



T.C.

**ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**GÖRÜNTÜ İŞLEME VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİYLE KARACİĞER TOMOGRAFİ
GÖRÜNTÜLERİNDE BENİGN VE MALİGN KİTLENİN TESPİTİ**

YÜKSEK LİSANS TEZİ

CANER KOCAMAZ

Tez Danışmanı

Dr. Öğr. Üyesi SAİT CAN YÜCEBAŞ

ÇANAKKALE – 2024



T.C.

ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**GÖRÜNTÜ İŞLEME VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİYLE KARACİĞER TOMOGRAFİ
GÖRÜNTÜLERİNDE BENİGN VE MALİGN KİTLENİN TESPİTİ**

YÜKSEK LİSANS TEZİ

Caner KOCAMAZ

Tez Danışmanı

Dr. Öğr. Üyesi SAİT CAN YÜCEBAŞ

ÇANAKKALE – 2024



T.C.
ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



Caner KOCAMAZ tarafından Dr. Öğr. Üyesi Sait Can YÜCEBAŞ yönetiminde hazırlanan ve **29/01/2024** tarihinde aşağıdaki jüri karşısında sunulan “**Görüntü İşleme ve Makine Öğrenmesi Yöntemleriyle Karaciğer Tomografi Görüntülerinde Benign ve Malign Kitlenin Tespiti**” başlıklı çalışma, Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim Enstitüsü **Bilgisayar Mühendisliği Anabilim Dalı**’nda **YÜKSEK LİSANS TEZİ** olarak oy birliği ile kabul edilmiştir.

Jüri Üyeleri

İmza

Dr. Öğr. Üyesi Sait Can YÜCEBAŞ

.....

(Danışman)

Dr. Öğr. Üyesi Ali AKMAN

.....

Dr. Öğr. Üyesi Yonca BAYRAKDAR YILMAZ

.....

Tez No : 10600308

Tez Savunma Tarihi : 29/01/2024

Prof. Dr. Ahmet Evren ERGİNAL
Enstitü Müdürü

.././2024

ETİK BEYAN

Çanakkale Onsekiz Mart Üniversitesi Lisansüstü Eğitim 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 taahhüt ve beyan ederim.

(İmza)

Caner KOCAMAZ

29/01/2024

TEŞEKKÜR

Bu tezin gerçekleştirilmesinde, çalışmam boyunca benden bir an olsun yardımlarını esirgemeyen saygı değer danışman hocam Dr. Öğr. Üyesi Sait Can YÜCEBAŞ'a, alan uzmanlığı bilgisi ile desteğini esirgemeyen Doç. Dr. Emre KARAKAYA'ya, çalışma süresince tüm zorlukları benimle göğüsleyen ve hayatımın her evresinde bana destek olan sevgili eşim Esen KOCAMAZ'a ve çocuklarım Azra, Ahu, Erdem ve Asudeciğime sonsuz teşekkürlerimi sunarım.

Caner KOCAMAZ
Çanakkale, Ocak 2024

ÖZET

GÖRÜNTÜ İŞLEME VE MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE KARACİĞER TOMOGRAFİ GÖRÜNTÜLERİNDE BENİGN VE MALİGN KİTLENİN TESPİTİ

Caner KOCAMAZ

Çanakkale Onsekiz Mart Üniversitesi

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Sait Can YÜCEBAŞ

29/01/2024,37

Bu çalışma, görüntü işleme ve makine öğrenme yöntemleri kullanarak. Karaciğer tomografi taramalarından elde edilen görüntülerde benign (iyi huylu) ve malign (kötü huylu) kitle tespitine yönelik yapay zeka araştırma modelini geliştirmek, radyologların hasta bakımının kalitesini ve güvenliğini iyileştirerek daha doğru ve verimli olmasını sağlamak, maliyetlerin ve gereksiz tıbbi prosedürlerin azaltılmasına yardımcı olabilmeyi amaçlamaktadır. Bu çalışmadaki deneylerde bir özel hastaneden alınan anonim hale getirilmiş iyi huylu tümör (FBNH, Hemanjiom, Kist) ve kötü huylu tümör (HCC, Kolanjiosellular, Metastaz) görüntü seti örnekleri kullanılmıştır. Görüntü işlemede OpenCV, özellik çıkarmada SqueezeNet ve sınıflandırmada Orange Data Mining (v.3.36.1) programı kullanılmıştır. Orange Data Mining programı kullanılarak ayrıca Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Orman ve Yapay Sinir Ağı modelleri kurulmuştur. Bu modeller doğruluk, kesinlik, geri çağırma ve F1 puanı gibi performans ölçütleri kullanılarak karşılaştırılmıştır Yapay Sinir Ağı, ortalama %98 doğruluk oranı, %88 kesinlik oranı ve %88 geri çağırma oranı ile başarı oranı en yüksek model olmuştur.

Anahtar Kelimeler: Görüntü İşleme, Bölütleme, Makine Öğrenmesi, Karaciğer Tümörleri.

ABSTRACT

DETECTION OF BENIGN AND MALIGN MASS IN LIVER TOMOGRAPHY IMAGES WITH IMAGE PROCESSING AND MACHINE LEARNING METHODS

Caner KOCAMAZ

Çanakkale Onsekiz Mart University

School of Graduate Studies

Master of Science Thesis in Computer Engineering Science

Advisor : Assistant Professor Sait Can YÜCEBAŞ

29/01/2024, 37

Using image processing and machine learning methods, this study aims to develop an artificial intelligence research model for Benign-Malignant mass detection in images obtained from liver tomography scans, enabling radiologists to be more accurate and efficient by improving the quality and safety of patient care, helping to reduce costs and unnecessary medical procedures. This study uses anonymized benign tumor (FBNH, Hemangioma, Cyst) and malignant tumor (HCC, Cholangiocellular, Metastasis) image set samples from a private hospital. OpenCV was used for image processing and Orange Data Mining (v.3.36.1) was used for feature extraction. Logistic Regression, Support Vector Machines, Random Forest and Artificial Neural Network models were built using SqueezeNet and Orange Data Mining for classification. These models were compared using performance measures such as accuracy, precision, recall and F1 score. Neural network was the best performing model with 98% accuracy , 88% precision and 88% rate.

Keywords: Image Processing, Segmentation, Machine Learning, Liver Tumors, Orange Data Mining

İÇİNDEKİLER

Sayfa No

JÜRİ ONAY SAYFASI.....	i
ETİK BEYAN.....	ii
TEŞEKKÜR.....	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
SİMGELER ve KISALTMALAR.....	viii
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ.....	x

BİRİNCİ BÖLÜM

GİRİŞ

1

İKİNCİ BÖLÜM

KURAMSAL ÇERÇEVE/ÖNCEKİ ÇALIŞMALAR

3

2.1. Genel Bilgiler.....	3
2.1.1. Karaciğer Kitlesi (Tümörleri).....	3
2.1.2. Bilgisayarlı Tomografi.....	3
2.1.3. Görüntü İşleme.....	4
Görüntü Histogramı.....	5
Histogram Eşitleme.....	6
Medyan Filtresi.....	7
2.1.4. Makine Öğrenmesi.....	8
Lojistik Regresyon.....	8
Destek Vektör Makineleri.....	9
Rastgele Orman.....	10
Yapay Sinir Ağı.....	12
2.1.5. Sınıflandırma Metrikleri.....	14
2.1.6. ROC Eğrisi.....	15
2.2. Önceki Çalışmalar.....	16

ÜÇÜNCÜ BÖLÜM	
MATERYAL YÖNTEM	24
3.1. Araştırma Yöntemi.....	24
3.2 Veri Seti.....	25
3.2.1. Görüntülerin İşlenmesi	25
3.2.2. Histogram Eşikleme ve Gri Seviye Segmentasyonu.....	26
3.2.3. Öznetelik Çıkarma.....	26
3.2.4 Sınıflandırma.....	27
DÖRDÜNCÜ BÖLÜM	28
ARAŞTIRMA BULGULARI	
BEŞİNCİ BÖLÜM	32
SONUÇ ve ÖNERİLER	
KAYNAKÇA	33
ÖZGEÇMİŞ.....	I

SİMGELER VE KISALTMALAR

BT	Bilgisayarlı Tomografi
AI	Artificial Intelligence (Yapay Zeka)
LR	Logistic Regression (Lojistik Regresyon)
SVM	Support Vector Machine (Destek Vektör Makinesi)
RF	Random Forest (Rastgele Orman)
ANN	Artificial Neural Network (Yapay Sinir Ağı)
ROI	Region of Interest(İlgilenilen Bölge)
CAD	Computer-Aided Diagnosis (Bilgisayar Destekli Teşhis)
LiTS	Liver Tumor Segmentation Benchmark
DSC	Dice Smilarity Coefficient Zar Benzerliği Katsayısı)
IOU	Intersection over Union(Birleşim Üzerindeki Kesişim)
Benign	İyi Huylu Tümör
Malign	Kötü Huylu Tümör

TABLÖLAR DİZİNİ

Tablo No	Tablo Adı	Sayfa No
Tablo 1	İyi huylu ve köyü huylu tümör çeşitleri	3
Tablo 2	Karmaşıklık matrisi	14
Tablo 3	Önceki çalışmaların özeti	22
Tablo 4	BT görüntü dosyalarının dağılımı	24
Tablo 5	Makine öğrenme yöntemleri varsayılan parametre değerleri	27
Tablo 6	FNHB için makine öğrenmesi performans değerleri	29
Tablo 7	HCC için makine öğrenmesi performans değerleri	29
Tablo 8	HEMANJIOM için makine öğrenmesi performans değerleri	30
Tablo 9	KIST için makine öğrenmesi performans değerleri	30
Tablo 10	KOLANJIOSELLULER için makine öğrenmesi performans değerleri	30
Tablo 11	METASTAZ için makine öğrenmesi performans değerleri	30

ŞEKİLLER DİZİNİ

Şekil No	Şekil Adı	Sayfa No
Şekil 1	Bilgisayarlı Tomografi cihazı ve elde edilen görüntü	4
Şekil 2	Görüntü histogramı örneği	5
Şekil 3	Histogram eşitleme işlemi öncesi görüntü ve histogram grafiği	5
Şekil 4	Histogram eşitleme işlemi sonrası görüntü ve histogram grafiği	6
Şekil 5	Medyan filtresi işlem adımı	7
Şekil 6	Hiper düzlem	9
Şekil 7	Rastgele Orman grafiği	11
Şekil 8	Biyolojik Nöron ve Yapay Sinir Ağı	12
Şekil 9	Örnek YSA modeli	13
Şekil 10	Sigmoid fonksiyonu grafiği	13
Şekil 11	ROC eğrisi test durumları (A, B, C, D)	16
Şekil 12	Tümör gruplarına ait örnek görüntüler	25
Şekil 13	Görüntü işleme adımları	26
Şekil 14	OTSU eşikleme ve bölütleme uygulanmış örnek görüntüler	26
Şekil 15	Orange Data Mining Squeezenet mimarisiyle özellik çıkarma	27
Şekil 16	Lojistik Regresyon algoritması için karmaşıklık matrisi	28
Şekil 17	Destek Vektör Makinesi algoritması için karmaşıklık matrisi	28
Şekil 18	Rastgele Orman için algoritması karmaşıklık matrisi	29
Şekil 19	Yapay Sinir Ağı algoritması için karmaşıklık matrisi	29
Şekil 20	ROC eğrileri	31

BİRİNCİ BÖLÜM

GİRİŞ

Karaciğer, vücudumuzun en büyük organıdır ve bir dizi önemli fonksiyondan sorumludur. Bu fonksiyonlar arasında protein sentezi, yağların (kolesterol) sentezi ve depolanması, karbonhidrat (glikoz) depolanması ve salınması, safranin üretimi, üre yapımı, çeşitli ilaç ve maddelerin vücuttan uzaklaştırılması gibi birçok temel işlev bulunmaktadır. (Taşdoğan,2021).

Karaciğer hastalıkları, vücudumuzdaki çeşitli sistemleri önemli ölçüde etkileyen ciddi sağlık sorunları arasında yer almaktadır. Bu organın normal işlevselliğini bozan herhangi bir hastalık, diğer vücut sistemlerinde önemli aksaklıklara neden olabilir. Erken aşamada doğru teşhis konulup uygun tedavilerle kontrol altına alınmazsa, hastalığın ilerleyen evrelerinde genel vücut fonksiyonlarında hızlı bir bozulma yaşanabilir ve sonuçta hayati tehlike arz eden sağlık sorunları ortaya çıkabilir. Karaciğer hastalıkları arasında hepatit, karaciğer yetmezliği, karaciğer kistleri, enfeksiyöz karaciğer hastalıkları ve karaciğer tümörleri öne çıkan başlıca sorunlardır (Florence,2023).

Birçok hastalığın teşhisinde kullanılan radyolojik değerlendirmeler genellikle radyolog veya radyoloji uzmanı tarafından icra edilir. BT, yani bilgisayarlı tomografi, radyoloji ana bilim dalında yaygın olarak kullanılan bir görüntüleme metodu olarak öne çıkar. BT, X-ışınları kullanarak vücudun belirli bölgelerini ayrıntılı ve katmanlı bir şekilde tarayarak detaylı bir resim oluşturur. Çekim esnasında vücut içindeki kemikler, yumuşak dokular, organlar ve damarlar gibi yapıları farklı açılardan tarar, böylece kesitler halinde ayrıntılı görüntüler elde edilir. Röntgenle kıyaslandığında, BT çok daha ayrıntılı, üç boyutlu ve kesitler halinde detaylı bilgiler sağlayarak hekimlere tedavi ve müdahale süreçlerine dair önemli bilgiler sunar (Bulut,2020).

Karaciğerde benign veya malign kitlelerin tespitinde, tomografi görüntülerinin radyologlar tarafından incelenmesi genellikle uzun sürer ve ara sıra bu değerlendirmelerde hatalı sonuçlar alınabilir. Karaciğer hastalıklarının erken aşamalarda teşhis edilebilmesi için yapay zeka yöntemlerinin kullanıldığı çalışmalar büyük önem taşımaktadır. Artan teknoloji ile bu hastalıkların teşhisinde etkili olan birçok farklı derin öğrenme modeli geliştirilmiştir.

Tezin amacı, Karaciğerin BT taramaları ile elde edilen görüntülerinden; görüntü filtreleme, kenar belirleme ve bölge tabanlı aktif kontur belirleme gibi çeşitli görüntü işleme teknikleri ile karaciğer bölütleme yapmaktır. Sonrasında elde edilen öz nitelikler karaciğer tümör tespitinde kullanılacaktır. Bu tespit için farklı makine öğrenme yöntemlerinin (Regresyon, Destek Vektör Makinesi, Rastgele Orman, Yapay Sinir Ağı) performansları karşılaştırılmıştır. Bu karşılaştırma sonucunda öne çıkan modelin klinikte potansiyel kullanımının tıbbi personele destek sağlayacağı, bu sayede klinik bakım kalitesi ve güvenilirliğinin artıracığı, teşhise giderken uygulanan tıbbi prosedür sayısını azaltarak maliyet etkinlik sağlaması ön görülmektedir.



İKİNCİ BÖLÜM

KURAMSAL ÇERÇEVE/ÖNCEKİ ÇALIŞMALAR

2.1. Genel Bilgiler

2.1.1. Karaciğer Kitlesi (Tümörleri)

Karaciğerde görülen kitle, tıbbi terimle tümör olarak adlandırılır. Bu kitleler genellikle organın kendi dokusundan kaynaklanan kötü huylu tümörlerdir, yani kanser hücreleri içerirler. Diğer yandan, farklı bir hastalığın neden olduğu durumlarda karaciğere sıçrayan iyi huylu kitleler de ortaya çıkabilir. "Benign," Latince kökenli bir terim olup "iyimser" veya "iyi huylu" anlamına gelir. Benign tümörler kanser olmayan, genellikle daha sınırlı büyüme potansiyeline sahip tümörleri ifade eder. "Malign" ise Latince kökenli bir terim olup "kötü huylu" anlamına gelir. Malign tümörler ise kanser potansiyeli taşıyan ve yayılma eğiliminde olan tümörleri ifade eder. Aşağıda, bazı iyi huylu ve kötü huylu tümör türlerini içeren Tablo 1 verilmiştir.

Tablo 1

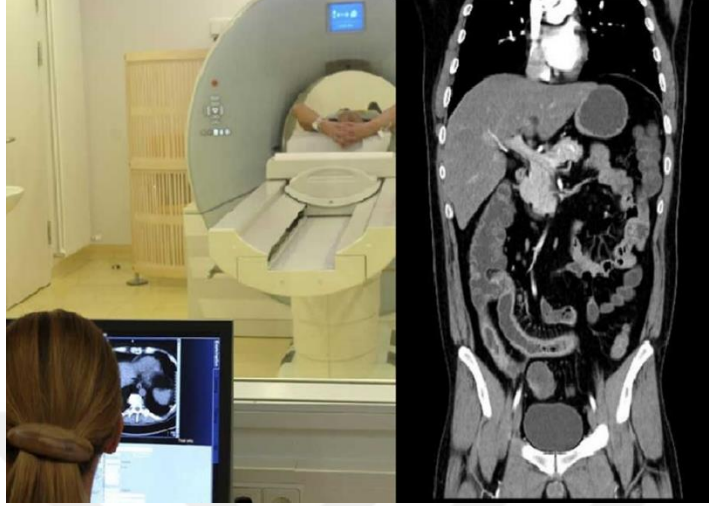
İyi huylu ve kötü huylu tümör çeşitleri

Tümör Adı	Tümör Grubu
FBNH	İyi huylu
HCC	Kötü huylu
HEMANJİOM	Kötü huylu
KIST	İyi huylu
KOLANJİOSELLÜLER	İyi huylu
METASTAZ	Kötü huylu

2.1.2. Bilgisayarlı Tomografi

Bir nesnenin iç yapısını üç boyutlu olarak gözlemlemek için, Şekil 1’de de görüldüğü üzere, çeşitli açılardan çok sayıda iki boyutlu X ışını görüntüsü alınır. Gri tonlamalı olarak sunulan bu X ışını görüntüleri, üç farklı dokunun derinliğini gösterir. Beyaz tonlu alanlar genellikle kemik veya kireç içeren yapıları temsil ederken, gri tonlar organlarda bulunan yumuşak dokuları ifade eder. Siyah bölgeler ise içinde hava bulunan bölgeleri gösterir. Bu

yöntem, yumuşak ve sert dokuların tespitinde sıkça kullanılır. Ancak, sağlık açısından zararlı olabilecek aşırı kullanımından kaçınılmalıdır (Yılmaz, 2007).



Şekil 1. Bilgisayarlı tomografi cihazı ve elde edilen görüntü (BT Enterografi,2023)

2.1.3. Görüntü İşleme

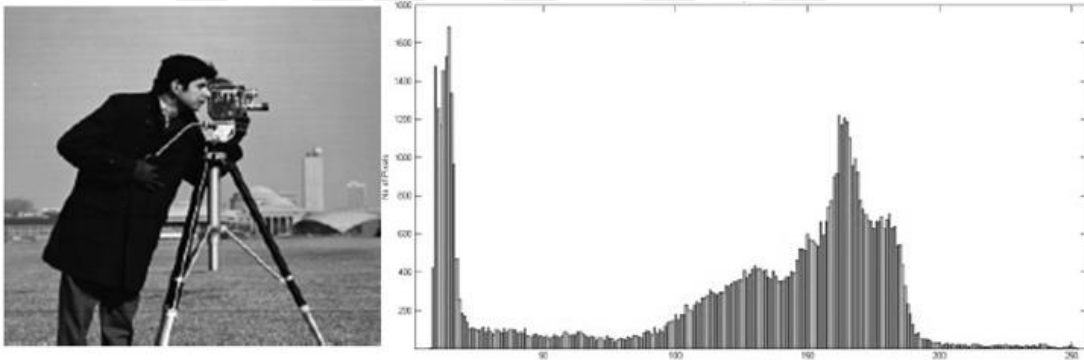
Dijital veya dijitalleştirilmiş görüntülerin belirli yöntemler kullanılarak analiz edilmesi, gerektiğinde bu görüntüleri iyileştirilmesi için özellik ve yapılarının değiştirilmesi ve/veya geliştirilmesi süreçlerini kapsayan teknoloji dalı görüntü işleme olarak adlandırılır. Modern teknoloji, herhangi bir görüntünün (fotoğraf veya video) girdi olarak kullanılarak istenilen özellikte başka bir görüntünün elde edilmesini veya girdi olarak kullanılan görüntü ile ilgili verilerin çıkarılmasını mümkün kılar. Görüntü işleme sayesinde bir görüntünün renk tonu, parlaklık, boyut, yapı gibi özellikleri çeşitli yazılımlar kullanılarak değiştirilebilir, geliştirilebilir ve analiz edilebilir. Bu yazılımlar, dijital ortama aktarılan görüntülerdeki bozuklukları düzeltebilir ve daha yüksek kalitede görüntüler elde etmek için kullanılabileceği gibi nesnelerin tanımlanması, hareketli ve hareketsiz nesnelerin ayrıştırılması gibi birçok amaç için de kullanılabilir. Görüntü işleme, farklı sektörlerde (güvenlik, astronomi, savunma sanayisi, kalite kontrolü gibi) çeşitli alanlarda kullanılabilecek esnek çözümler sunar (Yılmaz, 2007).

Görüntü işleme, dijital veya analog bir görüntünün üzerinde çeşitli teknikler kullanarak veriyi düzenleme ve iyileştirme sürecidir. Bu süreç, gürültü azaltma, boyutlandırma, normalleştirme ve kenar belirleme gibi işlemleri içerir. Hedef, elde edilen görüntüyü daha net ve anlamlı hale getirerek, görüntü analizi veya işleme için daha uygun hale getirmektir.

Görüntü Histogramı

Histogram, resim gibi iki boyutlu bir objenin içindeki sayısal verilerin grafik temsili biçimidir. Bir resmin histogramı, piksel sayısı ve bu piksellerin yoğunluk seviyeleri üzerine kuruludur. Bir resmin histogramını oluşturmak için, piksellerin yoğunluk değerleri veya gri seviye değerleri x eksenini (yatay eksen) boyunca çizilirken, piksellerin frekansı veya sayısı y eksenini (dikey eksen) boyunca çizilir.

Şekil 2’de görüldüğü üzere, 8 bpp (piksel başına bit) gri tonlamalı bir resimde 256 farklı gri seviyesi bulunmaktadır. Bu nedenle, yatay ekseninde 0’dan başlayıp 255’e kadar devam eder. Her aralık, resimdeki belirli bir gri seviye değerine sahip piksellerin sayısını içerir. Genellikle, aralıklar, piksel sayısının çizildiği gri tonlama değerleri aralığını temsil eden aynı genişlikte olur (Ghosh ve ark., 2015).



Şekil 2. Görüntü histogramı örneği (Ghosh ve ark.,2012)

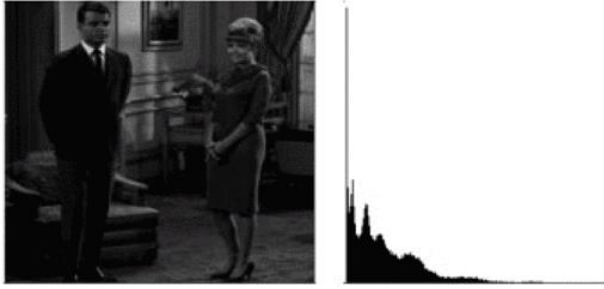
Histogram grafiğinden bir görüntünün ton dağılımı hakkında kısa bir fikir oluşturulabilir. Histogram kutuları, görüntünün türüne bağlı olarak çizimin farklı bölümlerine yoğunlaşmıştır. Daha koyu bir görüntü için histogram bölmeleri çoğunlukla gri tonlama değerlerinin alt tarafında yoğunlaşır. Daha açık bir görüntü için histogram bölmeleri çoğunlukla gri tonlama değerlerinin yüksek tarafında yoğunlaşır. Yüksek kontrastlı bir görüntüde histogram kutuları geniş bir gri tonlama değerleri aralığına dağıtılır. Benzer şekilde, düşük kontrastlı bir görüntüde histogram kutuları çok dar bir gri tonlama değerleri aralığına dağıtılır.

Bir görüntüdeki piksellerin uzaysal düzeni hakkında bilgi, histogramdan elde edilemez. Bu, bir görüntünün histogramını elde ettikten sonra histogramdan tam görüntüyü yeniden oluşturma mümkün olmadığı anlamına gelir.

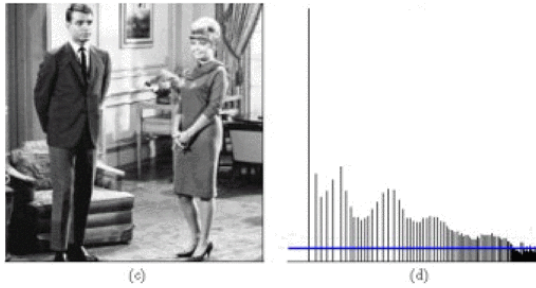
Histogram Eşitleme

Histogram eşitleme bir görüntüdeki renk değerlerinin belirli bir değer aralığı içerisinde kümelenmesinden kaynaklı renk bozukluğunu gidermek için kullanılan bir yöntemdir. Hesaplama zamanı düşük ve oldukça etkili sonuçlar üretebilen bir yöntemdir. Bununla birlikte ters çevrilebilir bir operatördür. Diğer bir ifadeyle, istenildiğinde görüntünün gerçek histogramına tekrar dönüşümü yapılabilir. Yöntemin zayıf noktası ise gürültü varlığı durumunda hem gerçek sinyalin hem de gürültü seviyesinin karşılıklarının birlikte artmasıdır. Bu nedenle, gürültü içeren görüntülere öncelikli olarak gürültü seviyesini azaltacak bir süzgecin uygulanması gereklidir (Arısoy ve Dikmen,2014).

Histogram eşitleme yönteminde ilk olarak gri görüntü seçilir ve renk histogramı hesaplanır. Histogram değerlerini tutacak 256 elemanlık matris tanımlanır. Görüntünün tüm pikselleri için histogram 1 artırılarak histogram matrisi bulunur. Renk aralıkları toplam boyuta bölünerek renk frekansları bulunur. Kümülatif dağılım fonksiyonu hesaplanır. Görüntüden alınan piksel değerleri ile kümülatif dağılım fonksiyonundan alınan değerler eşlenir (Pınar,2020). Histogram eşitleme işlemi öncesi ve sonrası görüntü ve histogram grafiği sırasıyla Şekil 3 ve Şekil 4'te gösterilmiştir.



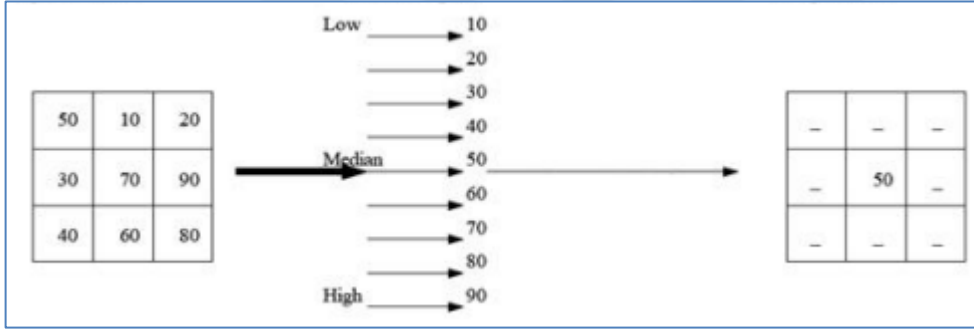
Şekil 3. Histogram eşitleme işlemi öncesi görüntü ve histogram grafiği (Nithyananda ve ark.,2016)



Şekil 4. Histogram eşitleme işlemi sonrası görüntü ve histogram grafiği Nithyananda ve ark.,2016)

Median (Medyan) Filtresi

Medyan filtre, görüntüdeki her pikseli sırayla filtreler ve yakındaki komşuları, çevresini temsil edip etmediğine karar vermek için kullanılır. Normalde piksel değerini komşu piksel değerlerinin ortalamasıyla değiştirmek yerine medyan filtresi, onu bu değerlerin ortancasıyla değiştirir. Yani çevre komşuluktan gelen değerler önce sayısal olarak sıralanır, ardından söz konusu pikselin değeri ortadaki (medyan) piksel değeriyle değiştirilir. Komşuluk burada pencere olarak ifade edilmektedir. Pencere, hedef piksele ortalanmış çeşitli şekillere sahip olabilir. Kare, iki boyutlu görüntüler için tanımlanan pencereler için seçilen tipik bir şekildir. Normal şartlarda medyan filtresinin tek sayıda piksel içeren bir pencere kullanılarak gerçekleştirildiğine dikkat edilmelidir. Söz konusu komşuluk çift sayıda pikselden oluşuyorsa, çıktı olarak seçilen medyan değeri ortadaki iki piksel değerinin ortalamasıdır. Şekil 5'te medyan filtre hesaplamasının pencerede nasıl gerçekleştirildiğine ilişkin bir örnek gösterilmektedir (Boateng ve ark.,2012).



Şekil 5. Medyan filtresi işlem adımı (Boateng ve ark.,2012)

Bu filtre, bir görüntüdeki her geçerli piksel için bir başlangıç pikseli etrafındaki piksellerin komşuluğunu analiz ederek çalışır. Burada çıktıyı hesaplamak için 3x3 piksellik bir pencere kullanılmaktadır. Görüntüdeki her piksel için komşu piksellerin penceresi bulunur. Yukarıdaki örnekte gösterildiği gibi penceredeki piksel değerleri artan sırada sıralanır ve medyan değeri seçilir. Bu durumda medyan değeri 50'dir. Daha sonra, giriş görüntüsündeki piksele karşılık gelen çıktı görüntüsündeki piksel, filtre sırası tarafından belirtilen değerle değiştirilir. Bu örneğe göre 70 ile 50 değiştirilir (Boateng ve ark.,2012).

2.1.4. Makine Öğrenmesi

Makine öğrenimi, veri ve algoritmaların entegrasyonu ile insan öğrenme süreçlerini taklit eden bir yapay zeka ve bilgisayar bilimi disiplini. Bu çalışmada, aşağıda belirtilen makine öğrenimi algoritmaları kullanılmıştır.

Lojistik Regresyon

Olasılık dağılımına dayalı olarak sınıf etiketini tahmin etmede kullanılan yöntemlerden birisidir. Özellikle binary sınıflandırma (iki sınıf arasında) problemlerinde yaygın olarak kullanılır, ancak çok sınıflı sınıflandırma problemlerine de genişletilebilir.

Lojistik regresyonun bir alt modeli olan ikili lojistik regresyon, bağımlı değişken kategorik ve iki sınıflı (binary) olduğunda kullanılır. Sınıf sayısı ikiden çok olduğunda "çok sınıflı lojistik regresyon" yöntemi kullanılır. Bu yöntem, bağımlı değişkenin birden fazla kategoriye ait olabileceği çok sınıflı sınıflandırma problemlerinde kullanılır.

“Lojistik Regresyon kullanarak sınıflandırma, bir dizi öngörücüye veya özelliğe dayalı olarak kategorik bir bağımlı değişkeni tahmin etmek için kullanılan makine öğrenimi yöntemlerinden biridir” (Behbudova,2023).

Lojistik fonksiyon, bağımlı değişkenlerin değerlerini tahmin etmek yerine bir olayın meydana gelme olasılığını sağlayan matematiksel bir fonksiyondur. Aşağıda Denklem 2.1 ile hesaplanır.

$$p = \frac{1}{1+e^{-z}} \quad (2.1)$$

Lojistik regresyon modelini formüle etmek için, genellikle incelenen olayın olasılığını etkileyebilecek faktörleri belirlemek ilk adımdır. Denklem 2.2’de p olasılık değerini, e Euler sayısını (yaklaşık 2.718), z ise aşağıdaki eşitliği gösterir (Denklem 2.2).

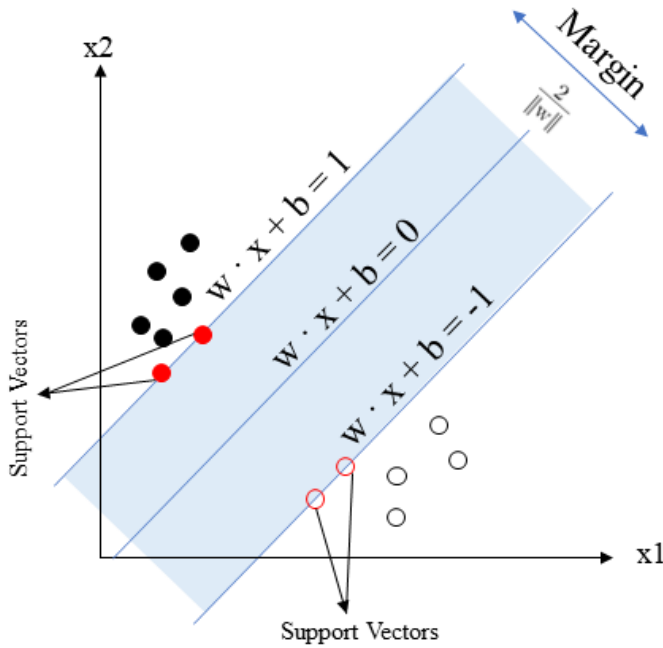
$$z_k = \beta_0 + \sum_{j=1}^J \beta_j x_{j,k} + u_k \quad (2.2)$$

k durum sayısını, β katsayıları, j bağımsız değişken sayısını, $x_{j,k}$ ise k durumu için bağımsız j değişkeninin özelliklerini ifade eder.

Destek Vektör Makineleri

Destek Vektör Makineleri (SVM), sınıflandırma ve kestirim için kullanılan ve literatürde yaygın olarak tercih edilen makine öğrenmesi yöntemlerinden biridir. Özellikle sınıflandırma problemlerinde başarılı olan SVM, veri noktalarını belirli sınıflara ayırmak ve bu ayrımı optimize etmek amacıyla tasarlanmıştır

Destek Vektör Makinelerinin sınıflandırma amacı, iki sınıf arasında optimal bir ayırıcı hiper düzlemi belirlemektir. Bu, genelleme hatasını en aza indirme ve marjı (margin) en üst düzeye çıkarma hedefine yönelik bir çabadır. Eğer herhangi iki sınıf, sonsuz sayıda hiper düzlemle ayrılabilirse, SVM, genelleme hatasını en aza indiren ve görünmeyen test desenleri için hatayı minimize eden hiper düzlemi belirlemeye çalışır (Şekil 6). (Pulat,2023).



Şekil 6. Hiper düzlem (Meriç ve Özer, 2023)

SVM algoritmasında veri noktaları ile hiper düzlem arasındaki marj maksimuma çıkarmaya çalışılır. Eğer destek vektörleri yanlış sınıflandırılmışsa veya kenar payı içerisinde ise hatalıdır. Bu kayba “menteşe” kaybı denmektedir (Demir,2020). Marjı maksimuma çıkarmaya yardımcı olan kayıp fonksiyonu menteşe kaybıdır (Denklem 2.3).

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \quad (2.3)$$

Tahmin edilen değer ile gerçek değer aynı işarete sahipse maliyet 0'dır. Değilse, kayıp değerini hesaplanır. Ayrıca maliyet fonksiyonuna bir düzenleme parametresi eklenir. Düzenleme parametresinin amacı marjin maksimizasyonunu ve kaybını dengelemektir. Regülasyon parametresi eklendikten sonra maliyet fonksiyonları Denklem 2.4'teki gibi görünür.

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y_i < x_i, w >)_+ \quad (2.4)$$

Kayıp fonksiyonu olduğuna göre, gradyanları bulmak için ağırlıklara göre kısmi türevler alınır. Gradyanlar kullanılarak ağırlıklar güncellenir (Denklem 2.5 ve 2.6)

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k \quad (2.5)$$

$$\frac{\delta}{\delta w_k} (1 - y_i < x_i, w >)_+ = \begin{cases} 0, & \text{if } y_i < x_i, w > \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases} \quad (2.6)$$

Yanlış sınıflandırma olmadığında, yani model veri noktasının sınıfını doğru bir şekilde tahmin ettiğinde, yalnızca gradyanı düzenleme parametresinde güncelleme gerekir (Denklem 2.7).

$$w = w - \alpha (2\lambda w) \quad (2.7)$$

Yanlış sınıflandırma olduğunda, yani model veri noktasının sınıfının tahmininde hata yaptığında, gradyan güncellemesi gerçekleştirilmesi için kayıp, düzenleme parametresiyle birlikte dahil edilir (Denklem 2.8).

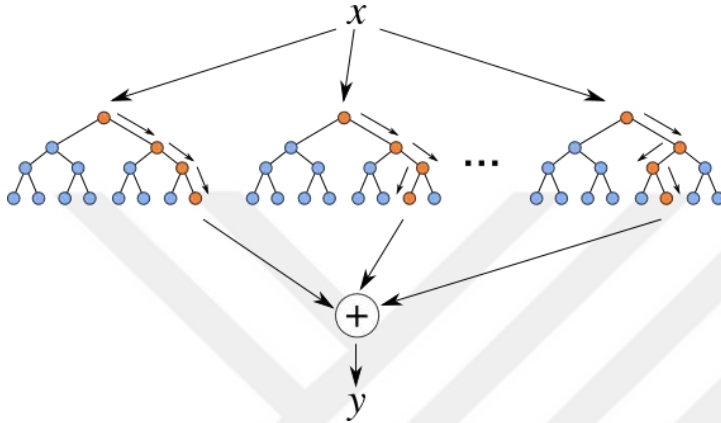
$$w = w + \alpha (y_i x_i - 2\lambda w) \quad (2.8)$$

Rastgele Orman

Rastgele Orman algoritması, giriş değerlerinden çıktı elde etmek amacıyla ağaç yapısını temel alarak yeni ağaçlar üreten bir tür karar ağacı algoritmasıdır. Algoritma, veri setini parçalara ayırır, her bir veri setiyle düğümler oluşturur ve ardından düğümlerden çıkan ağaçlarda rastgele özellik seçimi yaparak ağaçları geliştirmeye devam eder. Bu algoritmanın karar ağacı algoritmalarından en büyük farkı, budama işlemi uygulanmamasıdır (Akar ve Güngör,2012).

Rastgele Orman algoritmasında kullanıcıdan iki parametre istenir. Bunlar, ağaç sayısı (N) ve ağaç yapısının oluşturulması için her düğümden kullanılacak değişken sayısı

m'dir. Parametreler seçildikten sonra test için ayrı bir veri seti yoksa eğitim veri setinin 2/3'ü öğrenme verisi (inBag), 1/3'ü test verisi (Out of-Bag (OOB)) olarak belirlenir. Her ağaç için bootstrap tekniği kullanılarak önyüklemeli örneklem oluşturulur. Örneklemelerin inBag ve OOB verileri ayrılır. Sonra her bir örneklem için ağaç gelişimi başlar. Her düğümde tüm değişkenler arasından m sayıda rastgele seçilen değişkenler kullanılarak en iyi dallanma belirlenir. Bu işlem için CART (Classification and Regression Tree) algoritması kullanılır (Erdem ve ark.,2018).



Şekil 7. Rastgele Orman grafiği (Jibril ve ark., 2020)

Şekil 7’de Rastgele Orman oluşturan üç ayrı karar ağacı var. Rastgele Orman, toplu öğrenme olarak kabul edilir. Bu, sonuca varmak için birden fazla model kullanarak daha doğru sonuçlar oluşturmaya yardımcı olduğu anlamına gelir. Algoritma, kendi sonucuna varmak için her düğümün yapraklarını veya son kararlarını kullanır. Bu, birçok farklı karar ağacının sonuçlarına bakıp bir ortalama bulması nedeniyle modelin doğruluğunu artırır (Makine Öğrenimi için Rastgele Orman Algoritması,2019). Rastgele Orman algoritmasında verilerin nasıl dallara ayrıldığını belirlemek için ortalama kare hatası (MSE) kullanılır (Denklem 2.9).

Rastgele Orman algoritmasında verilerinizin nasıl dallara ayrıldığını belirlemek için ortalama kare hatasını (MSE) kullanılır..

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (2.9)$$

Sınıflandırma verilerine dayalı Rastgele Ormanlar gerçekleştirirken, genellikle Gini indeksi kullanılır. Denklem 2.10, bir düğümdeki her dalın Gini'sini belirlemek için sınıfı ve olasılığı kullanır ve hangi dalların oluşma olasılığının daha yüksek olduğunu belirler. Burada

p_i veri setinde gözlemlendiğiniz sınıfın göreceli frekansını, c ise sınıf sayısını temsil etmektedir.

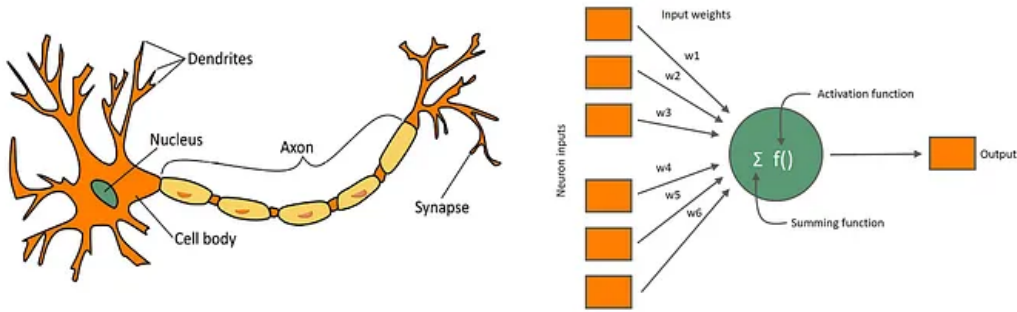
$$Gini = 1 - \sum_{i=1}^C (p_i)^2 \quad (2.10)$$

Bir karar ağacında düğümlerin nasıl dallandığını belirlemek için entropi(entropy) de kullanılabilir. Entropi, düğümün nasıl dallanması gerektiğine karar vermek için belirli bir sonucun olasılığını kullanır (Denklem 2.11).

$$Entropy = \sum_{i=1}^C -p_i * \log_2(p_i) \quad (2.11)$$

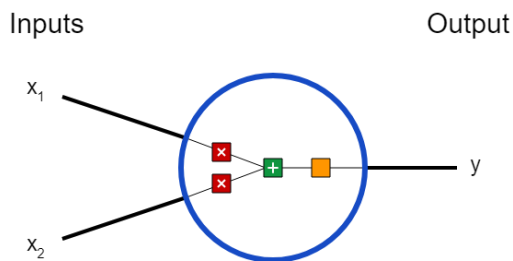
Yapay Sinir Ağı

Yapay Sinir Ağı (YSA), biyolojik sinir hücresini taklit eder. Bunu için de matematiksel yöntemlerden yararlanır. Ağın temelinde girdileri bir toplam fonksiyonu ile çıktı kararı verecek olan aktivasyon fonksiyonunu ileten sinir hücresi (nöron) bulunur (Şekil 8). Doğrusal olmayan sınıflama problemlerinin çözümünde yüksek başarımlı oranı ile literatürde sıklıkla tercih edilen bir yöntemdir (Demirci,2023).



Şekil 8. Biyolojik Nöron ve Yapay Sinir Ağı (Artificial Neural Networks,2023)

YSA matematiği basitçe Şekil 9'daki gibi modellenenebilir.



Şekil 9. YSA model (Ahirwar ve Behera, 2023)

Öncelikle YSA'ya giren değerler, rastgele oluşturulan ağırlıklar ile çarpılır (Denklem 2.12 ve 2.13).

$$x_1 \rightarrow x_1 * w_1 \quad (2.12)$$

$$x_2 \rightarrow x_2 * w_2 \quad (2.13)$$

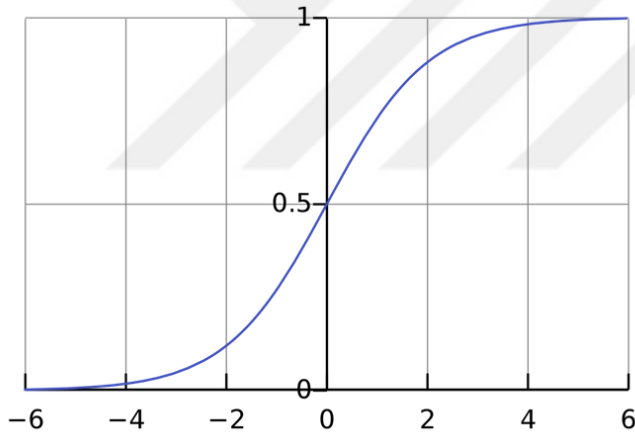
Daha sonra bu çarpımlar toplanır ve toplama bias değeri eklenir (Denklem 2.14).

$$(x_1 * w_1) + (x_2 * w_2) + b \quad (2.14)$$

Toplama sonucu bir aktivasyon fonksiyonundan (Denklem 2.15) geçirilir.

$$y = f(x_1 * w_1 + x_2 * w_2 + b) \quad (2.15)$$

Aktivasyon fonksiyonumuz Şekil 10'daki bir sigmoid fonksiyonu olabilir.



Şekil 10. Sigmoid fonksiyonu grafiği

Sigmoid fonksiyonu $(-\infty, +\infty)$ arasında $(0,1)$ değerlerini alır. Sinir Ağının öğrenme adımları aşağıdaki gibidir;

- Ağırlık değerlerini ve bias'ı başlat
- İleri yayılım (forward propagation)
- Hatayı kontrol et
- Geri yayılım (back propagation)
- Ağırlıkları ve bias ayarla
- Tüm eğitim örnekleri için tekrarla.

2.1.5. Sınıflandırma Metrikleri

Eğitilen model test veri seti üzerinde değerlendirilmiştir. Bu değerlendirmede Tablo 2'deki karmaşıklık matrisi üzerinde elde edilen sınıflama ölçütleri kullanılmıştır. Kullanılan ölçütlerin detayları Denklem 2.16 – Denklem 2.19 arasında verilmiştir

Tablo 2

Karmaşıklık matrisi

		Gerçek Değer	
		Pozitif	Negatif
Tahmini Değer	Pozitif	True Positive (TP)	False Positive (FP)
	Negatif	False Negative (FN)	True Negative (TN)

TP (True Positive, Doğru Pozitif): Gerçekte pozitif sınıfa ait olanlar içinde modelin de pozitif olarak tahmin ettikleridir.

FP (False Positive, Yanlış pozitif): Modelin pozitif olarak tahmin ettiği ancak gerçek durumda negatif olan verilerdir.

FN (False Negative, Yanlış negatif): Gerçekte pozitif sınıf etiketine sahip olmasına rağmen model tarafından negatif şekilde tahmin edilen verilerdir.

TN (True Negative, Doğru negatif): Gerçekte negatif sınıf etiketli olup, tahmin modelinin de negatif olarak etiketlediği verilerdir (Polat,2021).

Doğruluk (Accuracy), sınıflandırıcının doğru sonuçlara ulaşma oranını ifade eder (Denklem 2.16). Bu, doğru tahmin edilen gözlemlerin tüm gözlemlere oranını temsil eder. Eğer kullanılan model yüksek bir doğruluğa sahipse, genellikle modelin başarılı olduğu düşünülebilir. Ancak, modelin sınıflama kararında ne kadar hata yaptığını gösteren diğer metriklerde model performansını açıklamak adına önemlidir. Özellikle hassasiyet (Precision), duyarlılık (Recall veya True Positive Rate) ve F1 skoru gibi metrikler, bu tür durumlarda modelin performansını daha detaylı bir şekilde analiz etmeye yardımcı olabilir. (Şahin,2018).

$$\text{Doğruluk (Accuracy)} = (TP+TN) / (TP+TN+FN+FP) \quad (2.16)$$

Kesinlik (Precision, Prec), modelin yapmış olduğu doğru tahminlerin, gerçekte ne kadarının doğru olduğunu gösterir. Kesinlik, yanlış pozitiflerin sınıflandırıcının ne kadar hassas olduğunu gösteren bir ölçüdür (Denklem 2.17). Kesinlik değerindeki artış modelin pozitif tahminlerinin güvenilir olduğunu gösterir.

$$\text{Kesinlik (Precision)} = TP / (TP + FP) \quad (2.17)$$

Duyarlılık (Recall) Gerçek pozitiflerin oranıdır. Gerçek pozitiflerin, gerçek pozitif ve yanlış negatiflerinin toplamına oranıdır. (Denklem 2.18). Yüksek bir duyarlılık değeri, modelin gerçek pozitifleri kaçırmadaki durumunu gösterir.

$$\text{Duyarlılık(Recall)} = TP / (TP + FN) \quad (2.18)$$

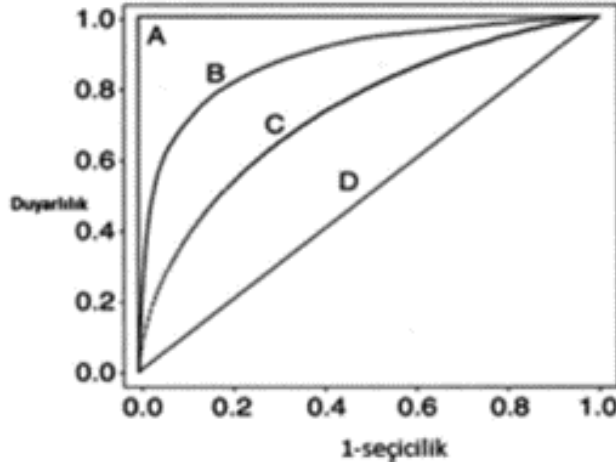
F1-Skoru, kesinlik ve duyarlılık arasındaki ilişkiyi Denklem 2.19’da verilen hesaplamalar ile gösterir. Sınıf etiketlerinin ve/veya öz niteliklerdeki dağılımın normal dağılıma uymadığı durumlarda F1 skoruna performans belirleyici olarak kabul edilebilir (Şahin,2018).

$$F\text{-Measure} = 2*Precision*Recall / (Precision + Recall) \quad (2.19)$$

2.1.6. ROC Eğrisi

ROC (Receiver Operating Characteristic) eğrisi, makine öğrenmesi modellerinin sınıflamam ve/veya kestirim karşılaştırılmasında sıklıkla kullanılan ölçütlerden birisidir. 0,5 ile 1 arasında değerlere sahip modellerden 1’e en yakını üstün olarak kabul edilir (Ertorsun ve ark., 2010).

Bu eğri, bir testin duyarlılık ve özgüllük performansını görsel olarak gösterir. ROC eğrisinin altındaki alan, çalışılan tanı testinin ayırma yeteneğinin istatistiksel olarak ne kadar önemli olduğunu ifade eder. ROC eğrisinin altındaki alanın büyüklüğü, bir tanı testinin pozitif ve negatif durumları ayırt etme yeteneğini ölçer.



Şekil 11: ROC eğrisi test durumları (A, B, C, D) (Kılıç,2013)

Şekil 11'e göre farklı dört sınıflama modeli için bir ROC eğrisi verilmiştir. Buna göre Model-A diğerlerine göre 1'e daha yakın sonuç verdiği için, en başarılı model olarak seçilmiştir (Kılıç,2013).

2.2. Önceki Çalışmalar

Makine öğrenmesi, büyük verileri analizindeki başarısı ile diğer birçok alanda olduğu gibi tıp alanında da sıklıkla tercih edilmektedir. Bu bölümde özellikle karaciğer rahatsızlıkları ile ilgili güncel makine öğrenme çalışmaları incelenmiştir.

Negassa ve ark. (2023), halka açık karaciğer tarama verilerinden kanserleri tespit etmenin yanı sıra, sonuçlarda mükemmel verimlilik sergileyen Gri kurt Optimizasyon (GWO)-Ekstrem Öğrenme Modeli (ELM) yaklaşımıyla yeni bir derin öğrenmeye dayalı segmentasyon sunmuştur. Karaciğer tümörü tespit sisteminin etkinliğini arttırmak için bu çalışmada GWO-ELM sınıflandırıcısını ve Haar dalgacık dönüşümünü uygulanmıştır. GWO-ELM, Sinir Ağı yapısına sahip bir Destek Vektör Makinesi gibi davranarak çoklu ve ikili sınıflandırma problemlerini çözebilir. Buna karşılık, Haar dalgacık dönüşümü en uygun özellikleri düşük boyutlulukla çıkarabilir. Sonuç olarak, GWO-ELM sınıflandırıcısı ve Haar dalgacık dönüşümü özellikleri, karaciğer tümörlerinden özelliklerin sınıflandırılması ve çıkarılması için yararlı bir yöntem sağlamak üzere kullanılmıştır. Sonuçlara göre önerilen GWO-ELM modeli çok iyi performans göstererek çok sınıflı bir veri seti için %99,41 doğruluk elde edilmiştir. Bu, GWO-ELM ve Haar dalgacık dönüşümünün karaciğer tümörlerini tanımlamak için güçlü bir sınıflandırıcı olduğunu ve çeşitli görüntü verisi türlerini işlemek için kullanılabileceğini ortaya koymuştur.

Aparna ve Libish (2023), Derin Evrişimli Sinir Ağları kullanılarak BT tarama görüntülerinde hepatik tümörlerin otomatik segmentasyonu ve sınıflandırılması için yeni bir yöntem açıklamıştır. Tümör segmentasyonu için değiştirilmiş bir Yoğun U-net modeli oluşturmuşlardır. Sınıflandırma çerçevesi, normal ve kötü huylu karaciğer tümörlerini ayırt etmek için önceden eğitilmiş bir VGG-16 ağına sahip yeni bir derin CNN'ye dayanmaktadır. Önerilen sistem, MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge veri setine göre değerlendirilmiş ve %95,40 Zar Skoru, %92 Jaccard İndeksi ve %92,60 segmentasyon doğruluğu ile en iyi sonucu elde etmiştir. Sınıflandırma modelinde ise, %96 doğruluk, %95,80 hassasiyet, %96,20 özgülük ve %95,80 kesinlik sonucu elde edilmiştir.

Patil ve ark. (2023), tıbbi görüntüleme uygulamalarında karaciğer tümörlerini ayırtmak için Selective Attention UNet adı verilen yeni bir mimari önermiştir. İyi bilinen UNet mimarisini temel alan önerilen mimari, ağın önemli görevlere odaklanmasını sağlarken gereksiz görevleri ortadan kaldıran seçici bir dikkat modülüne sahiptir. Kodlayıcı ve kod çözücü yolları arasındaki bağlantının atlanması, ağın hem düşük seviyeli hem de yüksek seviyeli nitelikleri kullanarak verileri etkili bir şekilde bölümlere ayırmasını sağlayan tasarımın başka bir ögesidir. Herkese açık LiTS veri kümesinde, önerilen mimarinin performansını değerlendirilmiş ve dört temel modelle karşılaştırılmıştır: FCN, UNet, UNet++ ve SegNet. Deneylerde elde edilen 0,89'luk Zar Benzerliği Katsayısı (DSC), ortalama 0,76 IOU, önerilen mimarinin doğruluk ve sağlamlık kriterleri açısından tüm temel modelleri geride bıraktığını göstermektedir.

Bilen (2022), çalışmasında veri girişi için arayüz tasarlamıştır. Arayüz aracılığı ile veriler veri tabanında toplamıştır. Bunun yanı sıra hastaların dosyasında bulunmayan bilgiler hastanenin bilgi işleminden talep edip hepsi tek bir veri havuzunda toplamıştır. Toplanan verilerin sınıflandırması için makine ve derin öğrenme algoritmalarından modelleri kullanmıştır. Veri setini %80'i eğitim, %20'si test olarak ayırmış olup modellerin eğitiminde ve test edilmesinde kullanmıştır. Aynı zamanda veri setinde ayırım yapılmadan k çapraz doğrulama yapmıştır. Eğitilen modellerin performans metriklerinin en iyi sonucun hangi modelde olduğunu göstermiştir. Veri ön işleme, modellerin oluşturulması, eğitilmesi ve test edilmesi Google Colaboratory bulut ortamında, Python programlama diliyle gerçekleştirmiştir. Modelleri oluşturmadan önce k çapraz doğrulama ve özellik çıkarıma yöntemlerinden LDA ve PCA yöntemlerini kullanmıştır. Veri setinde dört bağımlı değişken almıştır. Ex bağımlı değişkeni için modeller arasında en yüksek doğruluk oranı PCA yöntemi

ile %90 olmuştur. K çapraz doğrulama sonrası en yüksek doğruluk oranı LDA yöntemi ile %98 olmuştur. Nüks yeri bağımlı değişkeni için modeller arasında en yüksek doğruluk oranı her iki yöntem ile de %82 olmuştur. K çapraz doğrulama sonrası en yüksek doğruluk oranı LDA yöntemi ile %98 olmuştur. Tümör nüks bağımlı değişkeni için modeller arasında en yüksek doğruluk oranı her iki yöntem ile %79 olmuştur. K çapraz doğrulama sonrası en yüksek doğruluk oranı LDA yöntemi ile %99 olmuştur. Ölüm nedeni bağımlı değişkeni için modeller arasında en yüksek doğruluk oranı LDA yöntemi ile %86 olmuştur. K çapraz doğrulama sonrası en yüksek doğruluk oranı LDA yöntemi ile %99 olmuştur. Bağımlı değişkenlere k çapraz doğrulama uygulandıktan sonraki doğruluk oranlarının ortalaması sonucunda en iyi yöntem olarak LDA'yı seçmiştir.

Alici-Karaca ve ark. (2022), , radyasyonun neden olduğu karaciğer hasarını tespit etmek amacıyla histopatolojik görüntü sınıflandırması için yama tabanlı hafif bir evrişim sinir ağı önermiştir. Önerilen model, 555 histopatolojik görüntüden oluşan radyasyona bağlı karaciğer hastalığı veri kümesini sınıflandırmak için eğitilmiştir. Görünmeyen test setinde ikili sınıflandırmada %100 doğruluk ve çok sınıflı sınıflandırmada %87,57 doğruluk elde edilmiştir. Model, ResNet-50, AlexNet, Vgg16 ve GoogleNet dahil olmak üzere en son teknolojiye sahip modellerden daha iyi performans göstermiştir.

Saygın (2021), veri seti olarak UCI'de herkesin erişimine açık olan Indian Liver Patient Dataset (ILPD) kullanmıştır. Hastalık teşhis başarısını arttırmak amacıyla veri setlerinde bulunan öznelilikler üzerinde sarmal özellik seçme yöntemleri kullanarak teşhis başarısını arttıracak öznelilikler belirlemiştir. Belirlenen önemli öz nelilikler üzerinde makine öğrenmesi yöntemleri uygulanarak karaciğer hastalığı teşhis başarılarını ölçmüştür. Karaciğer hastalığının sınıflandırma işlemleri için makine öğrenmesi yöntemleri kullanılmıştır. Makine öğrenmesi yöntemlerinden Destek Vektör Makinası (SVM), Çok Katmanlı Algılayıcı (MLP), Karar Ağaçları (DT), Lojistik Regresyon (LR) ve Hafif Gradyan Güçlendirme Makinesi Sınıflandırıcısı (LGBM) yöntemleri uygulamıştır.

Dhahir (2021), AI ve tamamen evrişimli sinir ağı (FCNN) kullanarak otomatik karaciğer tümörü segmentasyonu ve sınıflandırması tasarlamış ve uygulamıştır. Görsel Geometri Grubu (VGG-16) önceden eğitilmiş modeli, R2019a Matlab'da çok büyük ölçekli veri setinde sorunsuz mimari olması nedeniyle veri setinin eğitimi için "LiTS veri seti"

kullanmıştır. Veri seti, eğitim için %70 ve test bölümü için %30 oranında bölmüştür. Tezinde, R2019a Matlab'da karaciğer tümör görüntülerinin tanınması için sağlamlıklarını geliştirmek gibi ağları eğitmek için bir yaklaşım sunmuştur. Bu eğitim stratejisi daha sonra tasarlanmış VGG-16 ağ mimarisi için değerlendirilmiştir. Çalışmanın sonucunda, eğitim algoritmasının, yüksek segmentasyon ve sınıflandırma doğruluğu ile hastalık görüntülerinin performansında bir artış pahasına, farklı karaciğer lezyon örneklerini tanımaya yönelik sağlamlığı artırabileceğiydi. Farklı mimari türlerinin avantajları değerlendirildiğinde, 10 çapraz kat doğrulama ile 28 saat süren 300 eğitim dönemi ile karaciğer lezyonu görüntüsüne göre karaciğer tümörünün doğruluğunun %91 civarında olduğunu bulmuştur. Araştırma, karaciğer tümörü segmentasyonu için VGG-16 ile tam otomatik FCN modelini yenilikçi bir şekilde birleştiriyor. İkinci olarak, tıbbi görüntülemeye etiket dengesizliği sorunu ile ilgili olarak, ROI'leri bulmak ve yerel segmentasyonları gerçekleştirmek amacıyla nesne algılama tekniklerini semantik segmentasyon çerçevesine entegre etmiştir. Bu lezyon lokalizasyonu adımı, segmentasyon ağlarının eğitilmesi için ek bir denetim düzeyi getirmiştir.

Özsoy(2021), çalışmasında Hint Karaciğer Hasta Kayıtları Hindistan'ın Kuzey Doğusundan toplanan hasta kayıtları ; ILPD (Hindistan Karaciğer Hasta Veri Seti) Veri Seti kullanarak veri setleri üzerinden hastanın; Yaşı ,cinsiyeti ,TB Toplam Bilirubin , DB Direkt Bilirubin , Alkphos Alkalın Fosfotaz, Sgpt Alamin Aminotransferaz , Sgot Aspartat Aminotransferaz , TP Toplam Protienler , ALB Albümin ,A / G Oranı Albümini ve Globulin Oranı , Verileri iki gruba ayırmak için kullanılan seçici alanı türleri gibi özelliklerle çeşitli makine öğrenmesi yöntemleri uygulamıştır. Çeşitli sınıflandırma algoritmaları: Lojistik regresyon, Naive Bayes Sınıflandırıcı, K-En Yakın Komşular, Karar Ağacı, Rastgele Orman, Destek vektör makinaları, Cosine Scikit-Learn algoritmaları karşılaştırmış, çıkan sonuçları değerlendirmiş ve farklı parametreler kullanarak karşılaştırmıştır.

Kavur (2020), segmentasyon yöntemleri için yeni karşılaştırma platformlarının oluşturulması, füzyon yöntemlerinin tıbbi görüntü segmentasyonu uygulamalarına analizi ve adaptasyonu ve yeni bir topluluk yönteminin tasarlanması ile ilgili yeni çalışmalar sunmuştur.

Khattab ve ark. (2022), üç farklı tipte atlama bağlantısına sahip istiflenmiş bir 2-U-Nets modeli önermiştir. Önerilen bağlantılar, ilk U-Net'in üst örnekleme yolu sırasında havuzlama ve düşük seviyeli özelliklerin kaybı nedeniyle ilk U-Net'in evrimsel yolundaki

yüksek seviyeli özelliklerin kaybını telafi etmek için çalışır. Atlama bağlantıları, önceki yollardan aynı seviyede üretilen tüm özellikleri, ikinci U-Net'in her iki yolundaki evrişimli katmanların girişlerine yoğun bir şekilde bağlı bir şekilde birleştirir. Tahmin edilen karaciğer bölgesi ile temel gerçek arasındaki Zar benzerliğini maksimuma çıkararak, her U-Net'in her seviyesinde farklı sayıda filtreye sahip modelin iki versiyonunu uygulanmıştır. Önerilen modeller, MICCAI 2007-2008 mücadelesi sırasında Şerit ve 3D karaciğer ve tümör segmentasyonu zorlukları için yayımlanan 3Dircadb genel veri seti ile eğitilmiştir. Deneysel sonuçlar, önerilen modelin orijinal U-Net ve 2-U-Nets varyantlarından daha iyi performans göstermiş ve son teknoloji ürünü mU-Net, DC U-Net ve Basamaklı UNE ile karşılaştırılabilir olduğunu göstermiştir.

Devi ve Seenivasagam (2020), otomatik bir CAD sistemini üç aşamalı olarak sunmuştur. İlk aşamada otomatik karaciğer segmentasyonu ve lezyon tespiti gerçekleştirilmiştir. Bir sonraki adımda özellikler/öznitelikler çıkarılmıştır. Son olarak yeni kontrast bazlı özellik farkı yöntemi kullanılarak karaciğer lezyonlarının kötü huylu ve iyi huylu olarak sınıflandırılması yapılmıştır. Çevresindeki normal karaciğer dokusuyla birlikte lezyon alanından çıkarılan özellikler yoğunluk ve dokuya dayanmaktadır. Lezyon tanımlayıcısı hem lezyon alanının hem de karaciğerin normal dokusunun özellikleri arasındaki fark dikkate alınarak elde edilmiştir. Son olarak karaciğer lezyonlarını kötü huylu veya iyi huylu olarak kategorize etmek için yeni tanımlayıcılar üzerinde yeni bir SVM tabanlı makine öğrenimi sınıflandırıcısı eğitilmiştir. Ayrıca öngörülen yaklaşım, demografik özellikler, görüntüleme cihazları, hastalar ve ortamlar arasındaki doku ve yoğunluk aralıklarına karşı duyarsızdır. Sınıflandırıcı, tümörü nispeten yüksek hassasiyetle ayırt etmiş ve radyoloğa daha sonraki bir görünüm sunmuştur.

Wu ve ark. (2020), karaciğerdeki tümörlerin segmentasyonu için kullanılabilecek çok aşamalı bir derin evrişimli sinir ağı (DCNN) modeli önermiştir. Nesne algılama modelinin doğruluğunu daha da artırmak için modelin performansını artıran ve tümörlerin daha doğru lokalizasyonunu sağlayan yinelemeli bir eğitim yöntemi önermiştir. Son olarak, önerilen yöntemin Zhujiang Hospital of Southern Medical University'den özel veri seti ve MICCAI-LiTS2017'den kamuya açık veri seti üzerinde değerlendirilmesi, modelin mevcut en gelişmiş modellerden üstün olduğunu ortaya koymuşlardır.

Avcı (2019), evrişimli sinir ağı modeli kurmuştur. İlgili çalışmasında hiper-parametre değişikliklerin sonuçlar üzerindeki etkisini incelemiştir. Bu parametrelerin biri olan farklı girdi ölçütlerinde en iyi sonucu beş yüz on ikilik kare matris ve U-NET modelinde elde etmiştir. Söz konusu modelin F metriği ile %95 üzerinde performans sergilemiştir. Klasik aktivasyon fonksiyonlarının (Sigmoid, Doğrusal vb.) başarımının ilgili modelde düşük olduğunu gözlemlemiştir.

Karaciğer teşhisi için literatürdeki çalışmalardan birisi de Karşlı (2019) tarafından yapılmıştır. Bu çalışmada farklı ağaç algoritmalarını kullanmış, bunların performansı kendileri ve klinikteki ölçütler ile karşılaştırılmıştır. İlgili çalışma özellikle yapılan ön işleme adımları ile önce çıkmaktadır. Eksik veriler için sınıf ortalaması, dengesiz dağılım için ise normalizasyondan yararlanılmıştır. Yine dağılım dengelemesi için elektronik veri üretimi yapılmıştır. Çapraz geçerlilik testi ile validasyonu yapılan modeller arasında Rastgele Orman %90 üzerinde ROC eğrisi altında kalan alan ile ilk sırada yer almıştır.

Borulday (2017), farklı karaciğer hastalıklarını kapsayan geniş bir veri tabanı oluşturmuştur. Karar Ağacı Algoritmalarını deneyerek en iyi sonuca ulaşılacak şekilde yazılım geliştirmiştir. Yapılan yazılım geliştirmelerinde klinikteki farklı süreçlere yönelik gruplamalar yapılmıştır. Bu sayede her farklı süreçte hekime kararlarında destek olacak bir yazılım gerçekleştirilmiştir.

Tablo 3, önceki çalışmaların temel bilgilerini ve ana sonuçlarını göstermektedir. Çalışmalardan eksik bilgiler olduğu için bu kısımları "Belirtilmemiş" olarak işaretlenmiştir.

Tablo 3

Önceki çalışmaların özeti

Çalışma (Yazar ve Yıl)	Kullanılan Yöntemler ve Teknolojiler	Veri Seti	Başarı Oranları
Negassa ve ark. (2023)	GWO-ELM, Haar dalgacık dönüşümü	Belirtilmemiş	%99,41 (Doğruluk)
Aparna ve Libish. (2023)	Evrişimli Sinir Ağları, Yoğun U-net, VGG-16	LiTS Challenge, LiTS	%95,40 (Zar Skoru), %92 (Jaccard İndeksi), %92,60 (Segmentasyon Doğruluğu)

Çalışma (Yazar ve Yıl)	Kullanılan Yöntemler ve Teknolojiler	Veri Seti	Başarı Oranları
Patil ve ark. (2023)	Selective Attention UNet	LiTS	0.89 (Zar Benzerliği Katsayısı), 0.76 (IOU)
Bilen (2022)	Arayüz Tasarımı, Makine ve Derin Öğrenme Algoritmaları	Google Colaboratory, Bilinmeyen	%90 (Ex), %98 (Nüks Yeri), %79 (Tümör Nüks), %86 (Ölüm Nedeni)
Alici-Karaca ve ark. (2022)	Hafif Evrişimli Sinir Ağı (555 görüntü)	Belirtilmemiş	%100 (İkili Sınıflandırma), %87,57 (Çoklu Sınıflandırma)
Saygın (2021)	Sarmal Özellik Seçme Yöntemleri, Makine Öğrenmesi Yöntemleri	ILPD	Belirtilmemiş
Dhahir (2021)	Tamamen Evrişimli Sinir Ağı (FCNN), VGG-16	LiTS	%91 (Karaciğer Tümörü Doğruluğu)
Özsoy (2021)	Seçici Alanlar, Makine Öğrenmesi Yöntemleri	ILPD	Belirtilmemiş
Wu ve ark. (2020)	Derin Evrişimli Sinir Ağı (DCNN)	Zhujiang Hospital veri seti, LiTS	Belirtilmemiş
Avcı (2019)	Evrişimli Sinir Ağları, Çeşitli Girdi Boyutları ve Aktivasyon Fonksiyonları	Belirtilmemiş	%97,58 (256x256), %94,97 (512x512)
Karşlı (2019)	Makine Öğrenmesi Algoritmaları, Çapraz Doğrulama, LDA, PCA	UCI Machine Learning Repository (Karaciğer Hastalığı veri seti)	%90 (Ex), %98 (Nüks Yeri), %82 (Tümör Nüks), %99 (Ölüm Nedeni)
Borulday (2017)	Karar Ağacı Algoritmaları	UCI Machine Learning Repository (Karaciğer Hastalığı veri seti)	Belirtilmemiş
Önerilen çalışma	Görüntü İşleme, Makine Öğrenmesi Yöntemleri	Hastane verileri	LR:%96, SVM:%95,RF:%90,NN:%98

Karaciğer tümör hastalıklarının teşhisi ve sınıflandırmasında daha çok evrişimli sinir ağlarının (CNN) tercih edildiği görülmektedir. Bu yöntemin tercih edilme nedeni çok sayıda görüntünün olması, karaciğerin ve tümörün otomatik olarak bölütlenmesindeki ve sınıflandırılmasındaki yüksek başarı olduğu değerlendirilmektedir.

Bu tez için yapılan çalışmanın üstün yanı görüntü işleme ile iyileştirme işlemlerin yapılması, bölütlemenin kontrol edilmesi ve diğer makine öğrenme yöntemlerinin testi edilmesi ile en iyi dört yöntemin seçilmesi olmuştur. Yapılan işlemin dezavantajı ise karaciğer bölütlemesinde görüntü işlemenin ileri seviyede bilgi gerektirmesi, bölütleme işlemlerinin çok uzun zaman alması, manuel maskeleme ile doğruluğunun teyit edilememesi ve alan uzmanlarının yoğun mesailerinden dolayı gerekli bilgi alışverişinin kısıtlı olması ve

buna baęlı olarak istenilen seviyede grntlerin sınıflandırmaya hazır hale getirilememesi gibi durumlardır.



ÜÇÜNCÜ BÖLÜM

MATERYAL VE YÖNTEM

3.1. Araştırma Yöntemi

Kullanılan veri kümesi hastalara ait tomografi görüntülerini içermektedir. Bu veriler hastane tarafından anonim hale getirildikten sonra (kişisel bilgilerden arındırıldıktan sonra) ilgili tez çalışması için paylaşılmıştır. Kullanılan bu veri setinin tümör adı ve tümör türüne göre hasta sayısı dağılımı ve toplam görüntüleme sayısı Tablo 4’te verilmiştir.

Tablo 4.

BT görüntü dosyalarının dağılımı

Tümör Adı/Klasör	Tümör Grubu	Hasta Sayısı	Görüntü Sayısı (JPG/TIFF)
FBNH	İyi huylu	22	2217
HCC	Kötü huylu	22	2499
HEMANJIOM	Kötü huylu	29	3140
KIST	İyi huylu	20	1930
KOLANJIOSELLÜLER	İyi huylu	19	2156
METASTAZ	Kötü huylu	21	2725

Tomografi görüntüleri ilaç kullanımına göre üçe ayrılmaktadır (kullanım öncesi – sonrası ve sonrası). Her bir görüntü anonim hasta numarasına göre ilaç kullanım durumu ve tümör şeklini belirtilen ayrı klasörler şeklinde saklanmıştır.

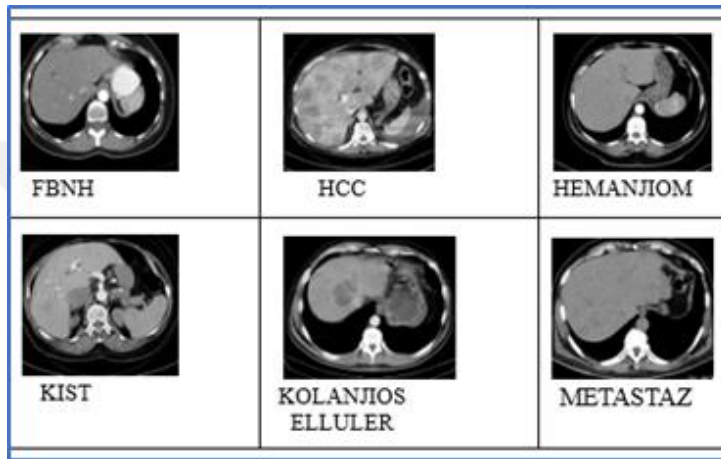
Görüntü dosyaları eğitim ve test kümelerine %80’e %20 oranı sağlanacak şekilde ayrılmıştır. Ön işleme adımında görüntüler yeniden boyutlandırılmış (512’lik kare piksel şeklinde) ve bölütleme için OTSU yöntemi (Otsu,1979) tercih edilmiştir. Bu sayede karaciğer görüntüleri ayrıştırılarak ayrı dosyalar olarak gruplandırılmıştır.

Özniteliklerin tespit edilebilmesi için Orange Data Mining ve SqueezeNet’den yararlanılmıştır. Sınıf etiketlerinin tespiti için Lojistik Regresyon (LR), Destek Vektör Makinesi (SVM), Rastgele Orman (RF) ve Yapay Sinir Ağı (ANN) modelleri kurularak bu

modellerin performansları karşılaştırılmıştır. Bu karşılaştırma için kesinlik, doğruluk, geri çağırma ve F1 skoru gibi farklı metrikler kullanılmıştır.

3.2. Veri Seti

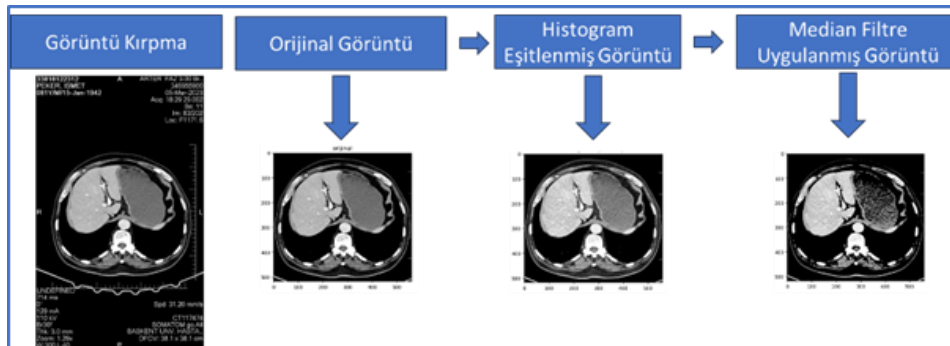
Karaciğer Bilgisayarlı Tomografi (BT) veri seti eğitim ve test olarak ikiye ayrılmıştır. Eğitim kümesinde 2622, test kümesinde 656 görüntü bulunmaktadır. Her bir görüntü üçer tane iyi ve kötü huylu tümörü temsil edecek şekilde ayarlanmıştır (Şekil 12).



Şekil 12. Tümör gruplarına ait örnek görüntüler

3.2.1. Görüntülerin İşlenmesi

Görüntü ön işleme adımında sırasıyla görüntü temizleme, görüntü dönüştürme ve görüntü hazırlama adımları uygulanmıştır. Bu sayede verinin analiz için en uygun haline getirilmesi amaçlanmıştır. Uygulanan adımlar Şekil 13'te görülmektedir.

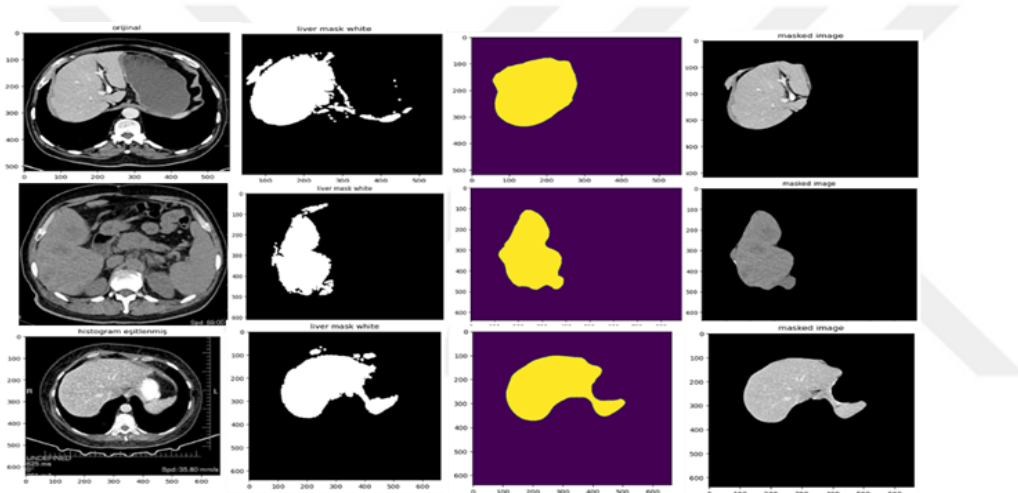


Şekil 13. Görüntü işleme adımları

Şekil 13'e göre önce görüntüler otomatik olarak çözünürlüklerine göre kırpma işleminde göre tabi tutularak gereksiz alanlar yok edilir. Daha sonra histogram eşitlemesi yapılarak renk dağılımı bozukluklarını giderilir. Medyan filtre kullanarak gürültünün temizlenmesi sağlanır.

3.2.2. Histogram Eşikleme ve Gri Seviye Segmentasyonu

OTSU eşikleme yöntemi (Otsu,1979) geliştirilen otomatik görüntü eşiklemede (Şekil 14) ilk yöntemlerden biridir. Bu yöntemde ön ve arka plandaki pikseller karşılaştırılır. Bu görüntü noktalarının normal dağılım içindeki değişimleri en aza indirgenerek en uygun eşik noktası hesaplanır (Karhan,2011).

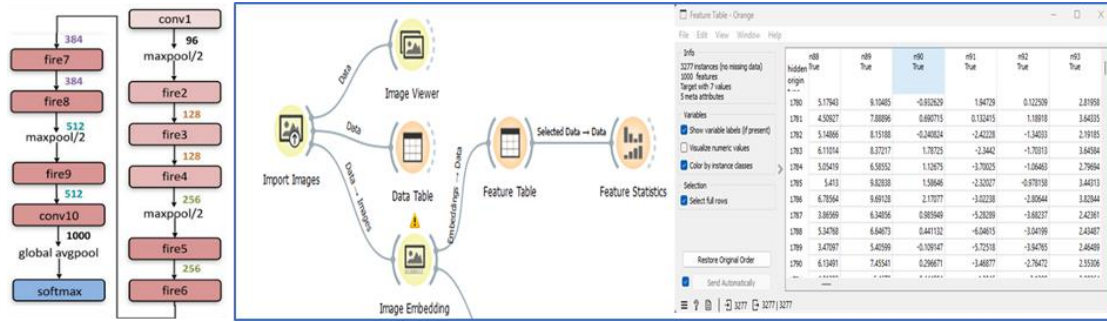


Şekil 14. OTSU Eşikleme ve bölütleme uygulanmış örnek görüntüler

3.2.3. Öznitelik Çıkarma

Görüntü sayısal hale çevrildiğinde tüm görüntüyü temsil edebilme yeteneği sahip alt özelliklerine (renk, şekil, doku, köşe noktalar vb.) öznitelik denir. Bunları tespit etme adımı ise öz nitelik çıkartımı olarak adlandırılır (Soysal,2016). Bu sayede eldeki problemle ilgili bilgi vermeyen veriler elenerek daha hızlı ve güvenilir sonuçlar elde edilebilir (Budak,2018).

Orange Data Mining programı genel amaçlı bir makine öğrenimi ve veri madenciliği aracıdır (Demsar ve Zupan,2013). İçerisinde önceden eğitilmiş SqueezeNet derin öğrenme modeli kullanılarak bir özellik matrisi elde edilmiştir. SqueezeNet mimarisinin amacı; daha az parametreye sahip bir sinir ağı oluşturmak ve 50 kat daha az parametre ile AlexNet düzeyinde doğruluk sağlamaktadır (Gökarp ve Aydın,2021). Yapılan çalışmada öznitelik çıkarımı için Orange Data Mining aracında bulunan SqueezeNet tercih edilmiştir (Şekil 15).



Şekil 15. Orange Data Mining Squeezenet mimarisiyle özellik çıkarma

İşlem sonucunda veri kümesindeki her görüntü için bin öz nitelikten oluşan bir öz nitelik havuzu elde edilmiştir.

3.2.4. Sınıflandırma

Orange Data Mining Programı üzerinde Bölüm 2’de ayrıntılı olarak açıklanan Lojistik Regresyon, Destek Vektör Makinesi, Rastgele Orman ve Yapay Sinir Ağı sınıflandırma algoritmaları varsayılan parametreleri ile kullanılmıştır (Tablo 5).

Tablo 5

Makine öğrenme yöntemleri varsayılan parametre değerleri

Sınıflandırma Algoritması	Vasayılan Parametreler
Lojistik Regresyon	Düzenleme: Ridge (L2), C(maliyet)=1, Sınıf Ağırlıkları: No
Destek Vektör Makinesi	SVM türü: SVM, C(maliyet)=1.0, $\epsilon=0.1$, Çekirdek: RBF, $\exp(-\text{auto} x-y ^2)$ Sayısal tolerans: 0.001 İterasyon sınırı 100
Rastgele Orman	Ağaç sayısı: 10 Maksimum öz nitelik sayısı: sınırsız Tekrarlanabilir eğitim: No Tek tek ağaçların derinliğinin sınırlanması: sınırsız Bölünebilecek en küçük alt küme: 5
Yapay Sinir Ağı	Gizli katman sayısı: 100 Aktivasyon fonksiyonu: ReLu Ağırlık optimizasyonu için çözücü: Adam Alpha (L2 cezası (düzenleme süresi) parametresi): 0.0001 Masimum iterasyon: 200 Tekrarlanabilir eğitim: Yes

DÖRDÜNCÜ BÖLÜM

ARAŞTIRMA BULGULARI

Bu çalışmada Jupyter Notebook ve Orange Data Mining programları kullanılmıştır. OpenCV kullanılarak görüntü işleme yöntemleri ile öncelikle karaciğer segmentasyonu için ön işlemler yapılmıştır. Daha sonra OTSU eşikleme yöntemi ile segmentasyon yapılmıştır. Öznitelik çıkarma işlemi için Orange Data Mining programında mevcut olan SqueezeNet modeli kullanılmıştır. Öznitelik seçiminde ise ReliefF algoritması kullanılmıştır. Sınıflandırma modelleri için detayları yukarıda açıklanan modellerin Orange Data Mining programındaki varsayılan parametreleri kullanılmıştır.

Tümör isimlerine göre makine öğrenmesi yöntemlerinin karmaşıklık matrisleri Şekil (16-19)'da gösterilmiştir.

		Predicted					
		FNHBT	HCCBTDINAMIK	HEMANJIOMBT	KISTBT	KOLANJIOSELLULERBT	METASTAZBT
Actual	FNHBT	74.8 %	3.2 %	5.4 %	5.5 %	3.4 %	1.0 %
	HCCBTDINAMIK	4.7 %	81.8 %	4.9 %	4.6 %	6.5 %	3.5 %
	HEMANJIOMBT	11.2 %	5.2 %	78.9 %	9.4 %	5.9 %	1.2 %
	KISTBT	4.0 %	2.6 %	4.6 %	74.0 %	4.0 %	1.3 %
	KOLANJIOSELLULERBT	3.6 %	4.3 %	4.4 %	4.4 %	79.6 %	0.5 %
	METASTAZBT	1.6 %	2.9 %	1.6 %	2.0 %	0.6 %	92.5 %

Şekil 16. Lojistik Regresyon algoritması için karmaşıklık matrisi

		Predicted					
		FNHBT	HCCBTDINAMIK	HEMANJIOMBT	KISTBT	KOLANJIOSELLULERBT	METASTAZBT
Actual	FNHBT	74.7 %	2.1 %	9.6 %	4.0 %	5.6 %	0.5 %
	HCCBTDINAMIK	4.9 %	85.1 %	6.0 %	5.1 %	8.4 %	2.8 %
	HEMANJIOMBT	12.1 %	6.3 %	69.9 %	6.8 %	18.9 %	0.6 %
	KISTBT	3.8 %	2.3 %	5.0 %	80.0 %	6.6 %	0.0 %
	KOLANJIOSELLULERBT	4.0 %	2.7 %	7.1 %	2.3 %	59.8 %	0.0 %
	METASTAZBT	0.4 %	1.5 %	2.3 %	1.8 %	0.7 %	96.0 %

Şekil 17. Destek Vektör Makinesi algoritması karmaşıklık matrisi

		Predicted					
		FNHBT	HCCBTDINAMIK	HEMANJIOMBT	KISTBT	KOLANJOSELLULERBT	METASTAZBT
Actual	FNHBT	62.0 %	5.1 %	11.5 %	9.4 %	6.6 %	1.4 %
	HCCBTDINAMIK	7.1 %	73.5 %	8.7 %	6.6 %	7.0 %	5.8 %
	HEMANJIOMBT	15.3 %	8.1 %	54.7 %	12.5 %	13.6 %	3.0 %
	KISTBT	6.2 %	3.8 %	9.5 %	62.7 %	9.0 %	1.9 %
	KOLANJOSELLULERBT	7.6 %	4.2 %	12.6 %	8.0 %	61.8 %	0.6 %
	METASTAZBT	1.7 %	5.3 %	2.9 %	0.9 %	2.0 %	87.3 %

Şekil 18. Rastgele Orman algoritması karmaşıklık matrisi

		Predicted					
		FNHBT	HCCBTDINAMIK	HEMANJIOMBT	KISTBT	KOLANJOSELLULERBT	METASTAZBT
Actual	FNHBT	86.4 %	1.8 %	4.2 %	2.1 %	1.3 %	0.9 %
	HCCBTDINAMIK	2.1 %	89.7 %	3.7 %	4.3 %	3.7 %	2.3 %
	HEMANJIOMBT	4.7 %	3.4 %	85.0 %	6.1 %	4.6 %	1.0 %
	KISTBT	3.3 %	1.7 %	2.6 %	83.9 %	1.8 %	0.4 %
	KOLANJOSELLULERBT	2.3 %	1.8 %	2.4 %	3.1 %	88.2 %	0.3 %
	METASTAZBT	1.3 %	1.5 %	2.1 %	0.5 %	0.4 %	95.1 %

Şekil 19. Yapay Sinir Ağı algoritması için karmaşıklık matrisi

Tümör isimlerine göre makine öğrenmesi yöntemlerinin değerlendirme sonuçları matrisleri Tablo (6-11)'de gösterilmiştir.

Tablo 6

FNHB için makine öğrenmesi performans değerleri

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.962	0.937	0.745	0.748	0.742	0.709	0.965	0.210
SVM	0.948	0.931	0.707	0.747	0.672	0.670	0.968	0.204
Random Forest	0.864	0.890	0.506	0.566	0.458	0.448	0.950	0.344
Neural Network	0.983	0.964	0.854	0.864	0.844	0.834	0.981	0.122

Tablo 7

HCC için makine öğrenmesi performans değerleri

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.956	0.924	0.814	0.818	0.810	0.767	0.954	0.276
SVM	0.948	0.926	0.809	0.851	0.772	0.765	0.965	0.234
Random Forest	0.898	0.892	0.729	0.745	0.714	0.662	0.937	0.427
Neural Network	0.980	0.954	0.885	0.897	0.873	0.856	0.974	0.167

Tablo 8

HEMANJIOM için makine öğrenmesi performans değerleri

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.953	0.908	0.785	0.789	0.782	0.727	0.943	0.301
SVM	0.907	0.864	0.671	0.699	0.645	0.587	0.924	0.336
Random Forest	0.859	0.808	0.605	0.542	0.684	0.486	0.842	0.402
Neural Network	0.981	0.938	0.857	0.850	0.865	0.818	0.958	0.186

Tablo 9

KIST için makine öğrenmesi performans değerleri

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.957	0.942	0.742	0.740	0.743	0.709	0.967	0.212
SVM	0.950	0.946	0.747	0.800	0.701	0.719	0.978	0.171
Random Forest	0.881	0.915	0.573	0.661	0.505	0.532	0.967	0.354
Neural Network	0.987	0.965	0.844	0.839	0.850	0.824	0.979	0.112

Tablo 10

KOLANJIOSELLULER için makine öğrenmesi performans değerleri

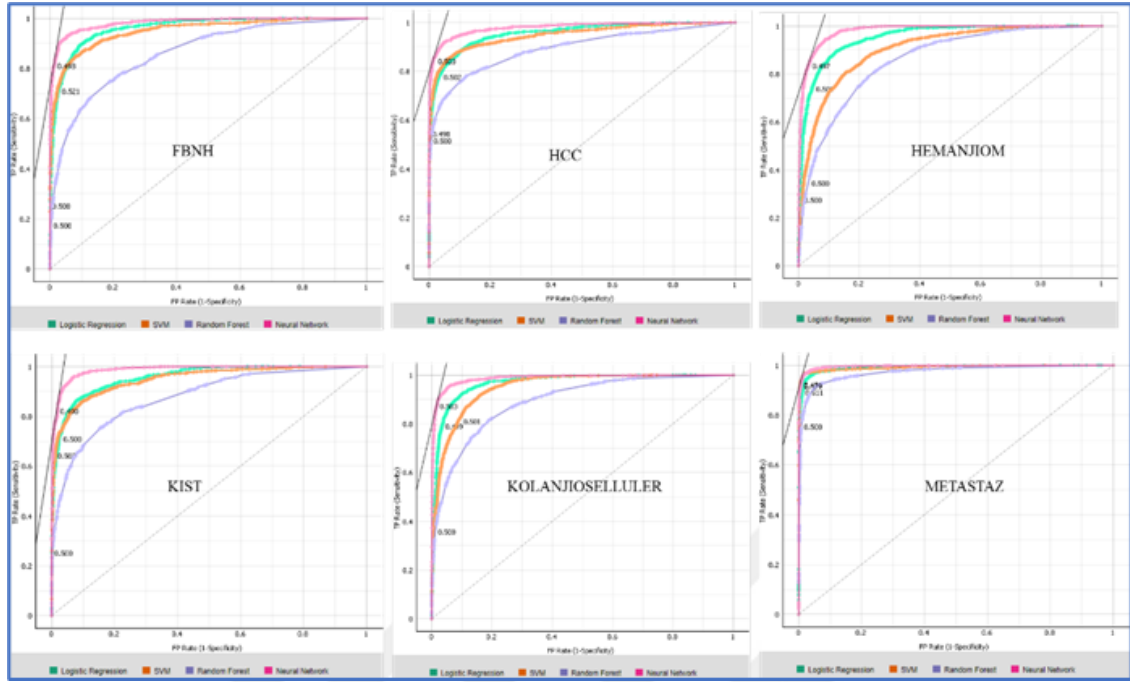
Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.968	0.937	0.810	0.796	0.825	0.773	0.959	0.244
SVM	0.952	0.882	0.699	0.598	0.841	0.643	0.890	0.259
Random Forest	0.891	0.873	0.614	0.608	0.620	0.537	0.922	0.352
Neural Network	0.987	0.965	0.893	0.882	0.905	0.872	0.976	0.127

Tablo 11

METASTAZ için makine öğrenmesi performans değerleri

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
Logistic Regression	0.992	0.972	0.922	0.925	0.918	0.905	0.984	0.111
SVM	0.992	0.982	0.949	0.960	0.939	0.938	0.991	0.076
Random Forest	0.970	0.950	0.862	0.863	0.861	0.831	0.970	0.238
Neural Network	0.996	0.981	0.947	0.951	0.943	0.935	0.989	0.067

Çalışmada değerlendirilen skorların ROC eğrileri Şekil 20’de sunulmuştur.



Şekil 20 ROC eğrileri

Bütün tümör çeşitlerinin sınıflandırmasında en yüksek tahminin Yapay Sinir Ağı modeli tarafından yapıldığı ROC eğrisinden görülmektedir. Lojistik Regresyon ikinci sırada, Destek Vektör Makinesi üçüncü ve Rastgele Orman modeli son sırada yer almaktadır. AUC ve F1 skorunun da aynı şekilde ROC eğrisine göre yapılan tahminlerde aynı sıralamada olduğu Tablo (6-11)’de görülmektedir.

BEŞİNCİ BÖLÜM

SONUÇ VE ÖNERİLER

İlgili çalışma görüntü işleme ve işlene görüntülerin etiketlenmesi ana adımlarından oluşmuştur. Görüntü işleme adımımda üçer tane farklı iyi ve kötü huylu kitlenin bölütlenmesi yapılmıştır. Sonrasında bu görüntülerin etiketlenmesi için farklı makine öğrenme yöntemlerinin başarımları performansları farklı metrikler kullanılarak karşılaştırılmıştır.

Çalışmada yukarıda sonuçları gösterilen 4 makine öğrenmesi sınıflandırma modeli (Lojistik Regresyon, Destek Vektör Makinesi, Rastgele Orman, Yapay Sinir Ağı) Orange Data Mining 3.36.1 programındaki varsayılan parametre değerleri ile kullanılmıştır. Çalışma sonucunda en iyi tahmin performansını Yapay Sinir Ağı modeli göstermiştir. Görüntü işleme adımımda, bölütleme kaynaklı hataların azaltılmasının modellerin tahmin performansını artıracakı ön görülmüştür.

Görüntü işleme ve makine öğrenmesi yöntemleri kullanarak Karaciğer tümörlerinin tespiti ve sınıflandırmasında teknik olarak en büyük zorluk standart bir görüntü formatının (çözünürlük, renk vb.) olmamasıdır. Bu nedenle farklı açık kaynak platformlarında görüntü işleme üzerine hazır olarak sunulan kütüphane ve uygulamalar mevcut veri seti ile çalıştığında bazı durumlarda tatmin edici performans değerlerine ulaşamamaktadır. Bunun yanında tez çalışma kapsamında en başarılı model olan yapay sinir ağı ile %98 doğruluk , %88 kesinlik ve %88 geri çağırma değerleri elde edilmiştir. Başarımların daha da artırılabilmesi için gelecek çalışmalarda alan uzmanlarının bulduğu ve doğruladığı BT görüntüleri üzerinde derin öğrenme ve optimizasyon algoritmaları kullanılarak hibrit modellerin geliştirilmesi ile karaciğer hastalıklarının daha yüksek tespit ve sınıflandırma oranına kavuşacakı değerlendirilmektedir.

KAYNAKÇA

- Ahirwar, M. and Behera,B.K.(2023). *Prediction of Shrinkage Behavior of Stretch Fabrics Using Machine-Learning Based Artificial Neural Network Textiles* 3, no. 1: 88-97. <https://doi.org/10.3390/textiles3010007>
- Akar, Ö. ve Güngör, O. (2012). Rastgele Orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi*, (106), 139-146. <https://doi.org/10.9733/jgg.241212.1t>
- Alici-Karaca, D., Akay, B., Yay, A., Suna, P., Nalbantoglu, O. U., Karaboga, D., Basturk, A., Balcioglu, E., Baran, M. (2022). *A new lightweight convolutional neural network for radiation-induced liver disease classification*. *Biomedical Signal Processing and Control*, 73. <https://doi.org/10.1016/j.bspc.2021.103463>
- Aparna, P.R. and Libish, T M. (2023). *Automatic segmentation and classification of the liver tumor using deep learning algorithms*. 2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), 334–339. <https://doi.org/10.1109/ACCESS57397.2023.10200900>
- Arısoy, M. ve Dikmen, Ü. (2014). Manyetik Belirti Haritalarının Histogram Eşitleme Yöntemi Kullanılarak İyileştirilmesi. *Yerbilimleri/ Earth Sciences*. 35. 175-184.
- Avcı, F. (2019). Derin Öğrenme Teknikleriyle Tomografi Görüntülerinde Karaciğer Bölütlemesi. Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir.
- Behbudova, S. (2023). Makine Öğrenmesi Algoritmaları ile Kardiyovasküler Hastalıkların Tahmin Edilmesi. Yüksek Lisans Tezi. İstanbul Üniversitesi-Cerrahpaşa Lisansüstü Eğitim Enstitüsü, İstanbul.
- Bilen, Z. (2022). Karaciğer Transplantasyonu Sonrası Makine ve Derin Öğrenme Yöntemleri ile Karaciğer Hastalıklarının Teşhisi. Yüksek Lisans Tezi, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Erzurum.
- Boateng,K, Weyori, B. ve Laar, D. (2012). Improving the Effectiveness of the Median Filter. *International Journal of Electronics and Communication Engineering*. 5. 85-97.

- Borulday, M.G. (2017). Karaciğer Hastalıklarının İşlemsel Zeka Yöntemleriyle Teşhisi. Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- BT Enterografi (2023). BT MR Enterografi. Erişim: 18.12.2023, <https://www.barisbakir.com/bt-mr-enterografi.aspx> (
- Budak, H. (2018). Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 21-31.
- Bulut, D. (2020). Tomografi (BT) nedir? Tomografi (BT) nasıl çekilir? Erişim : 18.12.2023, <https://www.medicalpark.com.tr/tomografi/hg-2049>.
- Demir, F. (2020). Konuşma olmayan sinyallerin evrimsel sinir ağları tabanlı yaklaşımlar ile sınıflandırılması. Doktora Tezi. Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ
- Demirci, İ. (2023). Görüntü İşleme ve Yapay Sinir Ağları Teknikleri ile Uçak Kanatlarında Kullanılan Karbon Elyaf/Bal Peteği Sandviç Kompozitlerde Düşük Hızlı Darbe Davranışları Üzerine Nano Grafenlerin Etkilerinin İncelenmesi. Doktora Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya
- Demšar, J. ve Zupan, B.. (2013). Orange: Data mining fruitful and fun - A historical perspective. *Informatica (Slovenia)*. 37. 55-60.
- Devi, R.M. and Seenivasagam, V. (2020). *Automatic segmentation and classification of liver tumor from CT image using feature difference and SVM based classifier-soft computing technique*. *Soft Computing- A Fusion of Foundations, Methodologies and Applications*, 24(24), 18591. <https://doi.org/10.1007/s00500-020-05094-1>
- Dhahir, I.A.D. (2021). Automated Liver Tumor Segmentation And Classification Using AI And Fully Convolutional Neural Network (CNN). İstanbul Altınbaş Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Erdem, F., Derinpınar, M. A., Nasırzadehdıazajı, R., Oy, S., Şeker, D. Z., Bayram, B., (2018). Rastgele Orman Yöntemi Kullanılarak Kıyı Çizgisi Çıkarımı İstanbul Örneği. *Geomatik* , vol.3, no.2, 74-81.
- Ertorsun, A.D., Bağ, B., Uzar, G., Turanoğlu, M.A.(Mayıs,2010). ROC (Receiver Operating Characteristic) Eğrisi Yöntemiyle Tanı Testlerinin Performanslarının

Değerlendirilmesi. *XII. Öğrenci Sempozyumu*, 27-28 Mayıs 2010, Başkent Üniversitesi, Ankara.

Florence Nightingale Hastanesi, (2023). Karaciğer Hastalıkları Nelerdir? Erişim: 18.12.2023, <https://www.florence.com.tr/karaciger-hastaliklari>.

Ghosh,S., Hazra,S., Maity,S.P. and Rahaman,H., "A new algorithm for grayscale image histogram computation," 2015 Annual IEEE India Conference (INDICON), New Delhi, India, 2015, pp. 1-6, doi: 10.1109/INDICON.2015.7443463.

Gökalp, S. ve Aydın, İ. (2021). Farklı Derin Sinir Ağı Modellerinin Duygu Tanımadaki Performanslarının Karşılaştırılması. *Muş Alparslan Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 2(1), 35-43.

Jibril, M., Algehyne, E., Usman, S. (2020). Predictive Supervised Machine Learning Models for Diabetes Mellitus. *SN Computer Science*. 1. 10.1007/s42979-020-00250-8.

Karhan, M. (2011). İmge İşleme Yöntemleri ile Kayıslarda Yaprak Delen Hastalığı Sonucu Oluşan Lekelerin Tespiti. Yüksek Lisans Tezi. Namık Kemal Üniversitesi, Fen Bilimleri Enstitüsü, Tekirdağ.

Karlı, Ö.B. (2019). Makine Öğrenme Yöntemleri ile Karaciğer Hastalığının Teşhisi. Yüksek Lisans Tezi. İbrahim Çeçen Üniversitesi, Fen Bilimleri Enstitüsü, Ağrı.

Kavur, A.E. (2020) Karaciğerde Başarımı ve Duyarlılığı Geliştirilmiş Bölütleme İçin Farklı Yöntemlerin Makina Öğrenmesi Tabanlı En İyiye Götürme. Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, İzmir.

Khattab, M.A., Liao,I.Y., Ooi,E.A., Chong,S.Y.(2022). *Compound W-Net with Fully Accumulative Residual Connections for Liver Segmentation Using CT Images*. *Computational and Mathematical Methods in Medicine*, 2022, 1–11.

Kılıç, S. (2013). Klinik Karar Vermede ROC Analizi. *Journal of Mood Disorders*, 3(3), 135-40. <https://doi.org/10.5455/jmood.20130830051624>

Meriç, E. ve Özer, Ç. (2023). Symptom Based Health Status Prediction via Decision Tree, KNN, XGBoost, LDA, SVM, and Random Forest. 10.1007/978-3-031-27099-4_15.

Negassa, W. G., Mishra, S., Bizuneh, H.D. (2023). *Liver Tumor Detection and Classification Using GWO-ELM Model*. 2023 International Conference in Advances in Power,

Signal, and Information Technology (APSIT), 526–531.
<https://doi.org/10.1109/APSIT58554.2023.10201687>

Nithyananda,C.R., Ramachandra,A.C. ve Preethi,A.C. "Histogram Eşitleme Tabanlı Görüntü Geliştirme Teknikleri Üzerine İnceleme", 2016 Uluslararası Elektrik, Elektronik ve Optimizasyon Teknikleri Konferansı (ICEEOT) , Chennai, Hindistan, 2016, s. 2512-2517, doi: 10.1109/ ICEEOT.2016.7755145.

Otsu,N. "A Threshold Selection Method from Gray-Level Histograms," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

Özsoy, Y.Y. (2021) Karaciğer Kanseri Tedavisinde Makine Algoritma Öğrenimi ve Yapay Zeka Teşhisi. Yüksek Lisans Tezi. İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Pınar,Z. (2020) Makine öğrenmesi yaklaşımları kullanılarak denim kumaşlarda kusur tespiti. Yüksek Lisans Tezi. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.

Pulat, M. (2023). Sınıflandırma Probleminde Makine Öğrenmesi Algoritmalarının Performansını Artırmak İçin Hibrit Bir Yöntem Önerisi. Doktora Tezi. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitü, İzmir.

Saygın, E. (2021). Karaciğer Hastalığı Teşhisinde Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi. Yüksek Lisans Tezi. Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.

Soysal, Ö.A. (2016). İçerik Tabanlı Görüntü Erişiminde Öznitelik Füzyonu. Yüksek Lisans Tezi. Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Şahin, E. (2018) Makine Öğrenme Yöntemleri ve Kelime Kümesi Tekniği ile İstenmeyen E-Posta / E-Posta Sınıflaması. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Taşdoğan, B.E. (2021). Karaciğer Nasıl Bir Organdır? Vücutta Ne Yapar? Erişim: 18 Aralık 2023, <https://www.tkad.org.tr/wp-content/uploads/2021/09/karaciger-nasil-bir-organdir-vucutta-ne-yapar.pdf>

Wu, Q., Zhao, L., Lin, C., Zhao, G. (2020). *A Multi-Stage DCNN Method for Liver Tumor Segmentation*. 2020 IEEE 3rd International Conference of Safe Production and Informatization ,500–505. <https://doi.org/10.1109/IICSPI51290.2020.9332353>

Yılmaz, A. (2007). Kamera Kullanılarak Görüntü İşleme Yoluyla Gerçek Zamanlı Güvenlik Uygulaması. Yüksek Lisans Tezi. Haliç Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.



ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

İsim SOYİSİM : [REDACTED]
Doğum Yeri : [REDACTED]
Doğum Tarihi : [REDACTED]

EĞİTİM DURUMU

Lisans Öğrenimi : Anadolu Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 1998
Yüksek Lisans Öğrenimi : Gazi Üniversitesi, Bilişim Enstitüsü, Bilgisayar Eğitimi Ana Bilim Dalı, 2008
Bildiği Yabancı Diller : İngilizce

BİLİMSEL FAALİYETLERİ

a) Bildiriler

Ulusal

-Kocamaz, C., Yücebaş, S.C., Karakaya, E. (2023) Görüntü İşleme ve Makine Öğrenmesi Yöntemleriyle Karaciğer Tomografi Görüntülerinde Benign ve Malign Kitlenin Tespiti. Ankara Bilim Üniversitesi 18. Uluslararası Bilimsel Araştırmalar Kongresi.

İŞ DENEYİMİ

Çalıştığı Kurumlar ve Yıl: [REDACTED]

İLETİŞİM

E-posta Adresi : [REDACTED]
ORCID : [REDACTED]