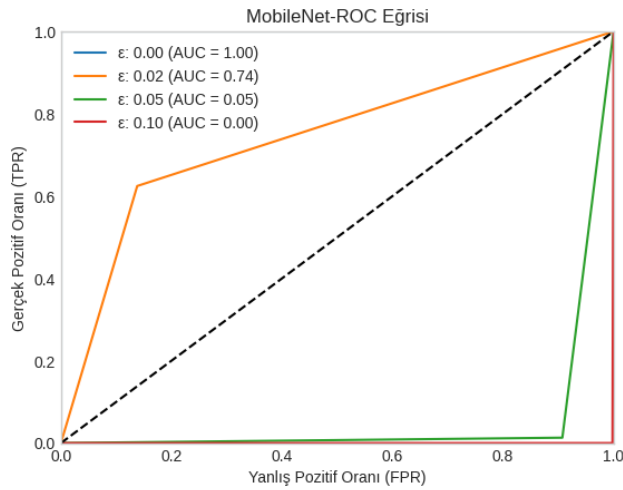
Şekil 6.167’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.167. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.167’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 86 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 70 örnek bulunmaktadır. Çizelge 6.5’ e göre %78,8 doğruluk, %67,14 kesinlik, %62,45 geri

çağırma, %64,71 F1 skoru ve %86,19 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 226 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 460 örnek bulunmaktadır. Çizelge 6.5' e göre %6,79 doğruluk, %0,65 kesinlik, %1,31 geri çağırma, %0,87 F1 skoru ve %9,27 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 506 örnek bulunmaktadır. Çizelge 6.5' e göre %0,14 doğruluk elde etmektedir. Şekil 6.168'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

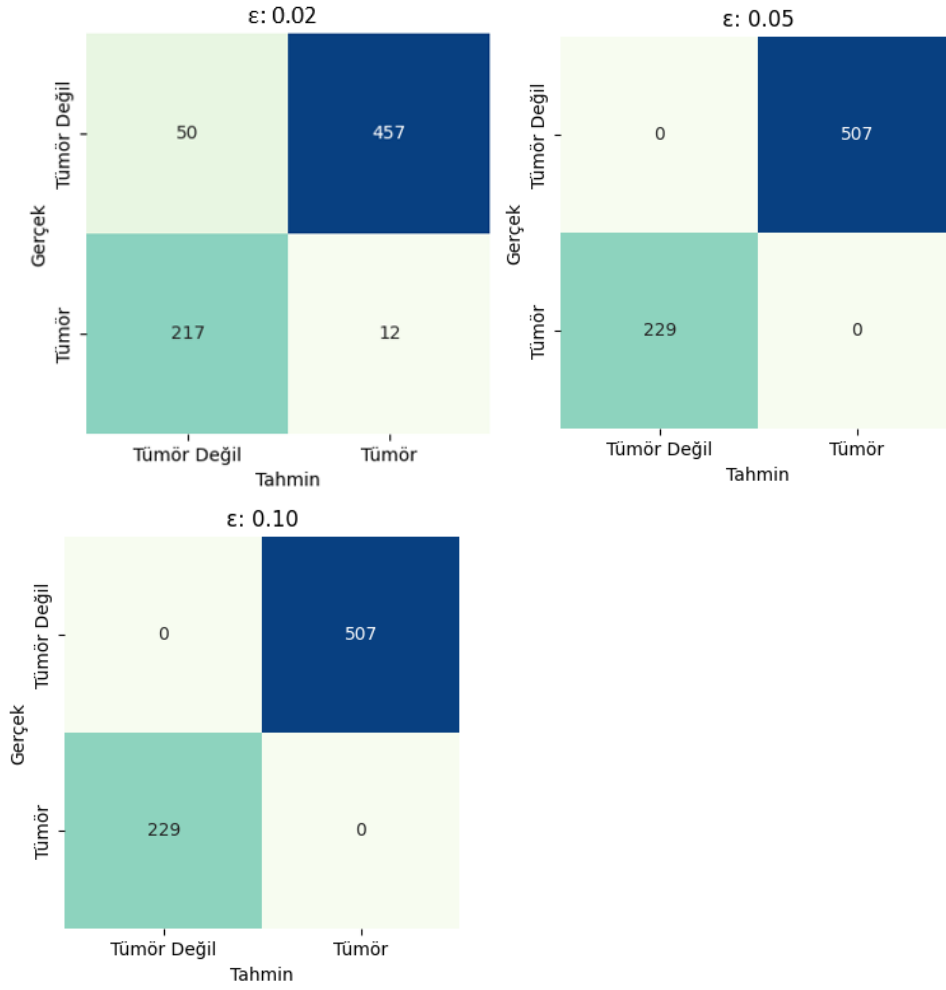


Şekil 6.168. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait ROC eğrileri

Şekil 6.168' e göre MobileNetV1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

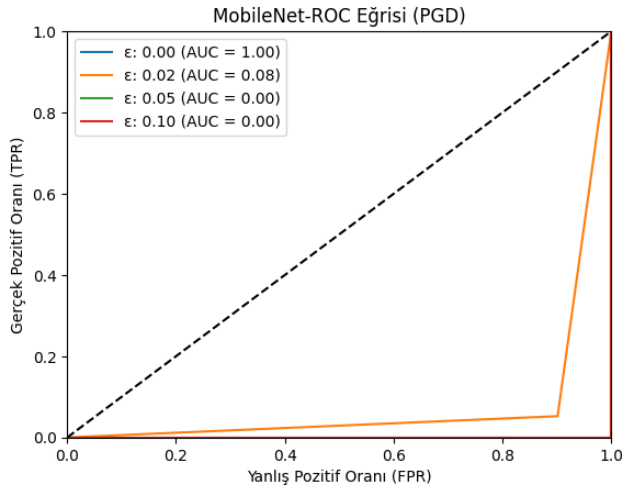
PGD saldırısı

Şekil 6.169' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.169. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait karmaşıklık matrisi

Şekil 6.169’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 217 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 457 örnek bulunmaktadır. Çizelge 6.6’ ya göre %8,42 doğruluk, %2,56 kesinlik, %5,24 geri çağırma, %3,44 F1 skoru ve %9,86 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.170’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



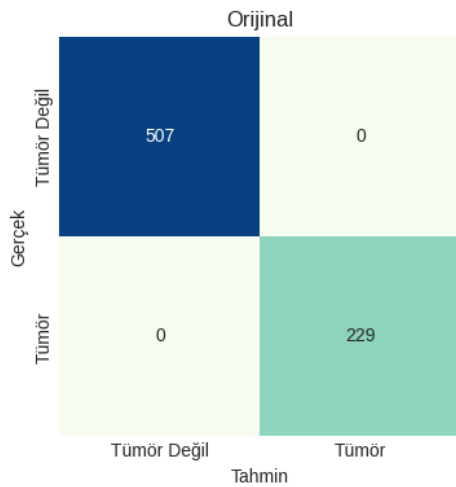
Şekil 6.170. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV1 modeline ait ROC eğrileri

Şekil 6.170' e göre MobileNetV1 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.15. MobileNetV2

Temiz

MobileNetV2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.171'de görülmektedir.

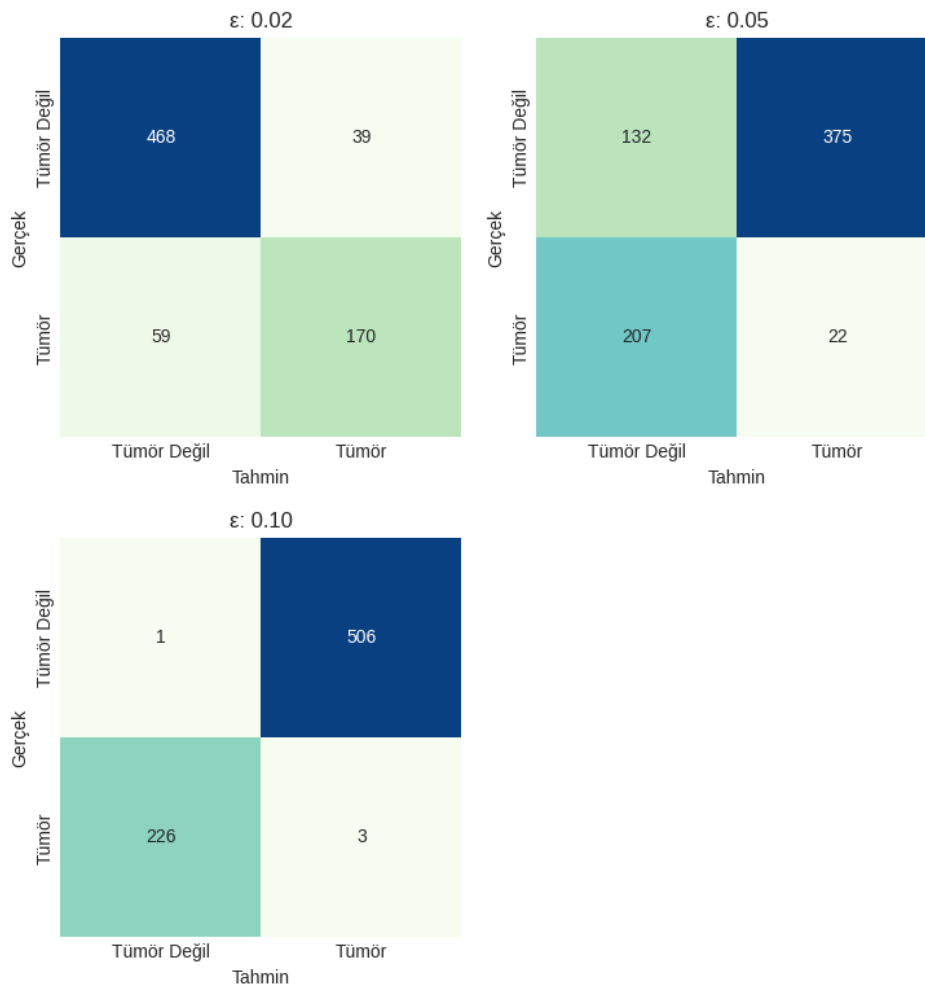


Şekil 6.171. Temiz Böbrek BT görüntüleri üzerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.171’ de görüldüğü gibi temiz test setinin MobileNetV2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

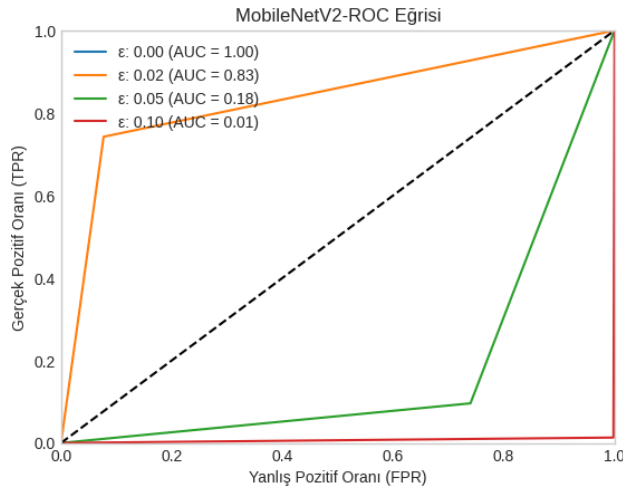
Şekil 6.172’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.172. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.172’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 59 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 39 örnek bulunmaktadır. Çizelge 6.5’ e göre %86,69 doğruluk, %81,34 kesinlik, %74,24 geri

çağırma, %77,63 F1 skoru ve %92,31 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 207 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 375 örnek bulunmaktadır. Çizelge 6.5' e göre %20,92 doğruluk, %5,54 kesinlik, %9,61 geri çağırma, %7,03 F1 skoru ve %26,04 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 226 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 506 örnek bulunmaktadır. Çizelge 6.5' e göre %0,54 doğruluk, %0,59 kesinlik, %1,31 geri çağırma, %0,81 F1 skoru ve %0,2 özgüllük elde etmektedir. Şekil 6.173'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

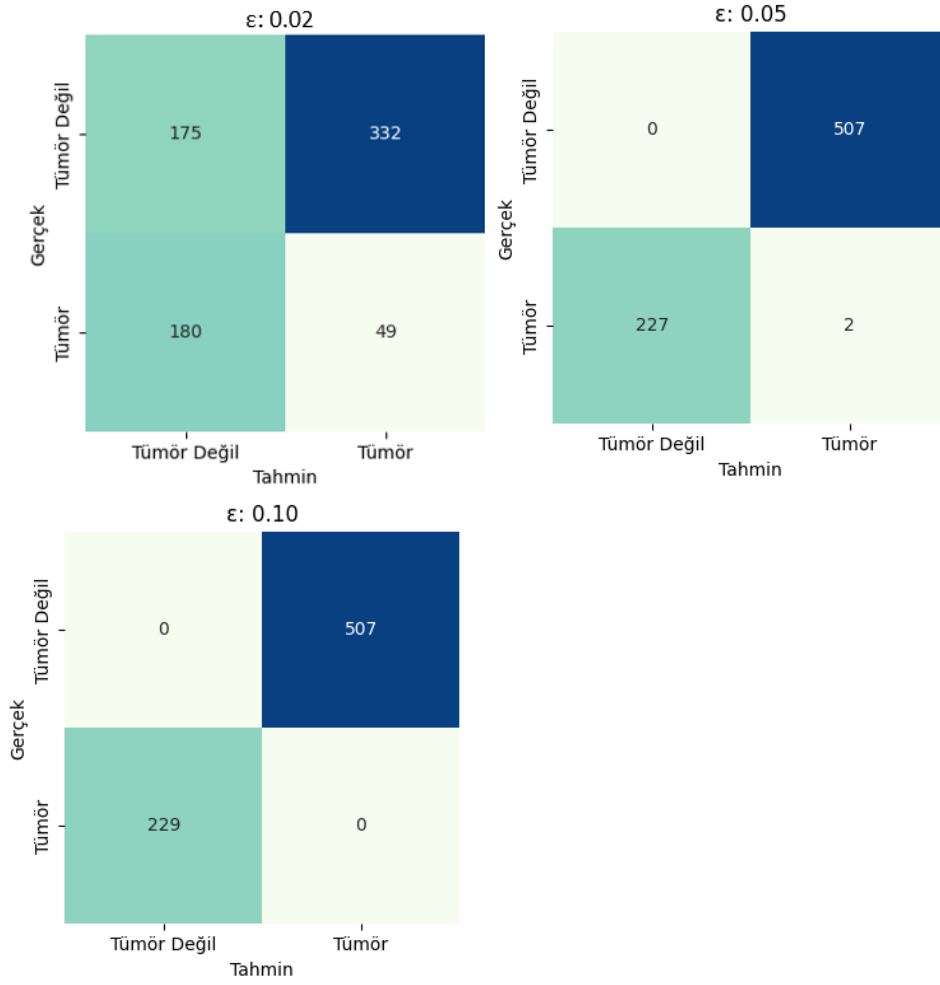


Şekil 6.173. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait ROC eğrileri

Şekil 6.173' e göre MobileNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %1'e kadar düşmektedir.

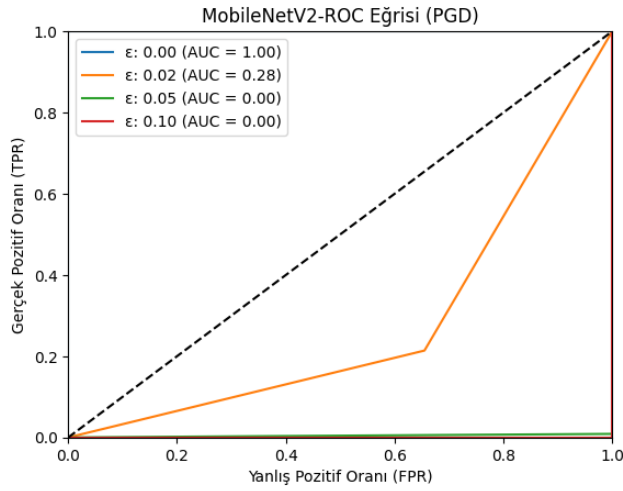
PGD saldırısı

Şekil 6.174' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.174. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait karmaşıklık matrisi

Şekil 6.174’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 180 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 332 örnek bulunmaktadır. Çizelge 6.6’ ya göre %30,44 doğruluk, %12,86 kesinlik, %21,4 geri çağırma, %16,07 F1 skoru ve %34,52 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 227 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 507 örnek bulunmaktadır. Çizelge 6.6’ ya göre %0,27 doğruluk, %0,39 kesinlik, %0,87 geri çağırma ve %0,54 F1 skoru elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.165’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



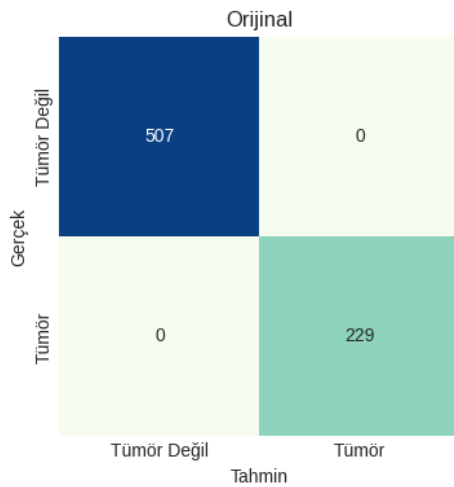
Şekil 6.175. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV2 modeline ait ROC eğrileri

Şekil 6.175' e göre MobileNetV2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.16. MobileNetV3Small

Temiz

MobileNetV3Small modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.176'da görülmektedir.

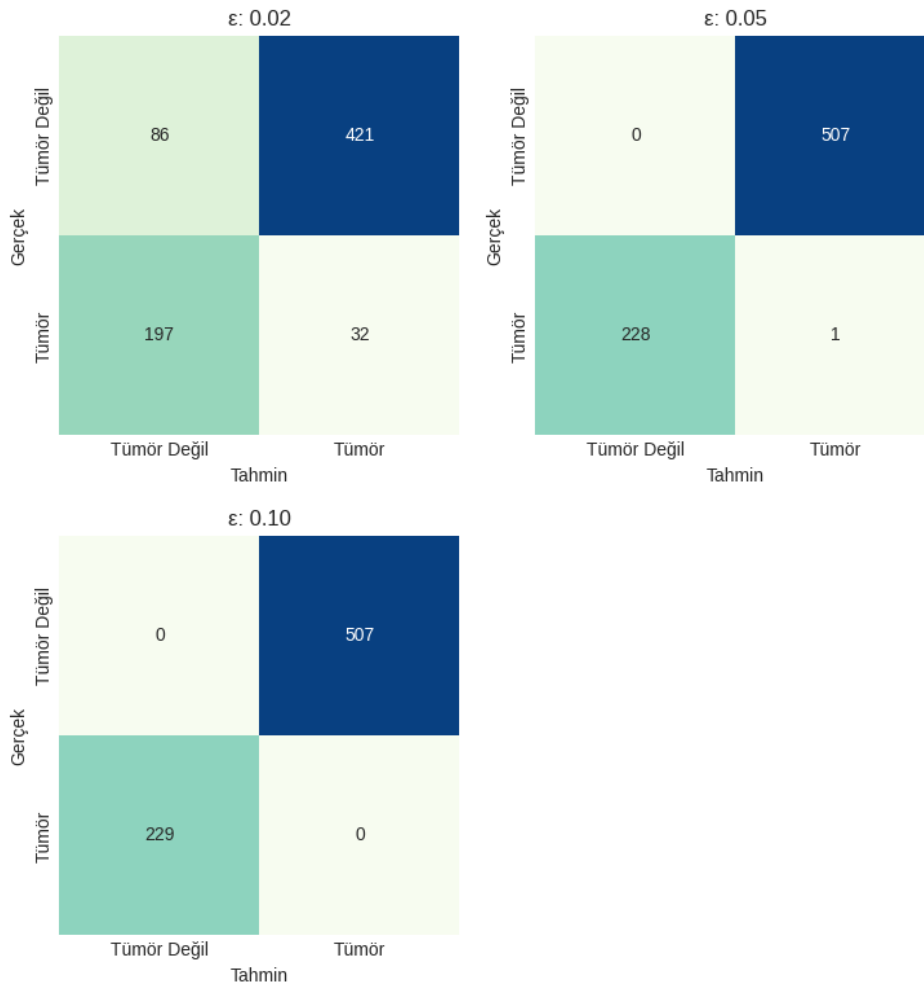


Şekil 6.176. Temiz Böbrek BT görüntüleri üzerinde MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.176’ da görüldüğü gibi temiz test setinin MobileNetV3Small modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

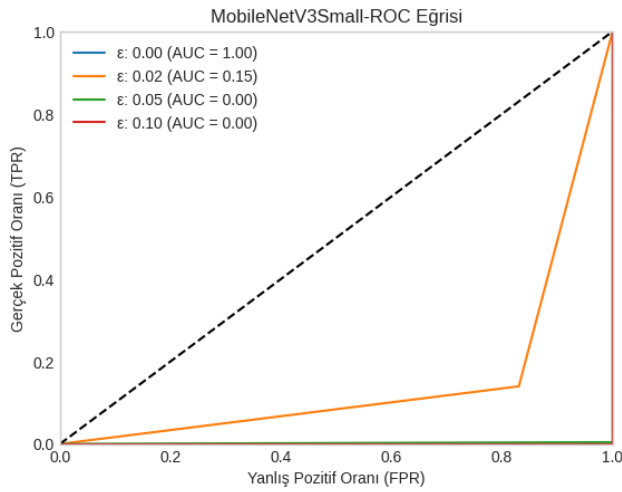
Şekil 6.177’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.177. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.177’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 197 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 421 örnek bulunmaktadır. Çizelge 6.5’ e göre %16,04 doğruluk, %7,06 kesinlik, %13,97

geri çağırma, %9,38 F1 skoru ve %16,96 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 228 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 507 örnek bulunmaktadır. Çizelge 6.5' e göre %0,14 doğruluk, %0,2 kesinlik, %0,44 geri çağırma ve %0,27 F1 skoru elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.178'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

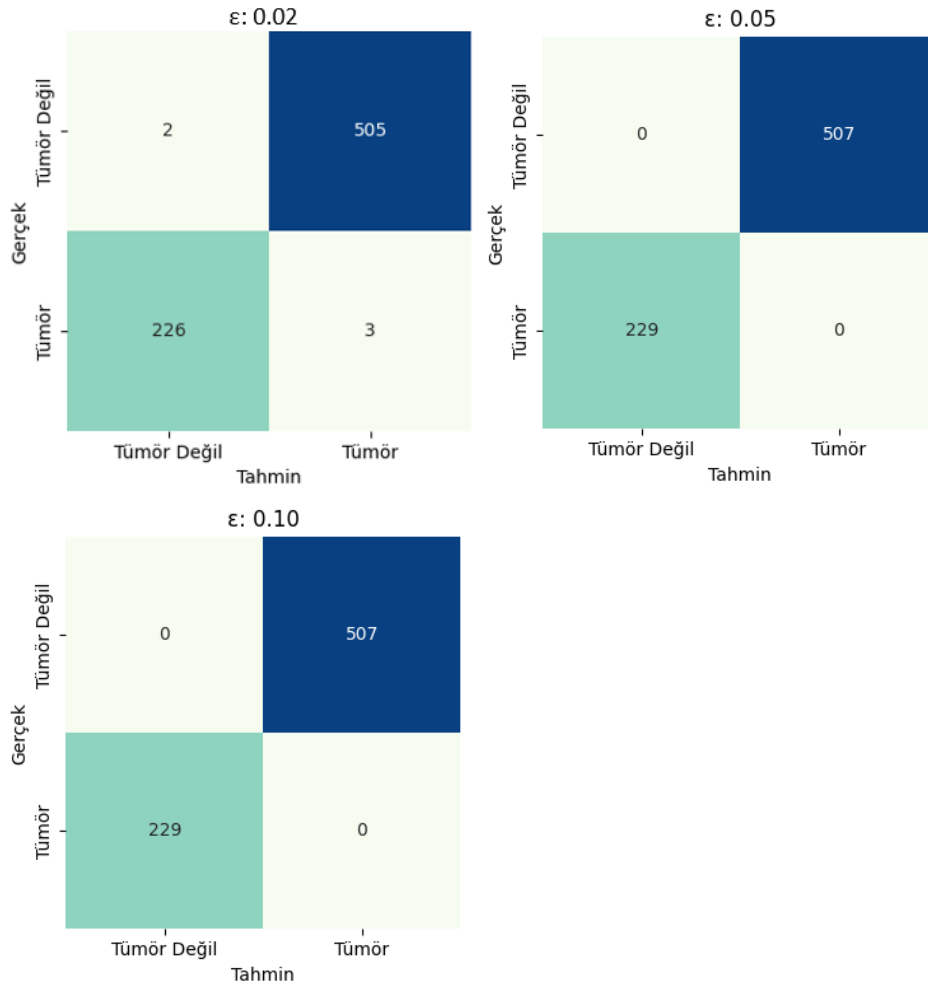


Şekil 6.178. FGSM saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait ROC eğrileri

Şekil 6.178' e göre MobileNetV3Small modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

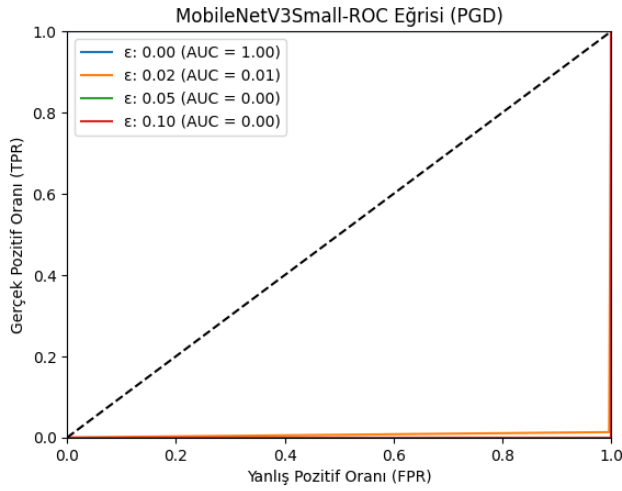
PGD saldırısı

Şekil 6.179' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.179. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait karmaşıklık matrisi

Şekil 6.179’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 226 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 505 örnek bulunmaktadır. Çizelge 6.6’ ya göre %0,68 doğruluk, %0,59 kesinlik, %1,31 geri çağırma, %0,81 F1 skoru ve %0,4 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.180’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



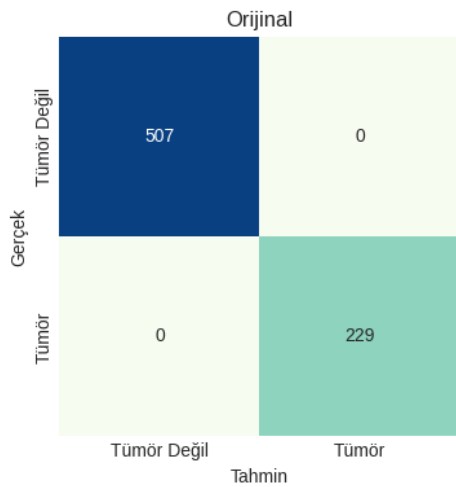
Şekil 6.180. PGD saldırısına uğrayan Böbrek BT görüntülerinde MobileNetV3Small modeline ait ROC eğrileri

Şekil 6.180' e göre MobileNetV3Small modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.17. NASNetMobile

Temiz

NASNetMobile modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.181'de görülmektedir.

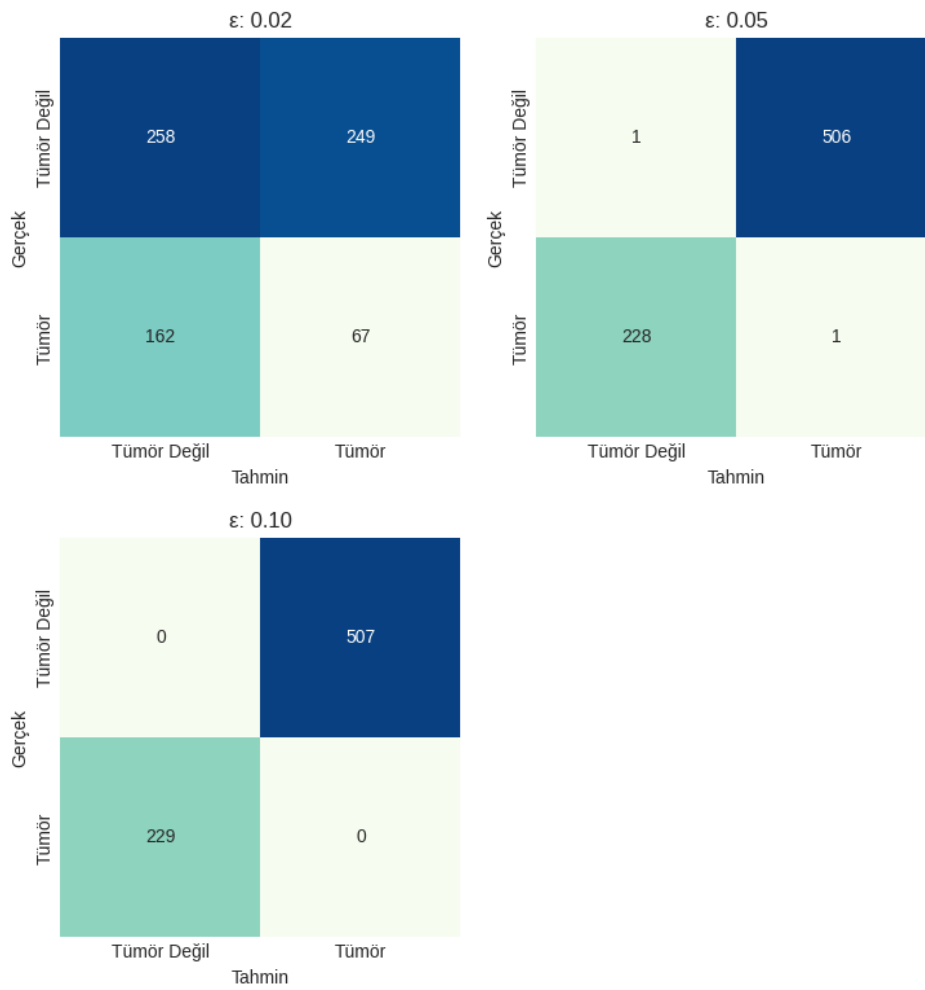


Şekil 6.181. Temiz Böbrek BT görüntüleri üzerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.181’ de görüldüğü gibi temiz test setinin NASNetMobile modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

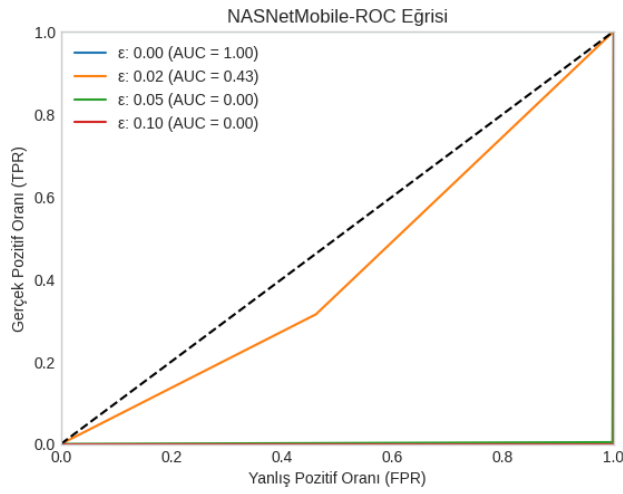
Şekil 6.182’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.182. FGSM saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.182’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 162 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 249 örnek bulunmaktadır. Çizelge 6.5’ e göre %44,16 doğruluk, %21,2 kesinlik, %29,26

geri çağırma, %24,59 F1 skoru ve %50,89 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 228 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 506 örnek bulunmaktadır. Çizelge 6.5' e göre %0,27 doğruluk, %0,2 kesinlik, %0,44 geri çağırma, %0,27 F1 skoru ve %0,2 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.183'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

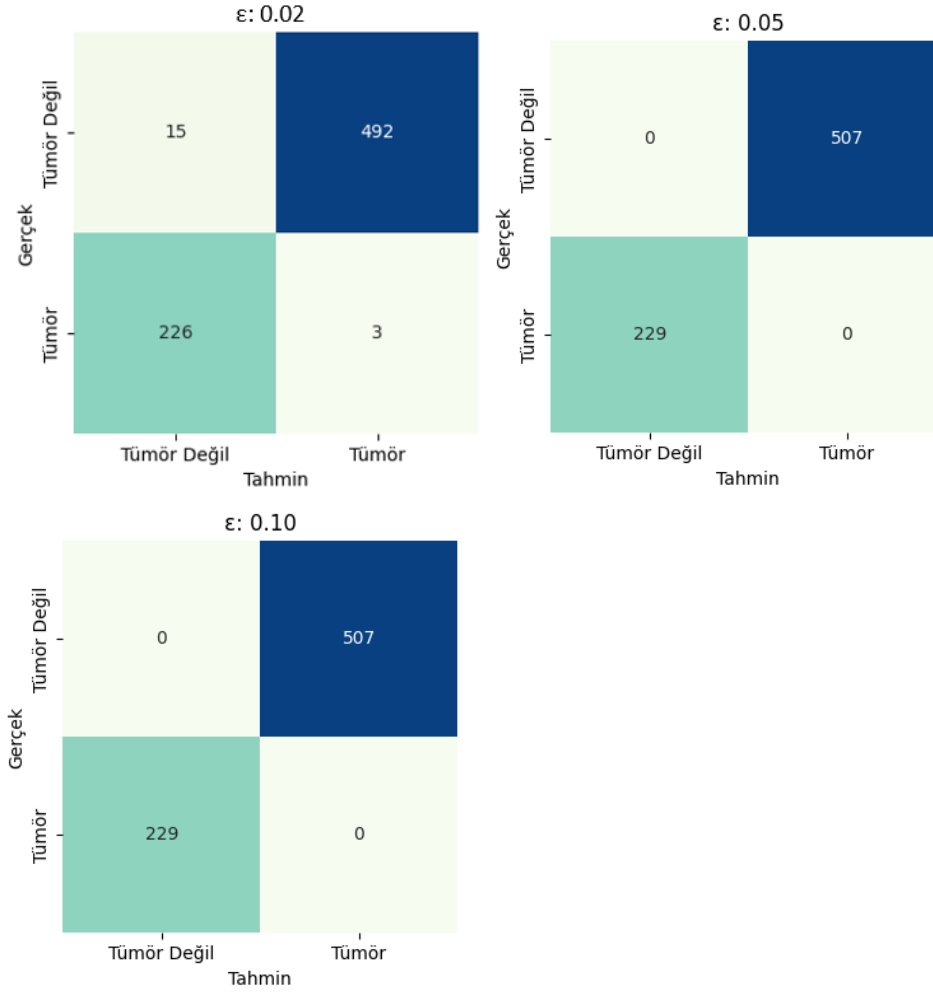


Şekil 6.183. FGSM saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait ROC eğrileri

Şekil 6.183.' e göre NASNetMobile modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

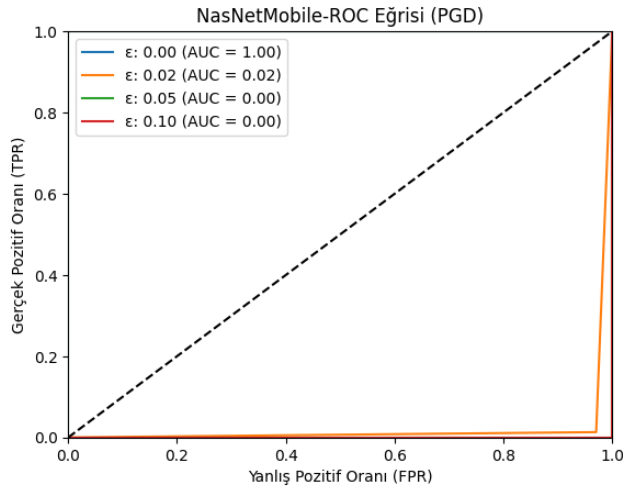
PGD saldırısı

Şekil 6.184' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.184. PGD saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait karmaşıklık matrisi

Şekil 6.184’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 226 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 492 örnek bulunmaktadır. Çizelge 6.6’ ya göre %2,45 doğruluk, %0,61 kesinlik, %1,31 geri çağırma, %0,83 F1 skoru ve %2,96 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.185’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



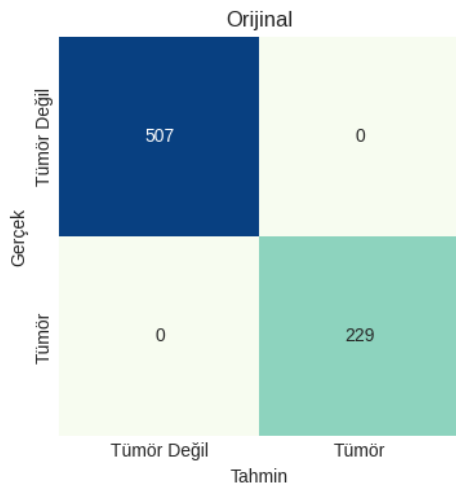
Şekil 6.185. PGD saldırısına uğrayan Böbrek BT görüntülerinde NASNetMobile modeline ait ROC eğrileri

Şekil 6.185' e göre NASNetMobile modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.18. ResNet50V2

Temiz

ResNet50V2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.186'da görülmektedir.

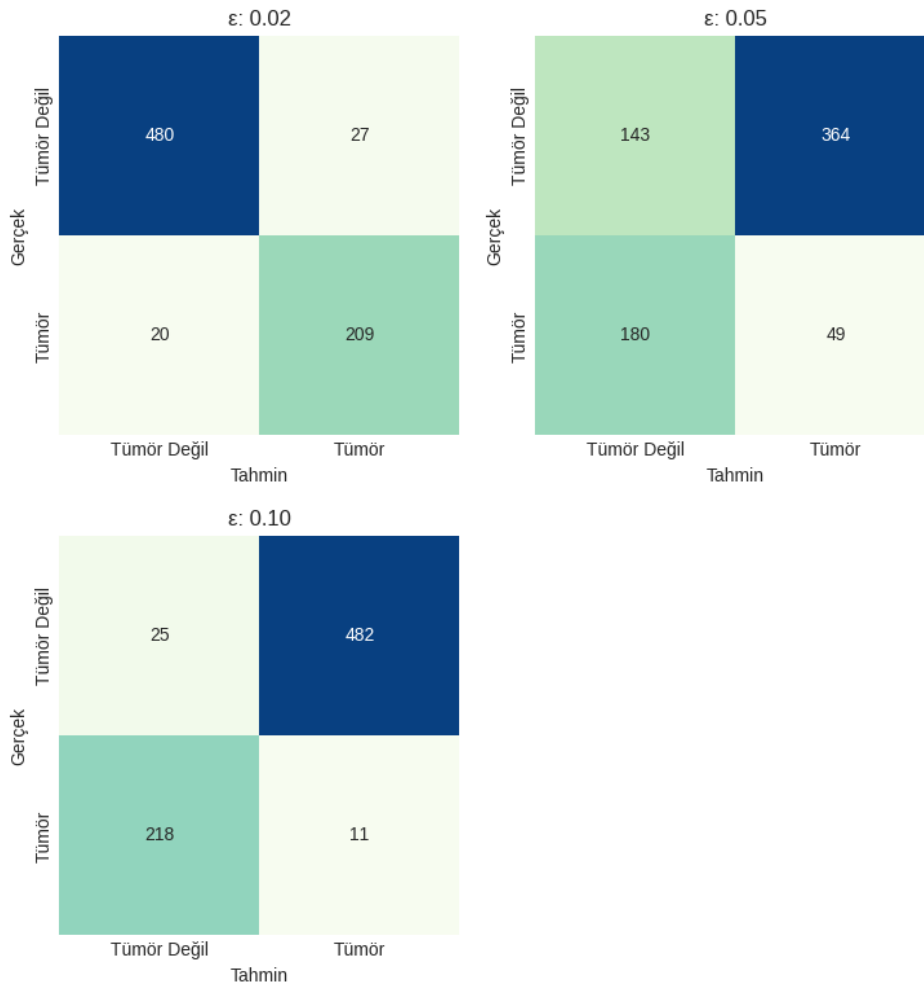


Şekil 6.186. Temiz Böbrek BT görüntüleri üzerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.186’ da görüldüğü gibi temiz test setinin ResNet50V2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

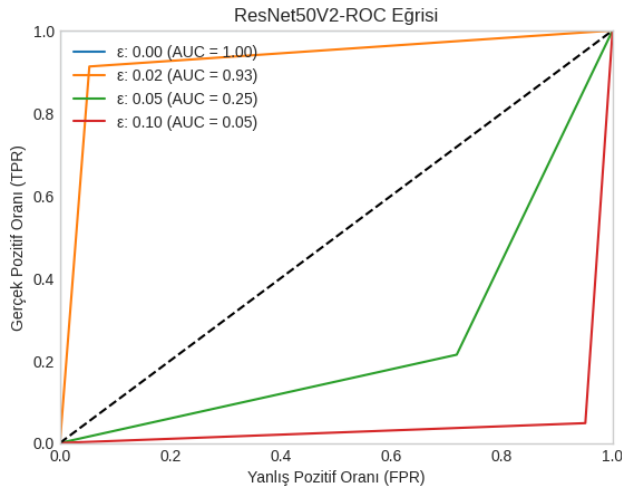
Şekil 6.187’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.187. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.187’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 20 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 27 örnek bulunmaktadır. Çizelge 6.5’ e göre %93,61 doğruluk, %88,56 kesinlik, %91,27 geri

çağırma, %89,89 F1 skoru ve %94,68 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 180 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 364 örnek bulunmaktadır. Çizelge 6.5' e göre %26,09 doğruluk, %11,86 kesinlik, %21,4 geri çağırma, %15,27 F1 skoru ve %28,21 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 218 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 482 örnek bulunmaktadır. Çizelge 6.5' e göre %4,89 doğruluk, %2,23 kesinlik, %4,8 geri çağırma, %3,05 F1 skoru ve %4,93 özgüllük elde etmektedir. Şekil 6.188'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

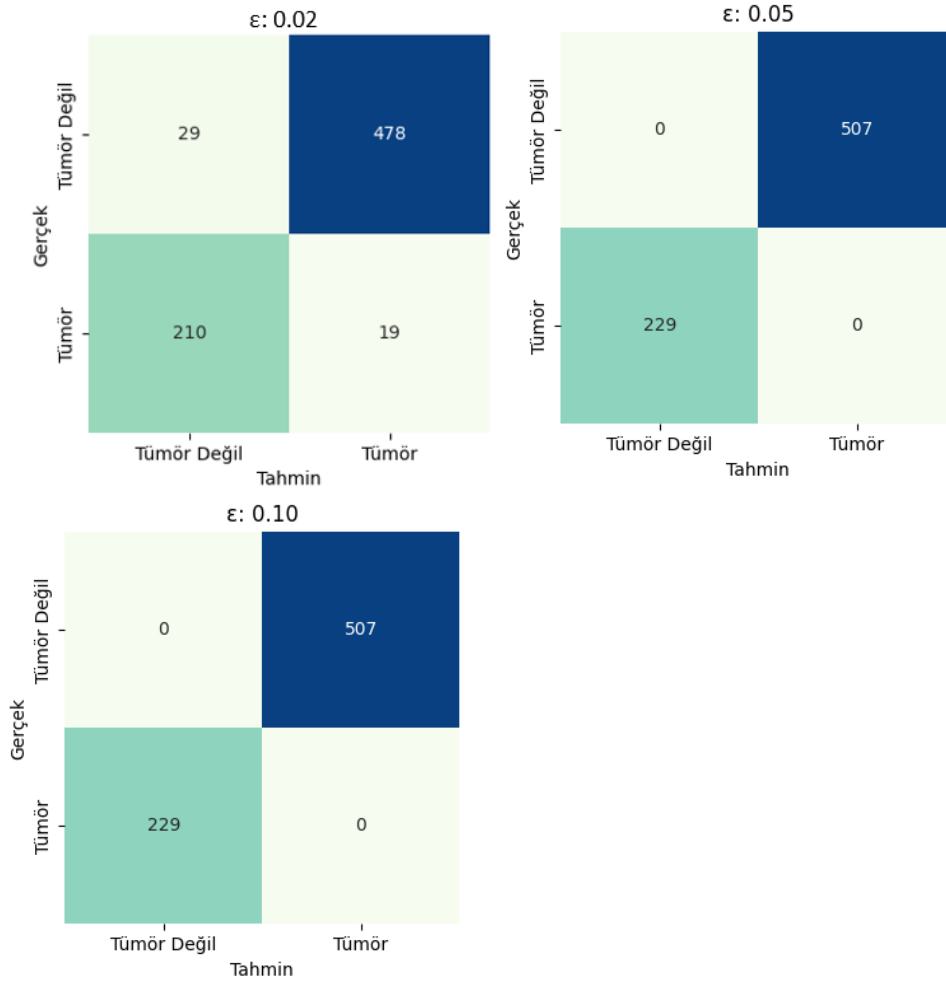


Şekil 6.188. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait ROC eğrileri

Şekil 6.188' e göre ResNet50V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %5'e kadar düşmektedir.

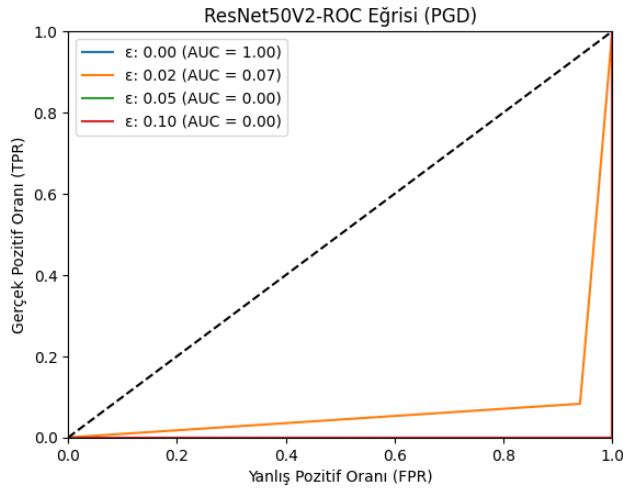
PGD saldırısı

Şekil 6.189' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.189. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait karmaşıklık matrisi

Şekil 6.189’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 210 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 478 örnek bulunmaktadır. Çizelge 6.6’ ya göre %6,52 doğruluk, %3,82 kesinlik, %8,3 geri çağırma, %5,23 F1 skoru ve %5,72 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.190’a pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



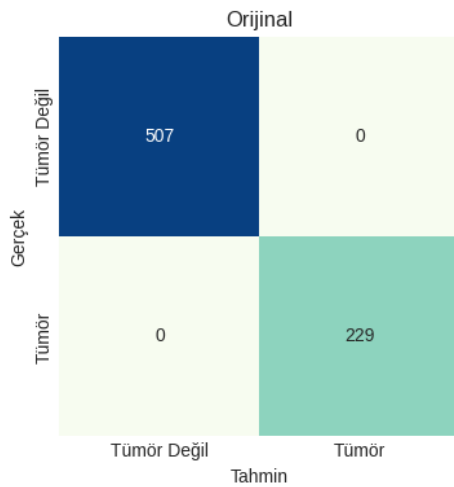
Şekil 6.190. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet50V2 modeline ait ROC eğrileri

Şekil 6.190' a göre ResNet50V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.19. ResNet101V2

Temiz

ResNet101V2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.191'de görülmektedir.

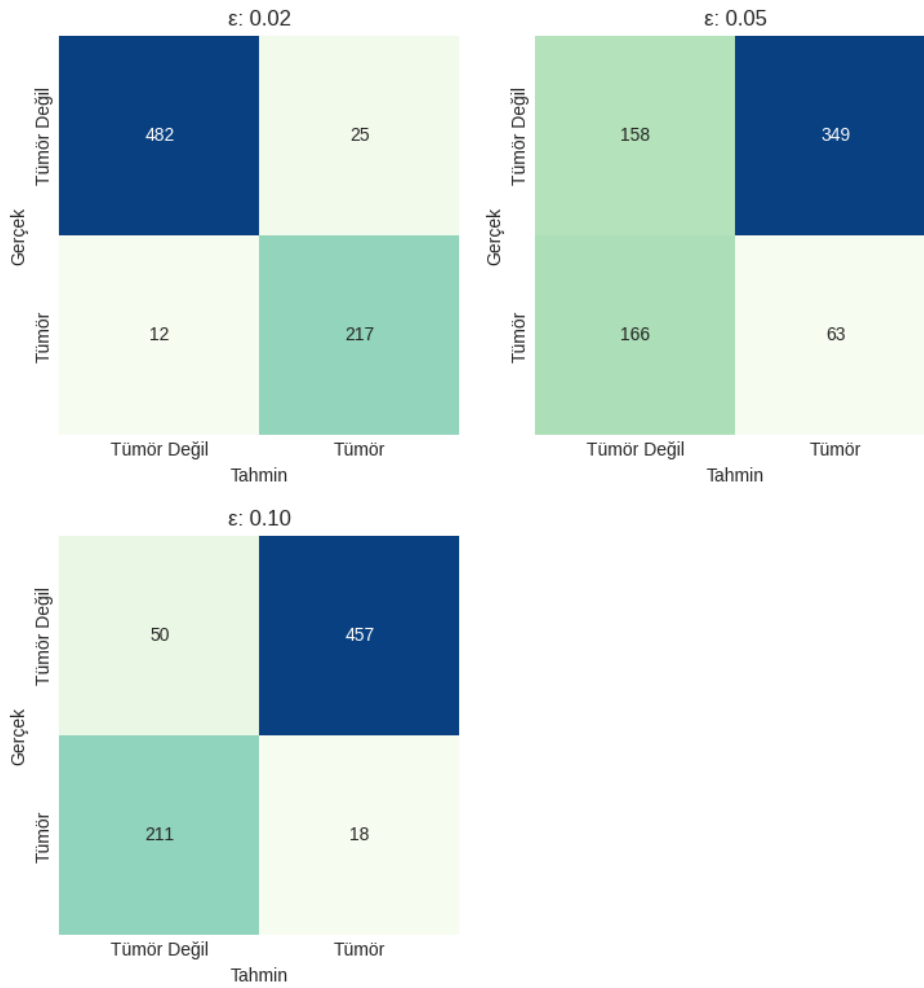


Şekil 6.191. Temiz Böbrek BT görüntüleri üzerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.191’ de görüldüğü gibi temiz test setinin ResNet101V2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

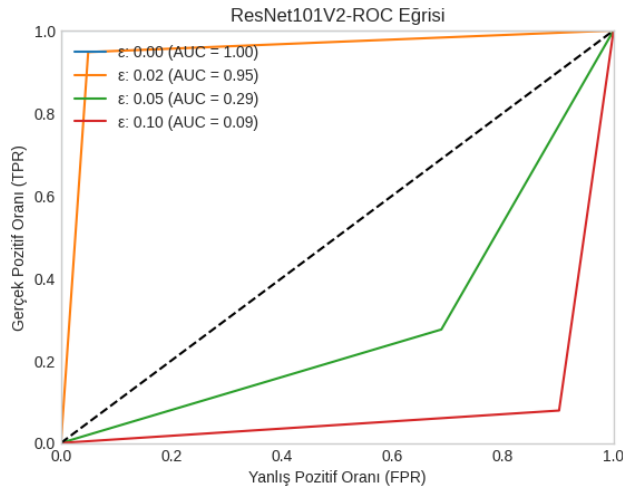
Şekil 6.192’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.192. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.192’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 12 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 25 örnek bulunmaktadır. Çizelge 6.5’ e göre %94,97 doğruluk, %89,67 kesinlik, %94,76 geri

çağırma, %92,14 F1 skoru ve %95,07 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 166 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 349 örnek bulunmaktadır. Çizelge 6.5' e göre %30,03 doğruluk, %15,29 kesinlik, %27,51 geri çağırma, %19,66 F1 skoru ve %31,16 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 211 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 457 örnek bulunmaktadır. Çizelge 6.5' e göre %9,24 doğruluk, %3,79 kesinlik, %7,86 geri çağırma, %5,11 F1 skoru ve %9,86 özgüllük elde etmektedir. Şekil 6.193'te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

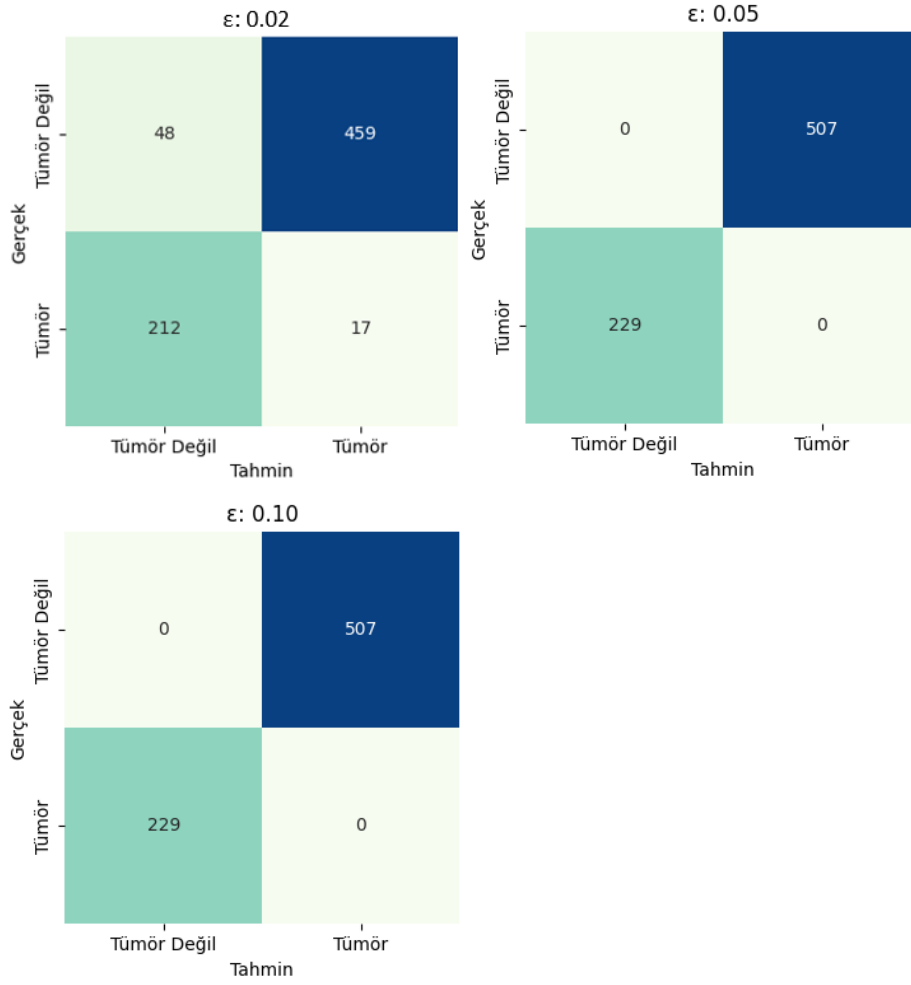


Şekil 6.193. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait ROC eğrileri

Şekil 6.193' e göre ResNet101V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %9'a kadar düşmektedir.

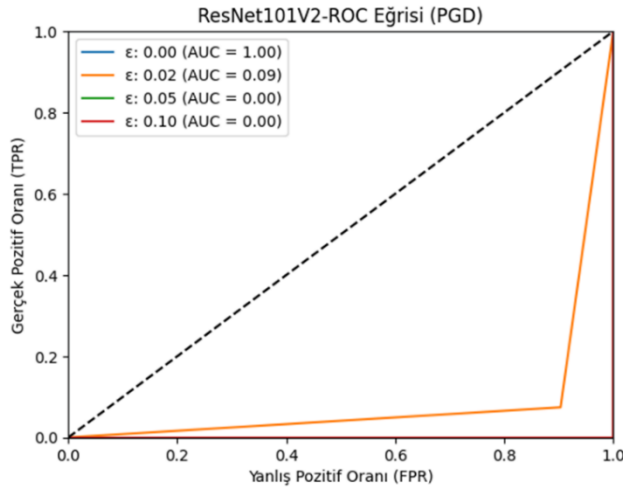
PGD saldırısı

Şekil 6.194' te farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.194. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait karmaşıklık matrisi

Şekil 6.194’ te görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 212 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 459örnek bulunmaktadır. Çizelge 6.6’ ya göre %8,83 doğruluk, %3,57 kesinlik, %7,42 geri çağırma, %4,82 F1 skoru ve %9,47 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 veya daha fazla olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.195’te pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



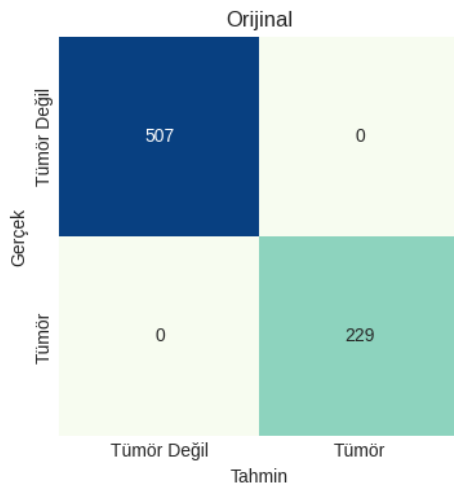
Şekil 6.195. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet101V2 modeline ait ROC eğrileri

Şekil 6.195' e göre ResNet101V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

6.2.20. ResNet152V2

Temiz

ResNet152V2 modeli ile saldırıya uğramamış orijinal Böbrek BT test setinin sınıflandırılmasına ait karmaşıklık matrisi Şekil 6.196'da görülmektedir.

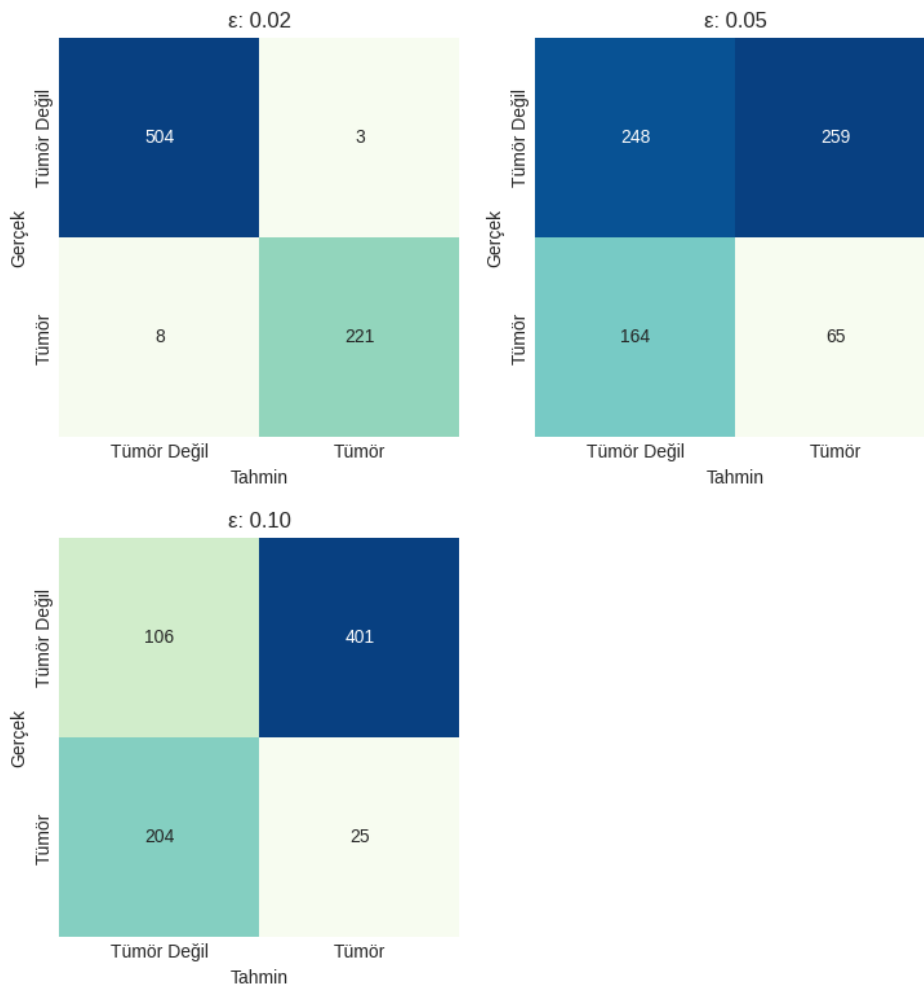


Şekil 6.196. Temiz Böbrek BT görüntüleri üzerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.196’ da görüldüğü gibi temiz test setinin ResNet152V2 modeli ile analizinde tüm görüntüler doğru sınıflandırılmıştır. Çizelge 6.4’ e göre doğruluk, kesinlik, geri çağırma, F1 skoru ve özgüllük değerleri %100’dür.

FGSM saldırısı

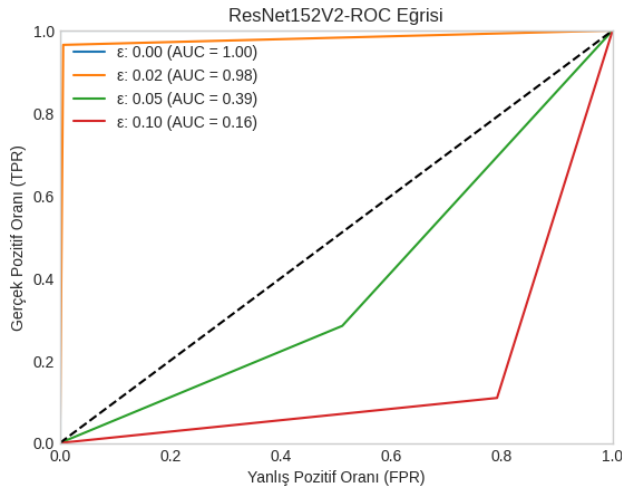
Şekil 6.197’ de farklı pertürbasyon değerlerinde FGSM saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.197. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.197’ de görüldüğü gibi 0,02 pertürbasyon düzeyinde FGSM saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 8 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 3 örnek bulunmaktadır. Çizelge 6.5’ e göre %98,51 doğruluk, %98,66 kesinlik, %96,51 geri

çağırma, %97,57 F1 skoru ve %99,41 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 164 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 259 örnek bulunmaktadır. Çizelge 6.5' e göre %42,53 doğruluk, %20,06 kesinlik, %28,38 geri çağırma, %23,51 F1 skoru ve %48,92 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 204 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 401 örnek bulunmaktadır. Çizelge 6.5' e göre %17,8 doğruluk, %5,87 kesinlik, %10,92 geri çağırma, %7,63 F1 skoru ve %20,91 özgüllük elde etmektedir. Şekil 6.198'de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.

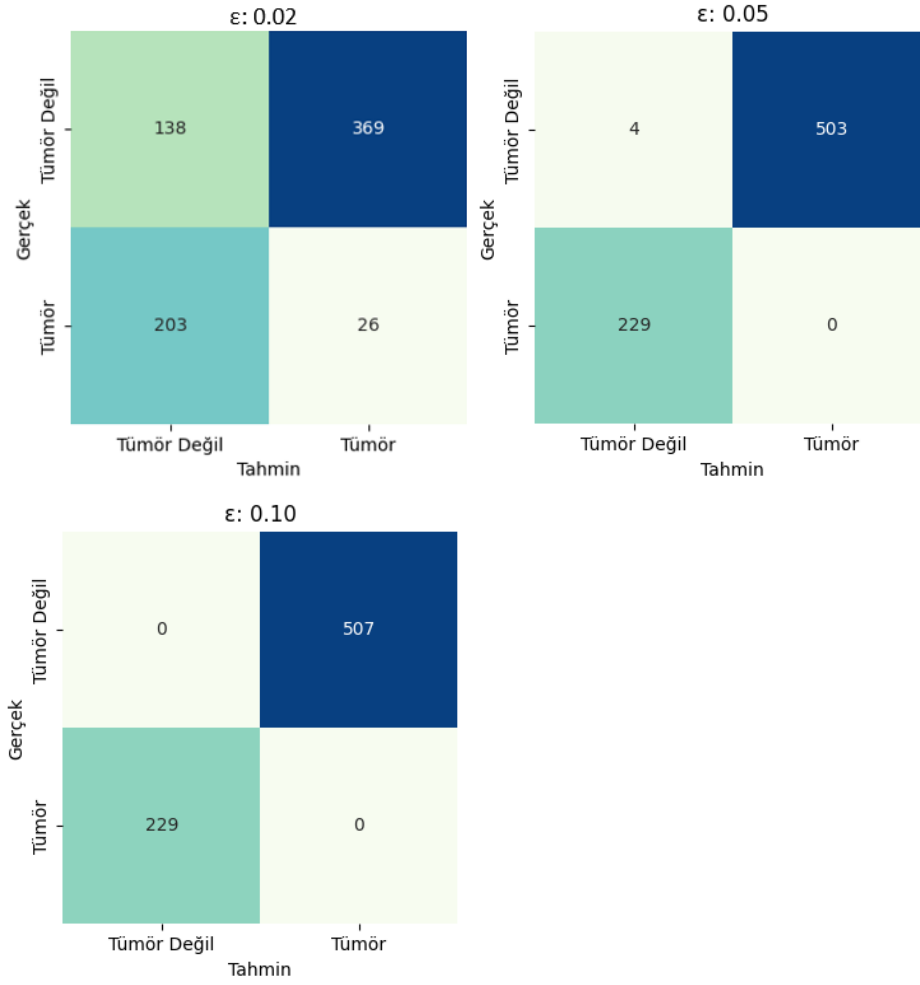


Şekil 6.198. FGSM saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait ROC eğrileri

Şekil 6.198' e göre ResNet152V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. FGSM saldırısındaki pertürbasyon artışı AUC değerinde kademeli azalmaya neden olmaktadır. AUC değeri %100' den %16'ya kadar düşmektedir.

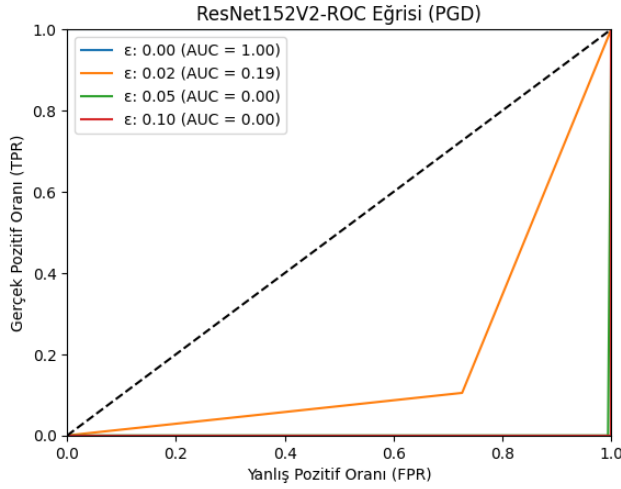
PGD saldırısı

Şekil 6.199' da farklı pertürbasyon değerlerinde PGD saldırısına uğrayan test setlerinin karmaşıklık matrisi görülmektedir.



Şekil 6.199. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait karmaşıklık matrisi

Şekil 6.199’ da görüldüğü gibi 0,02 pertürbasyon düzeyinde PGD saldırı sonucunda ortaya çıkan test setinin analizinde tümör vakalarının normal olarak sınıflandırıldığı (HN) 203 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 369 örnek bulunmaktadır. Çizelge 6.6’ ya göre %22,28 doğruluk, %6,58 kesinlik, %11,35 geri çağırma, %8,33 F1 skoru ve %27,22 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,05 olduğunda tümör vakalarının normal olarak sınıflandırıldığı (HN) 229 örnek bulunmaktadır. Normal BT görüntülerinin tümör olarak sınıflandırıldığı (HP) 503 örnek bulunmaktadır. Çizelge 6.6’ ya göre %0,54 doğruluk ve %0,79 özgüllük elde etmektedir. Pertürbasyon düzeyi 0,10 olduğunda tüm örnekler yanlış sınıflandırılmaktadır. Şekil 6.200’de pertürbasyon değerlerine bağlı ROC eğrileri yer almaktadır.



Şekil 6.200. PGD saldırısına uğrayan Böbrek BT görüntülerinde ResNet152V2 modeline ait ROC eğrileri

Şekil 6.200' e göre ResNet152V2 modeli ile temiz görüntülerin sınıflandırılma kabiliyeti %100 AUC değerine sahiptir. PGD saldırısındaki pertürbasyon artışı AUC değerinde azalmaya neden olmaktadır. AUC değeri %100' den %0'a kadar düşmektedir.

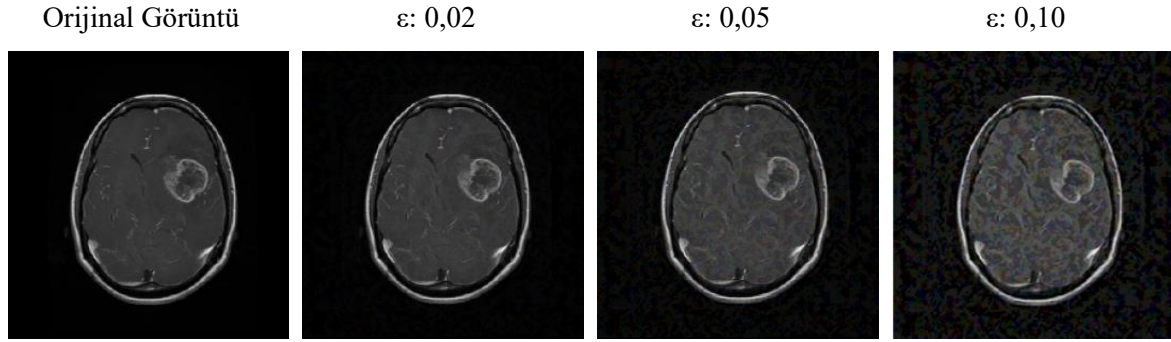
6.3. Görüntülerin Kalite ve Benzerlik Değerlendirmesi

Bu bölümde Beyin Tümör MR ve Böbrek BT görüntü modalitelerinin FGSM ve PGD saldırıları karşısında bozulma düzeyleri SSIM, MS-SSIM, MSE, PSNR, RMSE, UQI ve VIFP metrikleri kullanılarak değerlendirilmektedir. Kalite ve benzerlik metriklerinden SSIM, MS-SSIM, UQI ve VIFP için değerler 1' e yaklaştıkça bozulmanın daha az olduğu anlamı çıkarılmaktadır. Yani orijinal görüntüye benzerlik artmaktadır. MSE ve RMSE metriklerine göre değerlerin 0' a yaklaşması bozulmanın daha az olduğu anlamına gelmektedir. PSNR metriğine göre ise dinamik bir aralığı logaritmik desibel ölçeği cinsinden ifade etmektedir. Dolayısıyla sinyal-gürültü oranını belirten değer büyüdükçe görüntü kalitesinin ve orijinal görüntüye benzerliğin arttığı düşünülebilir.

6.3.1. Beyin Tümör MR görüntülerinin saldırılara karşı kalite ve benzerliklerinin değerlendirilmesi

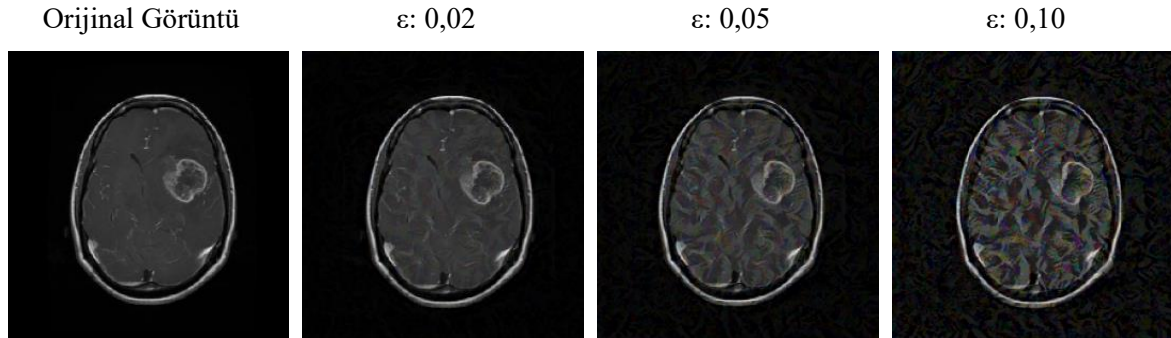
Beyin Tümör MR test seti üzerine uygulanan çekışmeli saldırılarda pertürbasyon artışının örneklerin yanlış sınıflandırılması oranını artırdığı görülmektedir. Örnek görüntüler üzerindeki pertürbasyona bağlı değişimler detaylı incelendiğinde, pertürbasyon artışının görüntülerin kalite ve orijinal görüntüye benzerliklerini düşürdüğü ortaya çıkmaktadır.

Şekil 6.201’ de bir Beyin Tümör MR görüntü örneğinin farklı pertürbasyon değerine sahip FGSM saldırıları karşısındaki değişimi görülmektedir. Pertürbasyon değeri büyüdükçe görüntüdeki bozulma daha belirgin hale gelmektedir.



Şekil 6.201. Orijinal Beyin Tümör MR görüntü örneği ve farklı pertürbasyona sahip FGSM saldırısına bağı olarak üretilen çekışmeli örnekler

Şekil 6.202’ de ise aynı Beyin Tümör MR görüntü örneğinin farklı pertürbasyon değerine sahip PGD saldırıları karşısındaki değişimi görülmektedir. Pertürbasyon değeri büyüdükçe görüntüdeki bozulma daha belirgin hale gelmektedir. Ancak FGSM saldırısına kıyasla bozulmanın daha büyük olduğı dikkat çekmektedir.



Şekil 6.202. Orijinal Beyin Tümör MR görüntü örneği ve farklı pertürbasyona sahip PGD saldırısına bağı olarak üretilen çekışmeli örnekler

FGSM ve PGD saldırılarına bağı olarak Beyin Tümör MR test görüntüleri üzerinde gerçekleşen bozulmalar nitel olarak değerdendirilebilmektedir. Beyin Tümör MR test görüntüleri ile üretilen çekışmeli örneklerin ortalama kalite ve benzerlik değeri Çizelge 6.7’ de görülmektedir. Her iki saldırı için de elde edilen ortalama değeri metriklerin doğası dikkate alınarak incelendiğinde, görüntülerde pertürbasyon derecesinin artışına bağı olarak kademeli bir bozulma meydana geldiğı nitel olarak doğrudanmaktadır.

Bozulma ile ifade edilmek istenen, üretilen görüntülerin orijinal görüntüye benzerliğinin azalması ve kalitesindeki düşüştür.

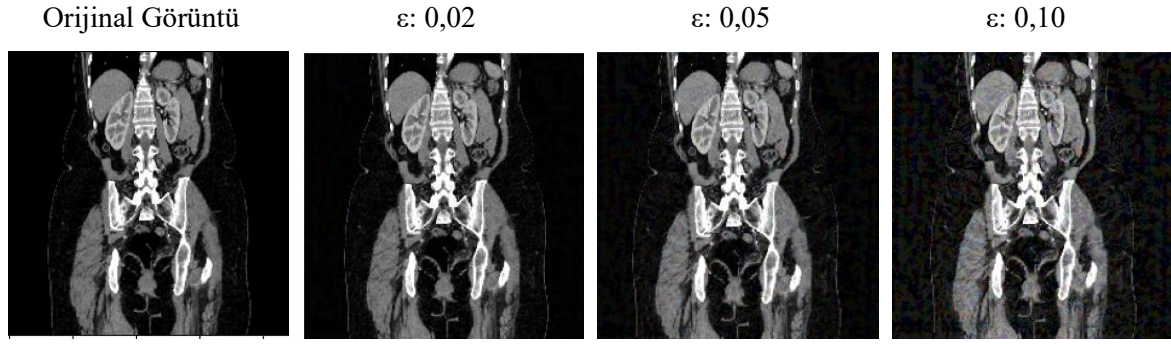
Çizelge 6.7.Beyin Tümör MR test görüntüleri üzerine uygulanan FGSM ve PGD saldırıları sonucunda elde edilen görüntülerin ortalama kalite ve benzerlik skorları

Metrikler	Orijinal	FGSM Saldırısı			PGD Saldırısı		
		0,02	0,05	0,10	0,02	0,05	0,10
SSIM	1,0	0,913	0,733	0,513	0,897	0,699	0,446
MSE	0,0	12,667	73,019	316,819	17,318	83,992	356,588
MS-SSIM	1,0	0,989	0,947	0,842	0,978	0,918	0,783
PSNR	inf	37,101	29,496	23,123	35,772	28,908	22,620
RMSE	0,0	3,561	8,545	17,799	4,155	9,155	18,872
UQI	1,0	0,770	0,627	0,528	0,774	0,646	0,527
VIFP	1,0	0,789	0,560	0,362	0,630	0,417	0,242

Çizelge 6.7' ye göre çok adımlı PGD saldırısı her bir pertürbasyon değeri için tek adımlı FGSM saldırısına göre daha fazla bozulmaya neden olduğu çeşitli metriklerce ortaya konulmaktadır. Örneğin, SSIM metriğine göre 0,02 pertürbasyonlu FGSM saldırısında 0,913 benzerlik değeri elde edilirken aynı pertürbasyona sahip PGD saldırısında 0,897 benzerlik değeri elde edilmektedir. SSIM metriğinde 1'e yaklaştıkça orijinal görüntüye benzerliğin arttığı kabul edilmektedir. Yalnızca UQI metriğinde FGSM ve PGD saldırılarının kalite ve benzerliğe etkisi dikkate alınmayacak derecede küçüktür.

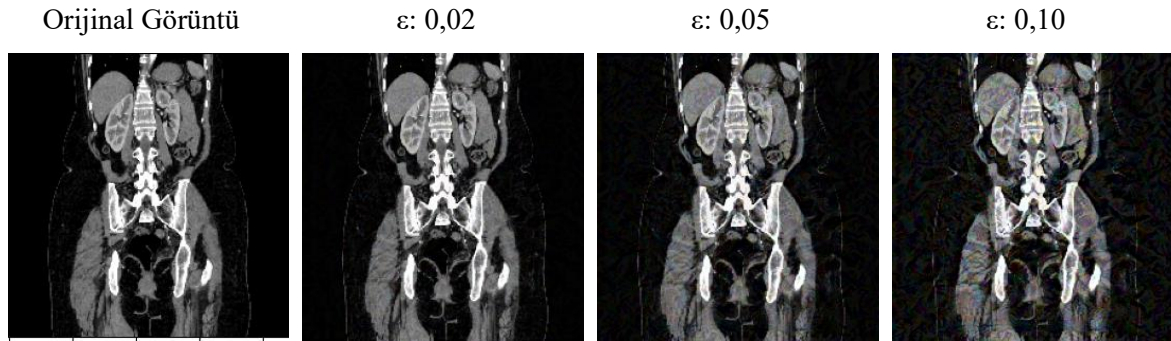
6.3.2. Böbrek BT görüntülerinin saldırılara karşı kalite ve benzerliklerinin değerlendirilmesi

Böbrek BT test seti üzerine uygulanan çekişmeli saldırılarda pertürbasyon artışının örneklerin yanlış sınıflandırılması oranını artırdığı görülmektedir. Örnek görüntüler üzerindeki pertürbasyona bağlı değişimler detaylı incelendiğinde, pertürbasyon artışının görüntülerin kalite ve orijinal görüntüye benzerliklerini düşürdüğü ortaya çıkmaktadır. Şekil 6.203' te bir Böbrek BT görüntü örneğinin farklı pertürbasyon değerine sahip FGSM saldırıları karşısındaki değişimi görülmektedir. Pertürbasyon değeri büyüdükçe görüntüdeki bozulma daha belirgin hale gelmektedir.



Şekil 6.203. Orijinal Böbrek BT görüntü örneği ve farklı pertürbasyona sahip FGSM saldırısına bağlı olarak üretilen çekişmeli örnekler

Şekil 6.204’ te ise aynı Böbrek BT görüntü örneğinin farklı pertürbasyon değerine sahip PGD saldırıları karşısındaki değişimi görülmektedir. Pertürbasyon değeri büyüdükçe görüntüdeki bozulma daha belirgin hale gelmektedir. Ancak FGSM saldırısına kıyasla bozulmanın daha büyük olduğu dikkat çekmektedir.



Şekil 6.204. Orijinal Böbrek BT görüntü örneği ve farklı pertürbasyona sahip PGD saldırısına bağlı olarak üretilen çekişmeli örnekler

FGSM ve PGD saldırılarına bağlı olarak Böbrek BT test görüntüleri üzerinde gerçekleşen bozulmalar nitel olarak değerlendirilebilmektedir. Böbrek BT test görüntüleri ile üretilen çekişmeli örneklerin ortalama kalite ve benzerlik değerleri Çizelge 6.8’ de görülmektedir. Her iki saldırı için de elde edilen ortalama değerler metriklerin doğası dikkate alınarak incelendiğinde, görüntülerde pertürbasyon derecesinin artışına bağlı olarak kademeli bir bozulma meydana geldiği nitel olarak doğrulanmaktadır. Bozulma ile ifade edilmek istenen, üretilen görüntülerin orijinal görüntüye benzerliğinin azalması ve kalitesindeki düşüştür.

Çizelge 6.8. Böbrek BT test görüntüleri üzerine uygulanan FGSM ve PGD saldırıları sonucunda elde edilen görüntülerin ortalama kalite ve benzerlik skorları

Metrikler	Orijinal	FGSM Saldırısı			PGD Saldırısı		
		0,02	0,05	0,10	0,02	0,05	0,10
SSIM	1,0	0,884	0,752	0,624	0,893	0,750	0,595
MSE	0,0	11,857	68,235	295,690	16,621	83,171	335,240
MS-SSIM	1,0	0,993	0,963	0,892	0,992	0,961	0,875
PSNR	inf	37,391	29,791	23,423	35,925	28,932	22,882
RMSE	0,0	3,444	8,260	17,196	4,077	9,119	18,305
UQI	1,0	0,597	0,579	0,542	0,597	0,578	0,540
VIFP	1,0	0,833	0,630	0,433	0,709	0,500	0,315

Çizelge 6.8' e göre çok adımlı PGD saldırısı her bir pertürbasyon değeri için tek adımlı FGSM saldırısına göre daha fazla bozulmaya neden olduğu çeşitli metriklerce ortaya konulmaktadır. Örneğin, SSIM metriğine göre 0,10 pertürbasyonlu FGSM saldırısında 0,624 benzerlik değeri elde edilirken aynı pertürbasyona sahip PGD saldırısında 0,595 benzerlik değeri elde edilmektedir. SSIM metriğinde 1'e yaklaştıkça orijinal görüntüye benzerliğin arttığı kabul edilmektedir. Yalnızca UQI ve MS-SSIM metriklerinde FGSM ve PGD saldırılarının kalite ve benzerliğe etkisi dikkate alınmayacak derecede küçüktür.

6.4. Veri Setlerinin Performans ve Benzerlik Karşılaştırması

Önceki bölümlerde Beyin Tümör MR ve Böbrek BT görüntü modalitelerinin FGSM ve PGD saldırılarına karşı verdiği reaksiyonlar ele alınmıştır. Bu bölümde, ön eğitilmiş modellerin saldırılar karşısında sınıflandırma performansları hakkında genel bir çıkarım yapılmaktadır. Ayrıca, her görüntü modalitesinin birtakım karakteristik özellik barındırdığı görüşüne dayanarak, saldırıların MR ve BT modaliteleri üzerindeki etkinliği kıyaslanmaktadır. Modellerin temiz test setlerini ve saldırıya uğramış test setleri üzerindeki sınıflandırma performansları önceki bölümlerde verilen ölçütler ve bu bölümde yer alan ek ölçütler kullanılarak değerlendirilmektedir. Veriler ışığında ön eğitilmiş modellerin senaryoya bağlı olarak farklı sonuçlar elde ettiği görülmektedir.

Temiz MR ve BT görüntülerinin sınıflandırma başarıları saldırı başarısını değerlendirmek için referans noktasını oluşturmaktadır. Çizelge 6.1' de elde edilen verilere göre Beyin Tümör MR temiz test seti üzerinde yapılan sınıflandırmada MobileNetV3Small %97,29 doğruluk oranı ile en düşük başarı elde eden modeldir. ResNet50V2 ise %99,72 ile en

yüksek performans gösteren modeldir. Çizelge 6.4' te yer alan Böbrek BT temiz test seti üzerinde yapılan sınıflandırmada ise 17 model %100 doğruluk oranı ile tüm örnekleri doğru sınıflandırmaktadır. EfficientNetV2B0, EfficientNetV2B1 ve InceptionResNetV2 modelleri ise %99,86 doğruluk oranı sağlamaktadır. Buna göre temiz BT görüntülerinin sınıflandırma başarısı temiz MR görüntülerinin sınıflandırma başarısından daha yüksektir.

Çizelge 6.2' de MR görüntülerinin sınıflandırılmasında FGSM saldırılarının etkinliği yer almaktadır. Elde edilen verilere göre 0,02 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, DenseNet169, DenseNet201 ve InceptionResNetV2 modelleri %94,02 doğruluk oranı ile en başarılı modellerdir. MobileNetV3Small modeli %51,71 doğruluk ile saldırıdan en çok etkilenen modeldir. 0,05 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, InceptionV3 modeli %71,94 doğruluk ile en başarılı modeldir. MobileNetV3Small %13,11 doğruluk oranı ile en düşük başarı elde eden modeldir. 0,10 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, ResNet152V2 modeli %49,15 doğruluk ile en başarılı modeldir. EfficientNetV2B0 %1,57 doğruluk oranı ile en düşük başarı elde eden modeldir.

Çizelge 6.3' te MR görüntülerinin sınıflandırılmasında PGD saldırılarının etkinliği görülmektedir. 0,02 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, EfficientNetV2B3 modeli %61,97 doğruluk oranı ile en başarılı modeldir. EfficientNetV2S modeli %7,85 doğruluk ile saldırıdan en çok etkilenen modeldir. 0,05 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, InceptionV3 ve InceptionResNetV2 modelleri %7,41 doğruluk ile en başarılı modellerdir. EfficientNetV2S ve MobileNetV1 modelleri %0,14 doğruluk oranı ile en düşük başarı elde eden modellerdir. 0,10 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, InceptionV3 modeli %1,71 doğruluk ile en başarılı modeldir. Bu senaryoda modellerin çoğunluğu tüm çekişmeli örnekleri yanlış sınıflandırmaktadır.

Çizelge 6.5' te BT görüntülerinin sınıflandırılmasında FGSM saldırılarının etkinliği yer almaktadır. 0,02 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, ResNet152V2 modeli %98,51 doğruluk oranı ile en başarılı modeldir. MobileNetV3Small modeli %16,03 doğruluk ile saldırıdan en çok

etkilenen modeldir. 0,05 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, DenseNet201 modeli %56,93 doğruluk ile en başarılı modeldir. MobileNetV3Small %0,14 doğruluk oranı ile en düşük başarı elde eden modeldir. 0,10 pertürbasyon değerine sahip FGSM saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, ResNet152V2 modeli %17,80 doğruluk ile en başarılı modeldir. MobileNetV3Small %0,14 doğruluk oranı ile en düşük başarı elde eden modeldir.

Çizelge 6.6’ da BT görüntülerinin sınıflandırılmasında PGD saldırılarının etkinliği görülmektedir. 0,02 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, EfficientNetV2B2 modeli %44,29 doğruluk oranı ile en başarılı modeldir. MobileNetV3Small modeli %0,68 doğruluk ile saldırıdan en çok etkilenen modeldir. 0,05 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, EfficientNetV2B3 modeli %2,72 doğruluk ile en başarılı modeldir. Bu senaryoda modellerin çoğunluğu tüm çekişmeli örnekleri yanlış sınıflandırmaktadır. 0,10 pertürbasyon değerine sahip PGD saldırısı ile üretilen çekişmeli örneklerin sınıflandırılmasında, tüm modeller çekişmeli örnekleri yanlış sınıflandırmaktadır.

Önceki bölümlerde verilen çizelgelere ve performans metriklerinin incelenmesine ek olarak orijinal test görüntülerinin doğru sınıflandırma oranı ile FGSM ve PGD saldırıları senaryoları sonucunda elde edilen doğru sınıflandırma oranları arasındaki fark ortaya konulmaktadır. Aşağıda yer alan Çizelge 6.9’ da Beyin Tümör MR görüntülerinin sınıflandırılmasında elde edilen başarı oranından saldırı senaryosunda elde edilen sınıflandırma başarısının çıkarılmasıyla elde edilen performans düşüşü gösterilmektedir. Benzer şekilde, Çizelge 6.10’ da Böbrek BT görüntülerinin sınıflandırılmasında elde edilen başarı oranından saldırı senaryosunda elde edilen sınıflandırma başarısının çıkarılmasıyla elde edilen performans düşüşü verilmektedir.

Çizelge 6.9. Beyin Tümör MR görüntüleri üzerinde orijinal görüntülerin sınıflandırılma başarısı ile FGSM ve PGD saldırı senaryolarının sınıflandırma başarıları arasındaki doğruluk oranı farkları

Modeller	FGSM			PGD		
	0,02	0,05	0,1	0,02	0,05	0,1
Xception	5,55	36,04	72,94	39,17	94,58	97,72
VGG16	7,97	55,69	91,02	58,40	97,15	98,29
VGG19	16,10	69,94	93,16	75,92	97,58	98,72
DenseNet121	9,68	50,57	89,74	68,37	97,15	99,0
DenseNet169	4,70	35,75	74,08	52,99	95,73	98,72
DenseNet201	4,84	40,31	83,33	52,99	95,58	98,72
EfficientNetV2B0	22,22	78,92	97,29	51,57	96,01	98,86
EfficientNetV2B1	16,09	75,50	94,44	50,28	95,15	99,0
EfficientNetV2B2	14,67	73,79	96,86	42,59	94,30	99,0
EfficientNetV2B3	12,25	63,11	96,16	37,18	95,16	99,15
EfficientNetV2S	25,36	74,79	85,61	91,44	99,15	99,29
InceptionResNetV2	4,98	30,77	52,28	39,31	91,59	98,14
InceptionV3	5,70	26,92	50,28	43,59	91,45	97,15
MobileNet	17,67	73,08	93,31	78,49	98,01	98,15
MobileNetV2	16,53	68,23	91,74	66,81	97,15	98,15
MobileNetV3Small	45,58	84,18	94,16	85,47	96,58	97,29
NasNetMobile	14,53	55,41	90,74	55,98	97,72	98,86
ResNet50V2	8,41	34,62	65,25	75,50	98,72	99,72
ResNet101V2	9,12	32,91	55,27	69,52	95,73	98,44
ResNet152V2	9,55	32,06	50,0	64,39	95,73	99,15

Çizelge 6.9'a göre temiz Beyin Tümör MR test setinin sınıflandırma başarısı ile 0,02 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarısı arasındaki farklar incelendiğinde DenseNet169 %4,70 ile en az başarı düşüşü yaşayan modeldir. MobileNetV3Small ise %45,58 ile en fazla düşüş yaşayan modeldir. 0,05 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarı düşüşleri incelendiğinde InceptionV3 %26,92 ile en az başarı düşüşü yaşayan modeldir. MobileNetV3Small ise %84,18 ile en fazla düşüş yaşayan modeldir. 0,10 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarı düşüşleri incelendiğinde ResNet152V2 %50 ile en az başarı düşüşü yaşayan modeldir. EfficientNetV2B0 ise %97,29 ile en fazla düşüş yaşayan modeldir. Temiz Beyin Tümör MR test setinin sınıflandırma başarısı ile 0,02 pertürbasyon ile üretilen PGD örneklerinin sınıflandırma başarısı arasındaki farklar incelendiğinde EfficientNetV2B3 %37,18 ile en az başarı düşüşü yaşayan modeldir. EfficientNetV2S ise %91,44 ile en fazla düşüş yaşayan modeldir. 0,05 pertürbasyon ile üretilen PGD

örneklerinin sınıflandırma başarı düşüşleri incelendiğinde InceptionV3 %91,45 ile en az başarı düşüşü yaşayan modeldir. EfficientNetV2S ise %99,15 ile en fazla düşüş yaşayan modeldir. 0,10 pertürbasyon ile üretilen PGD örneklerinin sınıflandırma başarı düşüşleri incelendiğinde ise InceptionV3 %97,15 ile en az başarı düşüşü yaşayan modeldir. ResNet50V2 ise %99,72 ile en fazla düşüş yaşayan modeldir. Elde edilen veriler önceki bölümlerde verilen çizelgeleri destekler niteliktedir.

Beyin Tümör MR seti üzerinde yapılan değerlendirme sonucunda, FGSM ve PGD saldırıları senaryolarına karşı modellerin dirençlerinin değiştiği görülmektedir. Genel anlamda, InceptionV3 modelinin saldırılara karşı daha istikrarlı bir direnç gösterdiği söylenebilir. Ancak, MobileNetV3Small ve EfficientNetV2S modelleri FGSM ve PGD saldırı senaryolarından en hızlı etkilenen modeller olmaktadır.

Çizelge 6.10. Böbrek BT görüntüleri üzerinde orijinal görüntülerin sınıflandırılma başarısı ile FGSM ve PGD saldırı senaryolarının sınıflandırma başarıları arasındaki doğruluk oranı farkları

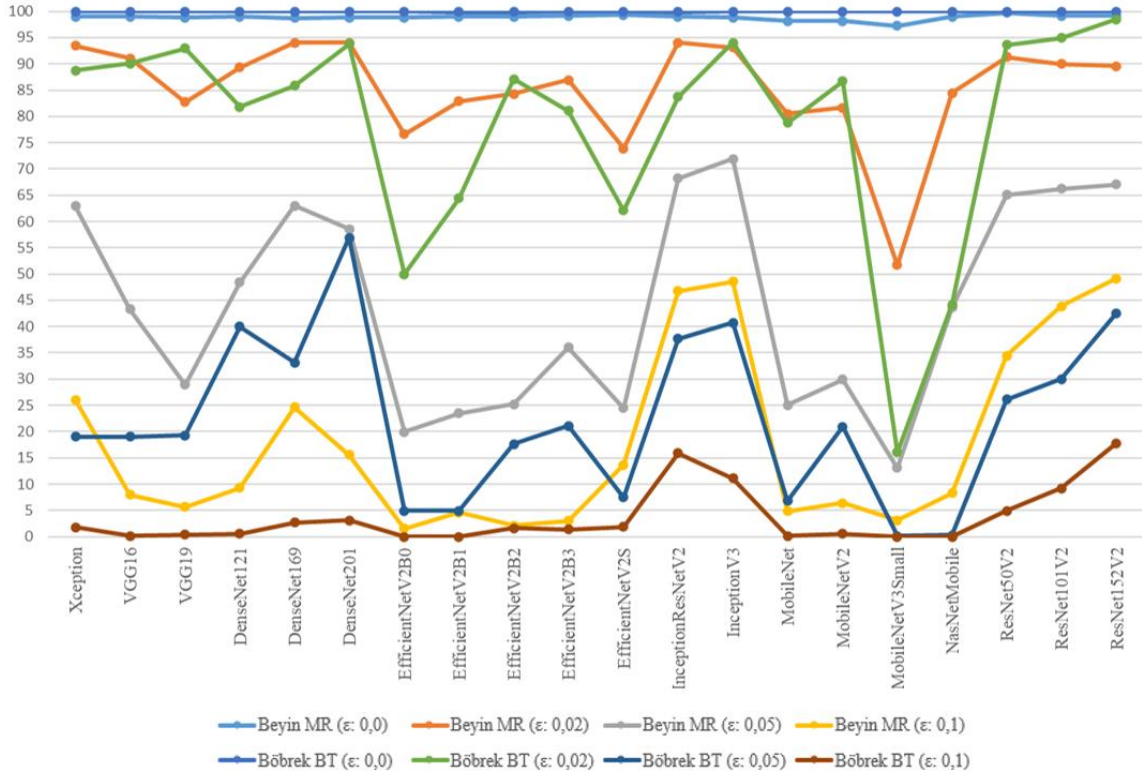
Modeller	FGSM			PGD		
	0,02	0,05	0,1	0,02	0,05	0,1
Xception	11,28	80,98	98,23	77,45	100,0	100,0
VGG16	9,92	80,98	99,86	62,36	100,0	100,0
VGG19	7,06	80,71	99,59	60,46	100,0	100,0
DenseNet121	18,21	60,05	99,46	84,37	100,0	100,0
DenseNet169	14,13	66,85	97,28	88,59	100,0	100,0
DenseNet201	6,25	43,07	96,87	70,38	100,0	100,0
EfficientNetV2B0	50,0	95,11	100,0	80,57	100,0	100,0
EfficientNetV2B1	35,46	94,97	99,86	77,85	99,86	99,86
EfficientNetV2B2	12,77	82,20	98,23	55,57	99,04	99,86
EfficientNetV2B3	18,89	78,94	98,64	58,29	97,28	100,0
EfficientNetV2S	37,91	92,53	98,10	92,25	100,0	100,0
InceptionResNetV2	16,16	62,22	83,96	64,94	98,64	99,86
InceptionV3	5,98	59,24	88,86	74,32	98,50	100,0
MobileNet	21,20	93,21	99,86	91,58	100,0	100,0
MobileNetV2	13,31	79,08	99,46	69,56	99,73	100,0
MobileNetV3Small	83,97	99,86	100,0	99,32	100,0	100,0
NasNetMobile	55,84	99,73	100,0	97,55	100,0	100,0
ResNet50V2	6,39	73,91	95,11	93,48	100,0	100,0
ResNet101V2	5,03	69,97	90,76	91,17	100,0	100,0
ResNet152V2	1,49	57,47	82,20	77,72	99,46	100,0

Çizelge 6.10'a göre temiz Böbrek BT test setinin sınıflandırma başarısı ile 0,02 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarısı arasındaki farklar incelendiğinde ResNet152V2 %1,49 ile en az başarı düşüşü yaşayan modeldir. MobileNetV3Small ise %83,97 ile en fazla düşüş yaşayan modeldir. 0,05 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarı düşüşleri incelendiğinde DenseNet201 %43,07 ile en az başarı düşüşü yaşayan modeldir. MobileNetV3Small ise %99,86 ile en fazla düşüş yaşayan modeldir. 0,10 pertürbasyon ile üretilen FGSM örneklerinin sınıflandırma başarı düşüşleri incelendiğinde ResNet152V2 %82,20 ile en az başarı düşüşü yaşayan modeldir. EfficientNetV2B0, EfficientNetV2B1, MobileNetV3Small ve NasNetMobile modelleri %100 düşüş yaşadığı görülmektedir. Temiz Böbrek BT test setinin sınıflandırma başarısı ile 0,02 pertürbasyon ile üretilen PGD örneklerinin sınıflandırma başarısı arasındaki farklar incelendiğinde EfficientNetV2B2 %55,57 ile en az başarı düşüşü yaşayan modeldir. MobileNetV3Small ise %99,32 ile en fazla düşüş yaşayan modeldir. 0,05 pertürbasyon ile üretilen PGD örneklerinin sınıflandırma başarı düşüşleri incelendiğinde EfficientNetV2B3 %97,28 ile en az başarı düşüşü yaşayan modeldir. 0,10 pertürbasyon ile üretilen PGD örneklerinin sınıflandırma başarı düşüşleri incelendiğinde çoğu modelin %100 düşüş yaşadığı dikkat çekmektedir. Elde edilen veriler önceki bölümlerde verilen çizelgeleri destekler niteliktedir.

Böbrek BT seti üzerinde yapılan değerlendirme sonucunda, ResNet152V2 modelinin FGSM saldırısına karşı daha dirençli olduğu görülmektedir. EfficientNetV2B2 ve EfficientNetV2B3 modelleri ise PGD saldırısına karşı daha başarılı sonuçlar elde etmektedir. MobileNetV3Small modelinin FGSM ve PGD saldırıları ile gerçekleştirilen tüm saldırı senaryolarından en çok etkilenen model olduğu görülmektedir.

Şekil 6.205' te yer alan grafikte FGSM saldırısı karşısında MR ve BT görüntülerinin doğruluk performansları görülmektedir. Her iki modalite de temiz test görüntülerini sınıflandırmada yüksek başarı sağlamaktadır. FGSM ile üretilen çekişmeli örneklerin sınıflandırma başarısı her iki modalitede düşmektedir. Saldırıda pertürbasyon parametresinin artışına bağlı olarak doğruluk azalmaktadır. BT görüntülerinin pertürbasyon artışına karşı daha duyarlı olduğu görülmektedir. Pertürbasyon değeri 0,02 olduğu durumda kullanılan derin öğrenme modeline bağlı olarak modalitelerin başarı sırası değişmektedir. Örneğin, VGG19 ve ResNet mimarileri ile saldırıya uğramış BT görüntülerinin MR görüntülerine kıyasla daha yüksek sınıflandırma başarısı elde ettiği

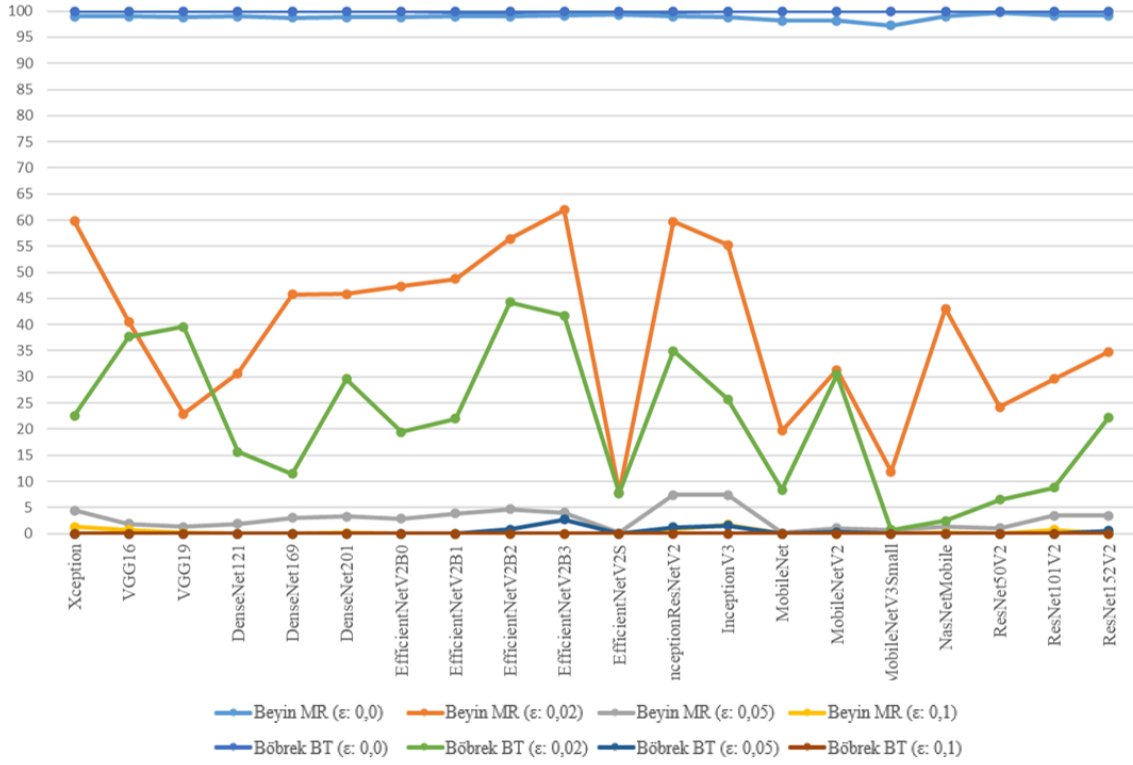
görülmektedir. Buna karşın EfficientNetV2B0 ve MobileNetV3Small modelleri tercih edildiğinde, saldırıya uğramış BT görüntülerinin tespiti MR görüntülerine kıyasla büyük bir farkla başarısız olmaktadır. Pertürbasyon değeri 0,05 ve 0,01 olduğu durumlarda ise kullanılan modelden bağımsız şekilde FGSM saldırısına uğramış MR görüntülerinin sınıflandırma başarısı daha yüksektir.



Şekil 6.205. FGSM saldırısı ile üretilen MR ve BT örneklerinin derin öğrenme modelleri bazında doğruluk performanslarının karşılaştırılması

Şekil 6.206' da yer alan grafikte PGD saldırısı karşısında MR ve BT görüntülerinin doğruluk performansları görülmektedir. Her iki modalite de temiz test görüntülerini sınıflandırmada yüksek başarı sağlamaktadır. PGD ile üretilen çekişmeli örneklerin sınıflandırma başarısı her iki modalitede hızlı bir düşüş yaşamaktadır. Saldırıda pertürbasyon parametresinin artışına bağlı olarak doğruluk azalmaktadır. Ancak, BT görüntülerinin PGD saldırısına karşı daha duyarlı olduğu görülmektedir. Pertürbasyon değeri 0,02 olduğunda, VGG19 modeli tercih edilmesi durumunda saldırıya uğramış BT görüntülerinin MR görüntülerine kıyasla daha yüksek sınıflandırma başarısı elde ettiği dikkat çekmektedir. Buna karşın, diğer derin öğrenme modelleri ile yapılan sınıflandırmada, saldırıya uğramış BT görüntülerinin tespiti MR görüntülerine kıyasla

başarısız olmaktadır. Pertürbasyon değeri 0,05 olduğunda saldırıya uğrayan birkaç MR görüntüsü doğru sınıflandırabiliyorken, BT görüntüleri için aynı durum söz konusu olmamaktadır. Pertürbasyon değeri 0,10 olduğu durumlarda ise kullanılan modelden bağımsız şekilde her iki modalitede PGD saldırısına karşı mağlup olmaktadır.

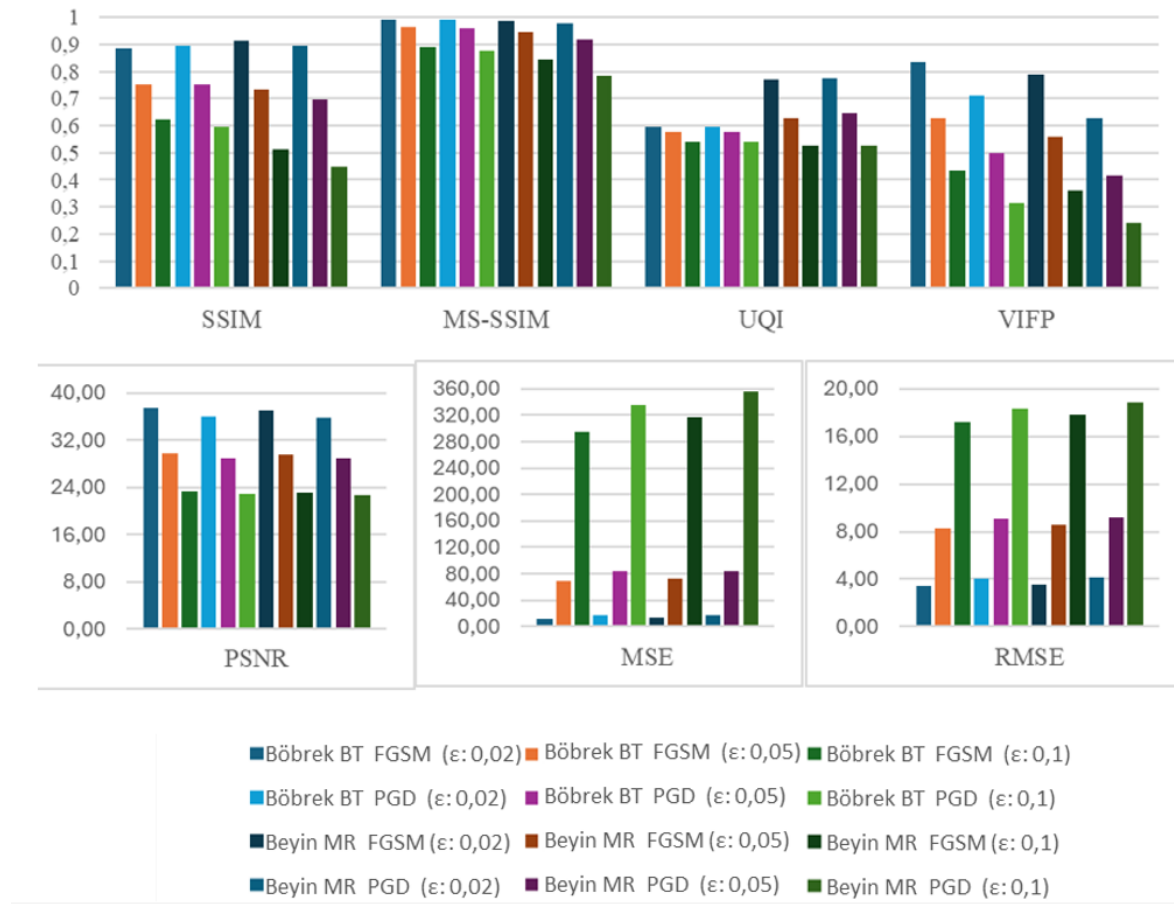


Şekil 6.206. PGD saldırısı ile üretilen MR ve BT örneklerinin derin öğrenme modelleri bazında doğruluk performanslarının karşılaştırılması

Bu çalışmada elde edilen sonuçlara göre MR görüntülerinin BT görüntülerine kıyasla FGSM ve PGD saldırılarına karşı daha dirençli olduğu ortaya konulmaktadır. Ancak bu çıkarımın sınırlı sayıda örnek içeren açık veri setleri kullanılarak yapıldığı göz önünde bulundurulmalıdır.

Şekil 6.207' de yer alan grafikler FGSM ve PGD saldırıları ile üretilen çekişmeli MR ve BT görüntülerinin orijinal görüntüye benzerlik değerlerini göstermektedir. SSIM, MS-SSIM ve VIFP metriklerine bakıldığında her iki modalitede de FGSM saldırısı ile üretilen örneklerin PGD saldırısı ile üretilen örneklerden daha kaliteli ve orijinal görüntüye benzer sonuçlar elde ettiği görülmektedir. Ancak, BT görüntüleri MR görüntüleri görüntülerine kıyasla daha yüksek benzerlik taşımaktadır. MSE, RMSE ve PSNR metriklerine

bakıldığında BT ve MR görüntülerinin benzer saldırı koşulları altında yakın kalitede çıktı ürettiğini göstermektedir. UQI metriğine göre ise modalitelerin saldırıdan bağımsız olarak belirli bir kalite standardı yakaladığı görülmektedir. MR görüntüleri BT görüntülerine kıyasla daha yüksek UQI skoru sağlamaktadır. Ancak, BT görüntülerinin farklı pertürbasyon değerlerine karşı kalitesini korumakta daha başarılı olduğu dikkat çekmektedir.



Şekil 6.207. FGSM ve PGD saldırıları ile üretilen MR ve BT örneklerinin orijinal görüntülere kıyasla ortalama kalite ve benzerliklerinin karşılaştırılması

7. SONUÇ VE ÖNERİLER

Teknolojide yaşanan gelişmeler ışığında sağlık alanında da yenilikler ve ilerlemeler meydana gelmektedir. Zamanında ve doğru şekilde tespit ve tanılama yapılmasını sağlamak için yapay zeka destekli sistemlerin kullanımı yaygınlaşmaktadır. Organ, hücre yapıları ve patolojik oluşumlar gibi çeşitli fiziksel ve anatomik bilgiler sağlayan medikal görüntüler, sağlık uzmanlarının yararlandığı önemi tıbbi veri kaynaklarından biridir. Medikal görüntülerin analizinde yapay zekanın alt dalı olan derin öğrenmeden yararlanılmaktadır. Derin öğrenme ile medikal görüntülerdeki anomalilerin tespiti, hastalıklı bölgelerin segmentasyonu, organ ölçümleri gibi görevler başarı ile gerçekleştirilmektedir. Ancak, yapılan araştırmalar derin öğrenme modellerinin çekişmeli saldırı adı verilen tehditlere açık olduğunu ortaya koymaktadır. Bu saldırılar gözle ayırt edilemeyecek düzeyde manipülasyonlarla modelin kararını olumsuz etkilemektedir. Bu güvenlik zafiyeti, derin öğrenme yöntemlerinin medikal görüntüler üzerindeki başarısına gölge düşürmekte ve sağlık alanında derin öğrenme modellerinin kullanımı hakkında soru işaretlerine neden olmaktadır. Bu durum çekişmeli saldırıların derin öğrenme tabanlı medikal görüntülerin analiz sistemleri üzerindeki etkisinin anlaşılmasını zorunlu kılmaktadır.

Bu tez çalışmasında, gradyan tabanlı çekişmeli saldırılar olan FGSM ve PGD saldırılarının MR ve BT görüntülerinin derin öğrenme modelleri ile analizine etkileri araştırılmaktadır. Çalışmada 20 ön eğitilmiş derin öğrenme modelinin saldırılara karşı direnci değerlendirilmektedir. Çekişmeli saldırıların derin öğrenme tabanlı medikal görüntü analiz modelleri üzerindeki etkisi çok yönlü olarak ele alınmaktadır. Pertürbasyon derecesi, medikal görüntü modaliteleri, mevcut derin öğrenme algoritmaları ve saldırıların iteratif özellikleri gibi değişkenlerin saldırı başarısı üzerindeki etkisi incelenmektedir. Çalışmada kullanılan FGSM ve PGD beyaz kutu saldırıları basit ve hızlı bir şekilde görüntüleri manipüle edebilen saldırılardır. Ayrıca, araştırmalarda en çok yararlanılan saldırılar olması nedeniyle bu çalışmada FGSM ve PGD saldırılarının kullanımı tercih edilmektedir. Beyin Tümör MR ve Böbrek BT setleri üzerinde yapılan deneyler iki sınıflı tümör tespit çalışması olarak gerçekleştirilmektedir. Her bir görüntü seti için temiz test setlerinde tümör tespiti gerçekleştirilmektedir. Ayrıca, üç farklı pertürbasyon değerine sahip (0,02; 0,05; 0,10) FGSM ve PGD saldırı senaryosu ile çekişmeli test setleri üretilerek tümör tespiti

gerçekleştirilmektedir. Belirlenen 20 ön eğitilmiş derin öğrenme modeli kullanılarak 7 senaryo analiz edilmektedir. Modellerin saldırı senaryolarına karşı medikal görüntülerde doğru tümör tespit başarısı çeşitli performans metrikleri ile değerlendirilmektedir. Ek olarak, çeşitli pertürbasyon derecelerine sahip FGSM ve PGD saldırıları ile üretilen örneklerin orijinal görüntüye benzerlikleri ve kaliteleri değerlendirilmektedir. Bu sayede, görüntülerdeki pertürbasyon derecesine bağlı bozulmanın saldırı başarısı ile ilişkisi ortaya konmaktadır.

Çalışma sonucunda, kullanılan 20 derin öğrenme modelinin temiz MR ve BT test görüntülerinin sınıflandırılmasında üstün başarı gösterdiği görülmektedir. Ancak saldırı senaryoları sonucunda belirgin düzeyde yanlış tespit gerçekleşmektedir. Gerçekleştirilen deneyler sonucunda, Beyin Tümör MR seti üzerinde yapılan değerlendirmede, InceptionV3 modelinin saldırılara karşı daha dirençli olduğu görülmektedir. MobileNetV3Small ve EfficientNetV2S modelleri ise FGSM ve PGD saldırı senaryolarından en çok etkilenen modeller olmaktadır. Böbrek BT seti üzerinde yapılan değerlendirmede, FGSM saldırısı altında ResNet152V2 modeli daha tutarlı sonuçlar elde ederken, PGD saldırı altında EfficientNetV2B2 ve EfficientNetV2B3 modelleri daha başarılı olmaktadır. MobileNetV3Small modeli ise tüm saldırı senaryolarından en çok etkilenen modeldir. Her senaryoda farklı bir model ön plana çıkmaktadır. Bu nedenle medikal görüntü analizinde gerçekleştirilebilecek potansiyel bir saldırının varlığında tek bir derin öğrenme modelinin kullanımı risk taşımaktadır. Pertürbasyon derecesindeki artış her iki saldırı için de yanlış sınıflandırma başarısını artırmaktadır. Kalite ve benzerlik metrikleri ile yapılan değerlendirme de pertürbasyon derecesindeki artışın görüntülerdeki bozulma oranını artırdığını göstermektedir. Bu sonuç pertürbasyon düzeyinin saldırı başarısında başlıca etki unsuru olduğunu doğrulamaktadır. Saldırıların iteratif özelliği de saldırı başarısını etkileyen bir diğer unsur olmaktadır. Çok adımlı bir saldırı olan PGD, tek adımlı FGSM saldırısına göre daha başarılıdır. PGD saldırısı daha küçük pertürbasyon değerleri ile daha büyük saldırı başarısı elde etmekte ve modellerin doğruluk oranlarını hızlıca düşürmektedir. Medikal görüntü modalitelerinin saldırılara karşı farklı dirençlere sahip olabileceği görüşü üzerine bir değerlendirme gerçekleştirilmektedir. Buna göre, MR görüntülerinin BT görüntülerine kıyasla çekışmeli saldırılara karşı daha dirençli olduğu ortaya konulmaktadır. Ancak, çalışmada sınırlı sayıda örnek içeren açık veri setleri kullanıldığı göz ardı edilmemelidir. Görüntü kalite ve benzerlik metriklerinin değerlendirmesine bakıldığında, mevcut çalışmanın, MR görüntülerinin BT

görüntülerinden daha dayanıklı olduğunu söylemek için tek başına yeterli olmayacağı düşünülmektedir.

Bu tez çalışması, derin öğrenme tabanlı medikal görüntü analizi sistemleri üzerinde çekişmeli saldırıların etkisine dair yapılacak gelecek çalışmalar için bir temel oluşturmaktadır. Saldırıların bu sistemler üzerindeki etkinliğinin ölçülmesi gerekli önlemlerin alınması ve uygun savunma yöntemlerinin geliştirilmesi için yarar sağlayacaktır. Çekişmeli saldırıların doğası ve saldırıların medikal görüntü analizinde kullanılan derin öğrenme modelleri üzerinde başarılı olmasına neden olan bazı temel faktörler araştırılmaktadır. Elde edilen bulgular ışığında araştırma sahası genişletilebilecektir. Çalışmadaki bir kısıt ise kullanılan ön eğitilmiş modellerin doğal görüntülerin analizinde kullanılmak üzere geliştirilen popüler modeller olmasıdır. İlerleyen çalışmalarda medikal görüntülerin analizi için özel olarak geliştirilen derin öğrenme modellerinin saldırılar karşısındaki sağlamlığının değerlendirilmesi faydalı olacaktır. Ayrıca, medikal görüntü modalitelerinin saldırılara direnci konusunda yapılan araştırmalar kısıtlıdır. Bu konuda yeni çalışmalar yapılması önem arz etmektedir.

KAYNAKLAR

1. Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M. K. (2018). Medical image analysis using convolutional neural networks: a review. *Journal of Medical Systems*, 42, 1-13.
2. Liu, Y., Cheng, H. D., Huang, J. H., Zhang, Y. T., Tang, X. L., Tian, J. W., and Wang, Y. (2012). Computer aided diagnosis system for breast cancer based on color Doppler flow imaging. *Journal of Medical Systems*, 36, 3975-3982.
3. Shen, D., Wu, G., and Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221-248.
4. Azam, M. A., Khan, K. B., Salahuddin, S., Rehman, E., Khan, S. A., Khan, M. A., Kadry, S., and Gandomi, A. H. (2022). A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Computers in Biology and Medicine*, 144, 105253.
5. Singh, B., Gautam, R., Kumar, S., Kumar, B. N., Nongthomba, U., Nandi, D., Mukherjee, G., Santosh, V., Somasundaram, K., and Umapathy, S. (2012). Application of vibrational microspectroscopy to biology and medicine. *Current Science*, 102(2), 232.
6. Bandon, D., Trolliard, P., Garcia, A., Lovis, C., Geissbühler, A., and Vallée, J.-P. (2004). *Building an enterprise-wide PACS for all diagnostic images*. In International Congress Series, Elsevier, 1268, 279-284.
7. Paré, G., and Trudel, M. C. (2007). Knowledge barriers to PACS adoption and implementation in hospitals. *International Journal of Medical Informatics*, 76(1), 22-33.
8. Tachibana, H., Omatsu, M., Higuchi, K., and Umeda, T. (2006). Design and development of a secure DICOM-Network Attached Server. *Computer Methods and Programs in Biomedicine*, 81(3), 197-202.
9. Schmidt, L., Santurkar, S., Tsipras, D., Talwar, K., and Madry, A. (2018). Adversarially robust generalization requires more data. *Advances in Neural Information Processing Systems*, 31, 5019-5031.
10. Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., and Lu, F. (2021). Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, 110, 107332.
11. Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., and Madry, A. (2019). Adversarial examples are not bugs, they are features. *Advances in Neural Information Processing Systems*, 32.
12. Internet: Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. Web: <https://arxiv.org/abs/1412.6572>, Son Erişim Tarihi: 02.05.2024.

13. Kaviani, S., Han, K. J., and Sohn, I. (2022). Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications*, 198, 116815.
14. Kovalev, V., and Voynov, D. (2019, 21-23 May). *Influence of control parameters and the size of biomedical image datasets on the success of adversarial attacks*. In Proceedings of the Pattern Recognition and Information Processing: 14th International Conference, Minsk, Belarus, Revised Selected Papers 14, 301-311.
15. Pal, B., Gupta, D., Rashed-Al-Mahfuz, M., Alyami, S. A., and Moni, M. A. (2021). Vulnerability in deep transfer learning models to adversarial fast gradient sign attack for covid-19 prediction from chest radiography images. *Applied Sciences*, 11(9), 4233.
16. Rodriguez, D., Nayak, T., Chen, Y., Krishnan, R., and Huang, Y. (2022). On the role of deep learning model complexity in adversarial robustness for medical images. *BMC Medical Informatics and Decision Making*, 22(2), 1-15.
17. İnternet: Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2013). Intriguing properties of neural networks. Web: <https://arxiv.org/abs/1312.6199> , Son Erişim Tarihi. 02.05.2024.
18. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., and Swami, A. (2017, 2-6 April). *Practical black-box attacks against machine learning*. In Proceedings of the 2017 ACM on Asia conference on computer and communications security, Abu Dhabi, United Arab Emirates, 506-519.
19. Qiu, S., Liu, Q., Zhou, S., and Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, 9(5), 909.
20. Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., and Frossard, P. (2017, 21-26 July). *Universal adversarial perturbations*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 86-94.
21. Machado, G. R., Silva, E., and Goldschmidt, R. R. (2021). Adversarial machine learning in image classification: A survey toward the defender's perspective. *ACM Computing Surveys (CSUR)*, 55(1), 1-38.
22. İnternet: Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. Web: <https://arxiv.org/abs/1706.06083> , Son Erişim Tarihi. 02.05.2024.
23. Chollet, F. (2021). *Deep learning with Python*. Baskı: Simon and Schuster.
24. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 1-74.
25. Suganyadevi, S., Seethalakshmi, V., and Balasamy, K. (2022). A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, 11(1), 19-38.

26. İnternet: Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Web: <https://arxiv.org/abs/1409.1556>, Son Erişim Tarihi. 02.05.2024.
27. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). *Going deeper with convolutions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 1-9.
28. Ioffe, S., and Szegedy, C. (2015, 7-9 July). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. In Proceedings of the International Conference on Machine Learning, Lille, France, 37, 448-456.
29. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016, 27-30 June). *Rethinking the inception architecture for computer vision*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2818-2826.
30. Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017, 4-9 February). *Inception-v4, inception-resnet and the impact of residual connections on learning*. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 4278-4284.
31. He, K., Zhang, X., Ren, S., and Sun, J. (2016, 27-30 June). *Deep residual learning for image recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 770-778.
32. He, K., Zhang, X., Ren, S., and Sun, J. (2016, 11-14 October). *Identity mappings in deep residual networks*. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings Part IV, 630-645 .
33. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017, 21-26 July). *Densely connected convolutional networks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2261-2269.
34. Chollet, F. (2017, 21-26 July). *Xception: Deep learning with depthwise separable convolutions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 1800-1807.
35. Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018, 18-23 June). *Learning transferable architectures for scalable image recognition*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 8697-8710.
36. İnternet: Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. Web: <https://arxiv.org/abs/1704.04861>, Son Erişim Tarihi: 02.05.2024.

37. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018, 18-23 June). *Mobilenetv2: Inverted residuals and linear bottlenecks*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 4510-4520.
38. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., and Adam, H. (2019, 27-30 October). *Searching for mobilenetv3*. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 1314-1324.
39. Tan, M., and Le, Q. (2019, 9-15 June). *Efficientnet: Rethinking model scaling for convolutional neural networks*. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 6105-6114.
40. Tan, M., and Le, Q. (2021, 18-24 July). *Efficientnetv2: Smaller models and faster training*. In Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 10096-10106.
41. Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612.
42. Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003, 9-12 November). *Multiscale structural similarity for image quality assessment*. In Proceedings of The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 2, 1398-1402.
43. Sheikh, H. R., and Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2), 430-444.
44. Sheikh, H. R., Bovik, A. C., and De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12), 2117-2128.
45. Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11), 3440-3451.
46. Wang, Z., and Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3), 81-84.
47. Paschali, M., Conjeti, S., Navarro, F., and Navab, N. (2018, 16-20 September). *Generalizability vs. robustness: investigating medical imaging networks using adversarial examples*. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2018, Granada, Spain, Proceedings Part I, 493-501.
48. İnternet: Finlayson, S. G., Chung, H. W., Kohane, I. S., and Beam, A. L. (2018). Adversarial attacks against medical deep learning systems. Web: <https://arxiv.org/abs/1804.05296> ,Son Erişim Tarihi: 02.05.2024.

49. Shah, A., Lynch, S., Niemeijer, M., Amelon, R., Clarida, W., Folk, J., Russell, S., Wu, X., and Abramoff, M. D. (2018, 4-7 April). *Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms*. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 1454-1457.
50. Yoo, T. K., and Choi, J. Y. (2020). Outcomes of adversarial attacks on deep learning models for ophthalmology imaging domains. *JAMA Ophthalmology*, 138(11), 1213-1215.
51. Anand, D., Tank, D., Tibrewal, H., and Sethi, A. (2020, 3-7 April). *Self-supervision vs. transfer learning: robust biomedical image analysis against adversarial attacks*. In Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020), Iowa City, IA, USA, 1159-1163.
52. Kovalev, V., Liauchuk, V., Voynov, D., and Tuzikov, A. (2021). Biomedical image recognition in pulmonology and oncology with the use of deep learning. *Pattern Recognition and Image Analysis*, 31, 144-162.
53. Yang, Z., Yang, Y., Xue, Y., Shih, F. Y., Ady, J., and Roshan, U. (2020, 16-19 December). *Accurate and adversarially robust classification of medical images and ECG time-series with gradient-free trained sign activation neural networks*. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 2456-2460.
54. Allyn, J., Allou, N., Vidal, C., Renou, A., and Ferdynus, C. (2020). Adversarial attack on deep learning-based dermatoscopic image recognition systems: Risk of misdiagnosis due to undetectable image perturbations. *Medicine*, 99(50), e23568.
55. Gongye, C., Li, H., Zhang, X., Sabbagh, M., Yuan, G., Lin, X., Wahl, T., and Fei, Y. (2020, 2-5 November). *New passive and active attacks on deep neural networks in medical applications*. In Proceedings of the IEEE/ACM International Conference On Computer Aided Design (ICCAD), San Diego, CA, USA, 1-9.
56. Zhou, Q., Zuley, M., Guo, Y., Yang, L., Nair, B., Vargo, A., Ghannam, S., Arefan, D., and Wu, S. (2021). A machine and human reader study on AI diagnosis model safety under attacks of adversarial images. *Nature Communications*, 12(1), 7281.
57. Bortsova, G., González-Gonzalo, C., Wetstein, S. C., Dubost, F., Katramados, I., Hogeweg, L., Liefers, B., van Ginneken, B., Pluim, J. P., and Veta, M. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*, 73, 102141.
58. Kong, F., Liu, F., Xu, K., and Shi, X. (2023). Why does batch normalization induce the model vulnerability on adversarial images? *World Wide Web*, 26(3), 1073-1091.
59. Hirano, H., Minagi, A., and Takemoto, K. (2021). Universal adversarial attacks on deep neural networks for medical image classification. *BMC Medical Imaging*, 21, 1-13.

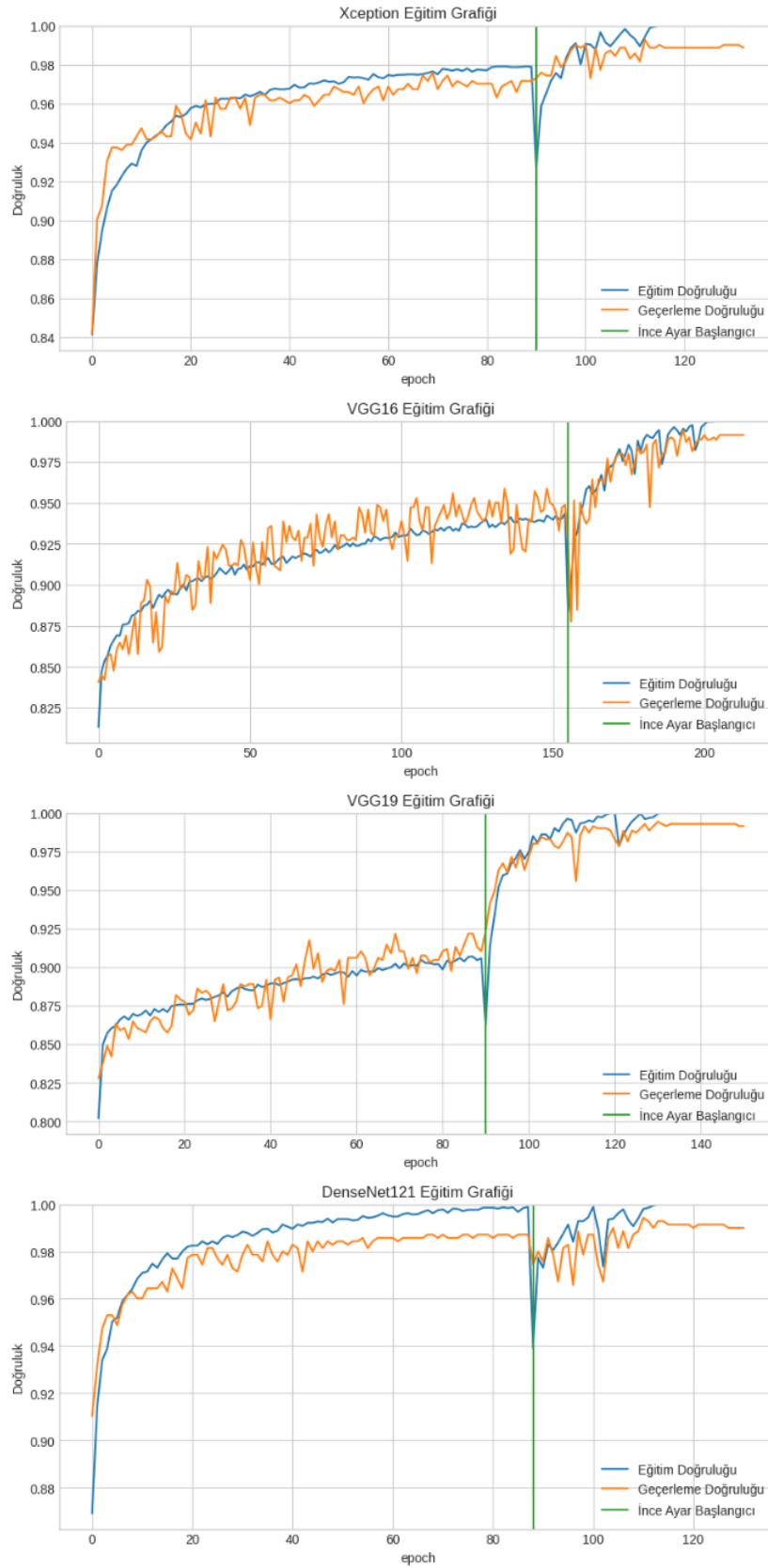
60. Minagi, A., Hirano, H., and Takemoto, K. (2022). Natural images allow universal adversarial attacks on medical image classification using deep neural networks with transfer learning. *Journal of Imaging*, 8(2), 38.
61. Diyasa, I. G. S. M., Wahid, R. R., and Amiruddin, B. P. (2021, 18-19 November). *Grasping Adversarial Attacks on Deep Convolutional Neural Networks for Cholangiocarcinoma Classification*. In Proceedings of the International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 1-4.
62. Rahman, A., Hossain, M. S., Alrajeh, N. A., and Alsolami, F. (2020). Adversarial examples-Security threats to COVID-19 deep learning systems in medical IoT devices. *IEEE Internet of Things Journal*, 8(12), 9603-9610.
63. Li, Y., and Liu, S. (2023). The Threat of Adversarial Attack on a COVID-19 CT Image-Based Deep Learning System. *Bioengineering*, 10(2), 194.
64. Joel, M. Z., Avesta, A., Yang, D. X., Zhou, J. G., Omuro, A., Herbst, R. S., Krumholz, H. M., and Aneja, S. (2023). Comparing Detection Schemes for Adversarial Images against Deep Learning Models for Cancer Imaging. *Cancers*, 15(5), 1548.
65. Byra, M., Styczynski, G., Szmigielski, C., Kalinowski, P., Michalowski, L., Paluszkiwicz, R., Ziarkiewicz-Wroblewska, B., Zieniewicz, K., and Nowicki, A. (2020, 7-11 September). *Adversarial attacks on deep learning models for fatty liver disease classification by modification of ultrasound image reconstruction method*. In Proceedings of the IEEE International Ultrasonics Symposium (IUS), Las Vegas, NV, USA, 1-4.
66. Mangaokar, N., Pu, J., Bhattacharya, P., Reddy, C. K., and Viswanath, B. (2020, 7-11 September). *Jekyll: Attacking medical image diagnostics using deep generative models*. In Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P), Genoa, Italy, 139-157.
67. Yao, Q., He, Z., Lin, Y., Ma, K., Zheng, Y., and Zhou, S. K. (2021). *A hierarchical feature constraint to camouflage medical adversarial attacks*. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021, Strasbourg, France, Proceedings Part III, 36-47.
68. Kulkarni, Y., and Bhambani, K. (2021, 21-24 June). *Kryptonite: An Adversarial Attack Using Regional Focus*. In Proceedings of the International Conference on Applied Cryptography and Network Security Workshops, Kamakura, Japan, 463-481.
69. Apostolidis, K. D., and Papakostas, G. A. (2022). Digital watermarking as an adversarial attack on medical image analysis with deep learning. *Journal of Imaging*, 8(6), 155.
70. Koga, K., and Takemoto, K. (2022). Simple black-box universal adversarial attacks on deep neural networks for medical image classification. *Algorithms*, 15(5), 144.
71. Maliamanis, T. V., Apostolidis, K. D., and Papakostas, G. A. (2022). How Resilient Are Deep Learning Models in Medical Image Analysis? The Case of the Moment-Based Adversarial Attack (Mb-AdA). *Biomedicines*, 10(10), 2545.

72. Vatian, A., Gusarova, N., Dobrenko, N., Dudorov, S., Nigmatullin, N., Shalyto, A., and Lobantsev, A. (2019, 8-12 April). *Impact of adversarial examples on the efficiency of interpretation and use of information from high-tech medical images*. In Proceedings of the 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 472-478.
73. Liu, S., Setio, A. A. A., Ghesu, F. C., Gibson, E., Grbic, S., Georgescu, B., and Comaniciu, D. (2020). No surprises: Training robust lung nodule detection for low-dose CT scans by augmenting with adversarial attacks. *IEEE Transactions on Medical Imaging*, 40(1), 335-345.
74. Taghanaki, S. A., Abhishek, K., Azizi, S., and Hamarneh, G. (2019, 15-20 June). *A kernelized manifold mapping to diminish the effect of adversarial perturbations*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 11332-11341.
75. Xue, F.-F., Peng, J., Wang, R., Zhang, Q., and Zheng, W. S. (2019, 13-17 October). *Improving robustness of medical image diagnosis with denoising convolutional neural networks*. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, Shenzhen, China, Proceedings Part VI, 846-854.
76. Daanouni, O., Cherradi, B., and Tmiri, A. (2022). NSL-MHA-CNN: a novel CNN architecture for robust diabetic retinopathy prediction against adversarial attacks. *IEEE Access*, 10, 103987-103999.
77. Li, X., and Zhu, D. (2020, 3-7 April). *Robust detection of adversarial attacks on medical images*. In Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020), Iowa City, IA, USA, 1154-1158.
78. Li, X., Pan, D., and Zhu, D. (2021, 13-16 April). *Defending against adversarial attacks on medical imaging AI system, classification or detection?* In Proceedings of the IEEE 18th International Symposium on Biomedical Imaging (ISBI 2021), Nice, France, 1677-1681.
79. Rizky, L. M. R., and Suyanto, S. (2022). Adversarial training and deep k-nearest neighbors improves adversarial defense of glaucoma severity detection. *Heliyon*, 8(12), e12275.
80. Xu, M., Zhang, T., Li, Z., Liu, M., and Zhang, D. (2021). Towards evaluating the robustness of deep diagnostic models by adversarial attack. *Medical Image Analysis*, 69, 101977.
81. Yu, S. T., and Zhou, H. Y. (2022). Position-Aware Anti-Aliasing Filters for 3D Medical Image Analysis. *IEEE Access*, 10, 100151-100159.
82. Chen, L., Zhao, L., and Chen, C. Y. C. (2021). Enhancing adversarial defense for medical image analysis systems with pruning and attention mechanism. *Medical Physics*, 48(10), 6198-6212.

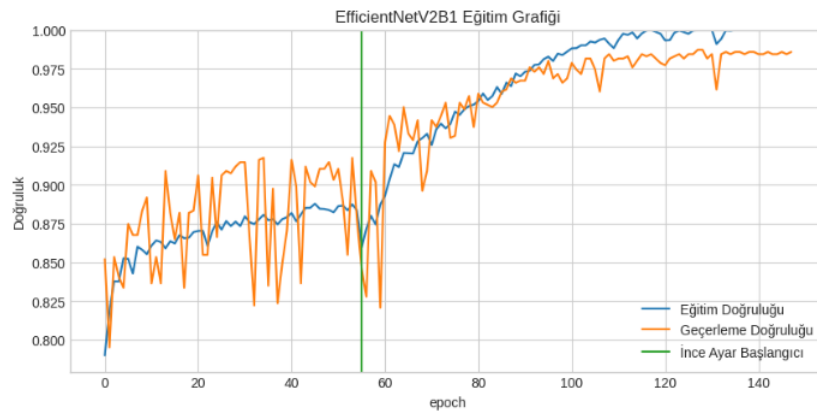
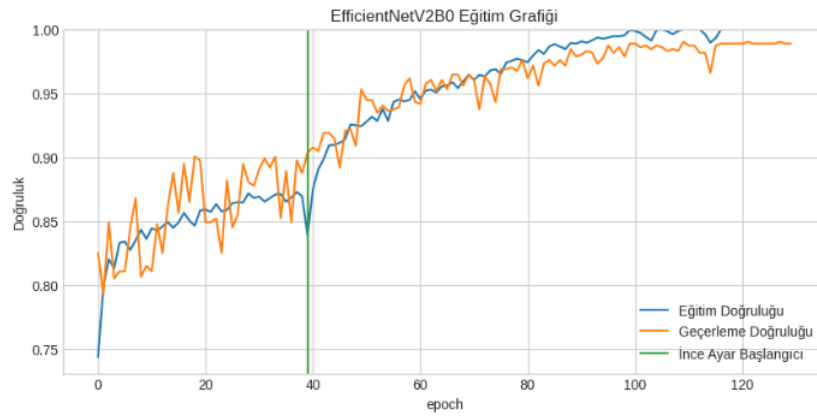
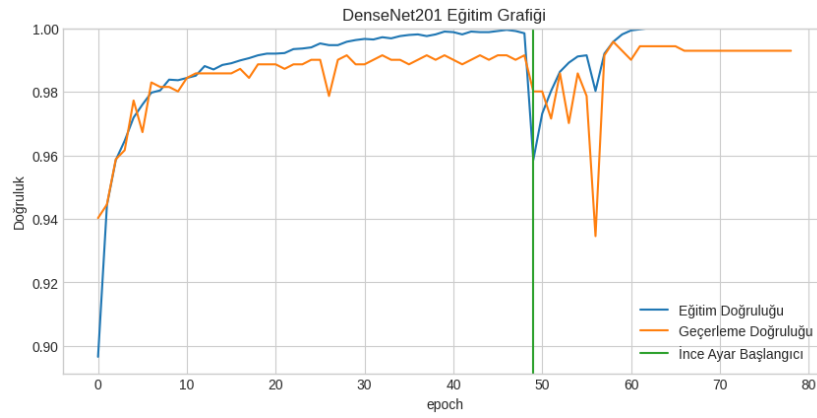
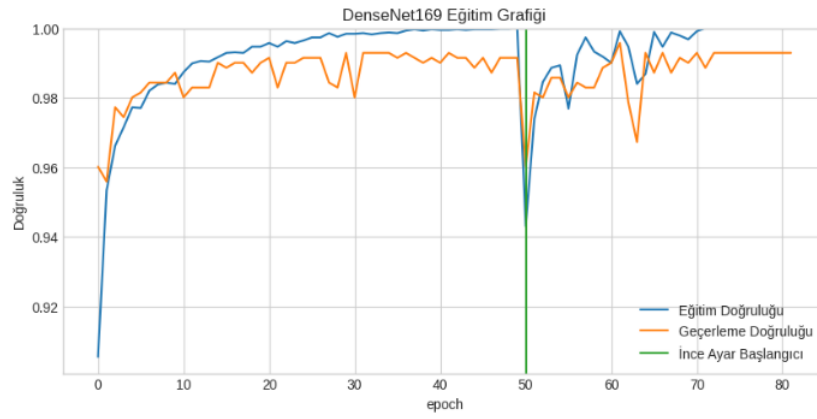
83. He, F., Chen, S., Li, S., Zhou, L., Zhang, H., Peng, H., and Huang, X. (2021, 13-16 April). *Colonoscopic image synthesis for polyp detector enhancement via gan and adversarial training*. In Proceedings of the IEEE 18th International Symposium on Biomedical Imaging (ISBI 2021), Nice, France, 1887-1891.
84. Ahmed, U., Lin, J. C. W., and Srivastava, G. (2022). Mitigating adversarial evasion attacks by deep active learning for medical image classification. *Multimedia Tools and Applications*, 81(29), 41899-41910.
85. Yang, Y., Shih, F. Y., and Roshan, U. (2022). Defense against adversarial attacks based on stochastic descent sign activation networks on medical images. *International Journal of Pattern Recognition and Artificial Intelligence*, 36(03), 2254005.
86. Sun, S., Xian, M., Vakanski, A., and Ghanem, H. (2022, 18-22 September). *MIRST-DM: Multi-instance RST with Drop-Max Layer for Robust Classification of Breast Cancer*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2022, Singapore, Proceedings Part IV, 401-410.
87. Kansal, K., Krishna, P. S., Jain, P. B., Surya, R., Honnavalli, P., and Eswaran, S. (2022). Defending against adversarial attacks on Covid-19 classifier: A denoiser-based approach. *Heliyon*, 8(10), e11209.
88. Xu, M., Zhang, T., and Zhang, D. (2022). Medrdfs: A robust and retrain-less diagnostic framework for medical pretrained models against adversarial attack. *IEEE Transactions on Medical Imaging*, 41(8), 2130-2143.
89. Shi, X., Peng, Y., Chen, Q., Keenan, T., Thavikulwat, A. T., Lee, S., Tang, Y., Chew, E. Y., Summers, R. M., and Lu, Z. (2022). Robust convolutional neural networks against adversarial attacks on medical images. *Pattern Recognition*, 132, 108923.
90. İnternet: Khalel, A. (2020). Sewar: A Python Package for Image Quality Assessment Using Different Metrics. Web: <https://github.com/andrewekhalel/sewar>, Son Erişim Tarihi: 02.05.2024.
91. İnternet: Brain Tumor MRI Dataset. Web: [https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset? resource =download](https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?resource=download), Son Erişim Tarihi: 02.05.2024.
92. İnternet: CT Kidney Dataset: Normal-Cyst-Tumor and Stone. Web: <https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone>, Son Erişim Tarihi: 02.05.2024.
93. Islam, M. N., Hasan, M., Hossain, M. K., Alam, M. G. R., Uddin, M. Z., and Soylu, A. (2022). Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Scientific Reports*, 12(1), 1-14.

EKLER

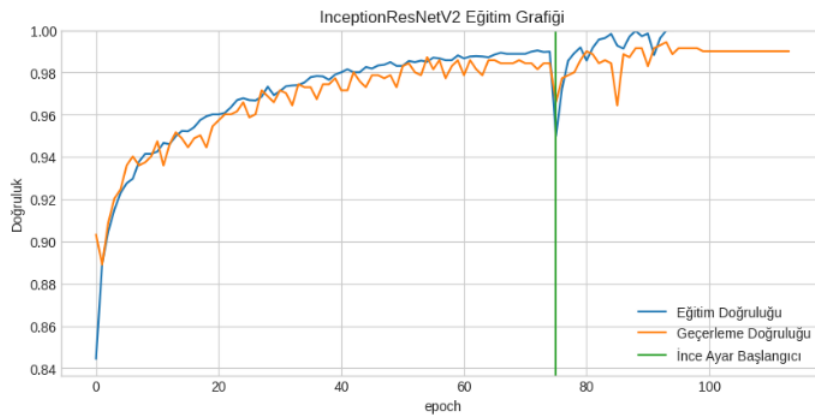
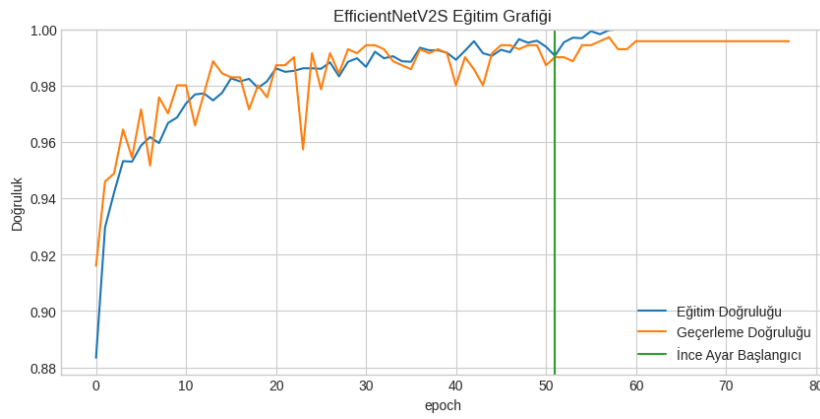
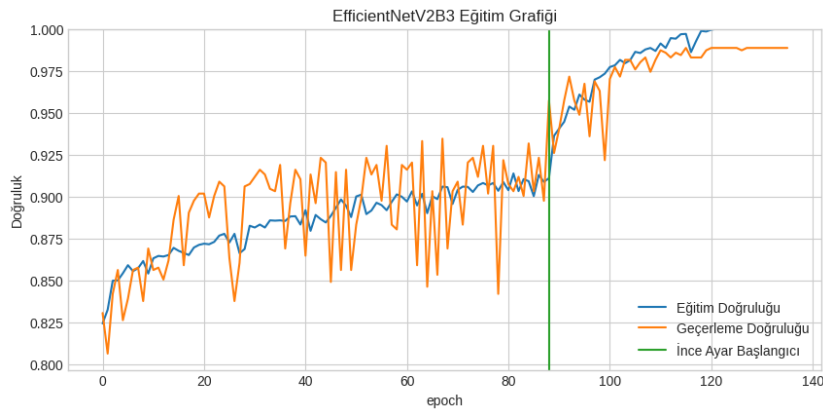
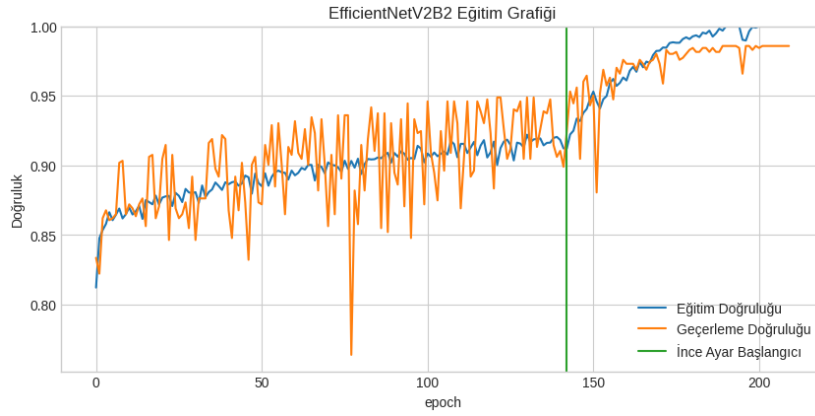
EK-1. Beyin Tümör MR Veri Seti İçin Modellerin Eğitim Grafikleri



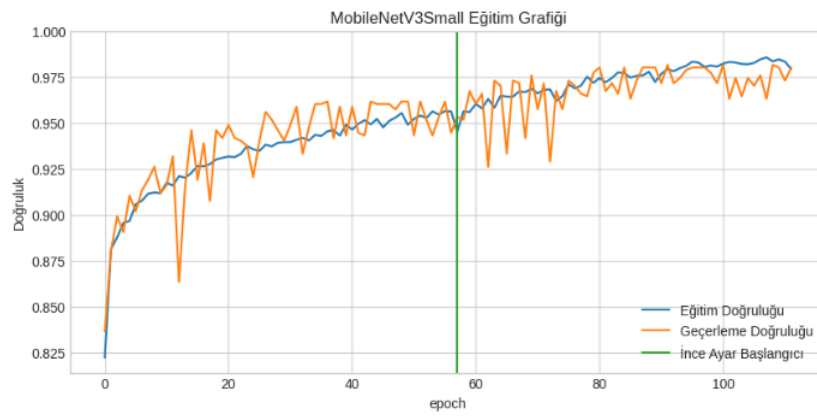
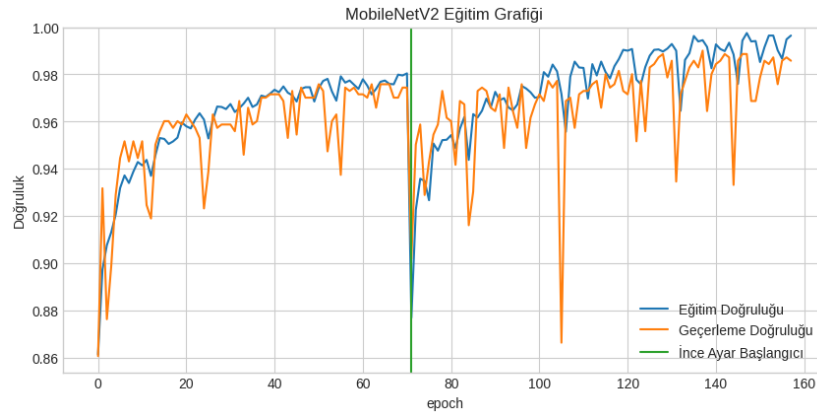
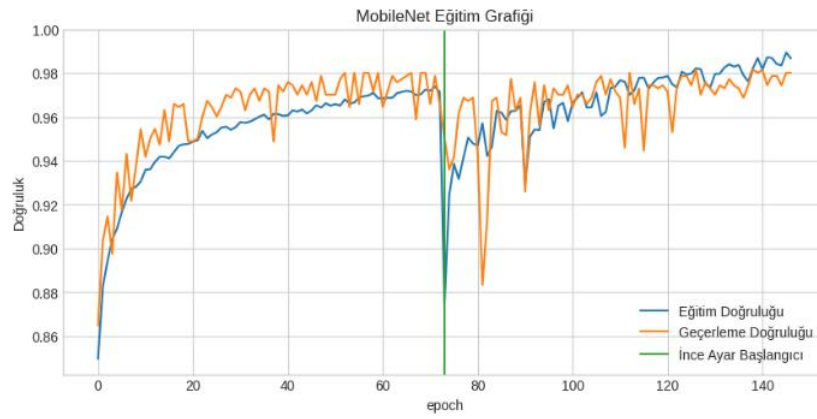
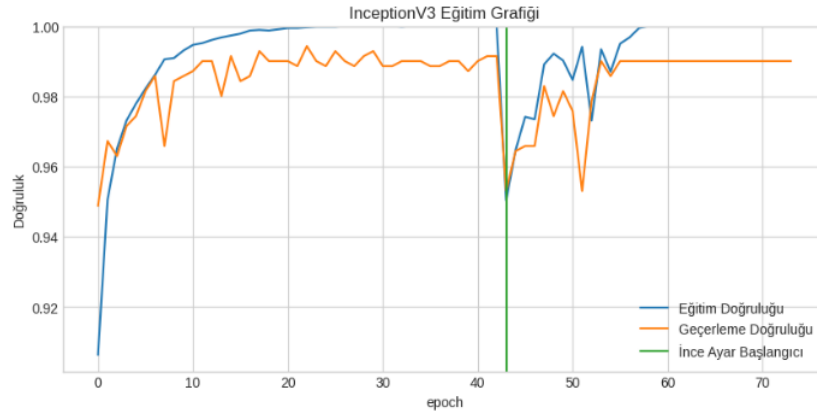
EK-1. (devam) Beyin Tümör MR Veri Seti İçin Modellerin Eğitim Grafikleri



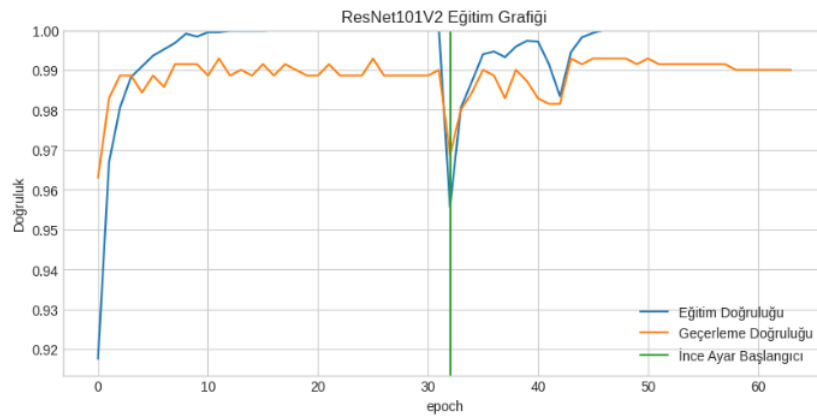
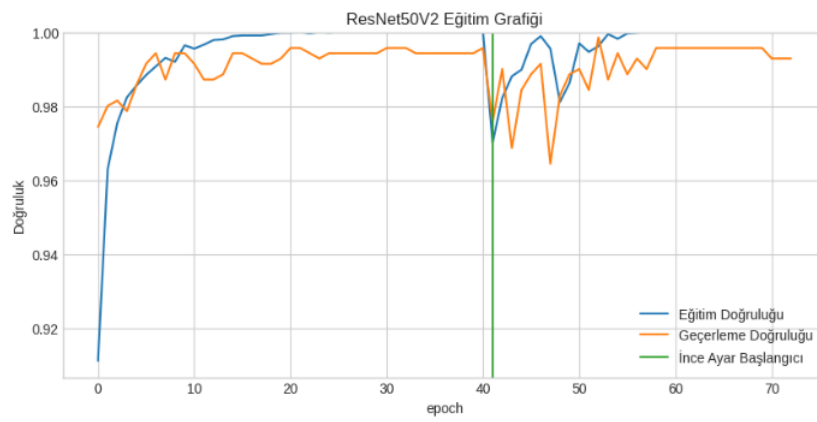
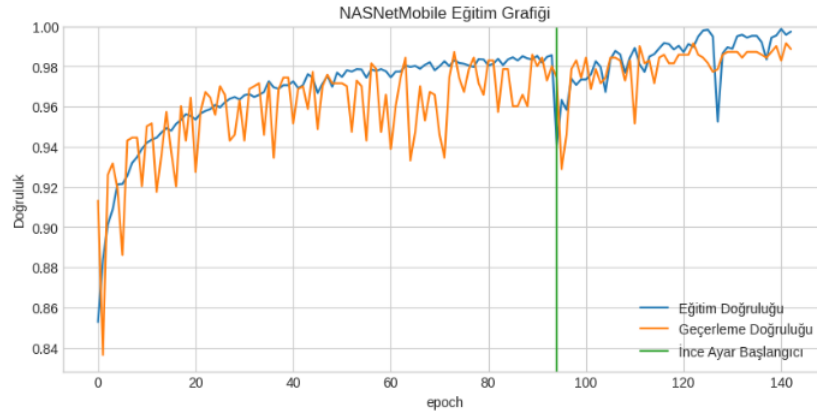
EK-1. (devam) Beyin Tümör MR Veri Seti İçin Modellerin Eğitim Grafikleri



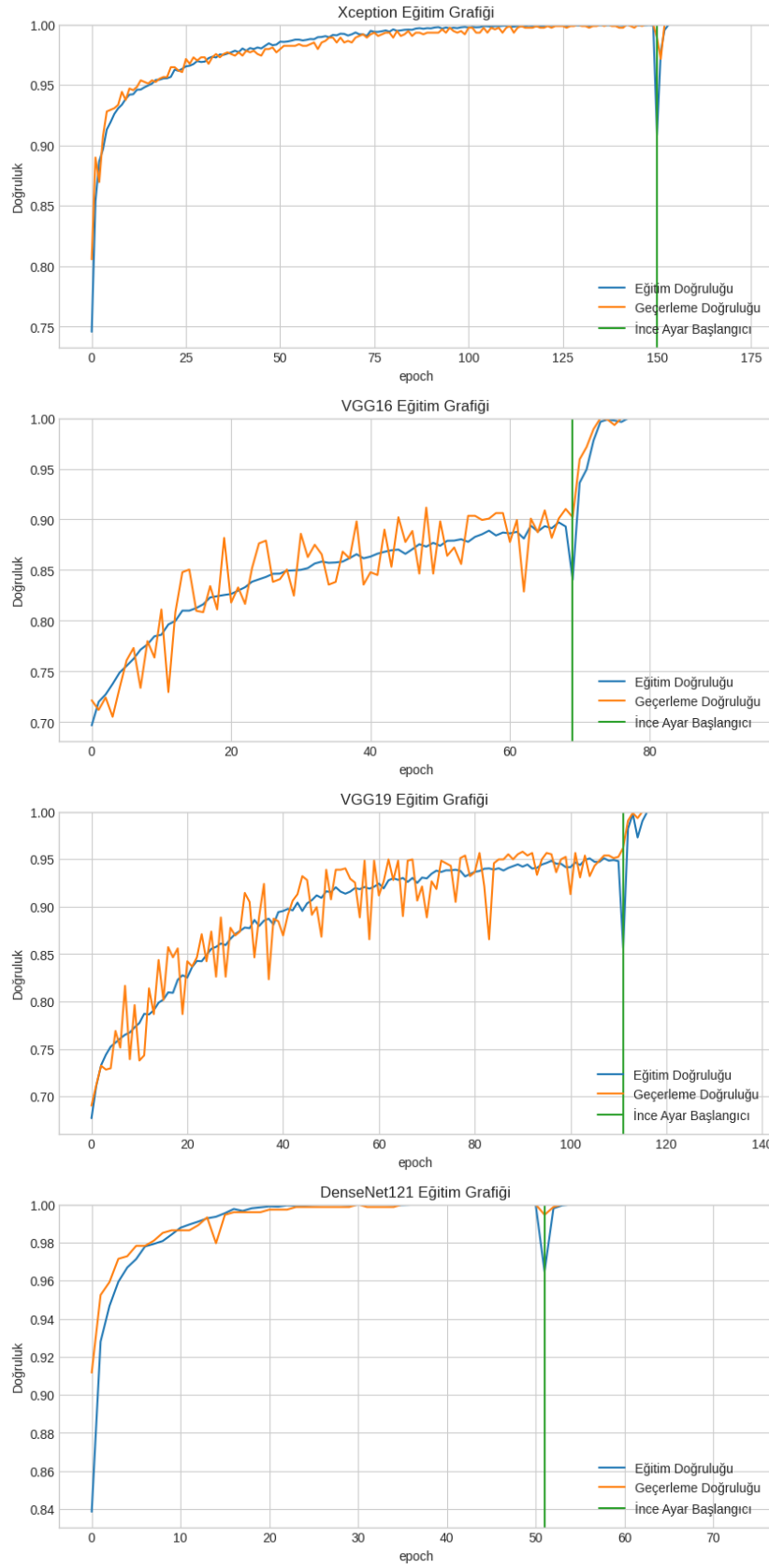
EK-1. (devam) Beyin Tümör MR Veri Seti İçin Modellerin Eğitim Grafikleri



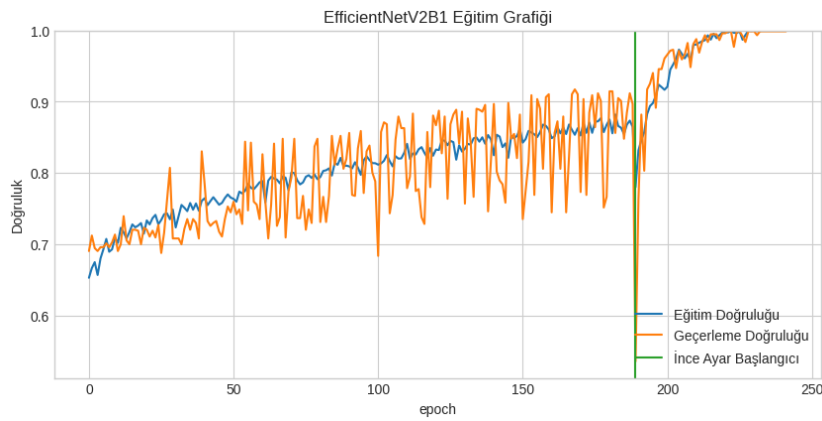
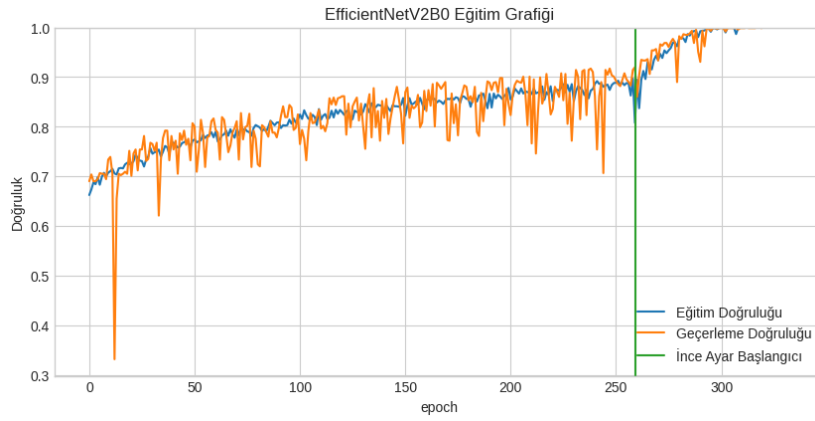
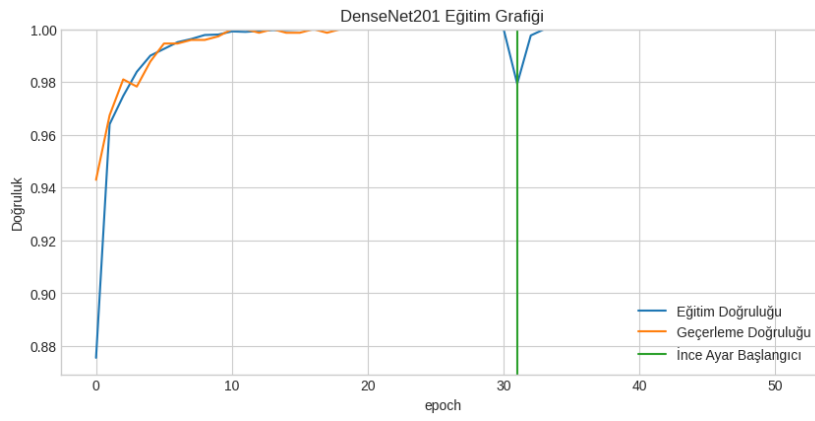
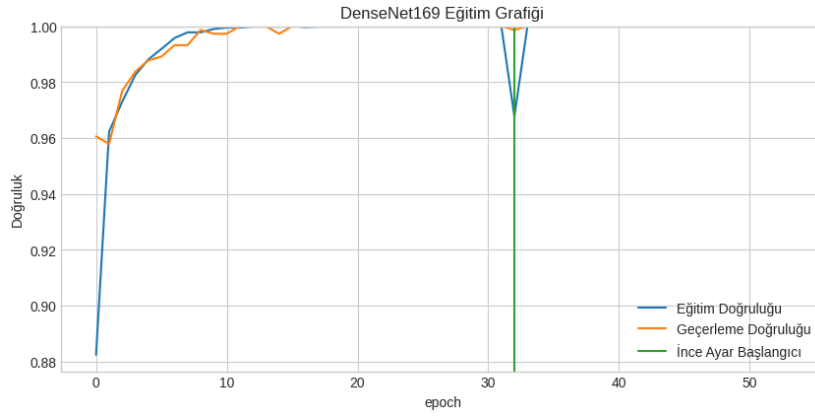
EK-1. (devam) Beyin Tümör MR Veri Seti İçin Modellerin Eğitim Grafikleri



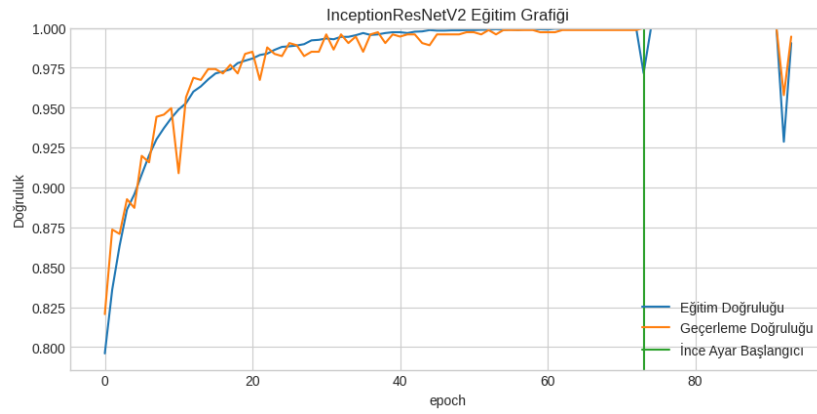
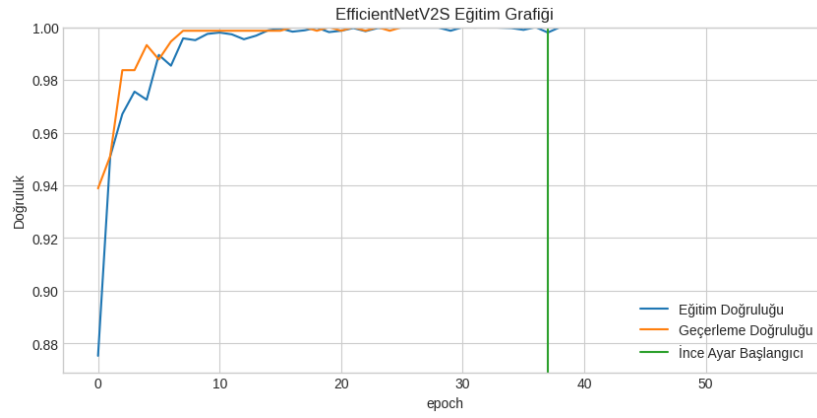
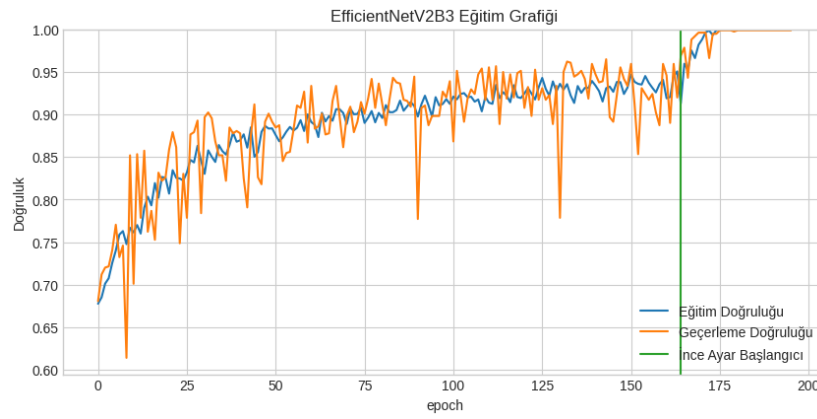
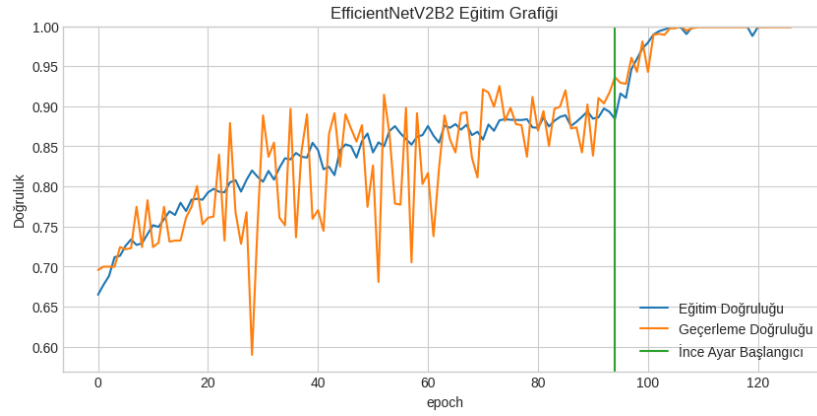
EK-2. Böbrek BT Veri Seti İçin Modellerin Eğitim Grafikleri



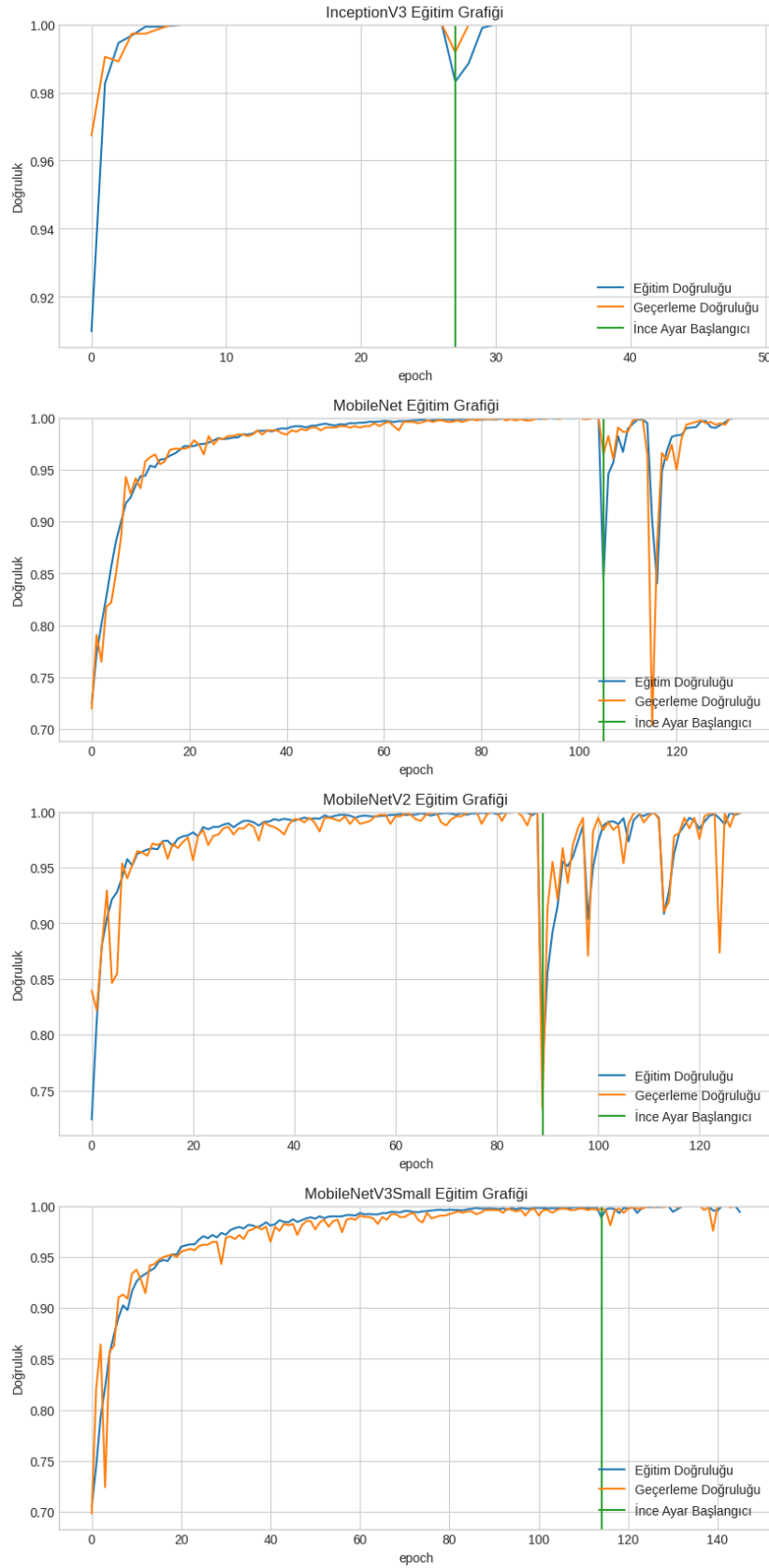
EK-2. (devam) Böbrek BT Veri Seti İçin Modellerin Eğitim Grafikleri



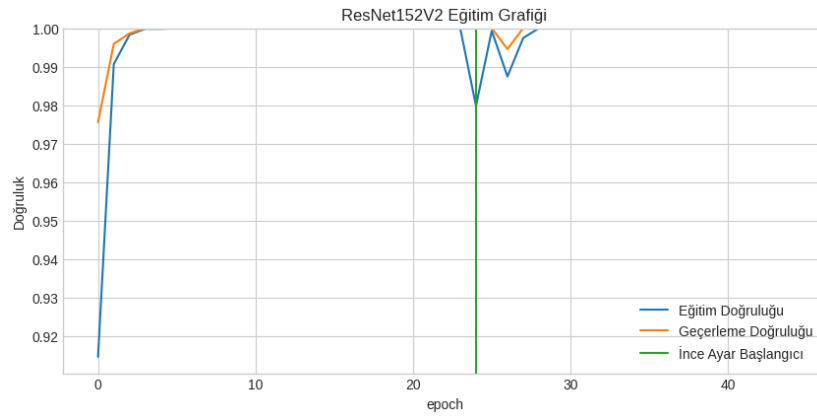
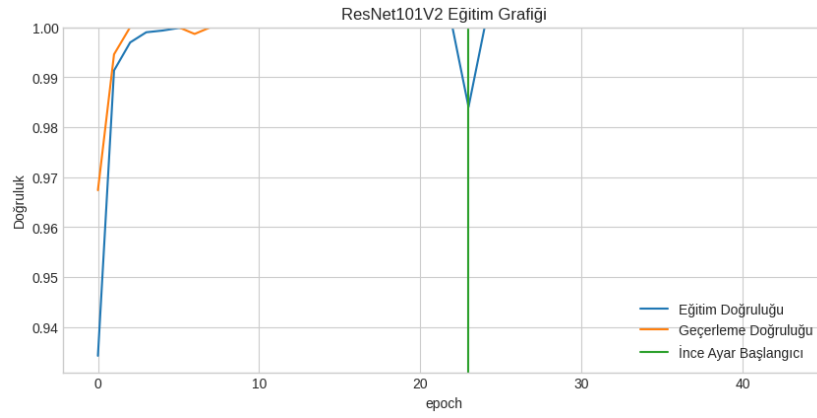
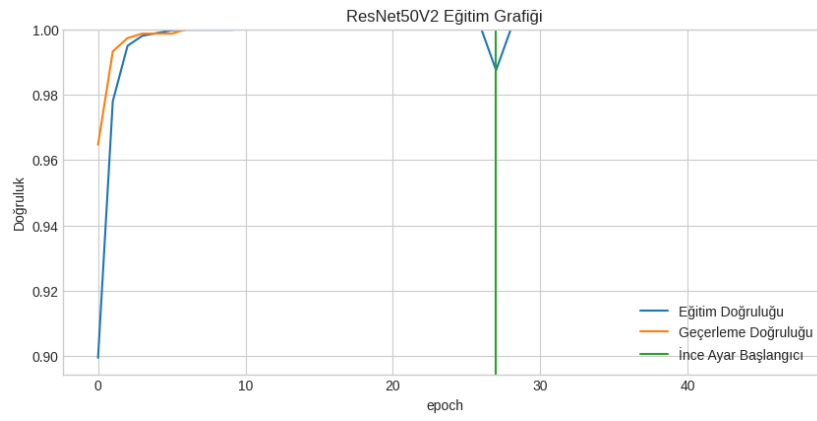
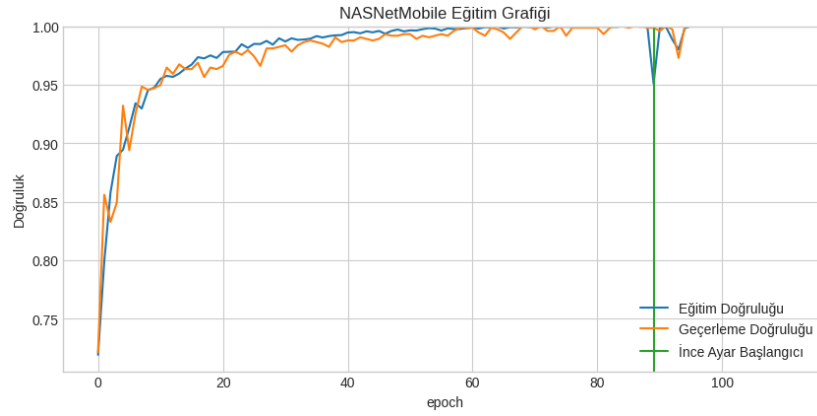
EK-2. (devam) Böbrek BT Veri Seti İçin Modellerin Eğitim Grafikleri



EK-2. (devam) Böbrek BT Veri Seti İçin Modellerin Eğitim Grafikleri



EK-2. (devam) Böbrek BT Veri Seti İçin Modellerin Eğitim Grafikleri





Gazili olmak ayrıcalıktır