
GAZİ ÜNİVERSİTESİ
Gazi
1926

DERİN ÖĞRENME YÖNTEMİ İLE DİFERANSİYEL MAHREMİYETLİ MEDİKAL GÖRÜNTÜ SINIFLANDIRMA

Şükriye AKKAYA

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AĞUSTOS 2021

ETİK BEYAN

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

- 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 çalışmasında 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.

Şükriye AKKAYA

11/08/2021

DERİN ÖĞRENME YÖNTEMİ İLE DİFERANSİYEL MAHREMİYETLİ MEDİKAL GÖRÜNTÜ SINIFLAMNDIRMA

(Yüksek Lisans Tezi)

Şükriye AKKAYA

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Ağustos 2021

ÖZET

Günümüzde derin öğrenme uygulamalarının gelişen kullanımı ile bu uygulamalarda kullanılan veri boyutu ve çeşitliliği oldukça artmıştır. Bu sebeple derin öğrenme uygulamalarında genellikle model başarımına odaklanan çalışmaların yanı sıra bu modellerin güvenliğinin sağlanması için de çeşitli çalışmalar yapılmaya başlanmıştır. Model eğitim verilerinin mahremiyetinin sağlanması için yapılan çalışmalarda, her probleme uygulanabilmesinden ve fayda mahremiyet dengesinin dikkate alınmasından dolayı diferansiyel mahremiyet uygulamaları öne çıkmaktadır. Bu tez çalışmasında, kişilerin sağlık durumu gibi hassas kişisel verileri içeren medikal görüntüler üzerinde diferansiyel mahremiyet korumalı bir derin öğrenme modeli geliştirilmiş ve modelin başarımı test edilmiştir. Model olarak çok ölçekli 3D evrişimli sinir ağı gradyan pertürbasyonuna dayalı DP-SGD optimizasyon algoritması ile kullanılmıştır. Geliştirilen modelde; beyin tümör veya kanseri olan hastaların tanısında mahremiyete saygı gösterilerek tümör derecelendirme adımında LGG ve HGG gliom sınıflandırması yapılmaktadır. Eğitim ve test seti olarak beyin MR görüntüleri içeren BraTS 2020 veri seti kullanılmış olup, mahremiyete duyarlı ve duyarsız derin öğrenme modellerinin başarım oranı test edilmiş ve model başarımlarının çok büyük oranlarda değişmediği görülmüştür. Ortalama %85 civarında model başarımının yakalandığı göz önüne alındığında veri seti boyutu arttığında daha yüksek doğruluğa ulaşılacağı ve mahremiyete saygı gösterilerek beyin MR görüntülerinin analiz edilebileceği, Türk Beyin Projesi kapsamındaki veri setlerine uygulanabileceği değerlendirilmektedir.

Bilim Kodu	:	92432
Anahtar Kelimeler	:	Medikal görüntü sınıflandırma, derin öğrenme, diferansiyel mahremiyet
Sayfa Adedi	:	77
Danışman	:	Prof. Dr. Şeref SAĞIROĞLU

DIFFERENTIAL PRIVACY MEDICAL IMAGE CLASSIFICATION WITH DEEP LEARNING METHOD

(M. Sc. Thesis)

Şükriye AKKAYA

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

August 2021

ABSTRACT

Nowadays, with the growing use of deep learning applications, the size and variety of data used in these applications have increased significantly. For this reason, in addition to the studies focusing on model performance, various studies also have been started to provide the security of deep learning models. In studies that provide the privacy of model training data, differential privacy applications come forward. Because differential privacy can be applied to every problem and, the tradeoffs between throughput and privacy is accounted. In this thesis, a differentially private deep learning model was developed on medical images which contain sensitive personal data such as the health status of individuals. After that, the performance of the model was tested. In this study, multiscale 3D convolutional neural network is used as a model with DP-SGD optimization algorithm based on gradient perturbation. In the developed model, diagnosis of patients with brain tumor or cancer was made with LGG and HGG glioma classification in the tumor grading step, at the same time respecting privacy. The BraTS 2020 dataset, which includes brain MR images, was used as the training and test set. The performance rate of privacy-sensitive and insensitive deep learning models was compared, and it was seen that the model performances did not change to a great extent. With an eye on that the model performance is around 85% on average, it is considered that higher accuracy will be achieved when the dataset size increases. With this, the brain MR images can be analyzed with respecting privacy and differential privacy can be applied to the datasets within the scope of the Turkish Brain Project.

Science Code	:	92432
Key Words	:	Medical image classification, deep learning, differential privacy
Page Number	:	77
Supervisor	:	Prof. Dr. Şeref SAĞIROĞLU

TEŞEKKÜR

Bu tez çalışması Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ve Gazi Üniversitesi iş birliği çerçevesinde gerçekleştirilen Türk Beyin Projesi kapsamında katkıda bulunmak amacı ile gerçekleştirilmiştir. Bu süre boyunca değerli katkıları ile beni yönlendiren, tez çalışmam boyunca gerekli olanakları sağlayan değerli tez danışmanım 'Gazi Üniversitesi' öğretim üyesi Sayın Prof. Dr. Şeref SAĞIROĞLU'na ve yüksek lisans eğitimim boyunca yanımda olan sevgili aileme teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ÇİZELGELERİN LİSTESİ ix

ŞEKİLLERİN LİSTESİ x

SİMGELER VE KISALTMALAR xii

1. GİRİŞ 1

2. SAĞLIKTA MAHREMİYET VE MAHREMİYET DUYARLI SİSTEMLER 5

2.1. Sağlık Hizmetlerinde Mahremiyetin Önemi 6

2.2. Derin Öğrenme Mahremiyet İhlalleri 9

2.2.1. Direk bilgi sızdırma 9

2.2.2. Dolaylı bilgi sızdırma 10

2.3. Mahremiyet Duyarlı Sistemler 14

3. DİFERANSİYEL MAHREMİYET 17

3.1. Temel Kavramlar 18

3.1.2. Küratör 18

3.1.3. Komşu veri tabanları 18

3.1.4. Mekanizma 18

3.1.5. Mahremiyet maliyeti parametresi (ε) 19

3.1.6. Hassasiyet 19

3.1.7. (ε, δ)-Diferansiyel mahremiyet 19

3.1.8. Laplace mekanizması 21

Sayfa

3.1.9. Gauss mekanizması 21

3.1.10. DM bileşik mekanizmalar 23

3.2. Derin Öğrenmede Diferansiyel Mahremiyet 23

3.2.1. Diferansiyel mahremiyet korumalı SGD optimizasyonu 25

3.2.2. Derin öğrenme diferansiyel mahremiyet uygulamaları 26

3.2.3. Diferansiyel mahremiyet uygulanmış modellerin saldırılardaki başarısı 28

4. DİFERANSİYEL MAHREMİYET UYGULANMIŞ BEYİN MR GÖRÜNTÜLERİNDE ANOMALİ VEYA TÜMÖR TESPİTİ 31

4.1. Beyin Tümörü Hastalığı ve Derin Öğrenme Çalışmaları 31

4.1.1. Beyin tümörü teşhis yöntemleri 32

4.1.2. Medikal beyin görüntüleri ile yapılmış derin öğrenme çalışmaları 34

4.2. Derin Öğrenme Temel Kavramlar ve Uygulanan Model 36

4.2.1. Derin öğrenme temel kavramlar 36

4.3. Uygulama Tasarımı 52

4.3.1. Veri seti 53

4.3.2. Veri önişleme 54

4.3.3. Eğitim, test ve doğrulama veri setlerinin oluşturulması 55

4.3.4. Model eğitimi ve hiper parametre seçimi 58

4.3.5. Sonuçlar 62

5. SONUÇ VE ÖNERİLER 67

KAYNAKLAR 69

ÖZGEÇMİŞ 77

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 3.1. Diferansiyel mahremiyet uygulanmış derin öğrenme çalışmaları 27

Çizelge 4.1. Sınıflara göre veri sayısı 56

Çizelge 4.2. Veri seti dağılımı 56

Çizelge 4.3. Uygulanan 3D çok ölçekli CNN ağı mimarisi..... 60

Çizelge 4.4. DP-SGD ile eğitilen modelin en yüksek başarımının metrik değerleri..... 64

Çizelge 4.5. DP-SGD ile eğitilen modelin en yüksek başarımının karmaşıklık matrisi..... 64

Çizelge 4.6. Batch size değerleri için hesaplanan epsilon değerleri 65

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 2.1. Derin öğrenme uygulamalarındaki tehditlerin sınıflandırılması..... 10

Şekil 2.2. Yang ve diğerleri çalışmalarından alınan bir sonuç örneği 13

Şekil 3.1. Laplace ve Gauss dağılımları..... 22

Şekil 3.2. Derin öğrenmede gürültü eklenebilecek adımlar..... 24

Şekil 3.3. DP SGD ve SGD algoritmalarının uygulanması 26

Şekil 3.4. (a) MNIST sonuçları (b) CIFAR-10 sonuçları 30

Şekil 4.1. Beyin MR görüntüleri kesit ve sekansları 34

Şekil 4.2. Beyin tümörü tanısı için önerilen genel akış şeması 35

Şekil 4.3. Biyolojik sinir ve yapay sinir hücreleri 37

Şekil 4.4. Makine öğrenmesi uygulamalarında temel sınıflandırma adımları..... 39

Şekil 4.5. Evrişimli ağın genel yapısı 40

Şekil 4.6. Evrişimli sinir ağının temel bileşenleri..... 41

Şekil 4.7. Evrişinli sinir ağlarında konvülasyon işlemi 41

Şekil 4.8. Ortaklama işlemi yapılmış girdi ve düşük çözünürlüklü çıktı..... 42

Şekil 4.9. Öğrenme modelinde modelin uygunluğu 43

Şekil 4.10. Yapay sinir ağlarında bırakma işlemi..... 43

Şekil 4.11. Veri çoğaltma yöntemi ile girdiden oluşturulan görüntüler 45

Şekil 4.12. Karmaşıklık matrisi 46

Şekil 4.13. Derin ağlarda öğrenme durumları..... 47

Şekil 4.14. Öğrenme katsayısı ve global minimum arasındaki ilişki..... 48

Şekil 4.15. Kayıp fonksiyonu gradyanı ve global minimum 50

Şekil 4.16. Birden çok lokal minimuma sahip kayıp fonksiyonu..... 51

Şekil **Sayfa**

Şekil 4.17. Kategori etiketleme 52

Şekil 4.18. Uygulama temel adımları 53

Şekil 4.19. Veri önişleme adımları akış şeması 54

Şekil 4.20. Veri setinde bulunan T1 ağırlıklı HGG gliom örneği 55

Şekil 4.21. HGG gliom görüntüsünde yapılan tümör işaretlemesi 55

Şekil 4.22. Veri setlerinin bölünmesi işlemi adımları 57

Şekil 4.23. Çok ölçekli CNN modelinin detaylı gösterimi 59

Şekil 4.24. Yığın (batch) sayısı 4 olarak seçilen modelin doğruluk kayıp grafikleri 62

Şekil 4.25. Yığın (batch) sayısı 8 olarak seçilen modelin doğruluk kayıp grafikleri 63

Şekil 4.26. Referans modelin, doğruluk kayıp grafikleri 63

Şekil 4.27. DP-SGD ile eğitilen modelde en yüksek başarımı gösteren doğruluk kayıp grafikleri 64

SİMGELER VE KISALTMALAR

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

Simgeler	**Açıklamalar**
ε	Mahremiyet maliyeti parametresi
δ	Hassasiyet
λ	Regülasyon katsayısı

Kısaltmalar	**Açıklamalar**
BDS	Bilgisayar destekli sistemler
BN	Batch normalizasyon
BOS	Beyin omurilik sıvısı
BT	Bilgisayarlı tomografi
CNN	Evrişimli sinir ağları (Convulational neural networks)
DICOM	Tıpta dijital görüntüleme ve iletişim (Digital imaging and communications in medicine)
DM	Diferansiyel mahremiyet
DNA	Deoksiribo nükleik asit
DP	Diferansiyel mahremiyet (Differential privacy)
DP-GAN	DM çekişmeli üretici ağ (DP generative adversarial networks)
DP-SGD	Differentially private stochastic gradient descent
FC	Tam bağlı (Fully connected)
FLAIR	Sıvı ile zayıflatılmış ters çevirme geri kazanımı (Fluid attenuated inversion recovery)
GPU	Grafik işlemci ünitesi (Graphics processing unit)
HGG	Yüksek dereceli gliom (High grade glioma)
LBS	Konum bazlı servis (Location-based service)
LGG	Düşük dereceli gliom (Low grade glioma)
LRP	Layer-wise relevance propagation

Kısaltmalar	Açıklamalar
LVQ	Öğrenmeli Vektör Kuantalama (Learning vector quantization)
MR	Manyetik rezonans
MRG	Manyetik rezonans görüntüleme
MS	Multiple skleroz
NIFTI	Neuroimaging informatics technology initiative
RNN	Yinelenen sinir ağı **(**Recurrent neural networks)
SGD	Stokastik dereceli azalma (Stochastic gradient descent)

1. GİRİŞ

Artan donanım gücü ve gelişen hesaplama teknikleri ile birlikte karmaşık hesaplamalar ve uzun eğitim süreci gerektiren makine öğrenmesi uygulamalarının kullanımı birçok alanda yaygınlaşmaktadır. Örneğin derin öğrenme algoritmalarının gelişimi görüntü işleme gibi daha karmaşık işlemler gerektiren problemlerin çözümünde kolaylık sağlamaktadır. Bunun yanında GPU'ların model eğitiminde kullanılmaya başlanması ile işlem gücü de artırılmıştır. Bu gelişmeler sayesinde makine öğrenmesi uygulamaları günlük hayatımızda yer almaya başlamıştır. Fakat gerçek uygulamalarda model başarımının yüksek olabilmesi için bunların yanında yüksek miktarda veriye ihtiyaç duyulmaktadır.

Model başarımı için ihtiyaç duyulan eğitim setleri farklı çeşitlerde ve yüksek boyutlarda birçok kamusal ve kişisel verinin toplanmasını gerektirir. Kişisel mahremiyeti ihlal edebilecek yüksek miktarda veri toplanması makine öğrenmesi uygulamalarında veri mahremiyeti problemini ortaya çıkarmaktadır. Örneğin literatürde bazı saldırı teknikleri ile model üzerinden eğitim veri setine ulaşılabildiğini gösteren çalışmalar mevcuttur. Modellere karşı üyelik çıkarımı, model tersine çevirme, eğitim veri seti bozma, özellik çıkarımı gibi saldırı teknikleri geliştirilmiştir.

Dolayısı ile makine öğrenmesi uygulamalarında model başarımının yanında daha önce üzerinde pek durulmayan veri güvenliği sorunu üzerinde durulması da artık bir gereklidir. Özellikle hassas kişisel veriler içeren alanlarda bu uygulamaların yaygınlaşması ile daha güvenli hale getirilmeleri ile ilgili çalışmalar artmıştır.

Derin öğrenmenin yaygın kullanıldığı alanlardan biri de sağlık sektörüdür. Sağlık hizmetleri kişisel veri mahremiyetinde en hassas alanlardan biridir. Bu konuda gelişmiş ülkelerde yasa ve yönetmelikler yayınlanmış, mahremiyet göz önüne alınmadan geliştirilen uygulamalara maddi yaptırımlar getirilmiştir. Eğer sağlık hizmetlerinde derin öğrenme tabanlı bir bilgisayar destek sistemi tasarlanacak ise bu konuda mahremiyet ihlallerine karşı güvenlik önlemleri geliştirilmelidir. Derin öğrenmede oluşabilecek mahremiyet ihlallerine Bölüm 2'de değinilmiştir.

Literatürde güvenliğin nasıl sağlanacağı konusunda da çalışmalarda mevcuttur. Bu saldırılara karşı alınan önlemler arasında şifreleme, anonimleştirme gibi temel yaklaşımların yanı sıra günümüzde oldukça popüler olan diferansiyel mahremiyet gibi yaklaşımlar da önerilmektedir.

Diferansiyel mahremiyet temel olarak pertürbasyon yöntemine dayalı, eklenen gürültünün mahremiyeti ne kadar sağladığının ölçüsünü veren matematiksel bir tanımdır. Bu tanımı sağlayan gürültü fonksiyonlarının mahremiyeti belirtilen oranda koruyacağı garanti edilir. İlk olarak veri analizinde kullanılmış, fakat kısa sürede birçok alanda yer bulmuştur. Makine öğrenmesi alanında veri setlerinin korunması da uygulama alanlarından biridir. Diferansiyel mahremiyet kullanılarak eğitilen bazı modellere yapılan saldırıların mahremiyetsiz yöntemlere göre daha başarısız olduğunu gösteren çalışmalar mevcuttur.

Bu tez çalışması kapsamında makine öğrenmesi uygulamalarına karşı oluşabilecek tehditler incelenmiş ve literatürde kullanımı gittikçe yaygınlaşan diferansiyel mahremiyet ile veri güvenliğinin sağlanabilmesi için bir model prototipi önerilmiştir.

Çalışmanın temel gerekçesi şu şekilde açıklanabilir; derin öğrenme modelleri hesaplamaların ve algoritmaların gelişmesi ile giderek daha başarılı sonuçlar elde etmektedir. Bu da karmaşık hesaplamalar gerektiren tıbbi görüntü işleme uygulamalarında gelecekte daha yaygın bir şekilde tercih edileceğini göstermektedir. Sağlık alanında bilgisayar destek sistemlerde kullanılan veriler, kişisel veri olduğundan ve 6698 nolu KVK Kanunu ile koruma altına alındığından yüksek hassasiyete sahiptir. Bu nedenle derin öğrenmenin bu alanda güvenle kullanılabilmesi için literatürde bulunan mahremiyet ihlallerine karşı çözümler üretilmelidir. Bu çalışma kapsamında veri seti mahremiyetini sağlamak için çözüm olarak diferansiyel mahremiyet yöntemi tercih edilmiştir.

Belirtilen gerekçe ile uygulamada amaçlanan, Gazi Üniversitesi ve Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ile ortaklaşa geliştirilen Türk Beyin Projesi kapsamında kullanılan beyin MR görüntülerinin mahremiyetinin sağlanması için bir ön çalışma yapmak, beyin görüntülerini mahremiyet-fayda prensibine saygı göstererek eğitmek ve medikal alanda sıklıkla kullanılmaya başlanan diferansiyel mahremiyet yaklaşımını ilk kez beyin MR'larına uygulamak, mahremiyete saygı göstererek verilerden faydalanmak için uygun modelleri belirlemek ve modellerin başarımlarını test etmektir.

Çalışmada ele alınan problemler aşağıdaki gibi sıralanabilir;

- Türk Beyin Projesi kapsamında; kullanılan veri setlerinin güvenliği için kolay uygulanabilir bir yöntem bulmak, sağlık alanında kullanılan veriler mahremiyet açısından hassas veriler olduğu için veri setleri ve model mahremiyetinin sağlanmasına katkı sağlamak, kullanılan modellere yapılacak bir saldırı ile hastaların kişisel bilgileri ve sağlık durumlarını içeren birçok verinin sızdırılma riskinden dolayı model güvenliğinin sağlanmasına katkıda bulunmak,Verilerden maksimum faydanın elde edilmesi için uygun bir model belirlemek ve bundan sonra yapılacak diğer çalışmalara da örnek olacak modeller geliştirmektir.
- Derin öğrenme alanında geliştirilen diferansiyel mahremiyet korumalı modellerin büyük çoğunluğu CIFAR-10 ve MNIST gibi basit veri setlerinde eğitilerek başarımları test edilmiştir. Bu çalışmada ele alınan diğer problem medikal görüntü sınıflandırma gibi önemli bir alanda, diferansiyel mahremiyet korumalı bir modelin karmaşık veri yapısına sahip girdiler üzerindeki başarımını test etmektir. Bunun için BRAST 2020 veri seti kullanılmıştır.

Mahremiyetin sağlanması için diferansiyel mahremiyet tercih edilmesinin sebepleri;

- Şifreleme ve anonimleştirme gibi yöntemlerin dezavantajlarını gidermesi,
- Mahremiyet korumasında mahremiyet fayda gibi zor bir probleme çözüm üretmesi,
- Kolay uygulanabilir olması,
- Düşük maliyetli olması,
- Makine öğrenmesi modellerine karşı yapılan çıkarım saldırıları için yapısı gereği doğal bir koruma yöntemi olması,
- Veri setinden bağımsız olarak modele uygulanabilmesidir.

Genel olarak, LGG ve HGG dereceli tümörler derin öğrenme modeli ile sınıflandırılarak izlenecek tedavi yönteminde uzmanlara fikir verebilen bir model prototipi tasarlanmıştır. Oluşturulan modelde diferansiyel mahremiyeti sağlayabilmek için gradyan pertürbasyonuna dayalı bir optimizasyon yöntemi kullanılmıştır. Yüksek özellik çıkarma yeteneğinden dolayı evrişimli ağlar tercih edilmiştir. Ayrıca beyin tümörü sınıflandırmada birçok özellik dikkate alındığı için daha detaylı özelliklerin çıkarımının yapılabilmesi için çok ölçekli 3 boyutlu

görüntü işleyebilen bir model tercih edilmiştir. Çok ölçekli evrişimli sinir ağlarının kullanılması ile farklı evrişim katmanlarından elde edilen özelliklerin hem yüksek çözünürlüklü hem de anlamsal olarak güçlü özellikler çıkarabilmesi hedeflenmiştir. Farklı katmanlardan elde edilen özellik çıkarımları birleştirilerek tam bağlı yapay sinir ağlarında sınıflandırılmıştır.

Bu tez çalışması 5 bölümden oluşmaktadır. Tezin ikinci bölümünde mahremiyet kavramı ve sağlık açısından önemi değerlendirilmiş ve mahremiyet duyarlı sistemlere değinilmiştir. Bölümün devamında derin öğrenme modellerinde oluşabilecek mahremiyet ihlalleri incelenmiştir. Bölüm 3'te diferansiyel mahremiyet açıklanmış ve derin öğrenme açısından ele alınmıştır. Bölüm 4'te oluşturulan mahremiyet duyarlı modelin test edileceği beyin tümörü hastalığı açıklanmış ve beyin tümörü MR görüntüleri ile yapılmış derin öğrenme çalışmaları incelenmiştir. Bölümün devamında derin öğrenme ile ilgili genel kavramlar açıklandıktan sonra tez çalışması için geliştirilen uygulama adımları ve yapısı alt başlıklar halinde sunulmuş ve sonuçlara değinilmiştir. Bölüm 5'te çalışmanın sonuçları genel hatları ile değerlendirilmiştir.

2. SAĞLIKTA MAHREMİYET VE MAHREMİYET DUYARLI SİSTEMLER

Mahremiyet, bireylerin rahatsız edici unsurlardan ari olduğu fiziksel alan arzusu ile kendisi hakkındaki kişisel bilgileri açıklama zaman ve tarzını kontrol edebilme yetisini ifade etmektedir. Mahremiyet kavramı doğrudan kişilikle ilgilidir ve bireyin kişiliğini, bağımsızlığını, onurunu ve bütünlüğünü korumaktadır. Mahremiyet hakkı ise sırları gizli tutma hakkını ve onları ancak özel ve bilinçli bir iradeyle paylaşmayı ifade etmektedir [1]. Mahremiyet kavramının, başkasının iletişiminin yasa dışı denetlenmemesi veya kişisel bilgilerinin yasal olmayan yollarla işlenmemesi ya da başka bir kimseyle paylaşılmaması gibi özellikleri bulunmaktadır.

Mahremiyet kavramının mekânsal mahremiyet, kişi mahremiyeti ve veri mahremiyeti olmak üzere üç çeşidi bulunduğu kabul edilmektedir. Mekânsal mahremiyet kişiyi çevreleyen yakın fiziksel alanın korunmasını, kişi mahremiyeti kişiyi haksız müdahalelere karşı korumayı, veri mahremiyeti ise kişisel verilerin toplanma, saklanma, işlenme ve iletiminin nasıl yapılacağının veya yapılmayacağının kontrol edilmesini ifade etmektedir.

Kurumların çoğu veri sorumlusu olarak hizmet verdiği kişilere (müşteri, hasta, kullanıcı, firma vb.) ait çeşitli verileri toplamakta ve bunları depolamaktadır. Bu veriler içerisinde bireyi doğrudan veya dolaylı olarak tanımlayabilecek kişisel veriler de yer alabilmektedir. Veri sorumluları, görevlerini yerine getirmek ve muhataplarına daha iyi hizmet sunmak (model ve örüntü çıkarmak, planlamalar yapmak, politikalar oluşturmak, karar verme mekanizmalarını geliştirmek vb.) amacıyla verileri işlemektedir. Bu süreçler boyunca, hakkında veri toplanan bireyin kişilik haklarının korunması ve bu alanda güvenin tesis edilmesi gerekmektedir.

Veri sorumlularının elektronik ortamlarda veri mahremiyetini sağlarken aynı zamanda veriden sağlanan fayda oranını da göz önünde bulundurmaları gerekir. Literatürde kullanılan mevcut mahremiyet koruma yöntemlerinde veri faydası ve mahremiyet derecesi arasında ters orantı bulunmaktadır. Yüksek mahremiyet koruması uygulandığında başarım düşmekte, yüksek başarım elde etmek için mahremiyet seviyesinin düşük tutulması gerekmektedir. Dolayısıyla veri mahremiyeti, mahremiyet seviyesi ile veri faydası arasındaki en iyi dengeyi bulmaya çalışan zor bir problemdir. Geliştirilen uygulamalarda mahremiyetin teknik açıdan

sağlanabilmesi için kullanılacak mahremiyet koruma yönteminden bu dengeyi fayda ve mahremiyet açısından kabul edilebilir oranlarda sağlaması beklenmektedir.

Bu tez çalışmasına konu olan derin öğrenme uygulamaları büyük çaplı kişisel verilerin kullanıldığı önemli alanlardan biridir. Bu uygulamalarda veri mahremiyetinin sağlanması için birçok teknik geliştirilmiştir. Bölüm 3'de derin öğrenme uygulamalarında mahremiyet tehditleri ve korunma yöntemleri konusu incelenmiştir. Yapılan çalışmada ise zor bir problem olan fayda mahremiyet dengesini matematiksel olarak ifade edebilen diferansiyel mahremiyet yöntemi kullanılmıştır. Geliştirilen model ve mahremiyet koruması kişiler için en önemli bilgileri içeren sağlık verileri üzerinde uygulanmıştır.

2.1. Sağlık Hizmetlerinde Mahremiyetin Önemi

Sağlık hizmetlerinde mahremiyet, doktora verilen bilgiler, doktorun aldığı notlar, alınan ilaçlar gibi bilgiler ile diğer tüm kişisel bilgilerin gizli tutulması anlamına gelir. Dünya Tabipler Birliğine göre kişisel sağlık bilgisi, "bireyin kimliğini ortaya koyan fiziksel ve duygusal sağlığı ile ilgili kayıtlı tüm bilgi" olarak tanımlanmış ve bu bilgilerin korunması gerektiği belirtilmiştir [2].

Kişilerin sağlık bilgileri ve tıbbi kayıtları sağlık uzmanları ile hastane, klinik gibi tüm sağlık kuruluşları tarafından güvenli ve gizli tutulmalıdır. Ülkemizde hastanın mahremiyetinin korunması gerekliliği Sağlık Bakanlığı Hasta Hakları Yönetmeliğinde belirtilmiştir. Ayrıca kişisel verileri koruma kanunu kapsamında "Kimliği belirli veya belirlenebilir gerçek kişiye ilişkin her türlü bilgi" olarak tanımlanan kişisel veri tanımı sağlık bilgilerini de kapsamaktadır. Ayrıca kanuna göre, "Kişilerin ırkı, etnik kökeni, siyasi düşüncesi, felsefi inancı, dini, mezhebi veya diğer inançları, kılık ve kıyafeti, dernek, vakıf ya da sendika üyeliği, sağlığı, cinsel hayatı, ceza mahkûmiyeti ve güvenlik tedbirleriyle ilgili verileri ile biyometrik ve genetik verileri özel nitelikli kişisel veridir". Bu veriler kişinin açık rızası olmada işlenemez. Tanımdan da anlaşıldığı gibi sağlık bilgileri kritik önem taşıyan özel nitelikli kişisel veriler içindedir. Dolayısıyla sağlık verilerinin mahremiyetinin korunması önemli bir husustur.

Sağlık verilerine erişim sağlık hizmetinin sağlanması, tedavi ve araştırma sürecinin doğru işlemesi için önemli bir gerekliliktir. Bu verilere erişim sırasında hasta verilerinin

mahremiyetini sağlamak hem sağlık personeli hem de veriyi işleyen bilgi teknolojileri uygulamalarının etik ve yasal olarak görevidir. Hastaya ait tıbbi bilgilere erişimin hekim ve uygulama açısından mümkün olduğunca kolaylaştırılması gerekirken, diğer yandan bilginin korunması için tedbir alınması gerekmektedir.

Bilgisayar destekli sağlık uygulamaları gibi hizmet kalitesini yükselten dijital sistemlerin hızla gelişmesi ve entegrasyonu sağlık hizmetlerinin daha hızlı ve kaliteli verilmesini sağlamaktadır. Fakat bu uygulamalar veriye daha çok erişim gerektirdiği için kişisel mahremiyetin ihlal edilme riskini artırmaktadır. Bu da sağlık alanında kişisel verileri koruma politikalarının önemini artırmaktadır. Bu uygulamaların geliştirilmesi ve kullanılması sırasında bilgi güvenliğini sağlayacak teknik önlemlerin alınması ve sistemlerin siber saldırılara karşı korunaklı hale getirilmesi gerekmektedir. KVK Kanununa göre özel nitelikli verilerin işlenmesi ile ilgili madde veri sorumlusuna hukuki sorumluluk yüklemektedir. Kanuna göre veri sorumlusu, verilerin hukuka aykırı olarak işlenmesini, erişilmesini engellemek ve muhafazasını sağlamakla yükümlüdür [3]. Dolayısı ile sağlık alanındaki teknolojik uygulamalarda veri güvenliğini sağlayacak teknik tedbirlerin alınması etik olmanın yanında kanuni bir zorunluluktur.

Sağlık hizmetleri uygulamalarının sahip olması gereken güvenlik tedbirlerinin uygulanmasını zorunlu kılan özel yasalar da mevcuttur. Amerika'da geçerli olan Sağlık Sigortası Taşınabilirlik ve Sorumluluk Yasası'na göre [4];

- Uygulamalarda iki faktörlü kimlik doğrulama kullanımı,
- Kişisel verilerin cihazda toplanırken ve sunucuya aktarılırken şifrelenmesi,
- Kullanıcı oturumunun belli bir süre işlem yapılmaması durumunda sonlandırılması,
- Düzenli test ve güncellemelerin tanımlanması,
- Cihaz kaybolduğunda otomatik veri geri yüklemenin tanımlanması gerekmektedir.

Sağlık verilerinin paylaşıldığı uygulamalarda uygulama geliştiriciler bu yasaya uymak zorundadır. Örneğin; Google Fit ve HealthKit gibi büyük platformlar, kullanıcıların kişisel verilerini doktorlarıyla paylaşmalarına izin verdiği için buna uymak zorundadır.

Avrupa birliği üye devletleri arasındaki veri akışını koordine etmek için 2018 yılında kabul edilen Genel Koruma Yönetmeliğine göre AB'de ikamet edenlerin verilerini toplayan sağlık uygulamalarının bu yönetmeliğe uyması zorunludur. Buna göre uygulama geliştiriciler hasta verilerinin güvenliğini ve gizliliğini sağlamalıdır [4]. Yönetmeliğe göre;

- Kullanıcılar veri toplama ve işleme hakkında bilgilendirilmelidir,
- Kullanıcılar veri işleme nedenleri hakkında bilgilendirilmelidir,
- Veriler işlemeden önce kullanıcılardan izin alınmalıdır,
- Toplanan kayıtlar anonimleştirilmelidir,
- Mahremiyet ihlali oluştuğunda kullanıcılar bilgilendirilmelidir,
- Veri aktarımı güvenli bir şekilde sağlanmalıdır.

Bu yönetmeliğe uyulmadığında sağlık uygulaması geliştiricilere 20 milyon Avro'ya varan para cezası uygulanabilmektedir. Bu yönetmeliklerden görüleceği üzere bilgi güvenliği kapsamında sağlık verilerinin korunması gelişmiş toplumlar tarafından hassasiyet gösterilen önemli bir konudur.

Bir sağlık hizmeti uygulamasında güvenliğin sağlanması aşağıdaki gibi gruplandırılabilir [5];

- Cihaz anonimliği: Bir tıbbi cihazın kimliğinin sistem tarafından bilinmemesi gerekir, yetkisiz kişiler cihaz türü, IP ve MAC adresleri gibi cihaz kimliklerini ele geçirememelidir.
- Veri güvenliği: Yetkisiz kullanıcıların bir kullanıcıyı ve kullanıcının hassas verilerini tanımlamasını önlemek gerekmektedir.
- İletişim anonimliği: Yetkisiz kişiler, kullanıcı ile sistem arasındaki bağlantıyı tanımlayamamalıdır.
- Bağlantısızlık: Gönderen ile alıcı arasındaki veri işlemlerini izleyen bir saldırgan, veri ile gönderen arasında bir ilişki kuramamalıdır.

Sağlık alanında, özellikle tıbbi görüntülemede karmaşık hesaplamaları yapabilme kabiliyeti sebebi ile derin öğrenme modelleri bilgisayarlı destek sistemlerinde sıkça kullanılmaya başlanmış, uzmanların karar verme ve teşhis koymalarına yardımcı olabilecek daha hızlı

hizmet veren uygulamalar geliştirilmeye çalışılmaktadır. Sağlık alanında makine öğrenmesi uygulamalarının kullanılmaya başlaması ile bu uygulamaların veri mahremiyetini ne ölçüde sağladığı daha çok üstünde durulması gereken bir konu olmuştur. Çünkü sağlık verileri en kritik kişisel verilerdendir. Bu nedenle derin öğrenme modellerinin sağlık alanında daha da yaygınlaşması ile bu uygulamaların güvenli olması önem taşımaktadır. Bölümün devamında bu uygulamalarda karşılaşılabilecek mahremiyet ihlalleri incelenmiştir.

2.2. Derin Öğrenme Mahremiyet İhlalleri

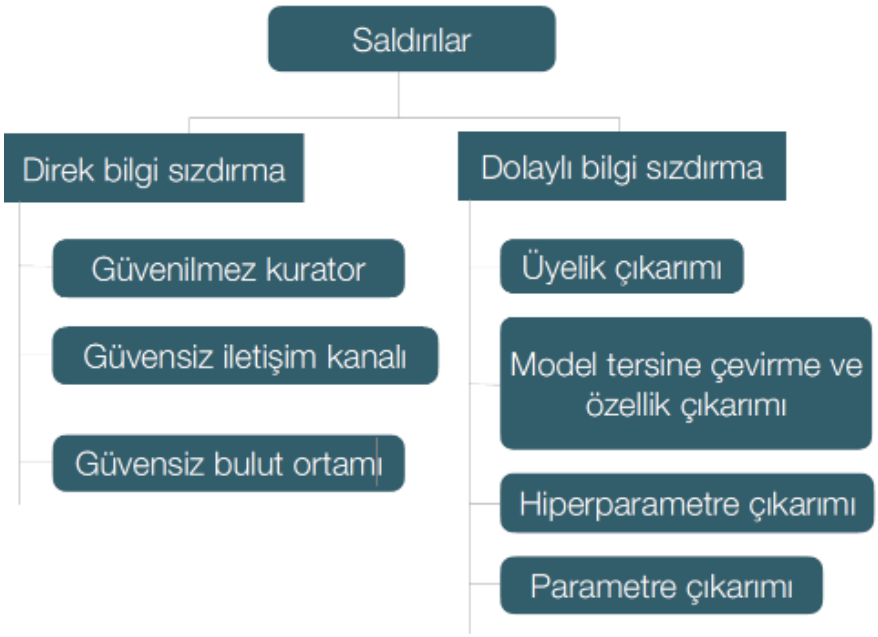
Derin öğrenme uygulamalarında model yapısı ile ilgili edinilen bilgiler veya veri seti hakkındaki ek bilgiler kullanılarak uygulanan saldırılar ile mahremiyet ihlalleri oluşabilir. Derin öğrenme modeline yapılan saldırılar saldırganın sahip olduğu bilgi miktarına göre kara kutu ve beyaz kutu saldırıları olarak ikiye ayrılır. Kara kutu yöntemlerinde saldırgan istemci modele tahmin sorguları göndererek elde ettiği sonuçlara göre çıkarım yapmaya çalışır. Beyaz kutu saldırılarında ise istemci model tanımını veya bazı model parametrelerini elde eder ve bu bilgileri kullanarak bir saldırı tekniği oluşturur. Literatürde beyaz kutu ve kara kutu için tasarlanmış saldırılar mevcuttur.

Derin öğrenme modellerinde bulunan zafiyetler erişim şekline göre direk ve dolaylı saldırılar olarak da ikiye ayrılabilir. Direk bilgi sızdırma zafiyetlerinde, gerçek bilgiye erişilerek veri ele geçirilir. Dolaylı bilgi zafiyetlerinde ise dış kaynaklardan toplanan bilgiler ile tahmin veya çıkarım yapılarak gerçek bilgiye erişilmeye çalışılır, gerçek bilgiye direk erişim yoktur. Makine öğrenmesi alanında oluşabilecek zafiyet çeşitleri [6] Şekil 2.1'de gösterilmiş ve alt başlıklarda şekilde verilen hususlar kısaca açıklanmıştır.

2.2.1. Direk bilgi sızdırma

Direk bilgi sızdırma zafiyetleri sadece derin öğrenme alanında değil genel veri güvenliği kapsamında oluşabilecek ihlallerdir. Hangi amaçla olursa olsun bütün veriler bu tarz yollarla sızdırılabilir. Bu saldırılarda kişi sisteme direk erişim sağlayarak verinin kendisine ulaşır. Mcafee tarafından yapılan bir çalışmada [7], kasti olmayan veri sızıntılarının yaklaşık %43'üne çalışanların sebep olduğu belirtilmiştir. Kurum dışından olan saldırganlar ise kimlik doğrulama aşamasını atlayarak, sistemin açık arka kapısını bulup sisteme sızarak veriye, modele veya parametrelere direk erişim sağlayabilir [7,8].

Şifresiz veya güvenilir bir şifreleme tekniği kullanmadan veri aktarımı yapmak, iletişim kanalı üzerinde oluşabilecek zafiyetlere sebep olabilir. Saldırganlar iletişim kanalına müdahale ederek verileri ele geçirebilir. Kaspersky Labs 2018 yılında yayınladığı bir raporda dört milyondan fazla Android uygulamasının şifresiz bir şekilde kullanıcı bilgi ve profillerini reklam sağlayıcılara aktardığını açıklamıştır [9].



Şekil 2.1. Derin öğrenme uygulamalarındaki tehditlerin sınıflandırılması [6]

Bulut sunucularda ise, servis olarak (as a service) kullanılan bazı derin öğrenme uygulamaları kullanıcılardan topladığı veriler ile ilgili işlemler bittiğinde veriye ne olacağı konusunda bilgi vermemektedir. Birçok uygulamada, bu verilerin başka bir şekilde kullanılıp kullanılmayacağı veya işlem bittikten sonra imha edilip edilmeyeceği hakkında bilgi yoktur. Dolayısıyla kullanıcıların mahremiyetlerinin nasıl korunduğuna dair bilgilendirilmemesi bu verileri elinde tutan kişilere karşı önyargı oluşturmaktadır.

2.2.2. Dolaylı bilgi sızdırma

Dolaylı zafiyetler, Şekil 2.1'de görüldüğü gibi üyelik çıkarımı, hiper-parametre çıkarımı, parametre çıkarımı, özellik çıkarımı ve modeli tersine çevirme olarak beş kategoriye ayrılmıştır. Bu ihlaller için hem beyaz kutu hem kara kutu yöntemler bulunmaktadır.

Üyelik çıkarımı

Eğitim seti üyelik çıkarımında, elde edilen bir veri örneğinin hedef modelin eğitim setinde kullanılıp kullanılmadığı öğrenilmeye çalışılır. Üyelik çıkarımı zafiyetinden ilk kez Shokri ve diğerleri [10] bahsetmiştir. Çalışmada, saldırganın hedef modele erişebilecek kara kutu sorgusuna sahip olduğu ve sorgulanan girdiler için güven skoru (olasılık vektörü) oluşturabildiği varsayılmıştır. Elde edilen verilerin hedef model eğitimine katılımı güven skoru ile ölçülür. Modelin eğitiminde kullanılan verinin, model tarafından daha önce görülmemiş bir veriden daha yüksek bir güven puanı alacağı varsayılır.

Üyelik çıkarımı saldırısında ilk olarak hedef modeli taklit eden gölge modeller oluşturulur [10]. Gölge modellerin eğitiminde tersine model çevirme, istatistik tabanlı sentez (eğitim setinin dağılımı hakkında yapılan varsayımlar ile) veya gerçek veriye gürültü ekleyerek elde edilen etiketlenmiş veri setleri kullanılır. Gölge modeller kullanılarak, verinin eğitim setinde bulunup bulunmadığını ayırt eden bir "saldırı modeli" geliştirilir. Saldırı modeli hedef modelin tahminlerini sınıflandıran bir sınıflandırıcı olarak düşünülebilir. Sonra hedef model sorgulanarak verilen her bir girdi verisi örneği için güven puanı hesaplanır ve girdinin hedef modelin eğitim setinin bir parçası olup olmadığına karar verilir.

Long ve diğerleri [11] belirli bir örneğin üyeliğini, eğitim veri seti içinde bulunup bulunmadığını, daha doğru şekilde test eden bir yaklaşım önermektedir. Gölge modeller hem verilen veri örneği ile hem de bu örnek olmadan ayrı ayrı eğitilip ardından Shokri ve diğerleri yaklaşımına benzer şekilde verinin hedef modeli eğitmek için kullanılıp kullanılmadığını test eder. Salem ve diğerleri [12] önceki çalışmalarda kullanılan birden çok gölge model eğitimi, hedef model yapısı hakkında bilgi sahibi olma ve hedef modelin eğitim verileriyle aynı dağılıma benzer bir veri kümesi elde etmek gibi gereksinimleri azaltabilecek daha genel bir teknik önermiştir. Bu tür yöntemlerin etkinliklerini düşürmeden daha düşük bir maliyetle de uygulanabileceği gösterilmiştir.

Yeom ve diğerleri [13], hedef modele beyaz kutu erişimin olduğu ve modelin ortalama eğitim kaybının bilindiği durumlar için bir üyelik çıkarım modeli önermektedir. Bu çalışmada, bir girdi verisi için model kaybı değerlendirilir ve eğer kayıp eşik değerden (eğitim setindeki ortalama kayıp) küçükse, girdi eğitim setinin bir parçası olarak kabul edilir.

Böylece gölge model yerine lineer bir sınıflandırıcının kaybı kullanılarak gölge model eğitimine gerek kalmamıştır.

Model ters çevirme ve öznitelik çıkarımı

Modelin tahminlerinden eğitim verileri hakkında bilgi edinmeyi amaçlayan model ters çevirme metodu hakkında birçok çok araştırma çalışması yapılmıştır. Araştırmalar incelendiğinde iki ana yaklaşım görülmektedir. Bunlardan birincisi, gradyan tabanlı optimizasyon kullanarak model ters çevirme yapar [14-16]. Ters çevirme işlemi, belirli bir sınıfa ait en uygun verileri bulmak için optimizasyon problemi olarak görülür. İkinci yaklaşımda ise, orijinal modelin tersi olarak hareket eden ikinci bir model eğitilerek model tersine çevrilir [17,18]. Eğitim tabanlı yaklaşımda, sinir ağının her katmanındaki özelliklerden görüntülerin yeniden oluşturulması hedeflenir. Görüntülerin maksimum düzeyde yeniden yapılandırılması için hedef modelin tahmin vektörleri, ikinci modeli eğitmek için kullanılır. Hedef modelin çıktıları saldırı modelinin girdisi olarak kullanılarak tersine öğrenme gerçekleştirilir.

Fredrikson ve diğerleri [15] kamuya açık yayınlanmış ilaç dozajı tahmini için kullanılan bir lineer regresyon modeli üzerinde model tersine çevirme saldırısı uygulaması içeren bir çalışma yapmıştır. Model çıktısı ve diğer önemsiz nitelikler (Örn. Boy, yaş, kilo) kullanılarak hasta hakkındaki genomik bilgilere ulaşılmıştır. Bu saldırıyı gerçekleştirmek için hedef modele kara kutu erişimi olduğu varsayılmaktadır.

Fredrikson ve diğerleri [19]'nin yaptığı bir başka çalışmada yapay sinir ağı modeline beyaz kutu erişimi olduğu varsayılarak, elde edilen model tahminlerinden eğitim verilerinin örneklerinin çıkarabileceği gösterilmektedir.

Yang ve diğerleri [20] ise eğitim temelli bir yaklaşım kullanmışlardır. Önceki çalışmalardan farklı olarak ikinci modeli eğitmek için kullanılan veri setinin orijinal veri ile aynı dağılımdan olması gerekmez. Çalışmada, kolayca elde edilebilecek arka plan bilgileri ile daha genel bir veri dağılımı kullanılarak elde edilen veriler kullanılmıştır. Örneğin, bir yüz tanıma sınıflandırıcısına karşı eğitim verilerinin dağılımı bilinmeden eğitim seti oluşturmak için internetten bulunan rastgele yüz görüntüleri kullanılabilir. Şekil 2.2'de çalışmadan alınan geri çevrilmiş bir örnek bulunmaktadır. Sağdaki eğitim veri setinden orijinal bir veri

iken solda orijinal sınıflandırıcının tahminini girdi olarak alan ters çevirme modelinden elde edilmiş yeniden yapılandırılmış resim bulunmaktadır.



Şekil 2.2. Yang ve diğerleri çalışmalarından alınan bir sonuç örneği [20]

Yeom ve diğerleri [21] üyelik çıkarım saldırısı için kullanılan yönteme dayanan bir öznitelik çıkarım saldırısı önermektedir. Farklı hassas öznitelik değerleri olan girdi verileri üzerinde model kaybı değerlendirilir. Orijinal veriler tarafından üretilen kayıp değerine benzer bir kayıp değeri veren girdinin seçilen özelliği hassas öznitelik olarak belirlenir. Salem ve diğerleri [22] çevrimiçi öğrenme için modelin gradyan güncellemesi öncesi ve sonrası durumuna bakan bir üretici bir ağ modeli geliştirmiştir.

He ve diğerleri [23], derin öğrenme ağının bölünerek farklı sunuculara dağıtıldığı işbirlikçi derin öğrenme sistemlerinde, test çıkarım sorgularını ortaya çıkarmak için yeni bir saldırı önermektedir. Çalışmada, sistemden veri sızdırmak isteyen bir sunucunun, diğer sunucuların verilerine veya hesaplamalarına erişim sağlayamamasına rağmen sistemi besleyen genel girdileri ele geçirebileceği gösterilmiştir.

Parametre ve hiper-parametre çıkarımı

Model parametrelerinin açığa çıkarılması eğitim verilerinin ele geçirilmesine neden olabilir. Model çalma saldırısında model parametreleri kara kutu erişimiyle ele geçirilmeye çalışılır.

Tramer ve diğerleri [24], tahminler ve güven skoru yoluyla model parametrelerini bulan bir saldırı yöntemi geliştirmiştir. Çalışma, girdi-çıktı çiftlerine dayanan denklem çözme yoluyla modelin parametrelerini bulmaya çalışır. Hiper-parametre çalma saldırıları, regülasyon

katsayısı [25] veya model mimarisi [26] gibi model eğitimi sırasında kullanılan hiper-parametreleri bulmaya çalışır.

Özellik çıkarımı

Bu saldırı türünde, hedef modelden belirli bilgi örüntüleri çıkarılmaya çalışılır. Bu saldırıların en önemli örneği, hedef modelin eğitim verilerinde hassas kalıplar bulmayı amaçlayan ezberleme saldırısıdır [23]. Bu saldırılar, Gizli Markov Modelleri (HMM) ile destek vektör makineleri (SVM) [27] ve sinir ağlarında [28] uygulanmıştır.

2.3. Mahremiyet Duyarlı Sistemler

Literatürde veri mahremiyeti konusunda pek çok çalışma bulunmaktadır. Genel olarak, mahremiyet koruması için geliştirilen sistemler şifreleme, anonimleştirme ve gürültü ekleme olarak üç başlık altında toplanabilir.

Şifreleme

Mahremiyeti korumak için veriye dışardan erişimi engellemek amacıyla kullanılan, hassas verinin gizlilik ve bütünlük ilkelerini korumak için yüksek seviyeli veri mahremiyeti sunan bir yaklaşımdır. Herhangi bir veri mahremiyet faydası oranı hesaplanamaması ve şifreleme tekniklerinin veri boyutu arttıkça oluşan yüksek maliyeti yaklaşımın dezavantajlarıdır [29].

Anonimleştirme

Anonimleştirme ise; mahremiyet seviyesi ile veri faydası arasında denge kurmayı sağlayan fayda temelli yaklaşımdır. Genelleştirme, baskılama, anatomi ve pertürbasyon gibi çeşitli operatörlerden yararlanarak paylaşılan veriyi ifşa saldırılarına karşı korur [30,31].

Literatürde bu konuda pek çok yöntem önerilmiştir. k-Anonimlik; bir verinin en az k-1 tane veriden ayırt edilememesini sağlayan mahremiyet koruma modelidir. Kimliği ifşa edilmek istenen kişinin yani kurbanın yarı tanımlayıcılarının değeri bilinse bile, o kişinin kaydı diğer k-1 tane kayıttan ayırt edilemez [32]. Bu şekilde kayıt bağlama saldırısı olarak da bilinen kimlik saldırısı engellenmiş olur. İlk bakışta basit bir problem olarak görünmesine rağmen

optimum k-Anonimliği sağlamanın NP-Zor bir problem olduğu çeşitli çalışmalarda [33] ispatlanmış ve gerek optimal gerekse de yakın optimal çözümler geliştirilmeye çalışılmıştır.

l-Çeşitlilik; k-Anonimlik, kimlik ifşası saldırılarına karşı koruma sağlarken hassas verilerin ifşasına karşı bir koruma sağlayamaz. Machanavajjhala ve ark. [34] k-Anonimlik modelinin bu sorununu vurgulayarak hassas özniteliklerin de mahremiyetini koruyan l-Çeşitlilik modelini önermiştir. Bu model, hassas verilerin ifşa edilmesini mümkün olabildiği kadar engellemek amacıyla hassas verilerin çeşitliliğinin en az l sayıda olmasını garanti eder.

t-Yakınlık, l-Çeşitlilik güçlü bir mahremiyet modeli olmasına rağmen, Li ve ark. [35] çarpık veri dağılımına sahip veri kümelerinde l-Çeşitlilik modelinin yetersiz olduğunu göstermiş ve t-Yakınlık modelini önermiştir. l-Çeşitlilik, hassas değerler arasındaki anlamsal yakınlıklara ve hassas değerlerin dağılımının genel dağılımdan önemli ölçüde farklı olmasına bağlı olarak yapılabilecek muhtemel çarpıklık saldırılarına karşı mahremiyet korumasında yetersiz kalır.

Arka plan bilgisine sahip bir saldırgan üyelik bilgisini de kullanarak veri bağlama yöntemleriyle yapacağı saldırılar ile kimlik ifşası gerçekleştirebilir. k-Anonimlik, l-Çeşitlilik ve t-Yakınlık modelleri kimlik ve öznitelik ifşalarına karşı koruma sağlarken üyelik ifşalarına karşı koruma sağlayamaz. Bu geleneksel yaklaşımların en önemli zafiyeti arka plan bilgisi saldırılarına karşı tam bir koruma sağlayamamalarıdır. Ancak diferansiyel mahremiyet modeli bu problemi gidererek yüksek mahremiyet garantisi sunar.

Gürültü ekleme

Veri mahremiyetini korumak için kullanılan diğer bir yaklaşım da veriye gürültü eklemek veya veri setini rastgele bozmaktır. Fakat bu teknik sezgisel olup eklenen gürültünün fayda mahremiyet analizi yapılamamaktadır. Gürültü ekleme veya pertürbasyon yöntemlerinde fayda mahremiyet analizi yapabilmeyi sağlayan diferansiyel mahremiyet yöntemi önerilmiştir.

Bu bölümde, artan veri boyutu, gelişmiş algoritmalar ve artan donanım gücü ile birlikte uygulama alanı genişleyen derin öğrenme modellerine karşı yapılan mahremiyet ihlalleri incelenmiştir. Ayrıca tez çalışmasında veri seti mahremiyetini sağlamak için kullanılan

diferansiyel mahremiyet yönteminin derin öğrenme uygulamalarında mahremiyeti nasıl sağladığı ve kullanım yöntemleri açıklanmıştır.

3. DİFERANSİYEL MAHREMİYET

Yapılan çalışmada zor bir problem olan mahremiyet fayda dengesine çözüm önermesi, kolay uygulanabilir olması, arka plan saldırılarına karşı diğer yaklaşımların önüne geçmesi sebepleri ile diferansiyel mahremiyet yöntemi tercih edilmiştir. Bu bölümde DM temel kavramları açıklanmış ve derin öğrenme uygulamalarında kullanımı ele alınmıştır.

Diferansiyel mahremiyet, mahremiyeti sağlarken aynı zamanda veri faydasını da göz önünde bulunduran, fayda-mahremiyet dengesini göz önünde bulundurarak önerilmiş bir yöntemdir. Gizlilik tanımını matematiksel olarak ifade ederek kullanıcıların gizliliğinin sağlandığını kanıtlamaktadır. Doğruluk ve tolare edilen mahremiyet kaybının ölçülmesini sağlar. Kişinin verilerini bir analizde kullanılmasına izin verdiğinde arka plan verileri ile yapılan saldırılardan etkilenmeyeceğini savunur. Doğru bir şekilde uygulandığında DM bireylerin mahremiyetini korurken belli bir popülasyonun özelliklerinin ortaya çıkarılmasına, toplu analizler yapılmasına olanak sağlar.

Diferansiyel mahremiyet kavramı Cynthia Dwork tarafından 2006 yılında tanıtılmıştır. Günümüze kadar DM tanımını ve kullanım alanını genişletmek için birçok çalışma yapılmıştır. İlk çalışmalar daha çok veri mahremiyetini sağlamak için gerekli gürültü miktarını hesaplamaya yönelik çalışmalardır. Bu çalışmalar sonucunda Laplace mekanizması tanımlanmıştır. [36,37] de yapılan çalışmalarda ise DM'yi sağlamak için eklenecek gürültü miktarını sınırlandırılması ele alınmıştır. Gürültü ekleme ile ilişkili olarak ε parametresinin en iyi şekilde nasıl seçilebileceğine yönelik çalışmalar da bulunmaktadır [38].

Diferansiyel mahremiyetle ilgili en kapsamlı çalışma Dwork ve Roth tarafından yayınlanan "The Algorithmic Foundations of Differential Privacy" kitabıdır [39]. Fakat orijinal DM tanımı komşu veri tabanları tanımı ile sınırlandırılmıştır. İlk yapılan tanım veri analiz sonuçlarına gürültü eklemek için ortaya atılmıştır. DM günümüzde birçok uygulama alanında kullanılmaktadır [40]. Kişiselleştirilmiş online reklamcılık, derin öğrenme, dağıtık zaman serileri verilerinin toplanması, blok zincir gibi uygulama alanları bulunmaktadır. Bu çalışmada DM derin öğrenme alanında kullanımı açısından ele alınmıştır. DM'nin derin öğrenme ile ilişkisi temel kavramlar bölümünden sonra açıklanmıştır.

3.1. Temel Kavramlar

Diferansiyel mahremiyet genel olarak pertürbasyon yöntemini kullanarak veri analizinde mahremiyet fayda oranı hesaplanabilen uygulamalar oluşturulmasına olanak sağlamaktadır. DM bir algoritma değil, mahremiyet fayda dengesini yönetmek ve ölçmek için kullanılan matematiksel bir tanımdır. Mahremiyeti korumak için bu tanımı sağlayan birçok algoritma kullanılabilir. DM bu algoritmaların gizlilik kayıplarını karşılaştırarak sabit doğrulukta hangisinin daha güvenli olduğunu kıyaslayabilir. Ayrıca pertürbasyon temelli olduğu için şifreleme ve anonimleştirme gibi koruma modellerinin zayıf kaldığı arka plan saldırılarına karşı kullanılabilir.

Bu bölümde DM konsepti için kullanılan terimler, diferansiyel mahremiyetin temel yapısı ve kavramları açıklanmıştır.

3.1.2. Küratör

Veriye sahip olan ve direk erişim sağlayan, M mekanizmasını kontrol edebilen kişidir. Global diferansiyel mahremiyet yapısında küratörün güvenilir olduğu varsayılır. Küratöre olan bağımlılık merkezi veri kaynakları için geçerlidir. Küratörün güvenilir olmadığı durumlar için lokal diferansiyel mahremiyet uygulamaları geliştirilmiştir. Bu uygulamalarda mahremiyet koruma yöntemi veri sahibinden veri toplanırken uygulandığı için küratör de gerçek veri hakkında bilgi sahibi olamaz.

3.1.3. Komşu veri tabanları

Sadece bir veri kaydı farklı olan veri tabanlarıdır. Diferansiyel mahremiyet kavramı bu komşu veri tabanları arasındaki farkın çıktı sonuçlarını önemli ölçüde değiştirmeyeceği fikrine dayanır.

3.1.4. Mekanizma

Veri kaynağı ile kullanıcı arasında bulunan bir ara yüzdür. Kullanıcıdan gelen sorguları veri kaynağına, uygulamadan gelen sonuçları da diferansiyel mahremiyet tanımını sağlayacak

şekilde gürültü ekleyerek kullanıcılara iletir. Böylelikle kullanıcı elindeki sonuçlardan veri kaynağı hakkında net bir bilgi edinememiş olur [39].

3.1.5. Mahremiyet maliyeti parametresi (ε)

Mahremiyet maliyeti parametresi, DM kullanılacak uygulamanın yapısına ve istenen gizlilik seviyesine göre seçilir ve diferansiyel mahremiyeti sağlayan mekanizmalar üretmek için kullanılır. Mekanizma, sorgu çıktısına mahremiyet maliyeti parametresine bağlı olarak rastgele gürültü ekler veya veri setindeki verileri rastgele değiştirir. Sonuca ne kadar çok gürültü eklenirse o kadar yüksek mahremiyet sağlanır, fakat yüksek gürültü sonuçların doğruluğunu düşürür. Dolayısıyla gizlilik ve doğruluk arasında dengeyi sağlayacak mekanizmaların kullanılması gerekir [5].

Diferansiyel mahremiyet tanımının sağlanması için kullanılan temel parametredir. Analizin doğruluğu ile mahremiyet seviyesi arasındaki dengeyi sağlamak için kullanılır. Uygulamada mahremiyetten ne kadar taviz verildiği bu parametre ile ölçülür. Diferansiyel mahremiyeti sağlayan mekanizmalar bu parametreye bağlı olarak pertürbasyon yapar. Dolayısıyla DM mekanizmalarında epsilon seçimi önemli bir faktördür. Epsilon büyüdükçe eklenen gürültü miktarı azalır, doğruluk oranı artar. Literatürde bu parametrenin nasıl seçilmesi gerektiğine dair çalışmalar [41] bulunsa da ε seçimi için kesin bir yöntem üzerinde uzlaşılmamıştır. Genellikle sezgisel olarak uygulama performansına göre seçilmektedir.

3.1.6. Hassasiyet

Komşu veri tabanları üzerinden alınan sorgu cevapları arasında en büyük fark değeridir. Örneğin, histogram hesaplayan bir fonksiyonun hassasiyeti 1'dir. Çünkü bir verinin eksik veya fazla olması histogramda en fazla bir birimlik değişim oluşturur. Hassasiyet mekanizmaların gürültü eklemek için kullandığı diğer bir parametredir. Hassasiyet ile eklenen gürültü miktarı doğru orantılıdır.

3.1.7. (ε, δ)-diferansiyel mahremiyet

DM'nin en yaygın kullanılan yaklaşımıdır. Bu yaklaşımda tanımlanan δ parametresi DM nin başarısız olmasına izin verilen olasılığın üst sınırını ifade eder.

K fonksiyonu rastgele gürültü ekleyen DM mekanizması; S olası çıktılar uzayı, epsilon mahremiyet maliyeti parametresi, D ve D' komşu veri tabanları olmak üzere; DM'nin sağlanması için seçilen mekanizma aşağıdaki eşitsizliği sağlamalıdır. Bu koşul sağlanıyor ise K mekanizması D, D' komşu veri tabanları için (ε, δ)-diferansiyel mahremiyetini sağlar denir;

$$\Pr[K(D) \in S] \leq \exp(\varepsilon)\, xPr\left[K\left(\acute{D}\right) \in S\right] + \delta \tag{3.1}$$

ε-diferansiyel mahremiyet ise (ε,0)-diferansiyel mahremiyetin kısaltması olarak kullanılır ve aşağıdaki şekilde ifade edilir;

$$\Pr[K(D) \in S] \leq \exp(\varepsilon)\, xPr\left[K\left(\acute{D}\right) \in S\right] \tag{3.2}$$

Mahremiyetin sağlanması için K fonksiyonunun, D ve D' den gelen sonuçları birbirine benzer olacak şekilde üretmesi gerekir. Örneğin ε'un sıfır olması durumunda, iki veri setinden gelen sonuçlardan birinin %50 seçilme olasılığı vardır. Böylece saldırgan ele geçirdiği arka plan verileri ile veri tabanında bulunan diğer tüm kişilerin bilgilerini bilse bile istediği kişinin veri tabanında olup olmadığını kesin olarak kestiremez. Fakat küçük ε değeri sonuçların doğruluğunu çok fazla bozabileceği için, eklenen gürültü ve doğruluk arasında istenen dengeyi sağlayacak oranda seçilmelidir. Kısaca mekanizma seçerken aşağıdaki koşulların göz önünde bulundurulması gerekir;

- Seçilen mekanizmanın aşağıdaki eşitsizliği sağlaması,

$$\frac{\Pr\left(K(D)\right)}{\Pr\left(K(\acute{D})\right)} \leq \exp\left(\varepsilon\right) \tag{3.3}$$

- Eklenen gürültüden sonra algoritmanın doğruluk oranı ölçümlerinin dikkate alınması gerekir.

DM'nin birçok alanda uygulanmaya başlanması ile birlikte bu koşulları sağlayan birçok mekanizma geliştirilmiştir. Laplace mekanizması [42] istatistiksel sorgular için geliştirilmiş ilk mekanizmadır. Literatürde, Üstel mekanizma [43], Medyan mekanizması [44],

Çarpımsal Ağırlıklar mekanizması [37], Geometrik mekanizma [45], Merdiven mekanizması [46], Gauss mekanizması [39] gibi mekanizmalar kullanılmaktadır.

3.1.8. Laplace mekanizması

Laplace mekanizması f sorgu fonksiyonunu hesaplayarak her koordinatına laplace dağılımına göre rastgele gürültü eklemektedir. Eklenen gürültü sorgunun hassasiyetinin ε'a oranı ölçeğinde kalibre edilir. Örneğin 1/ε ölçeği için $Lap(1/\varepsilon)$ dağılımı gürültü olarak eklenir. Laplace dağılımı üstel dağılımın simetrik şeklidir ve sorgu sonucuna devamlı olarak simetrik gürültü ekler [39].

Laplace dağılımı aşağıdaki gibi hesaplanır:

$$Lap(x|b)\frac{1}{2b}exp(-\frac{|x|}{b}) \tag{3.4}$$

$\mathbb{N}^{|x|}\rightarrow\mathbb{R}^{k}$ de tanımlı f fonksiyonu için Laplace mekanizması aşağıdaki şekilde ifade edilir;

$$M_L(x, f(.), \varepsilon) = f(x) + (Y_1, \ldots, Y_k) \tag{3.5}$$

Y_i, Laplace dağılımından üretilen i. Rastgele değerleri ifade etmektedir.

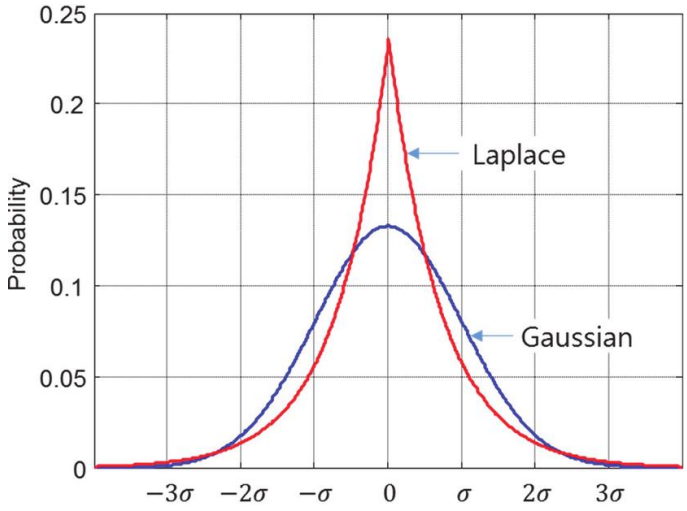
3.1.9. Gauss mekanizması

Gauss mekanizması Laplace mekanizmasına alternatif olarak geliştirilmiştir. Mekanizma pertürbasyon yaparken Laplace gürültüsü yerine Gauss gürültüsü ekler. Gauss mekanizması saf ε-diferansiyel mahremiyeti sağlamaz fakat (ε, δ)-diferansiyel mahremiyeti sağlamaktadır. Gauss gürültüsü Laplace gürültüsüne benzer şekilde eklenir. f: $\mathbb{N}^{|x|}\rightarrow\mathbb{R}^{k}$ de tanımlı sorgu fonksiyonu, F: mekanizma, s: fonksiyon hassasiyeti, $\mathcal{N}(\sigma^2)$: 0 merkezli σ^2 varyanslı Gauss dağılımı örneği olmak üzere Gaus mekanizması aşağıdaki şekilde ifade edilebilir [5];

$$F(x) = f(x) + \mathcal{N}(\sigma^2) \tag{3.6}$$

$$\sigma^2 = \frac{2s^2 \log(1{,}25/\delta)}{\varepsilon^2} \tag{3.7}$$

Laplace gürültüsü ile karşılaştırıldığında Şekil 3.1'deki grafikte Gauss gürültüsünün dağılımı daha basık ve yayılmış olarak görünmektedir. Dolayısıyla eklenebilecek gürültü aralığı daha geniştir. Bu durumda diferansiyel mahremiyet için Gauss gürültüsü kullanmak girdiye daha çeşitli gürültü örneği ekleme olanağı sağlayarak çıktının gerçekten daha çok uzaklaşmasını sağlar.



Şekil 3.1. Laplace ve Gauss dağılımları

Grafikte $\varepsilon = 1$ ve $\delta = 10^{-5}$ seçilerek hesaplanan Laplace ve Gauss dağılımları görünmektedir. ε-diferansiyel mahremiyeti sağlayamaması ve daha düşük doğruluk oranına sahip sonuçlar üretmesi Gauss mekanizmasının dezavantajlarıdır.

Laplace ve Gauss dağılımları $f: D \to \mathbb{R}^k$ vektör uzaylarında çıktı üretebilir. Örneğin histogramlar vektör uzay değerli fonksiyonlardır. Fakat iki dağılım arasında bir fark bulunmaktadır. Vektör değerli Laplace mekanizması L1 hassasiyete izin verirken, Gauss mekanizması hem L1 hem de L2 hassasiyete izin verir. L2 hassasiyet L1 hassasiyetten daha küçük değerler ürettiği için Gauss mekanizması daha küçük değerli gürültüler ekleyebilmektedir. Bu yönüyle Gauss mekanizması Laplace mekanizmasına göre daha avantajlı sayılmaktadır [47].

3.1.10. DM bileşik mekanizmalar

Bazı durumlarda sorgu basit bir istatistiksel veri tabanı sorgusu olmayabilir. Bu durumda problemin yapısına göre mahremiyetli sonuç birden çok mekanizma kullanılarak üretilmeye çalışılır. Bu bileşik/kompozit teori olarak adlandırılır [46-48]. Sistem için belirlenen ε parametresi bu algoritmalara paralel veya sıralı olarak dağıtılarak sistemin ε-diferansiyel mahremiyet yapısı korunur.

Sıralı Bileşim

Farklı n tane M diferansiyel mahremiyet mekanizması M_1, M_2, M_3, …, M_n; ε değerleri ε_1, ε_2, ε_3,…, ε_n olsun. Aynı veri seti için uygulanan M mekanizmaları ε-DM'i sağlar ve sistemin toplam ε'u bütün ε'ların toplamına eşit olur.

Paralel Bileşim

Veri seti D farklı parçaya bölünüp, her veri seti farklı birer ε-diferansiyel mahremiyetli M (M_1, M_2, M_3, …, M_n) mekanizmasında işlenirse, M mekanizması da ε-DM yi sağlar ve sistemin ε'u kullanılan M mekanizmaları arasında en yüksek ε'a sahip M'nin ε'udur.

Diferansiyel mahremiyet ilk olarak veri analizi yapılan verilere yönelik ortaya çıkmış olsa da literatürde mahremiyet korumalı veri yayınlama, federe öğrenme [49], makine öğrenmesi [50], bulut teknolojileri [47], sentetik veri üretimi ve blok zincir [51] gibi birçok alanda mahremiyeti korumak için bu yöntemi kullanan birçok çalışma bulunmaktadır. Bu tezde derin öğrenme alanında yapılan diferansiyel mahremiyet çalışmaları konu alınmıştır.

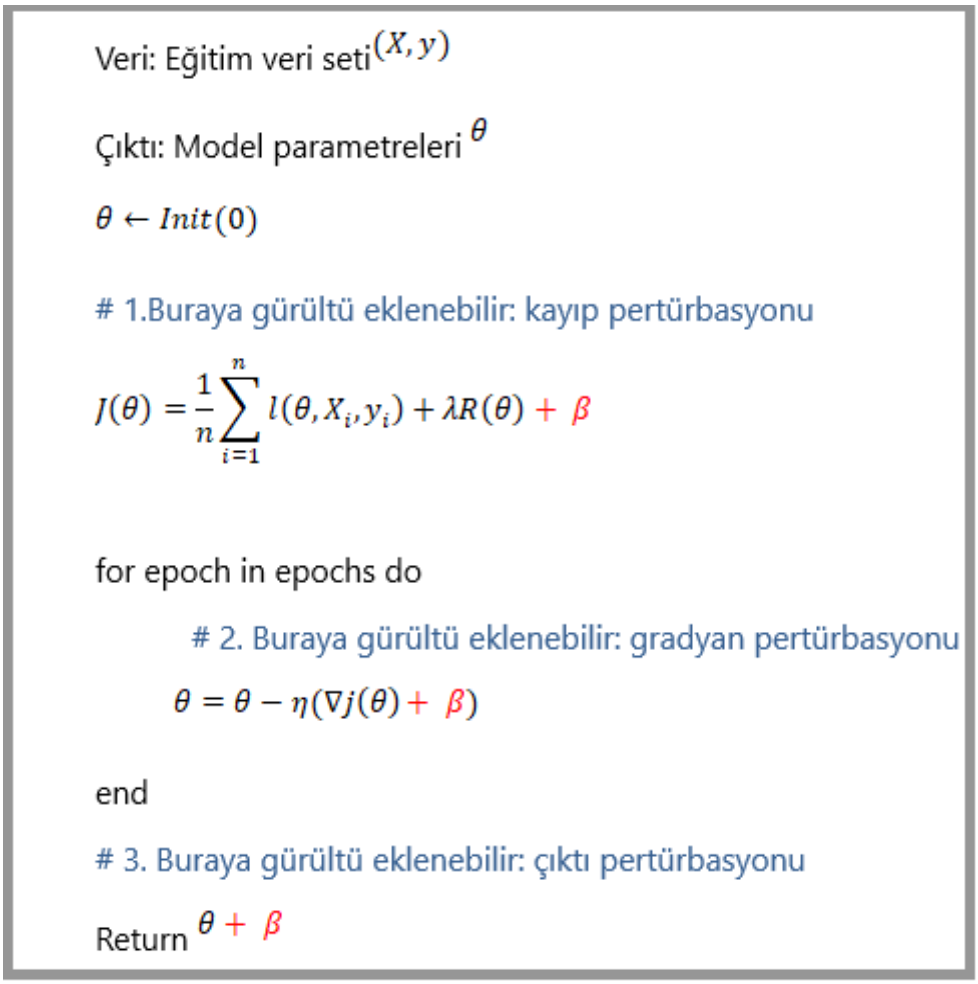
3.2. Derin Öğrenmede Diferansiyel Mahremiyet

Diferansiyel mahremiyet ilk olarak veri tabanlarında istatistiksel sorgulara gürültü ekleyerek mahremiyeti sağlamak amacıyla önerilmiştir. Fakat yöntemin eklenen gürültünün sağladığı mahremiyet derecesini ölçebilme ve kanıtlayabilme özelliği birçok alanda kullanımını cazip kılmıştır. Bu bölümde derin öğrenme alanında diferansiyel mahremiyetin uygulama biçimlerine değinilmiş ve yapılan son çalışmalar incelenmiştir.

Literatürde yapılan çalışmalar incelendiğinde diferansiyel mahremiyetin derin öğrenmede üç temel şekilde uygulandığı görülmektedir. Bunlar; veri toplama, eğitim ve çıkarım (test) safhası olarak sınıflandırılabilir.

Veri toplama aşamasında, veri seti eğitilmeden önce veriye gürültü ekleyerek mahremiyet sağlanır. Bu yöntem sadece derin öğrenme için değil bütün veri işleme yapan uygulamalarda kullanılabilir. Veriyi bozmak model performansını düşüreceği için makine öğrenmesi uygulamalarında genellikle tercih edilmemektedir.

Eğitim safhasında ise, modele birçok aşamada gürültü eklenebilir. Derin öğrenme ağında eğitim sırasında en çok kullanılan yöntem optimizasyon sırasında uygulanan gradyan pertürbasyonudur. Bunun yanında; kayıp fonksiyonu pertürbasyonu, çıktı ve etiket pertürbasyonu gibi gürültü ekleme yöntemleri de bulunmaktadır [6].



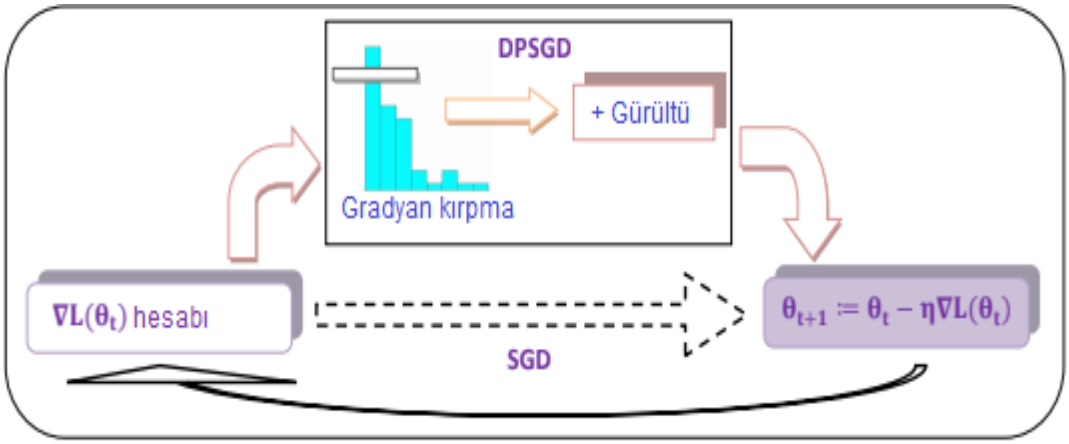
Şekil 3.2. Derin öğrenmede gürültü eklenebilecek adımlar

Kayıp fonksiyonu karıştırma ve çıktı karıştırma, konveks kayıp fonksiyonları üzerinde uygulanır. Örneğin lojistik regresyon modelinde, Chaudhuri ve diğerleri kayıp fonksiyonu karıştırmak için gürültünün $\frac{2}{v.\varepsilon}$ ölçeğinde örneklenmesi gerektiğini, çıktı karıştırmak için ise $\frac{2}{v.\lambda.\varepsilon}$ ölçeğinde ölçeklenmesini önermiştir. $\boldsymbol{\lambda}$ regülasyon katsayısı, n örnek sayısıdır [52]. Derin öğrenme modellerinde, kayıp fonksiyonunun konveks olmaması nedeniyle, fonksiyon hassasiyetinin (eklenen gürültünün yoğunluğunu belirlemek için gerekli) hesaplanması önemsiz hale gelir. Buna çözüm olarak, konveks olmayan kayıp fonksiyonunun yaklaşık bir konveks polinom fonksiyonu [53] ile değiştirilip ardından kayıp fonksiyonu pertürbasyonu uygulanmasıdır. Bu uygulamanın pratikteki zorluğu nedeniyle derin öğrenme uygulamalarında bu yöntem yerine genellikle gradyan pertürbasyonu kullanılmaktadır.

Yapılan çalışmada modelin veri seti mahremiyetini sağlamak için literatürde bulunan gradyan pertürbasyonuna dayalı DP-SGD algoritması kullanılmıştır. Bu algoritmanın tercih edilme sebebi hem kolay uygulanabilmesi hem de kompleks bir veri yapısında mahremiyet korumasının matematiksel olarak kanıtlanmış bir yöntem ile kullanılmasının daha doğru olacağı değerlendirilmesidir.

3.2.1. Diferansiyel mahremiyet korumalı SGD optimizasyonu

Derin öğrenmede model tahminlerine ve eğitim verisine gürültü eklemenin doğruluğu önemli ölçüde etkilemesi nedeniyle eğitim sırasında uygulanan mahremiyet koruma yöntemleri daha çok tercih edilmektedir. Bu yöntemlerde gürültü veriye değil eğitim sırasında modelin kendisine eklenmektedir. Literatürde bunlara verilecek en bilinen örnek Abadi ve diğerleri [54]'nin yapmış olduğu çalışmadır. Gradyan pertürbasyonuna dayalı bu çalışmada klasik SGD optimizasyon algoritmasının diferansiyel mahremiyet uygulanmış hali önerilmiştir. Yöntemde ilk olarak veri setinden rastgele L örnek seçilir. Seçilen batch (veri seti alt kümesi) için gradyan hesabı yapılır. Daha sonra hesaplanan gradyan l_2 formunda en seçilen eşik değer C kadar normalize edilir. Normalize edilen gradyana seçilen mahremiyet maliyeti parametresine (gürültü katsayısı) ve C'ye bağlı olarak Gauss gürültüsü eklenir. Son olarak SGD algoritmasının 'descent (azaltma)' adımı gerçekleştirilir. Çıktının ve eklenen gürültünün veri setindeki örneklerden etkilenmediği ve model doğruluğunun model yapısından çok ε ve batch boyutuna bağlı olduğu görülmüştür.



Şekil 3.3. DP SGD ve SGD algoritmalarının uygulanması [67]

DP-SGD yönteminde birçok kez gradyan hesabı yapıldığı için eğitim setine her erişimde mahremiyet kaybı yaşanabilir. Adım sayısı çok fazla olduğunda bileşik diferansiyel mahremiyet tanımına göre hesaplanan toplam ε çok büyük değerlere ulaşabilir. Bu durumda istenen mahremiyet derecesi sağlanamayabilir. Bu gibi seri bileşik algoritmalarda en büyük problem toplam mahremiyet kaybını takip edebilmektir. Abadi'nin önerdiği mahremiyet hesaplayıcısında toplam ε her adımda takip edilmiştir. Eklenen gürültü sınırlandırılarak ve gradyan kırpma yapılarak toplam mahremiyet korunmaya çalışılır. DP-SGD'nin bu özelliği çok katmanlı yüksek batch sayısına sahip modellerde daha yüksek doğruluk oranları ile mahremiyetin sağlanmasına olanak verir.

3.2.2. Derin öğrenme diferansiyel mahremiyet uygulamaları

Literatürde derin öğrenme modelleri için yapılan bazı diferansiyel mahremiyet çalışmaları Çizelge 3.1'de gösterilmiştir.

Çizelge 3.1'de özetlenen çalışmalara bakıldığında;

- He ve diğerleri [55] farklı türlerde (yatay, dikey, keyfi) bölünmüş veri setlerini eğiten sunucular için TransNet yöntemini önermiştir. Sunucuda, veriye gürültü ekleyen dönüştürücü bir katman bulunmaktadır. Veri sahipleri sunucuya veri yüklerken veriler bu katmanda dönüştürülerek ağa iletilir.
- Chen ve diğerleri [56] gerçek zamanlı güzergah bilgileri için RNN tabanlı bir diferansiyel mahremiyet şeması önermektedir. LBS uygulamalarına mahremiyet

sağlayan veriler önerilmiştir. Güzergâh bilgisi toplanan verilerin zaman bilgisine Laplace gürültüsü eklenerek işlenir. RNN ağı kullanılarak oluşturulan güzergahlar Laplace gürültüsü eklenerek yayınlanır.

Çizelge 3.1. Diferansiyel mahremiyet uygulanmış derin öğrenme çalışmaları

[55]	Letter, MNIST, Fashion, SVHN
[56]	Gerçek Araç Yol Bilgisi
[54]	MNIST, CIFAR-10
[57]	MNIST, SVHN
[58]	MNIST, CIFAR-10, SVHN
[59]	MNIST, Motion Tracking
[60]	MNIST, CDR, TRANSIT
[61]	UCI Dermotology, UCI WDBC
[62]	IoT system of the City of Antibes
[63]	MNIST, SVHN

- Papernot ve diğerleri [57] yaptıkları çalışmada, gradyan pertürbasyonu yerine etiket pertürbasyonu yapan bir model önermiştir. Ağ orijinal veriler ile eğitildikten sonra, sonuçlara üstel gürültü eklenir. Test sırasında mahremiyetin düşmesini engellemek için eğitim verilerine ve sonuçlara etiketsiz açık veriler eklenmiştir.
- Wang ve diğerleri [58] çıkarım için veri sıfırlama (null) ve diferansiyel mahremiyet uygulayarak gürültü ekleme mekanizması olan ARDEN'i önermiştir. Yöntem, sunucuların gizliliğini korumak için veri sıfırlama (null) ve gürültü ekleme kullanarak farklı sorguları ayırt edilemez hale getirir. Farklı katmanlara gürültü eklenerek tüm ağın yeniden eğitimi yapılır. Ağın her katmanında global hassasiyeti hesaplamak karmaşık bir iş olduğundan gürültü ekleme mekanizmasının girdisi, eğitim setinin herhangi bir üyesi tarafından üretilen olası en büyük değere yuvarlanır.
- Hammer ve diğerleri [59] sınıflandırma için kullanılan LVQ ağına DM uygulanmıştır. Mahremiyeti sağlamak için DPSGD algoritması LVQ ağında gradyan hesabı için kullanılır. Sınıf merkezleri her sınıftaki örneklerin toplanıp sınıf örnek sayılarına oranlanmasıyla hesaplanır. Sınıfların hesabında, bir elemanın yer değiştirmesi sınıflandırma sonuçlarını, histogramlara benzer şekilde bir birim değiştireceği için hassasiyeti 1 olan laplace mekanizması kullanılmıştır. DM modeli hem sınıfların

oluşturulmasında hem gradyan hesabında kullanıldığı için seçilen ε bu iki mekanizmaya bölüştürülür. Yüksek mahremiyet için ağın doğruluk oranının düştüğü gözlemlenmiştir.

- Acs ve diğerleri [60] çok boyutlu veriler üreten üretici yapay sinir ağları için diferansiyel mahremiyet korumalı bir yöntem önermiştir. Yöntemde veri seti diferansiyel mahremiyet uygulanmış kernel k-kümeleme yöntemi ile kümelere ayrılır. Daha sonra her küme ayrı üretici ağlarda eğitilir. Eğitim sırasında gradyan pertürbasyonu kullanılır. Ağın gerçekçi sentetik veriler ürettiği gözlemlenmiştir.
- Adesuyi ve Kim [61] yaptıkları çalışmayı geliştirerek LRP (Layer-wise Relevance Propagation) metodu ile diferansiyel mahremiyeti birleştirmiştir. Çalışmada Laplace gürültüsü nöronların ağa katılım oranlarını hesaplarken eklenmiştir.
- Xie ve diğerleri [62] mahremiyetli veri yayını yapabilmek için DPGAN üretici ağ modelini önermiştir. Önceki çalışmalardan farklı olarak bu yöntem zaman serileri, sürekli veriler ve ayrık verilerin üretimine kadar farklı kullanım alanlarına kolayca uygulanabilir.
- Shokri ve diğerleri [63] derin sinir ağlarının pertürbe edilmiş parametreler ile dağıtık bir şekilde eğitildiğini göstermiştir. Fakat bu çalışmada gerekli ε değeri oluşturulan modelin boyutu ile orantılı olmalıdır. Bu yüzden ε milyonu bulan değerler alabilmektedir ve mahremiyet derecesi düşmektedir.

3.2.3. Diferansiyel mahremiyet uygulanmış modellerin saldırılardaki başarısı

Mahremiyet korumalı makine öğrenmesi uygulamalarında genellikle istenilen doğruluğu sağlayacak keyfi bir mahremiyet maliyeti parametresi seçilir. Karmaşık modellerde bileşim teoremine göre ε değeri yükselir. Örneğin; Shokri ve diğerleri [63]'nin çalışmasında modelin boyutları ile doğru orantılı olarak seçilen parametre milyonları bulan değerleri alabilir. Fakat çok büyük ölçekli mahremiyet maliyeti parametresi seçmek hassasiyeti veya veri faydasını düşürür. Literatürde bu parametre için kesin bir eşik değer üzerinde uzlaşılmamıştır. Verimliliğin artması için genişletilmiş diferansiyel mahremiyet yöntemleri önerilmiştir [64-66]. Fakat bu yöntemlerin ne kadar mahremiyet sızıntısına izin verdiği tam olarak ölçülememiştir.

Diferansiyel mahremiyetin ne kadar koruma sağladığını ölçmek için literatürde bazı çalışmalar vardır. Bu çalışmalarda pratik olarak diferansiyel mahremiyetin saldırılara karşı

ne kadar koruma sağladığını test etmek için üyelik çıkarımı, özellik çıkarımı ve veri bozma gibi saldırılar uygulanmıştır.

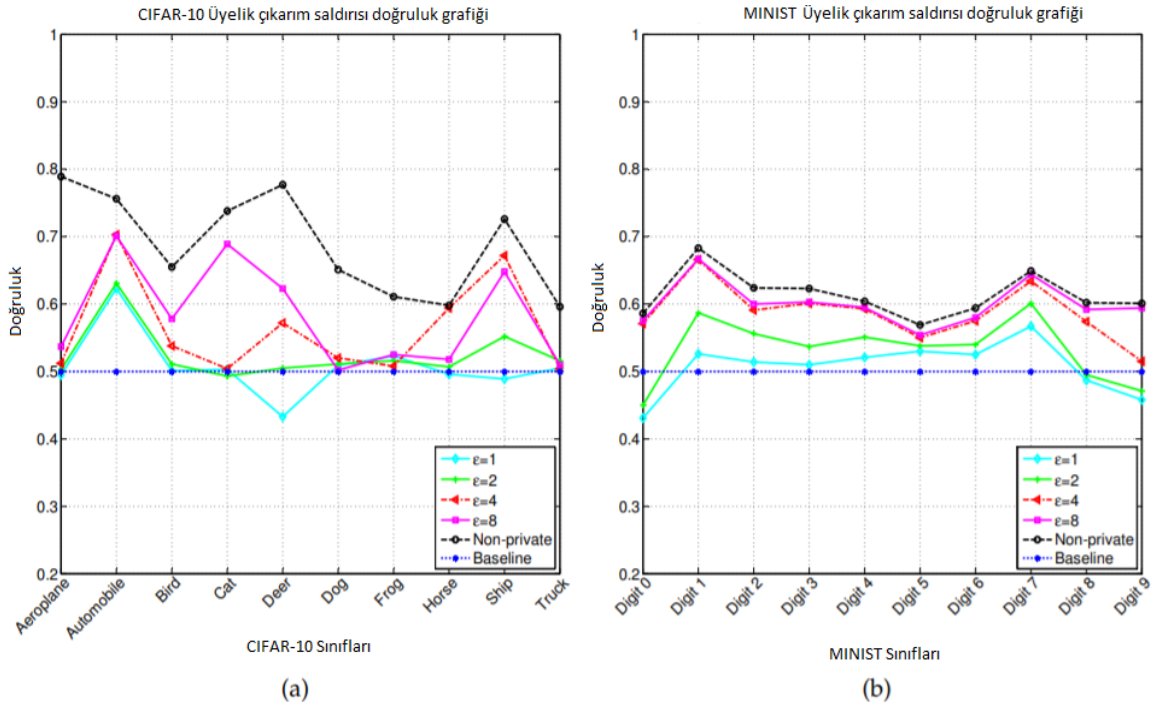
Önceki bölümde açıklanan üyelik çıkarımı saldırısı için diferansiyel mahremiyet doğal bir koruma yöntemi olarak görülebilir. Çünkü bu saldırı bir örneğin eğitim setinde olup olmadığını çıkarmaya çalışırken DM'de bir örneğin veri setindeki varlığını gizlemeye çalışır. Li ve diğerleri [67] yaptıkları çalışmada bu noktayı vurgulayarak DM ile üyelik çıkarımı saldırısı arasında doğrudan bağlantı kurmuştur. Backes ve diğerleri [68] yaptıkları microRNA çalışmasında DM'nin üyelik çıkarımı saldırısının başarısını test etmiş, bir miktar doğruluk kaybı göze alındığında saldırının başarısının düştüğünü gözlemlemiştir. Yeom ve diğerleri [74] üyelik çıkarımı saldırısı için dağılımdan rastgele seçtikleri bir verinin eğitim setinde olup olmadığını bulan bir model tasarlamıştır. Üyelik çıkarımı için saldırı modelinin TP ve FP oranlarına bakarak çıkarım yapılır. Çalışmada hedef model ε diferansiyel mahremiyet uygulanmış model eğitildiğinde üyelik çıkarımının $1 - e^{\varepsilon}$ ile sınırlandığı gözlemlenmiştir.

Özellik çıkarımı saldırısı ile diferansiyel mahremiyet bir kaydın veri setinde olup olmadığının çıktıyı önemli ölçüde değiştirmeme özelliği ile ilişkilendirilmiştir. Bir kaydın çıktı üzerindeki etkisi azaltıldığında girdinin sahip olduğu özelliklerinde çıktı üzerinde etkisinin düşeceği savunulur. Özellikler ve veri arasındaki bu bağlantı Yeom ve diğerleri [13]'nin çalışmasında incelenmiştir. Jayaraman ve diğerleri [69] bazı genişletilmiş DM yöntemleri kullanılmış lineer regresyon ve yapay sinir ağları modellerinde üyelik çıkarımı ve özellik çıkarımı saldırıları uygulayarak mahremiyetin ne kadar korunduğunu deneysel olarak ölçen bir çalışma yapmıştır. Çalışmada ε değeri arttıkça saldırıların daha başarılı olduğu gözlemlenmiştir.

Model ezberleme saldırısında ise yüksek kapasiteli modellerin bazı veri örüntülerini ezberlediği varsayılır ve buna yönelik saldırılar düzenlenir. Yapılan bir çalışmada çok düşük mahremiyet maliyeti parametresi ($\varepsilon = 10^9$) ile bu saldırının başarısının düştüğü gözlemlenmiştir [27].

Rahman [70] ve diğerleri DP-SGD ile optimize edilmiş bir modele gölge modeller eğiterek üyelik çıkarımı saldırısı yapmıştır. MNIST ve CIFAR-100 veri setlerinin kullanıldığı çalışmada aşırı öğrenme durumunda saldırının daha başarılı olduğu gözlemlenmiştir. Çünkü

aşırı öğrenme durumunda model eğitim seti hakkında daha fazla bilgi tuttuğu için daha çok bilgi sızdırılmasına neden olur. Çalışmada mahremiyetsiz modelde, mahremiyet uygulanmış modele göre daha çok aşırı öğrenme olduğu ve bu sebeple mahremiyet uygulanmış modelin saldırıya karşı daha dayanıklı olduğu gözlemlenmiştir. Ayrıca modelde ε değeri arttıkça saldırının daha başarılı sonuçlar elde ettiği görülmüştür. Şekil 3.4'te yapılan üyelik çıkarımı saldırısının farklı **ε** değerleri ile MNIST ve CIFAR-10 veri setleri üzerindeki başarım oranları gösterilmiştir.



Şekil 3.4. (a) MNIST sonuçları (b) CIFAR-10 sonuçları [70]

4. DİFERANSİYEL MAHREMİYET UYGULANMIŞ BEYİN MR GÖRÜNTÜLERİNDE ANOMALİ VEYA TÜMÖR TESPİTİ

Bu bölümde sağlık hizmetlerinde tıbbi görüntü işleyen bilgisayar destek sistemleri için veri seti mahremiyetini diferansiyel mahremiyet ile sağlayan bir uygulama geliştirilmiştir. Geliştirilen uygulamada derin öğrenme modeli MR görüntülerinden LGG ve HGG beyin tümörlerini sınıflandırmaktadır.

Bölüm akışında öncelikle beyin tümörü ve teşhis yöntemleri açıklanarak literatürde beyin MR görüntülerini sınıflandırmak için bulunan derin öğrenme çalışmaları incelenmiştir. Daha sonra derin öğrenme temelleri ile çalışma için kullanılacak model ve uygulama akışı açıklanmıştır.

4.1. Beyin Tümörü Hastalığı ve Derin Öğrenme Çalışmaları

Beyin tümörü beynin içinde veya omurilikte gelişir. Koordinasyon bozukluğu, tekrarlayan baş ağrıları, konuşmada değişiklikler, konsantrasyon bozukluğu, nöbetler ve hafıza kaybı gibi semptomlar gösterir.

Tümöre yol açan birçok etmen bulunmaktadır. Dünya üzerindeki kanserlerin %15'i virüs kaynaklıdır [71]. Günümüzde insanda kansere yol açan birkaç DNA virüsü tespit edilmiştir. Bunun yanında çevresel faktörler de kansere sebep olan diğer başlıca etmenlerdir. X-ray ışınlarına maruz kalmak, UV ışınları, tütün ürünleri, kirlilik, günlük kullanılan kimyasallardaki kanser yapıcı etkenler kansere sebep olabilir. Gün ışığı deri hücrelerini mutasyona uğratarak cilt kanserine, sigara kullanımı akciğer kanserine sebep olabilir [72].

Kanserin temel sebebi kontrolsüz hücre büyümesidir. DNA içinde bulunan genler istenmeyen sağlıksız hücrelerin ölümünü ve sağlıklı hücrelerin yeniden üretilmesini kontrol eder. DNA diziliminde bu genlerde meydana gelen bir mutasyon hücrenin istemsizce bölünmesine sebep olur. Mutasyon hayat tarzı, çevresel faktörler ve yeme davranışlarından kaynaklanabilir. Mutasyon hücre ölümünü yavaşlatarak [73], DNA'yı onaran genleri bozarak, hücre bölünmesi sinyalinin durdurulamamasına ve kontrolsüz hücre büyümesine sebep olur. Kanser bu mutasyonlarla başlayabileceği gibi kan damarları üzerinden metastaz yoluyla diğer organlardan da sıçrayabilir. Tümörler hayatta kalabilmek için besin ve kan

desteğine ihtiyaç duyar. Vücutta bulunan bazı genler hücrelerin bu ihtiyacını fark eder ve bir damar ağı kurarak destek sağlar. Böylece kanserli doku gelişerek çoğalmaya devam eder.

Beyin tümörleri yapısı, kaynağı, büyüme oranı ve ilerleme hızlarına göre sınıflandırılır [74,75]. Tümörler iyi huylu veya kötü huylu olarak da ikiye ayrılır. İyi huylu tümörler, çevre hücrelere yayılım göstermez ve kesin sınırlara sahiptir. Kötü huylu tümörlerde yayılım hızı yüksektir ve kesin bir sınırları yoktur. Kaynağına göre birincil ve ikincil tümörler olarak ayrılırlar. Birincil tümörler direk beynin içinde oluşum gösterirken ikincil tümörler diğer organlardaki tümörlerin metastaz yoluyla beyne sıçramasıyla oluşur. Dünya sağlık örgütü beyin tümörlerini büyüklüklerine göre dörde ayırmıştır. Dünya sağlık örgütünün belirlediği sınıflandırmada ilk iki evrede bulunan tümörler LGG, üç ve dördüncü evrede bulunan tümörler HGG olarak sınıflandırılır. Tümörler ilerleme hızına göre de dörde ayrılır. 0. evrede anormal tümör hücreleri görülür fakat yayılım yoktur. 1, 2 ve 3. evrelerde tümör hızla yayılır. 4. evrede kanser tüm vücuda yayılmaya başlar. İleri seviyelerdeki tümörlerin tedavileri çok zor olduğu için hayatta kalma oranı düşükken, erken evrelerde teşhiste birçok hastanın hayatı kurtarılabilmektedir. Tümör sınıflandırması uygulanacak tedavi açısından önemlidir.

4.1.1. Beyin tümörü teşhis yöntemleri

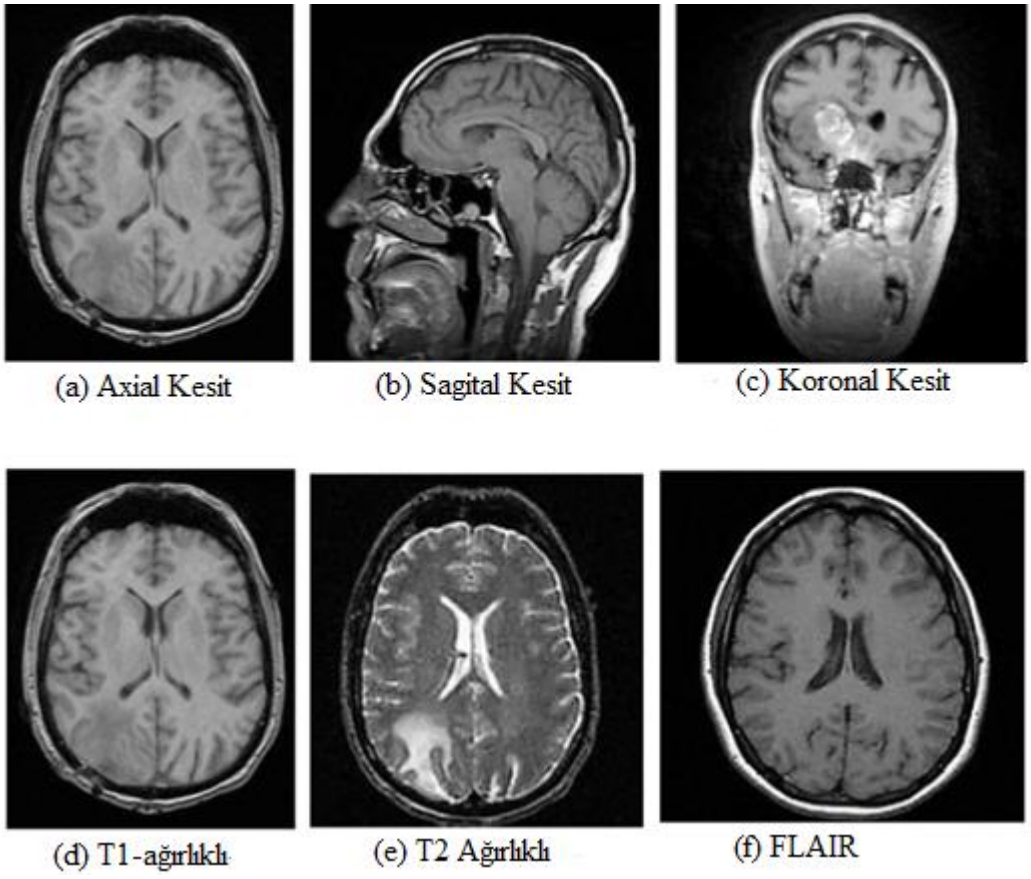
Beyin kanseri teşhisi biyopsi ve medikal görüntüleme teknikleri kullanılarak yapılır. Biyopside tümörden alınan örnek patoloji uzmanı tarafından incelenerek karar verilir. Görüntüleme teknikleri ise bilgisayarlı tomografi ve manyetik rezonans görüntüleme yöntemleridir. Bu yöntemler vücut bütünlüğüne direk müdahalede bulunulmadığı için biyopsiye göre daha hızlı ve güvenilirdir. Elde edilen görüntüler radyoloji uzmanı tarafından değerlendirilir.

Medikal görüntüleme teknikleri hücresel değişimleri görüntüleyebilir. Uzmanlara fiziksel müdahale yapmadan insan vücudu ve aktiviteleri hakkında bilgi verir. Kanser tanısında teşhis, aşama tahmini, tedaviye verilen cevap, cerrahi planlama ve yaşanabilecek zorluklar hakkında çıkarım yapmaya yardımcı olur. Beyin görüntüleme teknikleri yapısal ve fonksiyonel olmak üzere ikiye ayrılır [76]. Yapısal görüntüleme beyin bozuklukları, tümör yeri gibi beynin yapısı ile ilgili ölçümler yapar. Fonksiyonel görüntüleme metabolizma değişimleri, lezyonlar ve beyin aktiviteleri ile ilgili bilgi verir.

MRG radyasyon içermediği için BT'den daha güvenilirdir ve daha yüksek çözünürlüğe sahip olduğu için detaylı görüntüleme sağlar. MRG beyin görüntülemede aksial, sagital ve koronal olmak üzere üç temel düzlem bulunur.

MRG'de sekans terimi; görüntülenecek doku veya örneğin özelliklerini ortaya çıkaracak kontrasta sahip olacak şekilde uygulanan radyo frekansı darbeleri, gradyan alanları ve sinyal kayıt zamanlarının bütününü ifade eder. Bazı çekim parametreleri değiştirilerek farklı görünümlerde değişik sekanslar elde etmek mümkündür. MRG görüntüsünde siyah ile beyaz arasındaki gri tonlar kullanılarak görüntü oluşturulur [77].

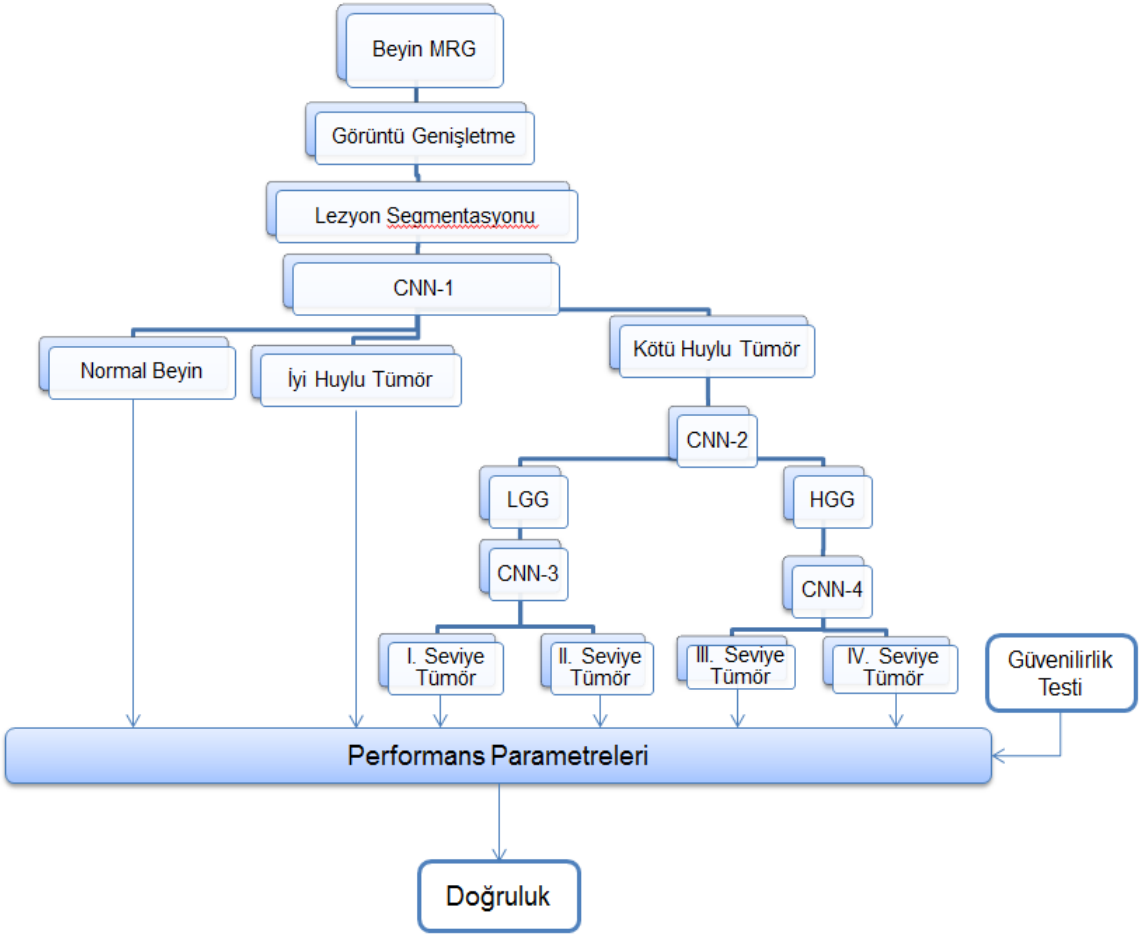
MRG'de T1-ağırlıklı, T2-ağırlıklı ve proton-ağırlıklı olmak üzere 3 temel dalga sekansı bulunmaktadır. T1-ağırlıklı görüntülerde anatomik detay yüksektir. T2-ağırlıklı sekanslar doku karakterlerinin ve patolojilerin saptanmasında daha etkilidir. T1-ağırlıklı sekansta BOS siyah, proton ağırlıklı çekimlerde gri, T2-ağırlıklı çekimlerdeyse beyaz olarak görülür. Az sayıda kullanım alanı olan proton-ağırlıklı görüntüler standart beyin incelemelerinden kaldırılmıştır. Temel sekansların yanında BOS gibi sıvıları baskılayıp, sıvıların koyu gri görünmesini sağlayan sekanslar mevcuttur (FLAIR gibi). Beyindeki özellikle periventriküler yerleşimli lezyonlar T2-ağırlıklı sekanslarda BOS'un beyazlığı ile karışabilir. FLAIR yöntemiyle sudan gelen sinyaller baskılanarak iskemi, MS plakları gibi demiyelinizan lezyonların saptanması kolaylaşır [77].



Şekil 4.1. Beyin MR görüntüleri kesit ve sekansları

4.1.2. Medikal beyin görüntüleri ile yapılmış derin öğrenme çalışmaları

Görüntüleme tekniklerinde kararı veren uzmanın teşhis koyma kabiliyeti önemlidir. Günümüzde uzmanlara karar vermeleri için yardımcı olan çeşitli bilgisayar destekli (CAT) sistemler bulunmaktadır. Artan hesaplama gücü ve donanım maliyetinin azalması ile birlikte CAT sistemlerin kullanımı yaygınlaşmaya başlamıştır. Özellikle makine öğrenmesi alanında yaşanan gelişmeler otomatik karar sistemlerinin geliştirilmesi için imkân sağlamış bu yönde çalışmalara hız verilmiştir. Derin öğrenme teknikleri MRG değerlendirmede en çok kullanılan makine öğrenmesi yöntemidir. Normal veya anormal beyin sınıflandırma, segmentleme, tümör sınıflandırma, alzaymır teşhisi gibi alanlarda çalışmalar bulunmaktadır. Aşağıda otomatik beyin tümörü teşhisi yapan ve sınıflandıran bir derin öğrenme sisteminin ana yapısı verilmiştir [78].



Şekil 4.2. Beyin tümörü tanısı için önerilen genel akış şeması

Literatürde derin öğrenme yöntemleri kullanılarak yukarıdaki temel yapının belirli kısımlarını gerçekleştirmeye çalışan uygulamalar bulunmaktadır. MR görüntüleri için yapılan çalışmalarda genellikle CNN mimarisinin tercih edildiği görülmüştür.

- Sunil ve diğerleri [79] aşırı öğrenmeyi azaltmayı amaçlayan beyin MR görüntüleri üzerinde ikili sınıflandırıcı önermiştir. Aşırı öğrenmeyi azaltmak için softmax aktivasyon fonksiyonu ve regülasyon kullanılmıştır. İyi huylu ve kötü huylu tümör tespitinden sonra tümör segmentleme yapılmıştır. Farklı veri setlerinde farklı doğruluk oranları gözlemlenmiştir.
- Neelum ve diğerleri [80] çok seviyeli özellik çıkarımı yapan bir ağ modeli önermiştir. Çıkarılan çoklu özellikler birleştirilerek sınıflandırma işlemi yapılmıştır. Inception-v3 ve DenseNet201 modelleri kullanılarak çıkarılan çoklu özellikler softmax aktivasyonu kullanılan bir sınıflandırıcıda sınıflandırılmıştır.

- Javaria ve diğerleri [81] gliom ve strok lezyon teşhisi için 14 katmanlı bir CNN modeli önermiştir. 5, 6 ve 7. katmanlarda yığın normalleştirme uygulamış, 11. katmanda ortalama ortaklama (average pooling) yapılmış ve softmax aktivasyonu kullanılan tam bağlı sinir ağları sınıflandırıcısı kullanılmıştır.
- Guotai ve diğerleri [82] otomatik beyin tümörü segmentleme için kademeli CNN modeli önermiştir. Önce tümörü, sonra tümör çekirdeğini daha sonra genişletilmiş tümör çekirdeği bulan üç aşamalı bir yapı kurulmuştur.
- Talo ve diğerleri [83] normal ve anormal beyin sınıflandırmak için CNN tabanlı ResNet34 modelini kullanmıştır. Model eğitiminde daha önce büyük bir veri seti tarafından eğitilmiş bir modelin parametreleri kullanılarak öğrenme transferi uygulanmıştır. Böylece küçük veri setinde daha başarılı sonuçlar elde edilmiş ve hesaplama yükü azaltılmıştır.
- Salem ve diğerleri [84] T2 ağırlıklı MR görüntülerinde lezyon teşhisi için CNN tabanlı U-Net mimarisini kullanmıştır.
- Nawab ve diğerleri [85] öğrenme transferi uygulayarak VGG19 modelinin parametreleri ile beyin tümörü sınıflandırması yapmıştır.
- Samikannu ve diğerleri [86] gliomları sınıflandırmak için öncelikle tümörlü ve tümörsüz beyin MR görüntülerini ayırt eden bir 5 katmanlı basit bir CNN modeli kullanmış ve tümörlü görüntülerde k-means yöntemi ile segmentleme yapmıştır.

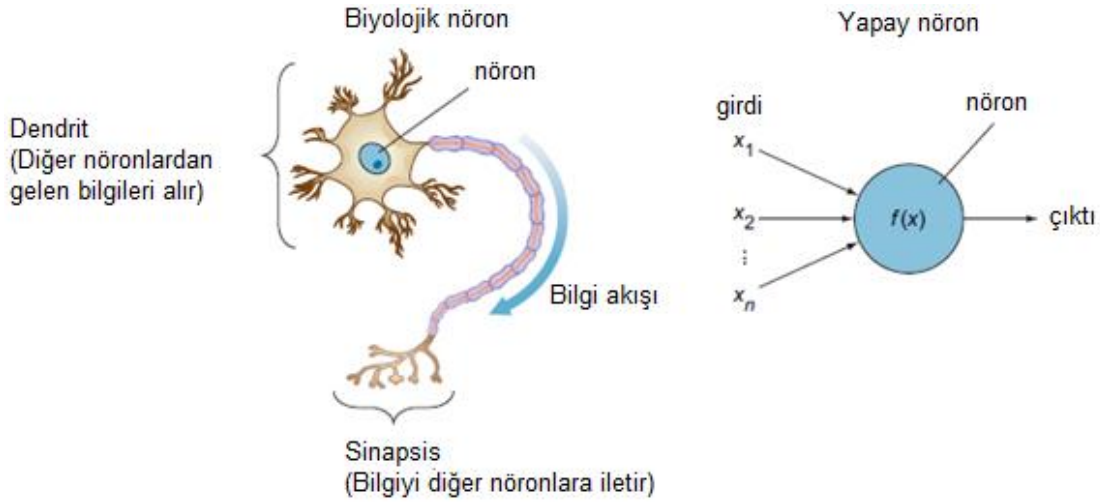
Bu tez çalışmasının uygulama bölümünde CNN ağı kullanılarak çoklu özellik çıkarımı yapan bir derin öğrenme modeli tasarlanmıştır. Modelin önceki bölümlerde bahsedilen mahremiyet ihlallerine karşı daha dirençli olması için eğitim sırasında diferansiyel mahremiyet uygulanmış optimizasyon algoritması DP-SGD kullanılmıştır. Sonraki bölümde ağ yapısı ve deney sonuçları hakkında detaylı bilgi verilecektir.

4.2. Derin Öğrenme Temel Kavramlar ve Uygulanan Model

4.2.1. Derin öğrenme temel kavramlar

Derin öğrenme makine öğrenmesinin bir alt dalıdır. Hesaplama karmaşıklığının artması ile birlikte veriyi tanımlayan özellikleri daha detaylı çıkarabilmek için çok katmanlı bir mimari kullanır. Hiyerarşik bir şekilde oluşturulan bu katmanlar nöronlardan oluşmaktadır. Girdi ve

çıktı katmanları arasında kalan katmanlara gizli katmanlar denir. ‘Derin’ ifadesi bu gizli katmanları temsil eder. Ağ mimarisi ise kökenini insan beynindeki sinir hücreleri bağlantılarını modelleyen bir çalışmadan aldığı için [87] kullanılan ağlara da yapay sinir ağları denir. Derin öğrenme kavramı çok katmanlı yapay sinir ağlarından oluşturulmuş modelleri tanımlamak için kullanılır.



Şekil 4.3. Biyolojik sinir ve yapay sinir hücreleri

Yapay sinir ağları girdi örneğinin özelliklerini öğrenir ve bu özelliklere göre çıkarım yapar. 2009 yılında yapılan bir çalışmada yapay sinir ağının girdi özelliklerini nasıl öğrendiği gösterilmiştir [88]. Çeşitli kategorilerde resimler kullanılarak eğitilen ağ farklı katmanlarda resmin farklı özelliklerini öğrenmektedir. Ağın ilk katmanlarında, kenar ve köşe gibi basit özellikler öğrenilirken, ilerleyen katmanlarda bu basit özellikler birleştirilerek daha karmaşık özellikler elde edilir. Daha sonra yapay sinir ağı bu özelliklere göre girdi verilerinden çıkarımlar yapar.

Alex Krizhevsky [89]’nin 2012 yılında yaptığı çalışma yapay sinir ağları alanında bir dönüm noktası olmuştur. ImageNet yarışması için geliştirilen modelde artan ağ derinliğinin model kaybını azalttığı gösterilmiş, ağ derinliği ve katmanlı mimari önem kazanmıştır.

Kullanılan yöntemlerin geliştirilmesi model doğruluğunun artması, regülasyonların hesaplama süresini kısaltması ile derin öğrenme teknikleri daha çok dikkat çekmeye başlamıştır. Bunların yanında donanım gücünün artması da karmaşık hesaplamaların yapılabilmesine olanak sağlayarak derin öğrenme uygulamalarını öne çıkarmıştır.

Regresyon gibi modeller çok sayıda girdi alsa da girdi ve çıktı katmanları arasındaki hesaplama adımları karmaşık özelliklerin çıkarımı için çok kısadır. Ayrıca bazı girdilerin sonuca katkı sağlayabilmesi için diğer girdiler ile etkileşimli olması gerekebilir. Derin sinir ağları ile hem hesaplama aralığı uzatılabilir hem de girdiler birbiri ile ilişki kurabilir. Bu yüzden resim gibi çok boyutlu veriler için daha çok tercih edilmektedir. Derin sinir ağlarının diğer yöntemlere göre en büyük avantajı ise karmaşık özellikleri eğitim verilerinden kendi kendine öğrenebilme yeteneğidir.

Günümüzde çeşitli derin öğrenme uygulamaları bulunmaktadır. Bu uygulamalar temel olarak birbirine bağlı çok katmanlı yapay sinir ağlarından oluşur. Birbirinden farklı özellikleri ağ mimarisi ve eğitilme şekilleridir;

Çok katmanlı perseptronlar, en az bir gizli katmanı olan tam bağlı ileri beslemeli yapay sinir ağlarıdır.

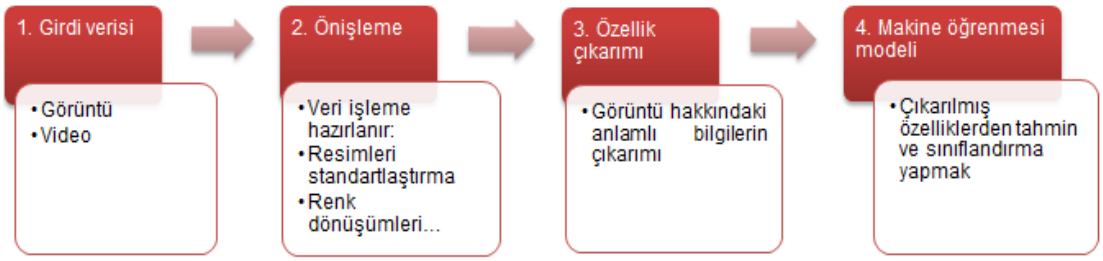
Evrişimli sinir ağları, ileri beslemeli farklı tiplerde özel katmanları olan sinir ağlarıdır. Girdi üzerinde filtre matrisleri gezdirilerek özellik çıkarımı yapılır. Evrişimli ağlar bu yönüyle beynin görme ile ilgili korteksine benzer bir yapıdadır.

Yinelemeli ağlarda, ağ önceki girdilerden elde edilen sonuçları tutan bir belleğe sahiptir. Bir girdi için yapılacak tahmini önceki girdinin sonuçları etkiler. Çıktılar sonraki işlem için girdi olarak kullanılır.

Otomatik kodlayıcılar, denetimsiz öğrenme sınıfındadır ve ağın girdi çıktı boyutları eşittir. Girdi özelliklerinin daha iyi öğrenilmesi amaçlanır.

Şekil 4.4’te makine öğrenmesi uygulamalarında kullanılan sınıflandırma adımları gösterilmektedir. Diğer yöntemlerde özellik çıkarımı adımı manuel olarak yapılıp makine öğrenmesi uygulamaları ile sınıflandırılır. Derin öğrenme ile birlikte sinir ağları hem özellik çıkarımı adımını hem de sınıflandırma adımını kendi kendine yapma yeteneği kazanmıştır. Fakat çok katmanlı ağlar resim gibi çok boyutlu girdileri işlerken veri kaybı yaşadığından girdi özelliklerini yeterince iyi belirleyememektedir. Evrişimli sinir ağlarının geliştirilmesi ile birlikte bu sorun büyük ölçüde giderilmiş, yapay sinir ağları özellik çıkarımında daha

kabiliyetli bir hale gelmiştir. Bu yönü ile görüntü işleme gibi karmaşık girdilere sahip uygulamalarda en çok tercih edilen makine öğrenmesi yöntemi CNN ağları olmuştur.



Şekil 4.4. Makine öğrenmesi uygulamalarında temel sınıflandırma adımları

Bu tez çalışmasının uygulama kısmında medikal görüntü sınıflandırma yapılacağı için derin öğrenme modellerinden evrişimli sinir ağları kullanılmıştır. Bölümün geri kalanında evrişimli sinir ağı mimarisi ve derin öğrenme uygulamalarının ortak parametreleri açıklanmıştır.

Evrişimli Sinir Ağları

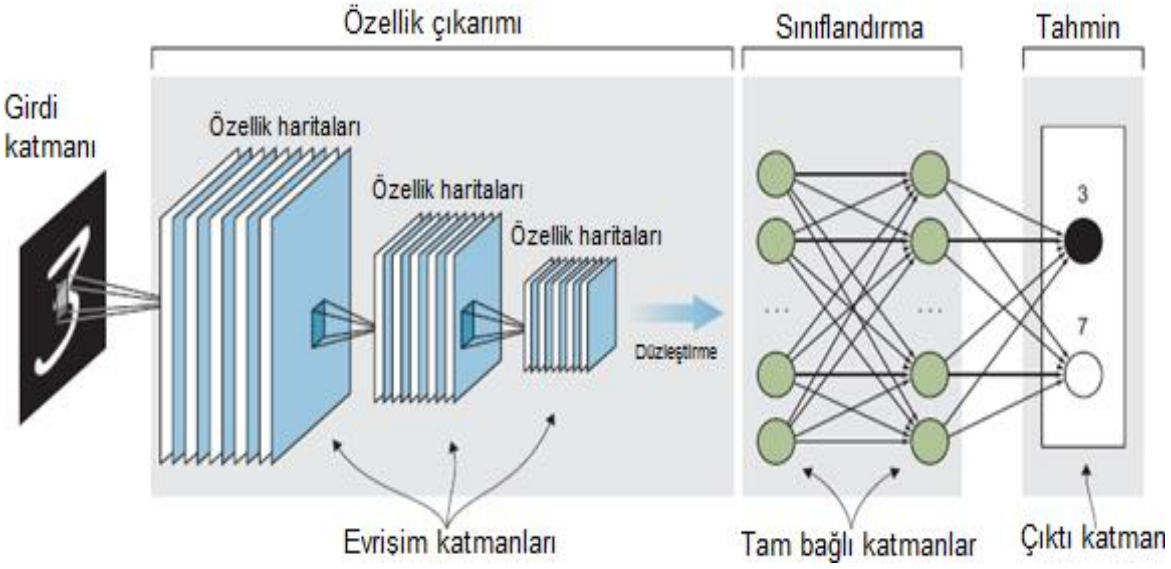
Evrişimli ağların eğitimleri, katmanlı mimari yapısı, ağırlıkları, hiper-parametre ve eğitim şekilleri çok katmanlı yapay sinir ağlarına benzemektedir. Genel olarak;

- Nöronlar katmanlar halinde birbirine bağlı olarak yerleştirilir.
- Ağırlıklar rastgele seçilir ve eğitim sırasında model doğruluğunu en yüksek yapacak şekilde güncellenir.
- Nöron çıktısının doğrusal yapısını bozmak için aktivasyon fonksiyonları kullanılır.
- Modeli değerlendirmek için kayıp fonksiyonu hesaplanır. Modeli iyileştirmek için ağırlıklar hata fonksiyonunu azaltacak şekilde güncellenir.

CNN ağının geleneksel çok katmanlı ağlardan farkı özellik çıkarımı için tam bağlı çok katmanlı ağlar yerine evrişim ağları kullanmasıdır.

CNN mimarisinde öncelikle girdi katmanı bulunur, girdi katmanından sonra evrişim ağları gelir. Evrişim ağları girdinin özellik haritalarını çıkararak ağın girdi hakkındaki özellikleri öğrenmesini sağlar. CNN ağlarında özellikler basitten karmaşığa doğru öğrenilir. Ağın ilk

katmanı köşe ve kenar gibi basit özellikleri öğrenirken sonraki katmanlarda daire ve kare gibi daha karmaşık özellikler çıkarılır. Son katmanlarda ise yüzün parçaları gibi nesnel özellikler belirlenir. Girdinin boyutu her katmanda azalırken özellik haritalarının derinliği artar ve son katmanda bir özellik dizisi elde edilir.

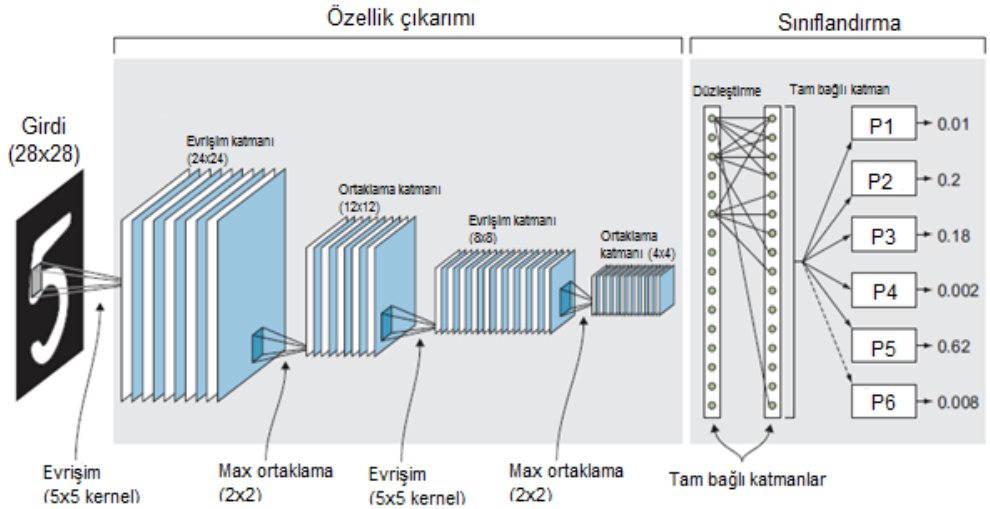


Şekil 4.5. Evrişimli ağın genel yapısı [90]

Daha sonra özellik haritalarından oluşan bu diziler düzleştirme (flatten) katmanında tek boyutlu vektörler haline getirilir. Vektör özellikler tam bağlı çok katmanlı sinir ağlarına gönderilerek sınıflandırma yapılır ve sonuç çıktı katmanına iletilir. Evrişim ağları özellik çıkarımı yaparken tam bağlı çok katmanlı sinir ağları sınıflandırma yapar. Basit bir CNN ağı Şekil 4.5'teki gibi görünmektedir.

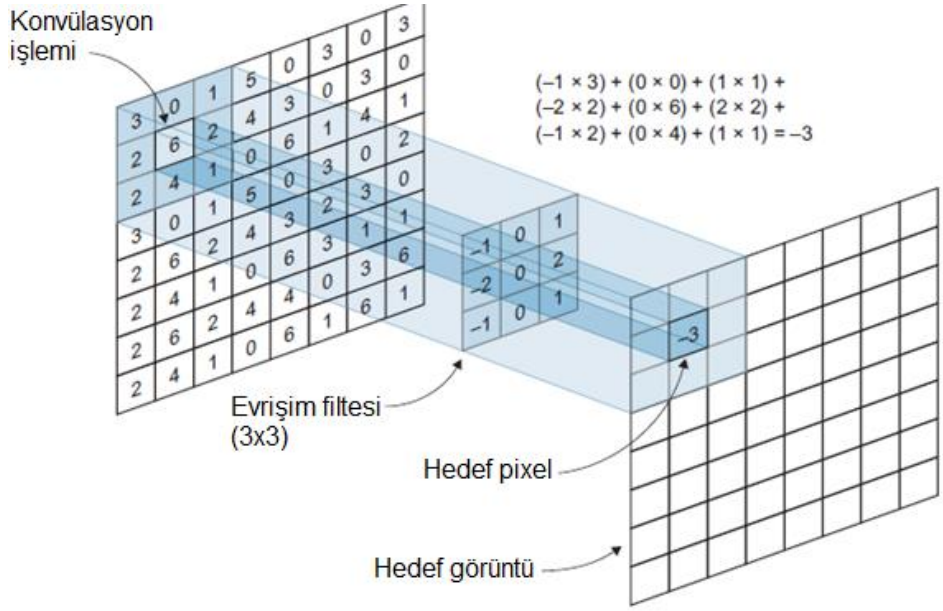
Evrişimli ağların temel bileşenleri

CNN ağ mimarisinin genel olarak üç temel bileşeni vardır: Evrişim katmanı, ortaklama (pooling) katmanı ve tam bağlı (fully connected) katman.



Şekil 4.6. Evrişimli sinir ağının temel bileşenleri [90]

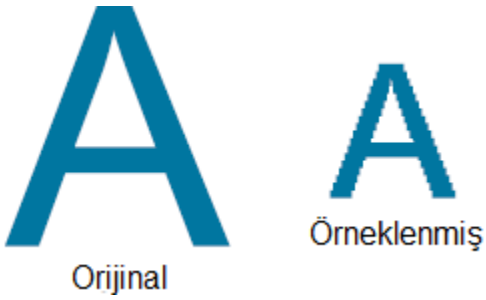
Evrişim katmanı resmi piKSel piKSel tarayarak anlamlı özellikler arar. Özellik çıkarımı işlemi girdileri anlamlı küçük parçalara ayırıp özellik vektörlerine dönüştürme olarak tanımlanabilir. Özellikler ‘kernel’ olarak isimlendirilen bir filtre yardımıyla bulunur. Kernel, CNN ağında ağırlıkları temsil eden bir matristir. Filtre girdi üzerinde gezdirilerek bazı matematiksel işlemler uygulanır ve hesaplanan yeni piksel değerleri ile özellik haritası oluşturulur. Filtrelenmiş çıktılara özellik haritası denmesinin sebebi çıkarılan özelliğin girdinin hangi bölgesinde bulunduğunu da tutmasıdır. Kernel değerleri ilk başta rastgele seçilir ve eğitim boyunca güncellenerek en uygun değerler bulunur.



Şekil 4.7. Evrişinli sinir ağlarında konvülasyon işlemi

Evrişim işleminde doldurma (padding) ve adım (stride) adı verilen iki parametre kullanılır., bir resim girdisi için fitrenin tek adımda kaç adım ilerleyeceğini ifade eder. Doldurma ise girdi ve çıktı arasındaki boyut farkını dengelemek için resmin etrafına sıfırlarla dolu bir çerçeve ekler.

Evrişim işleminden sonra genellikle ortaklama işlemi uygulanır. Ağda çok fazla evrişim katmanı varsa, parametre sayısı büyük boyutlara ulaşabilir. Bu da hesaplama karmaşıklığını ve eğitim için gerekli bekleme süresini arttırmaktadır. Bu durum ağda optimizasyon yaparak parametre sayısını azaltmayı gerektirir. Bunun için kullanılan en temel yöntem ortaklama (pooling) işlemidir. Özellik haritası max-min veya ortalama hesaplama gibi istatistiksel fonksiyonlar kullanılarak örneklenir. Böylece aynı derinlik seviyesine sahip daha küçük bir özellik haritası oluşturulmuş olur. Resmin çözünürlüğü düşerken sahip olduğu özellikler korunur. Literatürde parametre azaltma işlemi için ortaklama işlemi yerine daha büyük stride kullanılması da önerilmiştir [90].



Şekil 4.8. Ortaklama işlemi yapılmış girdi ve düşük çözünürlüklü çıktı

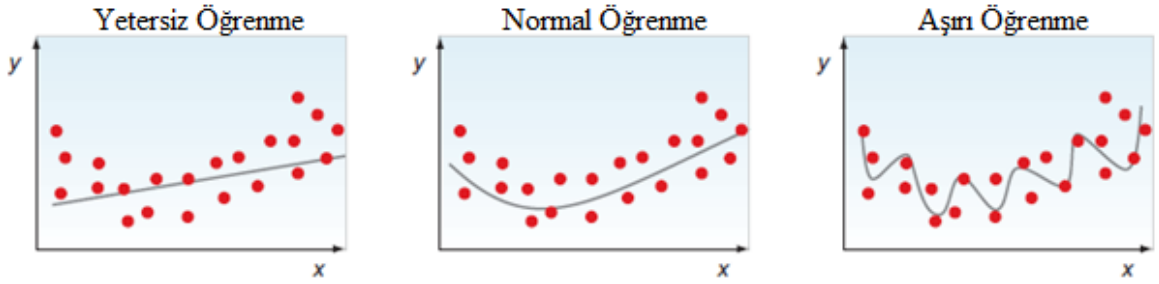
Yapılan özellik çıkarımı işleminden sonra, özellikler tam bağlı ağlara gönderilerek sınıflandırma yapılır. Tam bağlı çok katmanlı ağlar iyi sınıflandırıcılardır fakat özellik çıkarımında veri kaybı oluştuğundan özellik çıkarımı için evrişim ağları kullanılır.

Aşırı Öğrenme

Yapay sinir ağının mimarisi oluşturulurken modelin doğru çıkarımlar yapabilmesi için model eğitimi sırasında oluşabilecek aşırı öğrenme ve yetersiz öğrenme durumların önlenmesi gerekmektedir.

Yetersiz öğrenme, modelin eğitim veri seti ile uymaması ve doğru tahminler yapamamasıdır. Bu durumda model mimarisinin geliştirilmesi gerekir. Örneğin, ağa daha çok katman eklenebilir ya da kullanılan veri sayısı artırılabilir. Fakat ağa çok fazla katman eklenmesi aşırı öğrenmeye sebep olabilir.

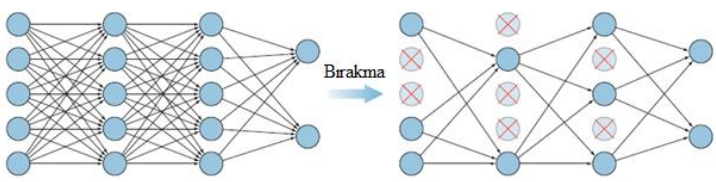
Aşırı öğrenme, modelin eğitim veri seti üzerindeki çıkarımlarında doğruluk oranı yüksekken test veri seti üzerindeki doğruluk oranının düşük olmasıdır. Model, eğitim setini mükemmel bir şekilde sınıflandırırken bilmediği bir veri üzerinde doğru çıkarım yapamamaktadır. Aşırı öğrenme durumundan kurtulmak için modelde bazı regülasyonlar yapılmalıdır. Bunlar; bırakma (dropout), L2 regülasyon ve veri çoğaltma olarak sıralanabilir.



Şekil 4.9. Öğrenme modelinde modelin uygunluğu

Bırakma

Bırakma işleminde ağdaki bazı nöronlar devre dışı bırakılarak eğitime dahil edilmez. Bu nöronların çıktıları ileri ve geri beslemede kullanılmaz. Böylece ağda baskın ağırlığa sahip nöronlar etkisiz hale getirilerek veri ezberleme engellenmiş olur. Baskın ağırlıklı nörondan gelen özellikler ile ezberleyen düşük ağırlıklı nöron kendi kendine öğrenmeye zorlanarak eğitime katkısı artırılır ve ağda denge sağlanır. Ayrıca bırakma işlemi baskın nöronun hata yapması durumunda ağı domine etmesini ve yanlış çıkarıma neden olmasını engeller. Bırakma işlemi sınıflandırma sırasında tam bağlı ağlarda gerçekleştirilir.



Şekil 4.10. Yapay sinir ağlarında bırakma işlemi

L2 regülasyon

Bu yöntemde hata (kayıp) fonksiyonuna L2 regülasyon değeri eklenir. Böylece hata miktarı gerçek değerine göre artırılmış olur ve optimizasyon sırasında ağırlık güncellenirken hesaplanan parçalı türevin değeri artar. Normalden daha küçük bir ağırlık elde edilmiş olur. Ağırlığının küçültülmesi ile nöronun öğrenmeye olan katkısı azalır, ağı domine etme ihtimali düşer ve aşırı öğrenme engellenmiş olur. λ: regülasyon parametresi; m: örnek sayısı, w: ağırlık olmak üzere;

$$l2\ regülasyon\ terimi = \frac{\lambda}{2m} x \sum \|w\|^2 \quad (4.1)$$

$$yeni\ hata\ fonksiyonu = eski\ hata\ fonksiyonu + L2\ regülasyon\ terimi \quad (4.2)$$

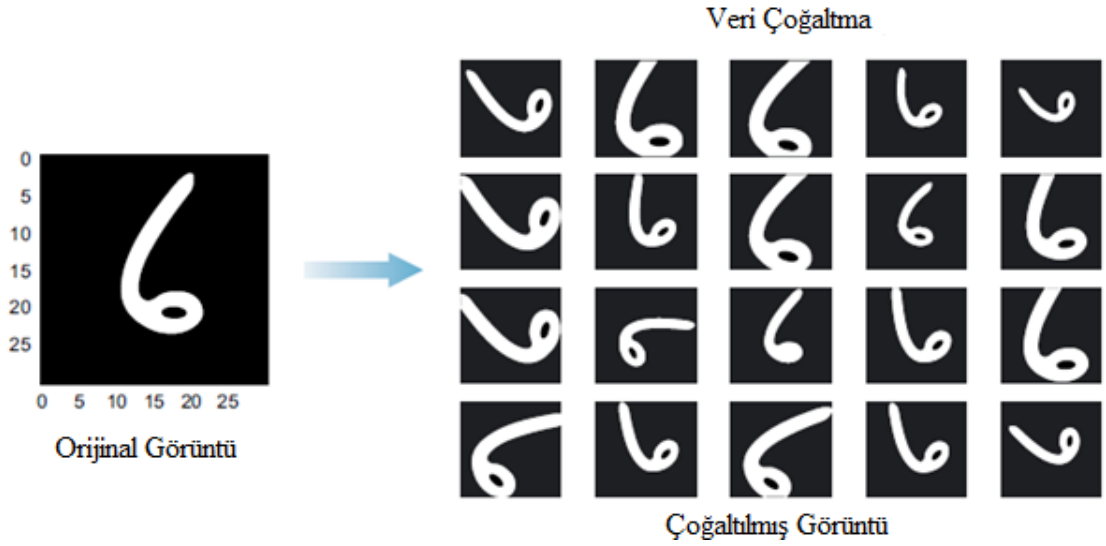
Ağırlık güncelleme aşağıdaki gibidir; α: öğrenme katsayısı, Hata: modelin hata fonksiyonu.

$$W_{yeni} = W_{eski} - \alpha \left(\frac{\partial Hata}{\partial W_x}\right) \quad (4.3)$$

Bırakma işlemi nöronları inaktif ederek aşırı öğrenmeyi engellerken, L2 regülasyon ağırlıkların değerini azaltarak eğitimi domine etmelerini engeller.

Veri çoğaltma

Aşırı öğrenmeyi engellemek için kullanılacak diğer bir yöntem de daha fazla veriye sahip olmaktır. Ancak her zaman daha fazla veri toplama olanağı olmayabilir. Bu gibi durumlarda veri çoğaltma yapılarak düşük maliyetle veri sayısı artırılabilir. Var olan girdi üzerinde çevirme, ölçekleme, yakınlaştırma, parlaklık değiştirme, kaydırma gibi işlemler uygulanıp veri çeşitlendirilebilir.



Şekil 4.11. Veri çoğaltma yöntemi ile girdiden oluşturulan görüntüler

Model Değerlendirme Metrikleri

Model mimarisi seçilirken birçok parametre göz önünde bulundurulmalıdır.

- Mimarinin kaç katmanlı olacağı,
- Her katmanda kaç filtre olacağı ve filtre boyutlarının ne olacağı,
- Öğrenme katsayısının değerinin ne olacağı,
- Hangi aktivasyon ve kayıp fonksiyonlarının kullanılacağı,
- Hangi optimizasyon yönteminin kullanılacağı,
- Diğer hiper-parametre ayarlarının nasıl seçileceği gibi birçok soru cevaplanmalıdır.

Bu gibi durumlarda literatür taraması yapılarak benzer veri setleri üzerinde başarım sağlamış modeller temel alınarak işe başlanabilir. Model performansına göre katman sayısı artırılıp azaltılarak ya da hiper-parametre ayarları değiştirilerek eğitim veri setine en uygun model oluşturulur.

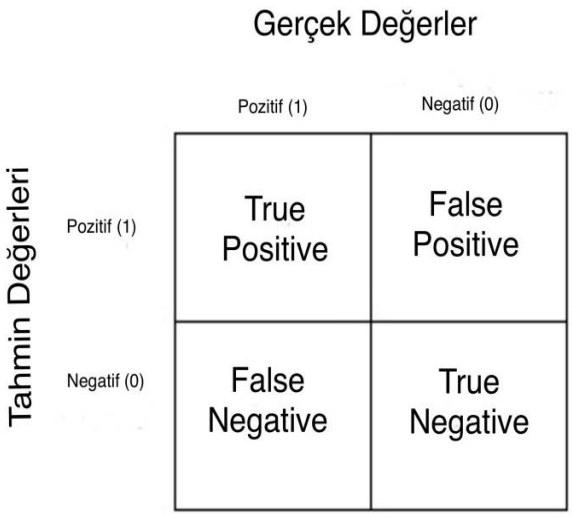
Bir ağda model mimarisini oluşturan bileşenlerin yanı sıra modeli etkileyen diğer bir unsur da hiper-parametrelerdir. Parametre, eğitim boyunca modelin başarımına göre güncellenen manuel müdahale edilemeyen değişkenlerdir. Hiper-parametre ise model programcısı tarafından manuel olarak belirlenen modelin doğru eğitilmesi için ayarlanan değişkenlerdir.

Model performansını ölçmek için bazı metrikler kullanılmaktadır bu metrikler modeli değerlendirip gerekli değişikliklerin yapılması için karar vermede tasarımcıya yardımcı olur.

Doğruluk, modelin eğitim veri üzerinde yaptığı doğru tahmin oranıdır. Fakat tek başına model ölçümü için yeterli değildir. Örneğin, milyonda bir görülen bir hastalık için tasarlanan bir modelde bütün çıktılar negatif olarak üretildiğinde model %99'a varan oranlarda doğruluk gösterebilir. Bu durumda asla hastalık teşhisi koyamayan bir modelin yüksek doğruluğa sahip olduğu görülür. Dolayısıyla ek performans metriklerine ihtiyaç duyulur.

$$doğruluk = \frac{doğru\ tahminler}{toplam\ örnek\ sayısı} \tag{4.4}$$

Karmaşıklık matrisi, sınıflandırma modellerini değerlendirmek için kullanılmaktadır. Tahmin sonuçları doğruluklarına göre kategorilere ayrılarak bir matrise yerleştirilir. Doğru pozitifler, modelin doğru tahmin sayısını, doğru negatifler, modelin doğru sınıflandırdığı yanlışların sayısını, yanlış pozitifler modelin doğru olarak sınıflandırdığı yanlışların sayısını, doğru yanlışlar ise modelin yanlış olarak sınıflandırdığı doğruların sayısını ifade eder. Bu matrise bakarak model başarımı değerlendirilebilir.



Şekil 4.12. Karmaşıklık matrisi

Eğer bir modelde yanlış negatiflerin oranı önemli ise duyarlılık (recall) değerine bakılır. Örneğin hastalık teşhisi yapan bir modelde, yanlışlıkla sağlıklı bireylerin hasta kabul edilmesi göz ardı edilebilir. Fakat gerçekten hasta bireylerin sağlıklı kabul edilmesi bireyin

hayatı için oldukça tehlikelidir. Bu nedenle bu modelde önemli olan yanlış negatiflerin düşük orana sahip olmasıdır.

$$duyarlılık = \frac{doğru\ pozitif\ sayısı}{doğru\ pozitif\ sayısı + yanlış\ negatif\ sayısı} \tag{4.5}$$

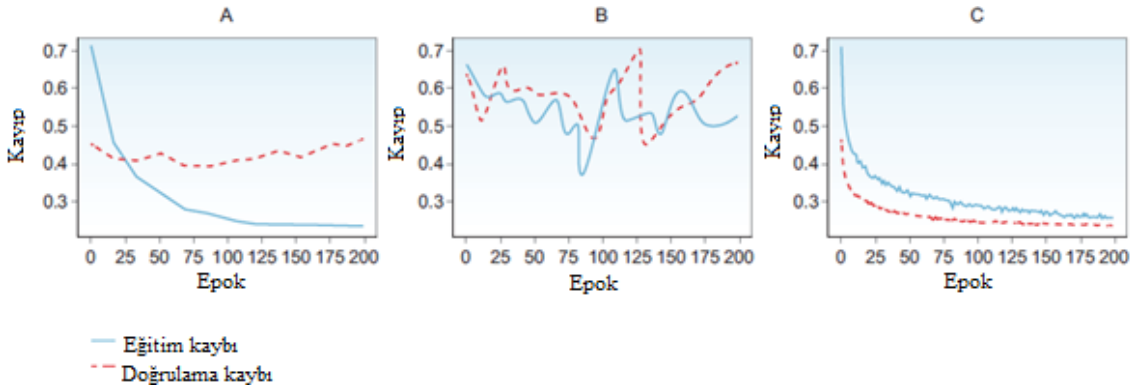
Kesinlik (precision) ise duyarlılığın tam tersidir. Ne kadar sağlıklı bireyin yanlışlıkla hasta olarak tahmin edildiği bilgisini verir. Yani yanlış doğruların önemli olduğu olaylarda kesinliğe bakılır.

$$kesinlik = \frac{doğru\ pozitif\ sayısı}{doğru\ pozitif\ sayısı + yanlış\ pozitif\ sayısı} \tag{4.6}$$

Bazı modellerde hem yanlış negatif sayısı hem de yanlış pozitif sayısı aynı oranda önemli olur. Bu durumda modelin F-puanlamasına bakılır. F-puanlama, duyarlılık (d) ve kesinlik (k)'ın harmonik ortalamasıdır.

$$F - puanlama = \frac{2dk}{d+k} \tag{4.7}$$

Ayrıca modelde aşırı öğrenme ve yetersiz öğrenme hakkında yorum yapabilmek için tek tek model hatalarını incelemek yerine kayıp eğrisi grafiğine bakılarak yorum yapılabilir. Grafikte modelin eğitim ve doğrulama veri setleri üzerinde uyumlu çıkarımlar yapması beklenir. Aşağıdaki şekilde A grafiği aşırı öğrenme olduğunu, B grafiği yetersiz öğrenme olduğunu, C grafiği ise modelin başarılı olduğunu gösterir.

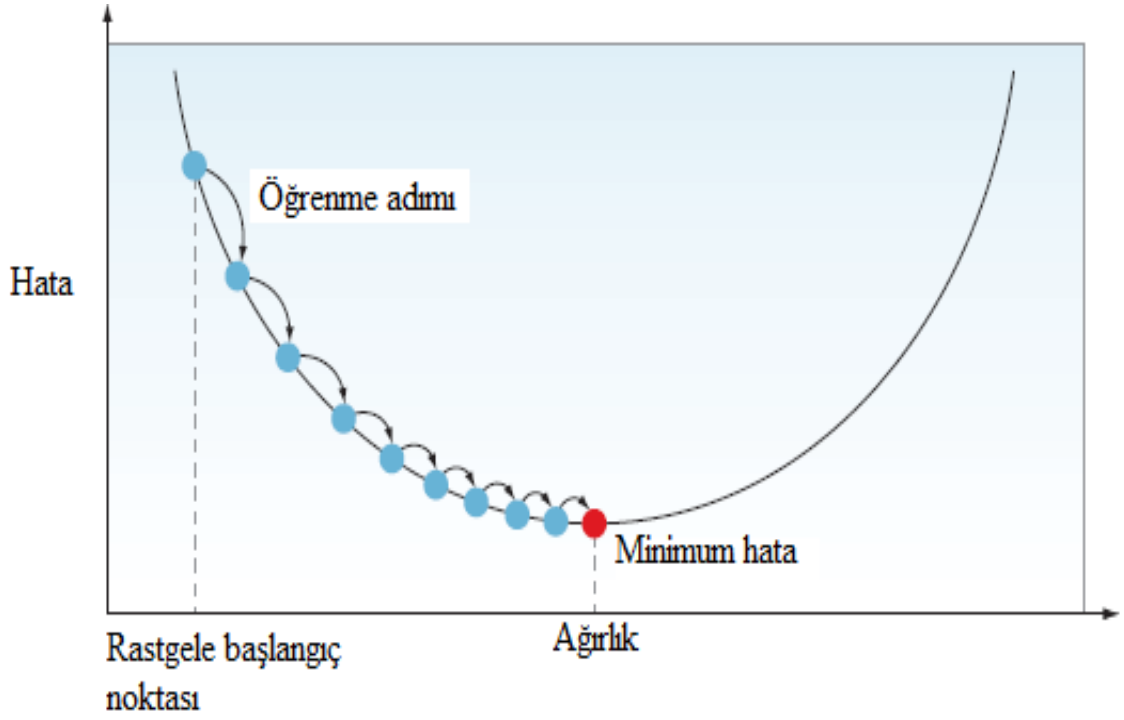


Şekil 4.13. Derin ağlarda öğrenme durumları

Model değerlendirme metriklerini tanımlamak gerekli bir adımdır. Bu değerler sistemi geliştirmek için tasarımcıya rehberlik eder. Tasarımcı hangi hiper-parametreyi ne yönde değiştirmesi gerektiği hakkında fikir sahibi olur.

Öğrenme Katsayısı

Derin ağların eğitiminde en önemli hiper-parametrelerden biri öğrenme katsayısıdır. Öğrenme katsayısı eğitim sırasında minimum hatayı verecek ağırlığın bulunması için ağırlık güncellemelerinde kullanılır. Uygulanan optimizasyon algoritmasının hızını-adım sayısını belirler.



Şekil 4.14. Öğrenme katsayısı ve global minimum arasındaki ilişki

Küçük öğrenme katsayısının lokal minimumu bulma ihtimali daha yüksektir fakat ağırlık değişimi çok yavaş olduğu için eğitim süresi uzundur. Öğrenme katsayısı çok büyük seçildiğinde de lokal minimumu isabet ettirme ihtimali düşer. Dengeli bir öğrenme katsayısı seçilmelidir. Öğrenme katsayısı statik bir değer olabileceği gibi eğitim sırasında güncellenerek değeri düşürülebilir. Örneğin her beş epoch'da yarıya indirilebilir ya da üssel olarak azaltılabilir.

Yığın (Batch) Normalleştirme

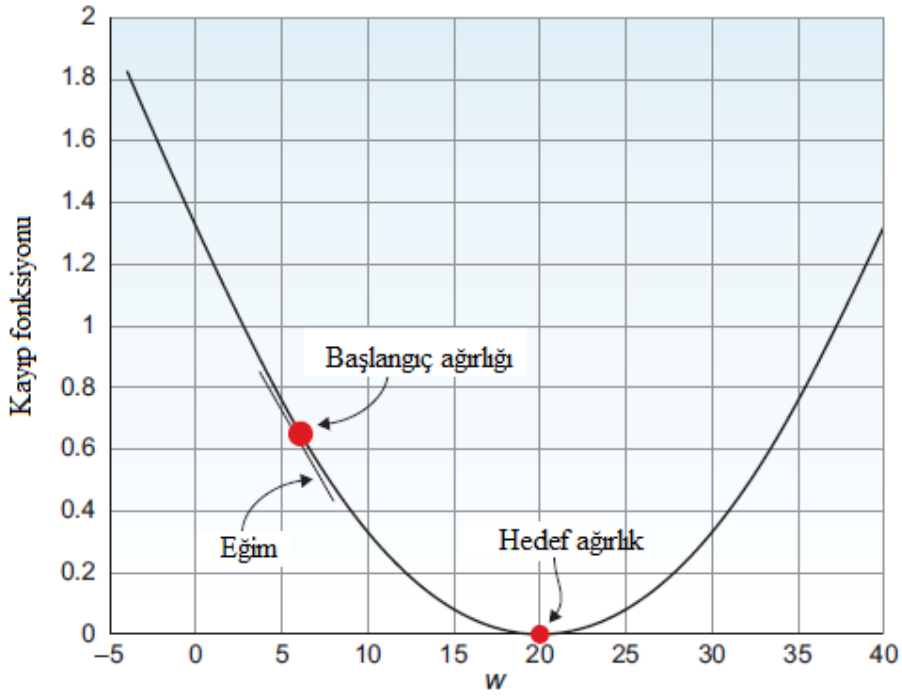
Modelin eğitim süresini artırmak için eğitimden önce veri normalleştirilir. Bu bakış açısından yola çıkarak öğrenme sırasında katman çıktılarının da normalleştirilerek eğitimin hızlandırılabileceği gösterilmiş, her katmanda aktivasyon kodundan önce yığın (batch) normalleştirme yapılmıştır [91]. Gizli katmanlara dağılımı sıfıra ortalanmış girdiler verilir. Gizli katman değerlerinin dağılım değişimlerinin dereceleri azaltılarak ağın daha stabil olması sağlanır.

Seçilen mini-yığın örneklerinin aritmetik ortalamaları hesaplandıktan sonra her girdiden bu ortalama değer çıkarılır. Çıkan değerlerin karesi alınıp toplanarak varyans değeri hesaplanır. Varyans mini-yığındaki veri sayısına bölünerek verilerin dağılımda bulunma aralıkları hesaplanır. Daha sonra normalleştirme işlemi yapılarak veri dağılımı sıfır seviyesinde ortalanır.

Model Optimizasyonu

Derin ağların eğitiminde tahmin edilen çıktı ile gerçek çıktı arasındaki fark en aza indirmeye çalışılır. Hata gerçek çıktı değeri ile tahmin edilen çıktı değeri arasındaki farktır. Ağ kabul edilebilir düşük hataya veya global minimuma eriştiğinde eğitim durdurulur. Eğitim durana kadar parametreler hatayı en aza indirecek şekilde güncellenerek eğitim tekrarlanır. Ağın toplam hatasını hesaplayan fonksiyona maliyet, kayıp veya hata fonksiyonu denir. Modelde kullanılacak hata fonksiyonu tasarımcı tarafından seçilir. Örneğin, Mean Squared Error, cross-entropy, exponential, hellinger distance, kullback-leibler divergence gibi hata fonksiyonları kullanılır.

Optimizasyon algoritmaları model parametrelerini güncelleyerek kayıp fonksiyonunu minimize etmeye ve eğitim hatasını düşürmeye çalışır. Gradyan azaltma kayıp fonksiyonunun minimum değerini bulmak için kullanılan optimizasyon yöntemlerinden biridir. Gradyan bir fonksiyonun eğimi olarak tanımlanabilir. Kayıp fonksiyonunu minimize etmek için gradyanın sıfıra yakınsaması istenir. Bunun için rastgele bir başlangıç değerinden başlayarak parametrelerin kayıp fonksiyonunu minimum yapan optimal değerleri bulunmaya çalışılır.



Şekil 4.15. Kayıp fonksiyonu gradyanı ve global minimum

Bir nöron ağırlığı için lokal hatanın eğimini ifade eden gradyan parçalı türev kullanılarak hesaplanır. Hesaplanan gradyan ağırlığı güncellemek için kullanılır. Ağırlık güncellemesinde gradyanı global minimuma gidecek yolun yönü, öğrenme katsayısını da adım sayısı olarak düşünebiliriz. Gradyan azaltma ise bir ağırlığın ağın toplam hatasını minimize edecek değerini bulmak için kullanılan bir optimizasyon yöntemidir. Ağırlıklar hata fonksiyonunun global minimumu bulunana kadar iteratif olarak güncellenir. Her adımda ağırlıktaki değişim aşağıdaki gibi hesaplanır; E: hata fonksiyonu, α: öğrenme katsayısı.

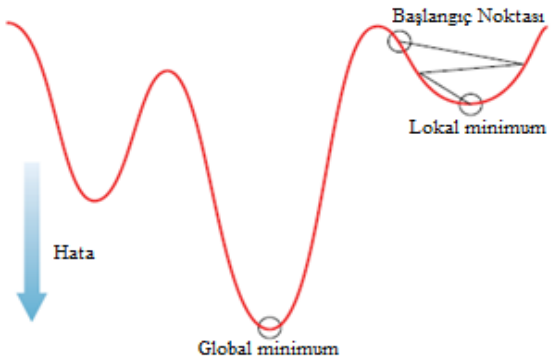
$$\Delta w_i = -\alpha \frac{dE}{dw_i} \tag{4.8}$$

Ağırlık güncellemesi aşağıdaki şekilde yapılır:

$$w_{yeni} = w_{şimdiki} + \Delta w \tag{4.9}$$

SGD Optimizasyon Algoritması

Temel gradyan azaltma algoritmasında ağırlıklar öncelikle rastgele başlatılır. Şekilde görüldüğü gibi bir başlangıç noktasının rastgele seçilmesi lokal minimumu global minimum gibi gösterebilir. Bu durumda minimum hataya ulaşılamamış olur. Ayrıca temel gradyan azaltma fonksiyonu ağırlık güncellemek için her adımda tüm veri setini kullanır. Bu da işlemlerin çok yavaş olmasına sebep olur. Bu problemleri çözmek için 'stokastik' gradyan azaltma yöntemi önerilmiştir.



Şekil 4.16. Birden çok lokal minimuma sahip kayıp fonksiyonu

'Stokastik' kelimesi SGD'nin rastgeleliğini temsil eder. Her iterasyonda veri setinden 'yığın (batch)' denen rastgele veriler seçilir ve gradyan bu veriler üzerinden hesaplanarak parametre güncellemesi yapılır. Her adımda bütün veri setini kullanmak yerine rastgele seçilen veri örnekleri kullanılır. SGD'de aynı anda birden çok rastgele başlangıç noktası seçilir. Böylece fonksiyon eğrisindeki her tepenin lokal minimumları hesaplanır. En küçük lokal minimum da global minimum olarak seçilir.

Veri Önişleme

Model doğruluğunu artırmak için veri seti eğitimden önce bazı önişleme adımlarından geçirilir ve probleme göre veri önişleme teknikleri uygulanır. Derin öğrenme teknikleri diğer makine öğrenmesi uygulamalarına göre daha az önişleme gerektirir.

Görüntü sınıflandırmada uygulamalarında önişleme adımı, resim boyutlandırma, etiket hazırlama, veri setini bölme, normalleştirme, resmi gri seviyeye dönüştürme, veri çoğaltma gibi adımları kapsar.

Evrişimli ağlarda işlenen her veri aynı boyutta olmalıdır. Bu yüzden girdiler arasında boyut farkı varsa yeniden boyutlandırma yapılmalıdır.
Kategorik veri sınıflandırma gibi işlemlerde sınıf kategorileri one-hot kodlama gibi tekniklerle etiketlenebilir. Bu işlem kategorileri ikili gösterimle ifade edebilmek için kullanılır. Kategoriler sadece o kategoriyi 1 ile işaretleyen diğer değerleri 0 olan vektörler ile ifade edilir.

Etiketleme

Yemek adı	Kategori	Kalori
Elma	1	95
Tavuk	2	231
Brokoli	3	50

→

One-Hot Kodlama

Elma	Tavuk	Brokoli	Kalori
1	0	0	95
0	1	0	231
0	0	1	50

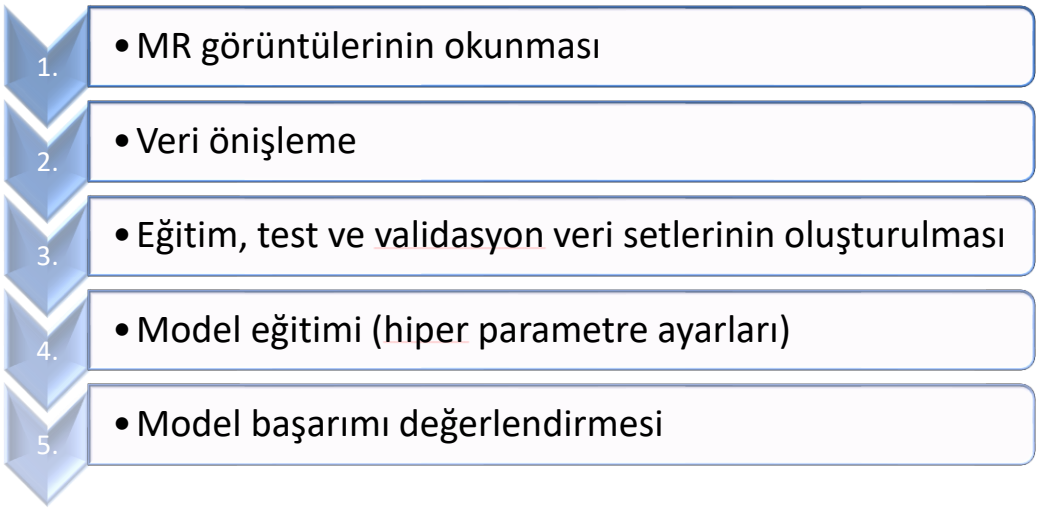
Şekil 4.17. Kategori etiketleme

Veri setini bölme işleminde ise veri eğitim, doğrulama ve test olarak bölünür. Eğitim veri seti modeli eğitmek için, doğrulama veri seti her adımdan sonra modeli değerlendirmek için kullanılırken, test veri seti de modelin bilmediği bir veri seti üzerindeki davranışını gözlemlemek için kullanılır.

Normalleştirme ise her girdi özelliğinin benzer dağılımda olması için uygulanır. Resim girdi verisi için her pikselden ortalama çıkarılıp standart sapmaya bölünerek normalleştirme yapılır. Bu işlem eğitim sürecini hızlandırır.

4.3. Uygulama Tasarımı

Uygulama verinin ön işlemesi, model için veri setlerinin hazırlanması, model mimarisinin oluşturulması, eğitim ve test aşamalarından oluşmaktadır. Aşağıdaki şekilde uygulamanın ana hatları belirtilmiştir.



Şekil 4.18. Uygulama temel adımları

Uygulamanın adımlarında ilk olarak veri setindeki verilerin kullanılan veri yapısına uygun olarak okunması sağlanmaktadır. NIfTI formatındaki veriler 3 boyutlu tensor'lar olarak okunur. Veri okuma işlemi tamamlandıktan sonra hem işlem maliyetini azaltmak hem de eğitimin daha yüksek doğruluk sağlamak amacıyla görüntü önişleme adımı gerçekleştirilir. Bu adımda veri setinde bulunan tümör maskeleri kullanılarak görüntülerdeki tümör bölgeleri işaretlenir ve kaydedilir. Eğitim sırasında segmentlenmiş olan bu görüntüler kullanılır. Daha sonra hem referans model ele alınarak hem de literatürde önerilen oranlar dikkate alınarak hiper parametre ayarları yapılır. Kullanılan model bu parametreler ile eğitilerek model başarımı değerlendirilir.

4.3.1. Veri seti

Literatürde taranan beyin MR görüntüleri ile yapılan çalışmalarda en çok BraTS veri setinin kullanıldığı görülmüş ve bu çalışmada da MR görüntülerinden oluşan BraTS 2020 yılı veri seti [92-94] kullanılmıştır.

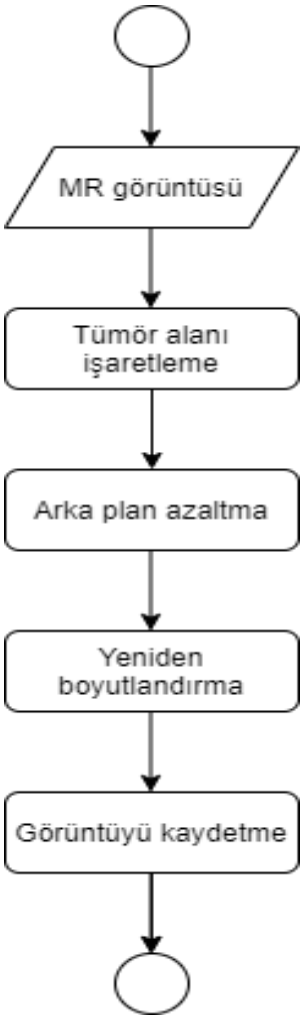
Veri setinde vaka başına (T1, T1c, T2 ve FLAIR olmak üzere) 4 adet üç boyutlu MR görüntüsü bulunmaktadır. Veriler farklı MR tarayıcıları kullanılmış farklı kurumlardan toplanmıştır. Görüntüler için doktorlar tarafından koyulan tümör tanıları da veri setinde bulunmaktadır.

4.3.2. Veri önişleme

Bu çalışmada veri setindeki NIfTI formatındaki 3 boyutlu T1-ağırlıklı görüntüler kullanılmıştır.

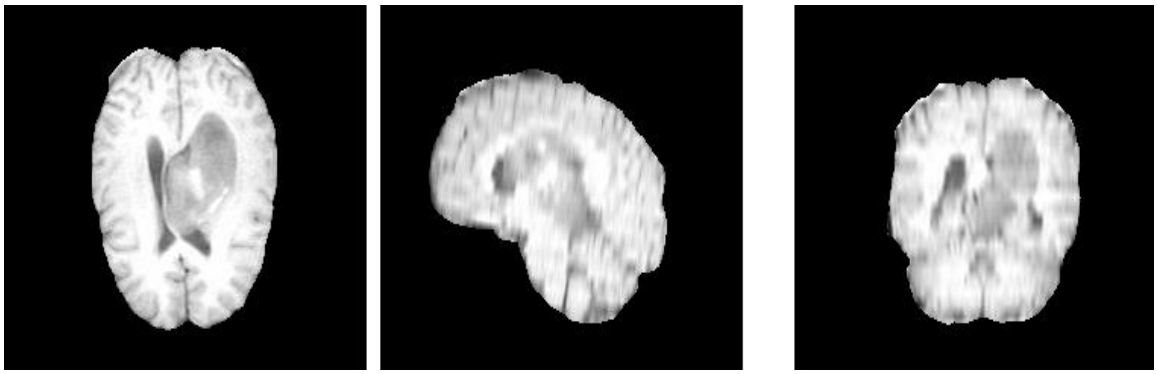
NIfTI formatı:

NIfTI (Neuroimaging Informatics Technology Initiative) nöro görüntüleme yapmak için kullanılan bir dosyalama formatıdır. NIfTI dosyaları, nörobilim ve hatta nöroradyoloji araştırmaları için görüntüleme bilişiminde çok yaygın olarak kullanılmaktadır. Klinik ortamda DICOM dosyaları standarttır, ancak bazı görüntüleme bilişim araçları DICOM dosyalarını otomatik olarak NIfTI formatına dönüştürebilmektedir. Tez kapsamında yapılan uygulamada kullanılan veri setindeki veriler bu formatta yayınlanmaktadır.

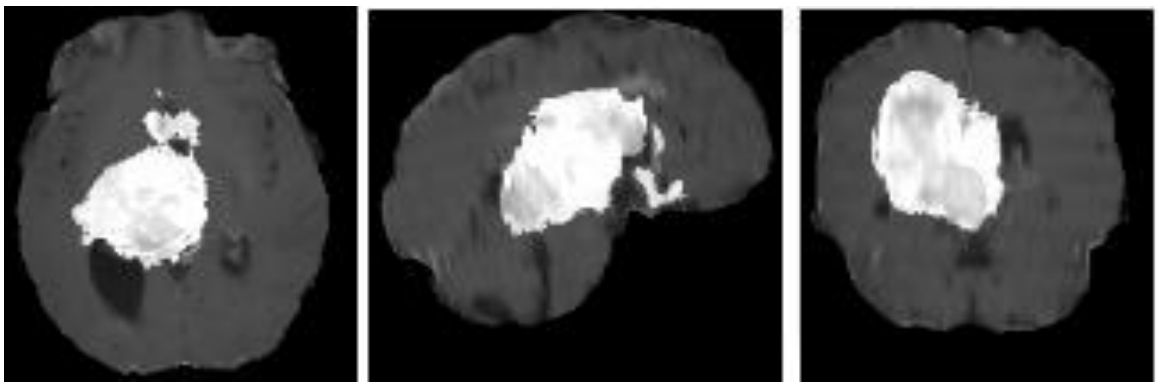


Şekil 4.19. Veri önişleme adımları akış şeması

Veri önişleme adımında öncelikle MR görüntüleri okunarak veri setinde bulunan tümör maskeleri kullanılarak tümör işaretlemesi yapılır. Tümör bölgesi işaretlenen görüntü segmentlenmiş görüntü olarak tekrar kaydedilir. Ayrıca, modelin tümör yapısını daha iyi algılamasını sağlamak için tümör alanı genişletilerek vurgulanmıştır. Daha sonra, hesaplama karmaşıklığını azaltmak için beyin boyutu sabit tutularak arka plan azaltma uygulanmıştır. Son olarak, CNN ağına verilecek girdilerin aynı boyutta olması ve yine hesaplama karmaşıklığını azaltmak için yeniden boyutlandırma yapılmıştır. Şekil 4.19'de veri önişleme adımları akış diyagramı içinde gösterilmiştir. Aşağıda ise veri seti içerisinde bulunan önişleme yapılmış bir veri örneği gösterilmektedir.



Şekil 4.20. Veri setinde bulunan T1 ağırlıklı HGG gliom örneği



Şekil 4.21. HGG gliom görüntüsünde yapılan tümör işaretlemesi

4.3.3. Eğitim, test ve doğrulama veri setlerinin oluşturulması

Veri setinde 136 adet HGG tümörlü, 76 adet LGG tümörlü hasta MR görüntüsü bulunmaktadır.

Çizelge 4.1. Sınıflara göre veri sayısı

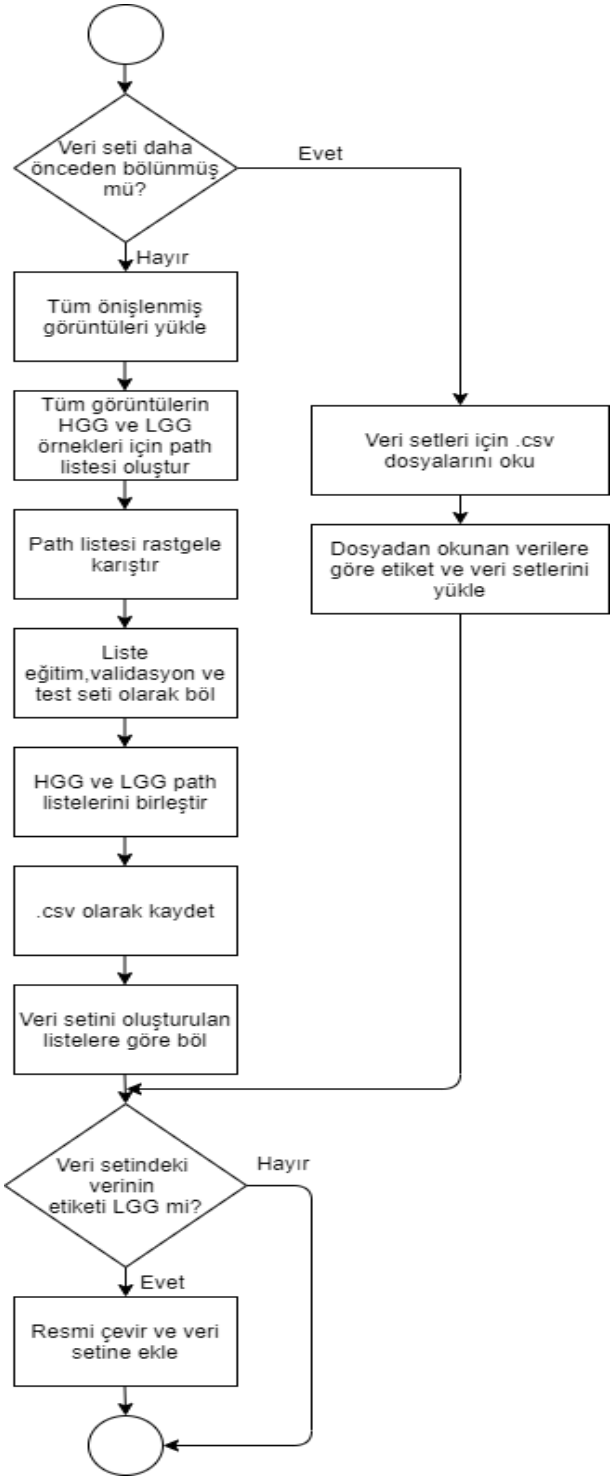
Sınıf	T1-MRG Sayısı
HGG	136 adet
LGG	76 adet (152)

Veri seti; eğitim (%60), doğrulama (%20) ve test (%20) olarak üç parçaya ayrılmıştır. LGG sınıfında HGG sınıfına göre daha az veri bulunduğu için LGG verileri üzerinde veri çoğaltma uygulanmıştır. Bunun için beynin sağ ve sol tarafları eşit kabul edilerek görüntüler ters çevrilmiştir. Bu yolla LGG veri seti iki katına çıkmış, iki sınıf arasındaki dengesizlik giderilmiştir. Veri çoğaltma ile aşırı öğrenme oranının düşürülmesi de hedeflenmiştir.

Çizelge 4.2. Veri seti dağılımı

Veri Seti	
Eğitim	174
Doğrulama	42
Test	42

Veri setleri bölünürken hangi görüntünün hangi veri setinde olduğu .csv uzantılı dosyalara kaydedilmiştir. CSV dosyasında hem görüntünün path (dosya yolu)'i hem de etiketi tutulmaktadır. Bu dosyalar bir kez oluşturulduğunda uygulama tekrar çalıştığında bu dosyalar okunarak veri seti yüklemesi yapılmaktadır. Şekil 4.22'de veri setlerinin bölünmesi işlemi akış diyagramı olarak verilmiştir.



Şekil 4.22. Veri setlerinin bölünmesi işlemi adımları

Burada program ilk olarak veri setinin daha önce bölünüp bölünmediğine bakılır. Veri seti daha önce bölünmüş ise eğitim, test ve doğrulama (validation) veri setlerinde hangi örneklerin olduğunu gösteren .csv dosyaları okunarak veriler ve etiketleri yüklenir.

Veri seti ilk kez bölünüyorsa, .csv dosyaları oluşturulur. Bunun için önişleme adımında işlenen veriler yüklenir, HGG ve LGG örnekleri için path listesi oluşturulur. Liste rastgele karıştırıldıktan sonra belirtilen oranlarda üçe bölünerek dosya isimleri .csv dosyalarına kaydedilir. Kaydetme işlemi yapılırken veri setinde daha az sayıda bulunan LGG verilerine veri çoğaltma uygulanır. Bunun için LGG verileri ters çevrilerek yeni veri olarak kaydedilir.

4.3.4. Model eğitimi ve hiper parametre seçimi

Uygulanan model mimarisi

Kullanılan model 3D beyin MR görüntülerini kullanarak gliom sınıflandırma yapmaktadır. Sınıflandırmada dünya sağlık örgütünün belirlediği evreler dikkate alınmıştır. Sınıflandırmada HGG ve LGG olmak üzere iki kategori bulunmaktadır.

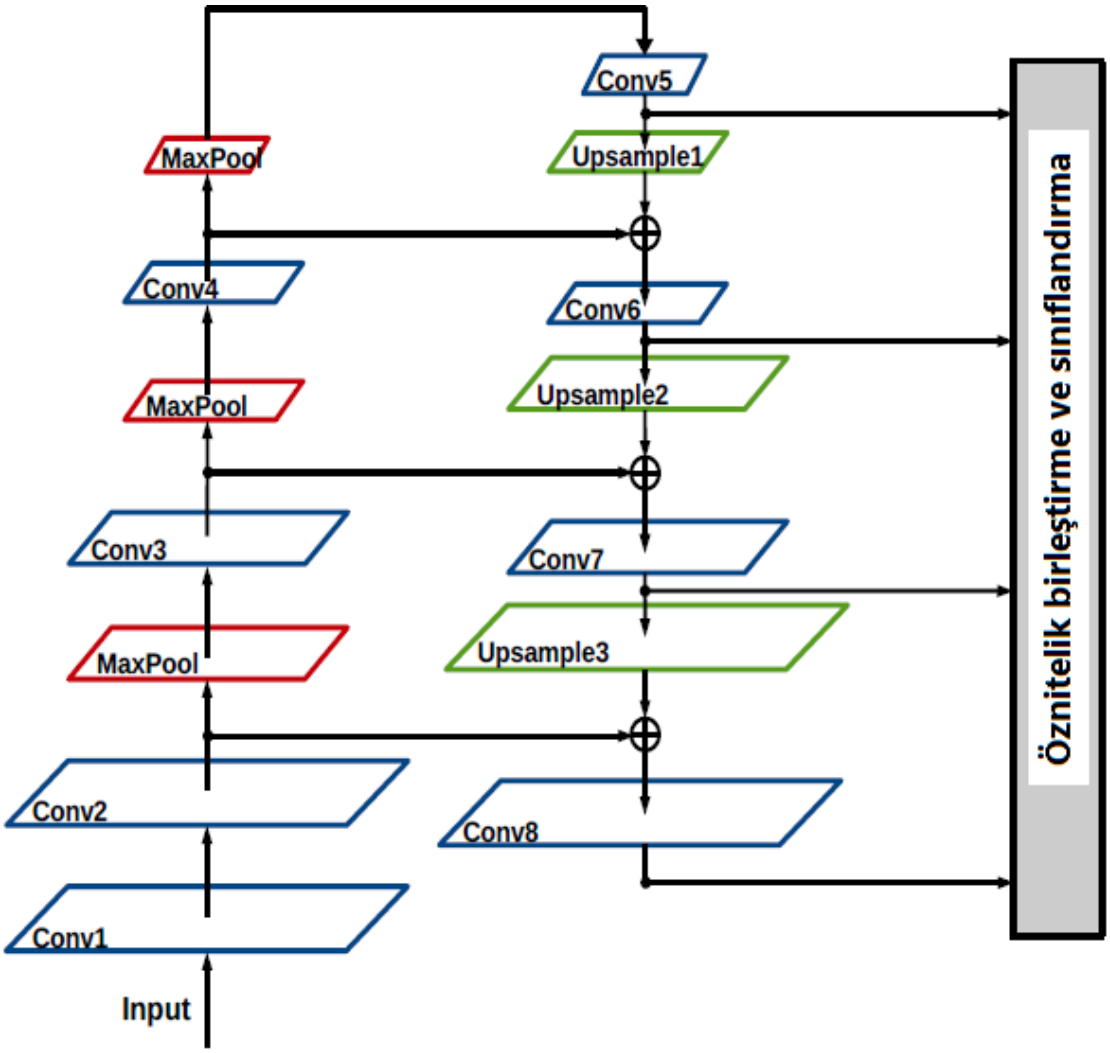
Model yapısı tasarımında görüntü üzerinde detaylı özellik çıkarımları yapabilen evrişimli sinir ağları kullanılmıştır. CNN ağının ara katmanlarındaki çıktılar özellik çıkarımına dahil edilir.

Çok ölçekli 3D evrişimli sinir ağları

Medikal görüntü işlemede genellikle tek boyutlu CNN ağları kullanılmaktadır. Fakat beyin MR görüntülerinde tümör yapısı, doku yoğunluğu, tümör konumu gibi farklı özellik yapıları bulunmaktadır. Beyin MR görüntülerinde böyle geniş çaplı özellikleri daha detaylı çıkarabilmek için çok ölçekli evrişim ağlarının kullanılması önerilmiştir [92]. Bu ağlarda birden çok özellik çıktısı bulunur.

Beyin MR görüntülerinde doku özelliklerinin iyi ayırt edilebilmesi için yüksek çözünürlüklü özelliklere ihtiyaç duyulur. Ağın ilk katmanlarında anlamsal özellikler zayıf, çözünürlük yüksektir. İlerleyen katmanlarda ise anlamsal özellikler güçlenirken çözünürlük düşmektedir. Önerilen modelde yüksek çözünürlüklü düşük anlamlı özellikler ile düşük çözünürlüklü yüksek anlamlı katmanlar toplanarak yüksek çözünürlüklü, güçlü anlamsal özellikler taşıyan özellik haritaları oluşturulur. Böylece her ölçek için zengin anlamsal özellikler taşıyan çıktılar oluşturulmuş olur.

Her boyutta toplama işleminden sonra oluşturulan özellik haritaları bir araya getirilir ve düzleştirme (flatten) uygulanarak özellik vektörleri oluşturulur. Oluşturulan özellik vektörleri tam bağlı yapay sinir ağlarında sınıflandırılır.



Şekil 4.23. Çok ölçekli CNN modelinin detaylı gösterimi [95]

Ağda standart sıralı evrişim ağı yapısından oluşan beş katman (Conv1-Conv5) ve birleştirilmiş üç katman (Conv6-Conv8) bulunmaktadır. İlk beş katmanda temel öznitelikler çıkarılırken, birleştirme katmanlarında daha detaylı özelliklerin çıkarılması hedeflenmiştir. Ağda her katman çıkışına yığın normalleştirme uygulanmış, aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Birleştirmede Conv5 en yakın komşuluğa göre 2 faktöründe yukarı örnekleme yapılarak çözünürlük artırılmış ve kendinden önceki katmanla birleştirilerek Conv6 katmanına girdi olarak verilmiştir. Ayrıca Conv5 katmanının çıktısı ilk özellik ölçeğini oluşturmaktadır. Diğer ölçeklerde aynı şekilde hesaplanmıştır. En son farklı

katmanlardan çıkarılan özellik ölçekleri aynı boyutta ortaklama uygulandıktan sonra düzleştirilerek tam bağlı katmanlarda sınıflandırılmıştır. Çizelge 4.3'te tasarlanan ağın katman yapısı listelenmiştir.

Çizelge 4.3. Uygulanan 3D çok ölçekli CNN ağı mimarisi

CNN Ağ Katmanı	Kernel	Çıktı Boyutu
Conv1+ReLU+BN	5*5*5*32	56*48*48*32
Conv2+ReLU+MaxPool+BN	3*3*3*64	28*24*24*64
Conv3+Relu+MaxPool+BN	3*3*3*128	14*12*12*128
Conv4+ReLU+MaxPool+BN	3*3*3*256	7*6*6*256
Conv5+ReLU+BN	3*3*3*256	7*6*6*256
UpSample1+MergeConv4+BN	-	14*12*12*256
Conv6+ReLU	3*3*3*128	14*12*12*128
UpSample2+MergeConv3+BN	-	28*24*24*128
Conv7+ReLU	3*3*3*64	28*24*24*64
UpSample3+MergeConv3+BN	-	56*48*48*64
Conv8+ReLU	3*3*3*32	56*48*48*32
FC1-1	-	256
FC1-2	-	256
FC1-3	-	256
FC1-4	-	256
Concatenation+BN	-	1024
FC2	-	256
FC3	-	2

Uygulamada üç boyutlu CNN ağları kullanılmıştır. 3D evrişim ağlarında veri setine üç boyutlu filtreler uygulanır. Uygulanan filtre düşük seviyeli özellik haritalarını hesaplamak için üç yönlü (x, y, z) hareket eder. Özellik çıktıları küp veya küboid gibi üç boyutlu bir hacim alanına sahiptir. Videolarda, tıbbi görüntülerde vb. kullanılır [96]. Medikal görüntü sınıflandırma uygulamalarında kullanılan çoğu yaklaşım sadece 2D görüntüler işleyebilirken, klinik uygulamalarda kullanılan çoğu medikal veri 3D hacimlerden (volume) oluşmaktadır. 2D örüntülerin çoğu giriş verilerini dilim dilim (slice by slice) işleyen farklı ağ derinliklerine sahip evrişim ağlarına dayanmaktadır. 3D CNN ağı ise genel olarak üç

boyutlu bağlamsal özellik çıkarma yeteneğine sahiptir. 3D CNN modelleri, dilim dilim analiz etmenin aksine girdi dilimlerini hacimsel olarak analiz eder ve girdi dilimleri üzerinde bütünsel özellikler oluşturur. Bu nedenle 3D CNN modelleri doğruluğu 2D CNN modellerinin doğruluğundan daha iyidir [97]. Çiçek ve diğerleri [98] tarafından, gelişen GPU olanakları da kullanılarak başarılı bir 3D CNN ağı modeli medikal görüntülerde uygulanmıştır.

Model mahremiyetini sağlamak için gradyan tabanlı diferansiyel mahremiyet uygulanmıştır. Bunun için optimizasyon algoritması olarak DP SGD algoritması kullanılmıştır.

Hiper parametre seçimi

Model eğitiminde kullanılan parametre seçimleri hem mahremiyetsiz referans model [95] hem de literatürde bulunan öneriler dikkate alınarak yapılmış, sonuçlara göre parametrelerde ayarlamalar yapılarak model tekrar tekrar eğitilmiştir. Temel ve en iyi başarımı gösteren parametre ayarları aşağıdaki gibidir:

- girdi boyutu: 112x96x96
- öğrenme katsayısı: 0,003
- epok sayısı: 100
- yığın (batch) sayısı: 4
- düzleştirme (flatten) katmanına ve birleştirme katmanlarına uygulanan bırakma (dropout) oranı: 0,5
- çıktı katmanı hariç tüm katmanlara uygulanan L2 regülasyon parametresi: 0,00005
- yığın normalleştirme (batch normalization) momentumu: 0,9
- tüm katmanlar ‘glorot-uniform’ metodu ile başlatılmıştır.
- modeli değerlendirmesi için cross-entropy kaybı ve doğruluk metrikleri göz önüne alınmıştır.

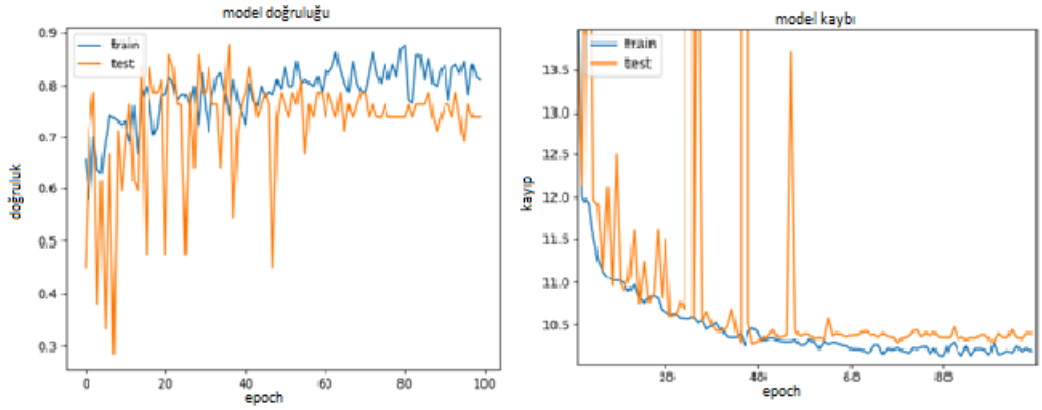
DP-SGD ile optimizasyon sırasında modele eklenen gürültü katsayısı parametresi ‘noise_multiplier’ ve ‘l2_norm_clip’ parametreleri [99]’ de önerildiği gibi sırasıyla 1,3 ve 1,5 olarak seçilmiştir. ‘l2_norm_clip’ parametresi her mini yığın üzerinde hesaplanan

gradyanın maksimum öklid normunu vererek gradyan hassasiyetini sınırlandırırken ‘noise_multiplier’ eklenen gürültü miktarını belirlemektedir.

4.3.5. Sonuçlar

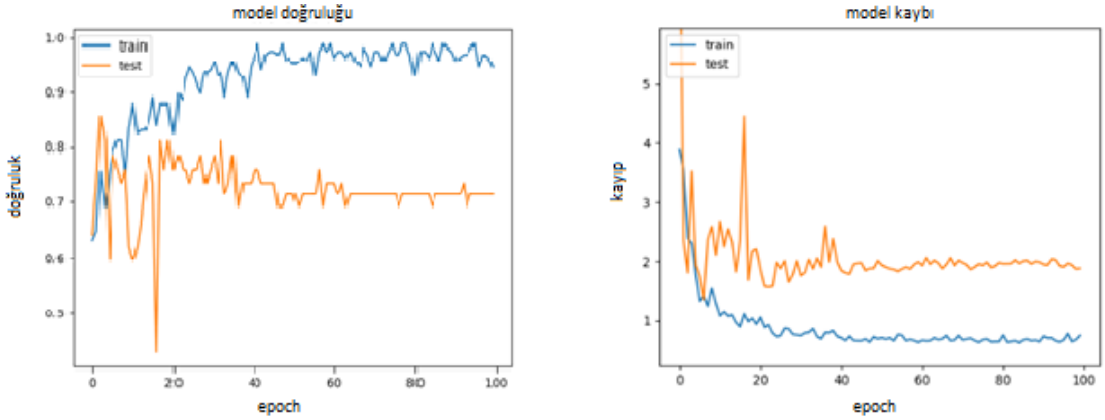
Uygulama IntelCore-i7 3.50 GHz işlemci, 16 GB Ram ve NVIDIA GeForce GTX 970 ekran kartına sahip Windows 10 işletim sistemi kurulu bir bilgisayarda çalıştırılmıştır. Açıklanan model mimarisi Tensorflow altyapısı ile Keras kütüphanesi kullanılarak gerçekleştirilmiştir.

İlk olarak yığın (batch) boyutunun eğitime etkisni görmek için model diğer parametreler sabit tutularak 4 ve 8 yığın sayıları ile ayrı ayrı eğitilmiştir. Yığın sayısının artırılmasıyla doğruluk ve kayıp grafiklerine bakıldığında eğitim seti üzerindeki başarımın arttığı fakat test verileri üzerindeki başarımın düşüp kayıp oranının arttığı gözlemlenmiştir. Yığın sayısının artırılması aşırı öğrenmeye sebep olmuştur. Eğitim sonuçları Şekil 4.24 ve şekil 4.25’teki gibidir.

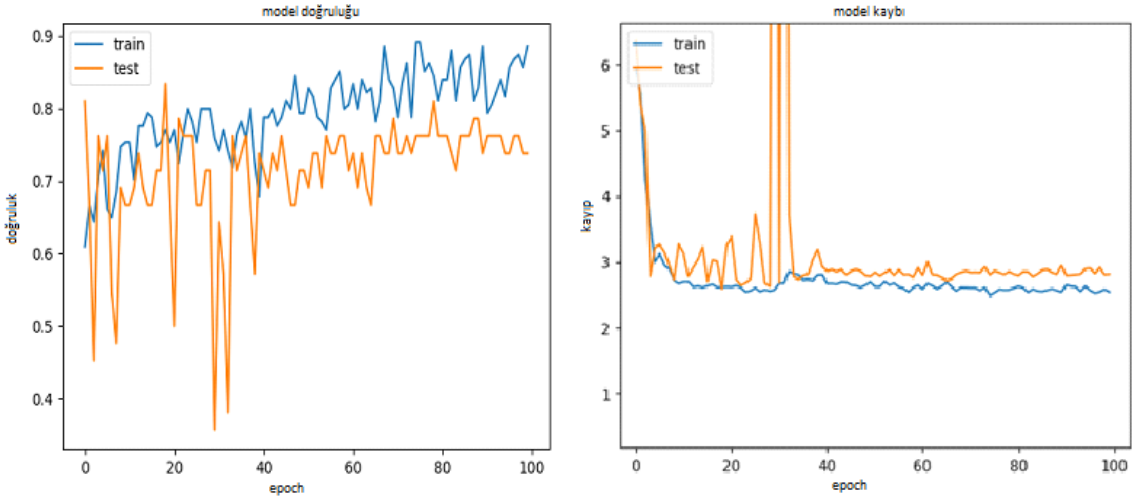


Şekil 4.24. Yığın (batch) sayısı 4 olarak seçilen modelin doğruluk kayıp grafikleri

Yığın (batch) sayısı 4 olarak seçilen modelin doğruluğu referans model ile karşılaştırıldığında ise kayıp oranının daha yüksek olduğu gözlemlenmiştir. Kayıp oranının düşürülmesi için batch sayısı 4 seçilip diğer parametreler değiştirilerek model tekrar eğitilmiştir. Referans modelin çıktıları Şekil 4.26’de verilmiştir.

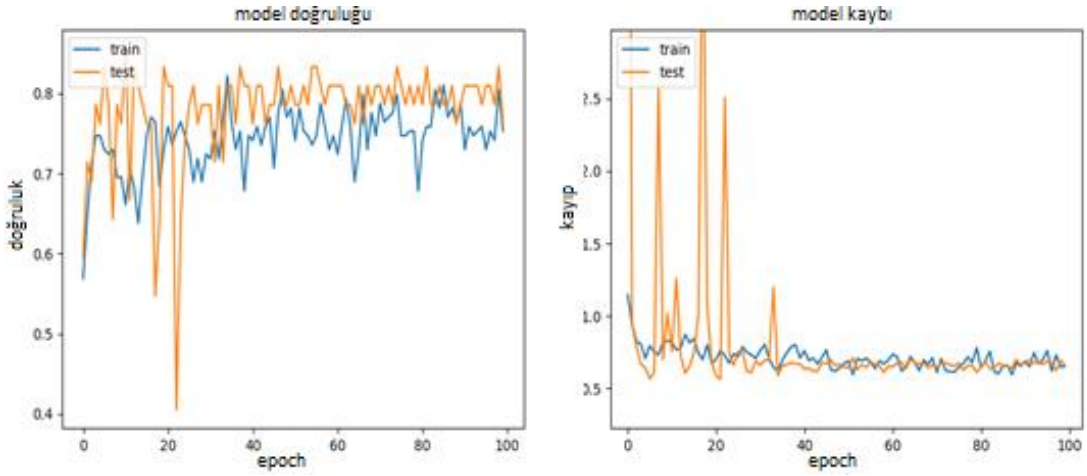


Şekil 4.25. Yığın (batch) sayısı 8 olarak seçilen modelin doğruluk kayıp grafikleri



Şekil 4.26. Referans modelin, doğruluk kayıp grafikleri

Model doğruluğunu artırmak için sonraki adımda öğrenme katsayısı 0,25'ten 0,003'e düşürülmüştür. Aynı zamanda 1.3 seçilen noise_multiplier gürültü katsayısı 0.1'e l2_norm_clip parametresi de 1.0'a düşürülmüştür.



Şekil 4.27. DP-SGD ile eğitilen modelde en yüksek başarımı gösteren doğruluk kayıp grafikleri

Çizelge 4.4. DP-SGD ile eğitilen modelin en yüksek başarımının metrik değerleri

Metrik İsmi	Değer
Doğruluk	0.844827586
HGG Doğruluk	0.951219512
LGG Doğruluk	0.75
Kayıp	0.349930659
HGG Kayıp	0.151407086
LGG Kayıp	0.526875584
HGG Precision	0.772277228
LGG Precision	0.94520548
HGG recall	0.951219512
LGG recal	0.75

Çizelge 4.5. DP-SGD ile eğitilen modelin en yüksek başarımının karmaşıklık matrisi

TP:78	FP:23
FN: 4	TN:69

Modelin mahremiyet seviyesi, yani diferansiyel mahremiyetin sağladığı mahremiyet garantisini ölçmek için kullanılan hassasiyet ve mahremiyet maliyeti parametreleri hesaplanarak modelin farklı yığın boyutlarındaki mahremiyet oranları da karşılaştırılmıştır. Temel olarak veri setindeki eleman sayısının tersinden daha küçük olması önerilmiştir [95]. Bu çalışmada hassasiyet 0.001 olarak seçilmiştir. ε ise gizlilik garantisinin ne kadar sağlam olduğunu gösterir. Epsilonun küçük olması mahremiyetin ve eklenen gürültünün yüksek olduğunu gösterir. Yapılan çalışma için 4 ve 8 yığın boyutları için hesaplanan epsilon değerleri Çizelge 4.6'daki gibidir.

Çizelge 4.6'da görüldüğü gibi modelin yığın boyutu arttıkça ε değeri büyümektedir. Diferansiyel mahremiyet tanımına göre yüksek ε değeri mahremiyet derecesini düşürdüğü için yığın boyutu arttıkça mahremiyetin düştüğü söylenebilir.

Çizelge 4.6. Batch boyutu değerleri için hesaplanan epsilon değerleri

Batch boyutu	Epsilon
4	11.76
8	18.11

Model sonuçları genel olarak değerlendirildiğinde DP SGD ile eğitilen model ve mahremiyetsiz referans model karşılaştırıldığında DP SGD yönteminde eğitim seti üzerindeki kayıp oranının bir miktar arttığı gözlemlenmektedir.

5. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, büyük önem taşıyan kişisel mahremiyetin sağlık gibi önemli bir alanda sağlanabilmesi için bir derin öğrenme modeli prototipi uygulanmıştır. Uygulamanın temel amacı Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ve Gazi Üniversitesi iş birliğinde gerçekleştirilen Türk Beyin Projesi kapsamında kullanılan verilerin mahremiyetini sağlamak için bir çözüm sunmaktır.

Derin öğrenme modelinin mahremiyetini sağlamak için literatürdeki mahremiyet koruma yöntemleri incelenmiş ve mahremiyet fayda dengesini dikkate alan diferansiyel mahremiyet yöntemi uygulanmıştır. Diferansiyel mahremiyet derin öğrenme modellerine çeşitli şekillerde uygulanabilmektedir. Bu tez çalışmasında ise gradyan tabanlı pertürbasyon uygulanmıştır. Bunun için literatürde matematiksel olarak kanıtlanmış DP SGD optimizasyon algoritması tercih edilmiştir.

Model seçiminde literatürde önerilen çok ölçekli 3 boyutlu CNN ağı tercih edilmiştir. Çok ölçekli ağ daha anlamsal ve çözünürlük olarak yüksek öznitelikler çıkarabilmektedir. Ağın ara katmanlarında çıkarılan özellikler ile son katmanlarda çıkarılan özellikler birleştirilerek daha kaliteli özellik haritaları elde edilmektedir. 3 boyutlu bir ağ seçilmesinin nedeni ise veri seti örneklerinin NIfTI formatında olmasıdır. Üç boyutlu olan NIfTI formatı üç boyutlu tensor yapısını karşılamaktadır.

Model başlangıç parametreleri referans model ve literatürdeki öneriler dikkate alınarak seçilmiştir. Girdi boyutu tüm veriler için 112x96x96 boyutunda seçilmiştir. Eğitim epoch sayısı 100, düzleştirme (flatten) katmanına ve birleştirme katmanlarına uygulanan bırakma (dropout) oranı 0,5, çıktı katmanı hariç tüm katmanlara uygulanan L2 regülasyon parametresi 0,00005, yığın normalleştirme (batch normalization) momentumu: 0,9 olarak seçilmiş tüm katmanlar ‘glorot-uniform’ metodu ile başlatılmıştır.

Model parametrelerinin seçimi DM açısından değerlendirildiğinde yığın (batch) sayısı arttığında epsilon değerinin büyüdüğü ve eklenen gürültü miktarının arttığı gözlemlenmiştir. Bundan dolayı model başarımı düşmekte, kayıp oranı artmakta ve aşırı öğrenme oluşmaktadır. Bu model için en uygun yığın (batch) sayısının 4 olduğu gözlemlenmiştir.

Daha sonraki adımda öğrenme katsayısı 0,003'e düşürülüp ve eklenen gürültü miktarı azaltılarak daha yüksek başarım sağlanmıştır.

Kullanılan veri seti 3 boyutlu verilere sahip olup T1 sekansındaki görüntüler kullanılmıştır. Veri önişleme adımında yeniden boyutlandırma ve arka plan azaltma yapılarak eğitim sırasındaki maliyetler düşürülmüştür. T1 görüntüleri veri setinde bulunan tümör maskeleri ile eşleştirilerek tümör işaretleme yapılmıştır. Eğitimde bu şekilde segmentlenmiş görüntüler kullanılmıştır. Bu veri setinin seçilmesinin temel sebebi Türk Beyin Projesinde kullanılan gerçek veri formatı ile aynı formatta olmasıdır.

Genel olarak elde edilen sonuçlar mahremiyetsiz model ile karşılaştırıldığında bir miktar kayıp olduğu gözlemlenmiş ve bu kayıp gürültü miktarı azaltılarak kapatılmaya çalışılmıştır. Elde edilen sonuçlarda iki modelin de doğruluk ve kayıplarının birbirine yakın olduğu gözlemlenmiştir.

Tez çalışması sırasında karşılaşılan en büyük problem modelin RAM'i yüksek güçlü bir bilgisayarda çalışması gerekliliğidir. Modelin eğitilebilmesi için en az 16GB RAM'e sahip bir bilgisayar gerekmektedir.

Sonuç olarak daha güçlü donanım ve daha fazla veri örneği ile bu prototipin Türk Beyin Projesi kapsamında veri seti mahremiyetini korumak için uygulanabileceği değerlendirilmektedir.

Modelin gerçek verilerde test edilmesi ve örnek sayısı artırılarak başarımının test edilmesi gelecek çalışmalar için önerilmektedir. Ayrıca mahremiyet korumalı modele karşı bir üyelik çıkarımı veya model tersine çevirme saldırısı tasarlanarak matematiksel olarak kanıtlanan mahremiyet derecesi pratik olarak da incelenmelidir.

KAYNAKLAR

1. Kılınç, D. (2012). Anayasal Bir Hak Olarak Kişisel Verilerin Korunması. *Ankara Üniversitesi Hukuk Fakültesi Dergisi*, 61(3), 1089 - 1169.

2. İnternet: Tabipler Birliği Tıp Etiği El Kitabı. Web: https://www.ttb.org.tr/kutuphane/tip_etigi.pdf adresinden 2 Ocak 2021'de alınmıştır.

3. İnternet: KVK Kanunu Web: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6698&MevzuatTur=1&MevzuatTertip=5 adresinden 2 Ocak 2021'de alınmıştır.

4. İnternet: Privacy and Security in Healthcare: A Must-read for Healthtech Entrepreneurs. Web: https://steelkiwi.com/blog/privacy-and-security-in-healthcare/ adresinden 2 Ocak 2021'de alınmıştır.

5. Iqridar Newaz, A., Sikder, A.K., Ashiqur Rahman, M., and Selcuk Uluagac, A. (2020). A Survey on Security and Privacy Issues in Modern Healthcare Systems: Attacks and Defenses. *arXiv:2005.07359*, 1-40.

6. Mireshghallah, F., Taram, M., Vepakomma, P., Singh, A., Raskar, R., and Esmaeilzadeh, H. (2020). Privacy in Deep Learning: A Survey. *arXiv:2004.12254*, 1-24.

7. İnternet: Mcafee. Data exfiltration study: actors, tactics, and detection. Web: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-data-exfiltration.pdf adresinden 2 Ocak 2021'de alınmıştır.

8. Lipp, M., Schwarz, M., Gruss, D., Prescher, T., Haas, W., Horn, J., Mangard, S., Kocher, P., Genkin, D., Yarom, Y., and Strackx, R. (2020). Meltdown: Reading Kernel Memory from User Space. *Communications of the Association for Computing Machinery*, 63(6), 46–56.

9. İnternet: Unuchek, R. Leaking Ads is User Data Truly Secure? Web: https://published-prd.lanyonevents.com/published/rsaus18/sessionsFiles/8161/ASEC-T08-Leaking-Ads-Is-User-Data-Truly-Secure.pdf adresinden 2 Ocak 2021'de alınmıştır.

10. Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). *Membership inference attacks against machine learning models*. 2017 IEEE Symposium on Security and Privacy, 3-18, San Jose, USA.

11. Long, Y., Bindschaedler, V., and Gunter, C.A. (2017). Towards Measuring Membership Privacy. *arXiv:1712.09136.*

12. Salem, A., Zhang, Y., Humbert, M., Berrang, P., Fritz, M., and Backes, M. (2018). ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models. *arXiv:1806.01246*.

13. Yeom, S., Giacomelli, I., Fredrikson, M., and Jha, S. (2017). Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting. *arXiv:1709.01604*.

14. Hayes, J., Melis, L., Danezis, G., and De Cristofaro, E. (2017). LOGAN: Membership Inference Attacks Against Generative Models. *arXiv:1705.07663*.

15. Fredrikson, M., Jha, S., and Ristenpart, T. (2015). *Model inversion attacks that exploit confidence information and basic countermeasures*. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 1322-1333, Denver, Colorado, USA.

16. Lee, S., and Kil, R.M. (1994). Inverse Mapping of Continuous Functions Using Local and Global Information. *IEEE Transactions on Neural Networks,* 5(3), 409-423.

17. Linden, A., and Kindermann, J. (1989). *Inversion of multilayer nets*. International 1989 Joint Conference on Neural Networks, 425-430, Washington, USA.

18. Nash, C., Kushman, N., and Williams, C.K.I. (2018). Inverting Supervised Representations with Autoregressive Neural Density Models. *arXiv:1806.00400*.

19. Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D., and Ristenpart, T. (2014). *Privacy in pharmacogenetics: an end-to-end case study of personalized warfarin dosing*. USENIX Security Symposium 2014, 17-32, San Diego, USA.

20. Yang, Z., Chang, E.-C., and Liang, Z. (2019). Adversarial Neural Network Inversion via Auxiliary Knowledge Alignment. *arXiv:1902.08552*.

21. Yeom, S., Giacomelli, I., Menaged, A., Fredrikson, M., and Jha, S. (2019). Overfitting, Robustness, and Malicious Algorithms: A Study of Potential Causes of Privacy Risk in Machine Learning. *Journal of Computer Security,* 28, 1-36.

22. Salem, A., Bhattacharya, A., Backes, M., Fritz, M., and Zhang, Y. (2019). Updates-Leak: Data Set Inference and Reconstruction Attacks in Online Learning. *arXiv:1904.01067*.

23. He, Z., Zhang, T., and Lee, R.B. (2019). *Model inversion attacks against collaborative inference*. Proceedings of the 35th Annual Computer Security Applications Conference, 148-162, Puerto Rico, USA.

24. Tramèr, F., Zhang, F., Juels, A., Reiter, M.K., and Ristenpart, T. (2016). Stealing Machine Learning Models via Prediction APIs. *arXiv:1609.02943*.

25. Ateniese, G., Felici, G., Mancini, L.V., Spognardi, A., Villani, A., and Vitali, D. (2013). Hacking Smart Machines with Smarter Ones: How to Extract Meaningful Data from Machine Learning Classifiers. *arXiv:1306.4447*.

26. Yan, M., Fletcher, C., and Torrellas, J. (2018). Cache Telepathy: Leveraging Shared Resource Attacks to Learn DNN Architectures. *arXiv:1808.04761*.

27. Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., and Song, D. (2018). The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. *arXiv:1802.08232*.

28. Ganju, K., Wang, Q., Yang, W., Gunter, C.A., and Borisov, N. (2018). *Property inference attacks on fully connected neural networks using permutation invariant representations*. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 619-633,Toronto, Canada.

29. Fang, W., Wen, X.Z., Zheng, Y., and Zhou, M. (2017). A Survey of Big Data Security and Privacy Preserving. *IETE Technical Review*, 34(5), 544-560.

30. Xu, Y., Ma, T., Tang, M., and Tian, W. (2014). A Survey of Privacy Preserving Data Publishing using Generalization and Suppression. *Applied Mathematics & Information Sciences*, 8, 1103-1116.

31. Ye, Y., Wang, L., Han, J., Qiu, S., and Luo, F. (2017). *An anonymization method combining anatomy and permutation for protecting privacy in microdata with multiple sensitive attributes*. 2017 International Conference on Machine Learning and Cybernetics, 404-411, Ningbo, China.

32. Sweeney, L. (2002). K-anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 557–570.

33. Meyerson, A., and Williams, R. (2004). *On the complexity of optimal k-anonymity*. Proceedings of the Twenty-Third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 223-228, Paris, France.

34. Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). *L-diversity: privacy beyond k-anonymity*. 22nd International Conference on Data Engineering, 24, Atlanta, USA.

35. Li, N., Li, T., and Venkatasubramanian, S. (2010). Closeness: A New Privacy Measure for Data Publishing. *IEEE Transactions on Knowledge and Data Engineering*, 22(7), 943-956.

36. De, A. (2011). Lower Bounds in Differential Privacy. *arXiv:1107.2183*.

37. Hardt, M., Ligett, K., and McSherry, F. (2010). A Simple and Practical Algorithm for Differentially Private Data Release. *arXiv:1012.4763*.

38. Lee, J., and Clifton, C. (2011). *How much is enough? choosing ε for differential privacy*. Information Security 14th International Conference, 325-340, Xi'an, China.

39. Dwork, C., and Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407.

40. Ji, Z., Lipton, Z.C., and Elkan, C. (2014). Differential Privacy and Machine Learning: a Survey and Review. *arXiv:1412.7584*.

41. Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B.C., and Roth, A. (2014). *Differential privacy: an economic method for choosing epsilon*. 2014 IEEE 27th Computer Security Foundations Symposium, 398-410, Vienna, Austria.

42. Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). *Calibrating noise to sensitivity in private data analysis*. (First Edition). Berlin: Springer Berlin Heidelberg, 265-284.

43. McSherry, F., and Talwar, K. (2007). *Mechanism design via differential privacy*. 48th Annual IEEE Symposium on Foundations of Computer Science, 94-103, Providence, USA.

44. Roth, A., and Roughgarden, T. (2010). *Interactive privacy via the median mechanism*. Proceedings of the Forty-Second ACM Symposium on Theory of Computing, 765-774, Cambridge, Massachusetts, USA.

45. Ghosh, A., Roughgarden, T., and Sundararajan, M. (2009). *Universally utility-maximizing privacy mechanisms*. Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, 351-360, Bethesda, MD, USA.

46. Geng, Q., and Viswanath, P. (2016). Optimal Noise Adding Mechanisms for Approximate Differential Privacy. *IEEE Transactions on Information Theory*, 62(2), 952-969.

47. Nie, X., Yang, L.T., Feng, J., and Zhang, S. (2020). Differentially Private Tensor Train Decomposition in Edge-Cloud Computing for SDN-Based Internet of Things. *IEEE Internet of Things Journal*, 7(7), 5695-5705.

48. İnternet: Lundmark, M., and Dahlman, C. (2017). Differential privacy and machine learning: calculating sensitivity with generated data sets. Web: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-209481 adresinden 5 Eylül 2021'de alınmıştır.

49. Rodríguez-Barroso, N., Stipcich, G., Jiménez-López, D., Ruiz-Millán, J.A., Martínez-Cámara, E., González-Seco, G., Luzón, M., Veganzones, M., and Herrera, F. (2020). Federated Learning and Differential Privacy: Software Tools Analysis, the Sherpa.ai FL Framework and Methodological Guidelines for Preserving Data Privacy. *Information Fusion,* 64, 270-292.

50. Manogaran, G., Shakeel, P.M., Hassanein, A.S., Kumar, P.M., and Babu, G.C. (2019). Machine Learning Approach-Based Gamma Distribution for Brain Tumor Detection and Data Sample Imbalance Analysis. *IEEE Access*, 7, 12-19.

51. Liu, X., Zhou, P., Qiu, T., and Wu, D.O. (2020). Blockchain-Enabled Contextual Online Learning Under Local Differential Privacy for Coronary Heart Disease Diagnosis in Mobile Edge Computing. *IEEE Journal of Biomedical and Health Informatics*, 24(8), 2177-2188.

52. Chaudhuri, K., Monteleoni, C., and Sarwate, A.D. (2009). Differentially Private Empirical Risk Minimization. *arXiv:0912.0071*.

53. Phan, N., Wu, X., and Dou, D. (2017). Preserving Differential Privacy in Convolutional Deep Belief Networks. *arXiv:1706.08839*.

54. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep Learning with Differential Privacy. *arXiv:1607.00133*.

55. He, Q., Yang, W., Chen, B., Geng, Y., and Huang, L. (2020). TransNet: Training Privacy-Preserving Neural Network over Transformed Layer. *Proceedings of the VLDB Endowment,* 13(12), 1849–1862.

56. Chen, S., Fu, A., Shen, J., Yu, S., Wang, H., and Sun, H. (2020). RNN-DP: A New Differential Privacy Scheme Base on Recurrent Neural Network for Dynamic Trajectory Privacy Protection. *Journal of Network and Computer Applications,* 168, 102736.

57. Papernot, N., Abadi, M., Erlingsson, Ú., Goodfellow, I., and Talwar, K. (2016). Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data. *arXiv:1610.05755*.

58. Zhao, L., Wang, Q., Zou, Q., Zhang, Y., and Chen, Y. (2018). Privacy-Preserving Collaborative Deep Learning with Unreliable Participants. *arXiv:1812.10113*.

59. Brinkrolf, J., Göpfert, C., and Hammer, B. (2019). Differential Privacy for Learning Vector Quantization. *Neurocomputing,* 342, 125-136.

60. Acs, G., Melis, L., Castelluccia, C., and De Cristofaro, E. (2017). Differentially Private Mixture of Generative Neural Networks. *arXiv:1709.04514*.

61. Adesuyi, T.A., and Kim, M. (2020). A Neuron Noise-Injection Technique for Privacy Preserving Deep Neural Networks. *Open Computer Science,* 10, 137-152.

62. Xie, L., Lin, K., Wang, S., Wang, F., and Zhou, J. (2018). Differentially Private Generative Adversarial Network. *arXiv:1802.06739*.

63. Shokri, R., and Shmatikov, V. (2015). *Privacy-preserving deep learning*. 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 909-910, Monticello, IL, USA.

64. Adesuyi, T.A., and Kim, B.M. (2019). *Preserving privacy in convolutional neural network: an ϵ-tuple differential privacy approach*. 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention, 570-573, Seoul, South Korea.

65. Bun, M., and Steinke, T. (2016). Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. *arXiv:1605.02065*.

66. Dwork, C., and Rothblum, G.N. (2016). Concentrated Differential Privacy. *arXiv:1603.01887*.

67. Li, N., Qardaji, W., Su, D., Wu, Y., and Yang, W. (2013). *Membership privacy: a unifying framework for privacy definitions*. Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, 889-900, Berlin, Germany.

68. Backes, M., Berrang, P., Humbert, M., and Manoharan, P. (2016). *Membership privacy in microRNA-based studies*. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 319-330, Vienna, Austria.

69. Jayaraman, B., and Evans, D. (2019). Evaluating Differentially Private Machine Learning in Practice. *arXiv:1902.08874*.

70. Rahman, M.A., Rahman, T., Laganière, R., Mohammed, N., and Wang, Y. (2018). Membership İnference Attack Against Differentially Private Deep Learning Model. *Transactions on Data Privacy*, 11, 61-79.

71. McLaughlin-Drubin, M.E., and Munger, K. (2008). Viruses Associated with Human Cancer. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1782(3), 127-150.

72. Golemis, E.A., Scheet, P., Beck, T.N., Scolnick, E.M., Hunter, D.J., Hawk, E., and Hopkins, N. (2018). Molecular Mechanisms of The Preventable Causes of Cancer in The United States. *Genes & Development*, 32(13-14), 868-902.

73. Evan, G.I., and Vousden, K.H. (2001). Proliferation, Cell Cycle and Apoptosis in Cancer. *Nature*, 411(6835), 342-348.

74. İnternet: Brain Tumor Diagnosis. Web: https://www.cancer.net/cancer-types/brain-tumor/diagnosis adresinden 2 Ocak 2021'de alınmıştır.

75. İnternet: Society, A.C. Web: www.cancer.org/cancer.html adresinden 2 Ocak 2021'de alınmıştır.

76. Amyot, F., Arciniegas D., Brazaitis M., Curley, C., Diaz-Arrastia, R., Gandjbakhche, A., Herscovitch, P., Hindsll, S., Manley, G., Pacifico, A., Razumovsky, A., Riley, J., Salzer, W., Shih, R., Smirniotopoulos, J., and Stocker, D. (2015). A Review of the Effectiveness of Neuroimaging Modalities for the Detection of Traumatic Brain Injury. *Journal of Neurotrauma*, 32(22), 1693-1721.

77. İnternet: Barburoğlu, M., Sencer, S. , Aydın, K. Nöroradyolojik Görüntüleme ve Tedavi Yöntemleri. Web: http://www.itfnoroloji.org/nororad/nororad.htm adresinden 2 Ocak 2021'de alınmıştır.

78. Tandel, G., Biswas, M., Kakde, O., Tiwari, A., Suri, H., Turk, M., Laird, J., Asare, K., Ankrah, A., Khanna, N., Madhusudhan, B., Saba, L., and Suri, J. (2019). A Review on a Deep Learning Perspective in Brain Cancer Classification. *Cancers*, 11, 111.

79. Maharjan, S., Alsadoon, A., P.W.C, P., Kadhm, M., and Alsadoon, O. (2019). A Novel Enhanced Softmax Loss Function for Brain Tumour Detection using Deep learning. *Journal of Neuroscience Methods*, 330, 108520.

80. Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., and Shoaib, M. (2020). A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor. *IEEE Access*, 8, 55135-55144.

81. Amin, J., Sharif, M., Anjum, M.A., Raza, M., and Bukhari, S.A.C. (2020). Convolutional Neural Network with Batch Normalization for Glioma and Stroke Lesion Detection Using MRI. *Cognitive Systems Research,* 59, 304-311.

82. Wang, G., Li, W., Ourselin, S., and Vercauteren, T. (2019). Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation. *Frontiers in Computational Neuroscience,* 13, 13-56.

83. Talo, M., Baloglu, U.B., Yıldırım, Ö., and Rajendra Acharya, U. (2019). Application of Deep Transfer Learning for Automated Brain Abnormality Classification Using MR Images. *Cognitive Systems Research,* 54, 176-188.

84. Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., Rovira, A., and Lladó, X. (2020). A Fully Convolutional Neural Network for New T2-w Lesion Detection in Multiple Sclerosis. *NeuroImage: Clinical,* 25, 102149.

85. Swati, Z.N.K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., and Lu, J. (2019). Brain Tumor Classification for MR Images Using Transfer Learning and Fine-Tuning. *Computerized Medical Imaging and Graphics,* 75, 34-46.

86. Samikann, R., Rohini, R., Diarra, B., and M, S. (2020). An Efficient Image Analysis Framework for the Classification of Glioma Brain Images using CNN Approach. *Computers, Materials & Continua,* 63, 1133-1142.

87. McCulloch, W.S., and Pitts, W. (1943). A Logical Calculus of The Ideas immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics,* 5(4), 115-133.

88. Lee, H., Grosse, R., Ranganath, R., and Ng, A.Y. (2009). *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations*. Proceedings of the 26th Annual International Conference on Machine Learning, 609-616, Montreal, Quebec, Canada.

89. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.R. (2012). Improving Neural Networks by Preventing Co-adaptation of Feature Detectors. *arXiv:1207.0580*.

90. Elgendy, M. (2020). *Deep Learning for Vision Systems*. New York: Manning Publications, 92-158.

91. Ioffe, S., and Szegedy, C. (2015). *Batch normalization: accelerating deep network training by reducing internal covariate shift*. Proceedings of the 32nd International Conference on International Conference on Machine Learning, 37, 448-456 Lille, France.

92. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Russel, T., Christoph, B., Ha, M., Martin, R., Marcel, P., and Menze, B. (2018). Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv:1811.02629*.

93. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J., Farahani, K., and Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI Collections with Expert Segmentation Labels and Radiomic Features. *Sci Data*, 4, 170117.

94. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., and Leemput, K.V. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993-2024.

95. Ge, C., Qu, Q., Gu, I.Y.H., and Jakola, A.S. (2018). *3D multi-scale convolutional networks for glioma grading using MR images*. 2018 25th IEEE International Conference on Image Processing, 141-145, Athens, Greece.

96. İnternet: 3D Convolutions: Understanding + Use Case - Drug Discovery. Web: https://www.kaggle.com/shivamb/3d-convolutions-understanding-use-case adresinden 24 Ağustos 2021'de alınmıştır.

97. Alalwan, N., Abozeid, A., ElHabshy, A.A., and Alzahrani, A. (2021). Efficient 3D Deep Learning Model for Medical Image Semantic Segmentation. *Alexandria Engineering Journal*, 60(1), 1231-1239.

98. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., and Ronneberger, O. (2016). *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2016*. (First Edition). Cham: Springer International Publishing, 424-432.

99. İnternet: Machine Learning with Differential Privacy in TensorFlow. Web: http://www.cleverhans.io/privacy/2019/03/26/machine-learning-with-differential-privacy-in-tensorflow.html adresinden 2 Ocak 2021'de alınmıştır.



GAZİ GELECEKTİR...