

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

TIBBİ GÖRÜNTÜ İŞLEME İLE TANI KOYMADA VERİ
MADENCİLİĞİ VE DERİN ÖĞRENME YÖNTEMLERİNİN
PERFORMANSLARININ İNCELENMESİ

Hanife AVCI

Biyostatistik Programı
YÜKSEK LİSANS TEZİ

ANKARA

2021

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

TIBBİ GÖRÜNTÜ İŞLEME İLE TANI KOYMADA VERİ
MADENCİLİĞİ VE DERİN ÖĞRENME YÖNTEMLERİNİN
PERFORMANSLARININ İNCELENMESİ

Hanife AVCI

Biyostatistik Programı
YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI
Doç. Dr. Jale KARAKAYA KARABULUT

ANKARA
2021

HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
TIBBİ GÖRÜNTÜ İŞLEME İLE TANI KOYMADA VERİ MADENCİLİĞİ VE DERİN
ÖĞRENME YÖNTEMLERİNİN PERFORMANSLARININ İNCELENMESİ
Hanife AVCI

Danışman: Doç. Dr. Jale KARAKAYA KARABULUT

Bu tez çalışması 29/07/2021 tarihinde jürimiz tarafından “Biyoistatistik Programı” nda yüksek lisans tezi olarak kabul edilmiştir.

Jüri Başkanı: *Prof. Dr. Erdem KARABULUT*
(Hacettepe Üniversitesi)

Tez Danışmanı: *Doç. Dr. Jale KARAKAYA KARABULUT*
(Hacettepe Üniversitesi)

Üye: *Doç. Dr. Beyza DOĞANAY ERDOĞAN*
(Ankara Üniversitesi)

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

02 Ağustos 2021

Prof. Dr. Diclehan ORHAN
Enstitü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 6 ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir.

/ /
Hanife AVCI

¹"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

(1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.

(2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ay aşmamak üzere tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir. Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, Do. Dr. Jale KARAKAYA KARABULUT danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sađlık Bilimleri Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

Hanife AVCI

TEŞEKKÜR

Yüksek lisans eğitimim ve tez dönemim boyunca beni bu alanda çalışmaya yönlendirerek ufkumun genişlemesini sağlayan, çalışmalarımnda beni teşvik eden ve her yönden destekleyen, bu çalışmanın her aşamasında ihtiyaç duyduğum her an ulaşabildiğim, ayrıca akademik hayatımın ilerlemesinde büyük katkıları olan, emeklerinin karşılığını asla ödeyemeyeceğim çok değerli danışman hocam Doç. Dr. Jale KARAKAYA KARABULUT'a sonsuz teşekkürlerimi sunarım.

Tez savunma sınavında yer alarak önerilerinden ve katkılarından dolayı Prof. Dr. Erdem KARABULUT'a ve Doç. Dr. Beyza DOĞANAY ERDOĞAN'a, tıbbi görüntülerin bölütleme aşamasında MedPic arayüz yazılımını kullanmam için bana ayırdığı vakit ve destek için Prof. Dr. Recep DEMİRCİ'ye çok teşekkür ederim.

Tez süresince desteğini esirgemeyen Biyoistatistik Anabilim Dalında bulunan tüm hocalarım ve asistan arkadaşlarıma çok teşekkür ediyorum.

Son olarak bu günlere gelmemde büyük katkıları olan, yalnız tez çalışması süresince değil tüm yaşamım boyunca bana her konuda ve her zaman güvenen sevgili annem ve babama çok teşekkür ediyorum.

ÖZET

Avcı, H., Tıbbi Görüntü İşleme ile Tanı Koymada Veri Madenciliği ve Derin Öğrenme Yöntemlerinin Performanslarının İncelenmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Yüksek Lisans Tezi, Ankara, 2021. Görüntü işleme, bilgisayara aktarılan görüntülerden faydalı bilgiler elde etmek amacıyla farklı matematiksel algoritmalar uygulanan yöntemler bütünüdür. Görüntü işlemede, birincil hedef olarak görüntülerin kalitesinin iyileştirilmesi amaçlanmaktadır. Görüntü işlemenin bir diğer olumlu yanı da görüntüleri tanımlayan sayısal değerlerin (öznitelikler) elde edilmesidir. Görüntü işleme yöntemi ile görüntülerden elde edilen öznitelikler, görüntüleri sınıflama amacıyla kullanılmaktadır. Çeşitli hastalıklara erken tanı koymada görüntüleme yöntemlerinin gelişmiş olmasına rağmen istatistiksel yöntemler ve çeşitli görüntü işleme algoritmaları da önemli olabilir. Bu çalışmada tıbbi görüntü işleme teknikleri kullanılarak mamografi görüntüleri sayısallaştırılıp öncelikle normal-anormal lezyon saptamada, sonrasında malign-benign lezyonları ayırmadaki başarısı incelenmiştir. Uygulanan beş farklı görüntü ön işleme algoritmasının, farklı iki bölütleme yönteminin ve meme doku tipinin sonuçlara olan etkisi araştırılmıştır. Ayrıca kullanılan farklı görüntü işleme tekniklerinde veri madenciliği ve derin öğrenme algoritmalarının da sınıflama performansları karşılaştırılmıştır. Bunun için açık kaynaklı MIAS veri tabanındaki 209 normal, 61 benign ve 52 malign olan 322 mamografi görüntüsü kullanılmıştır. Görüntü işleme içerisinde oluşturulan beş farklı ön işleme ve iki farklı bölütleme algoritmaları karşılaştırıldığında, görüntüler üzerinde uygulanan ön işleme ve görüntü bölütleme tekniklerinin sınıflama performans ölçülerini etkilediği görülmüştür. Genel olarak bu veri setinde klasik veri madenciliği yöntemlerinden DVM, RO ve YSA diğer algoritmalarından daha iyi performans göstermiştir. Derin öğrenme yöntemleri de klasik veri madenciliği yöntemlerine benzer sonuçlar vermiştir.

Anahtar Kelimeler: Bilgisayar destekli tanı, görüntü ön işleme yöntemleri, görüntü bölütleme yöntemleri, veri madenciliği yöntemleri, sınıflandırma performansları.

ABSTRACT

Avci, H., Investigation of the Performance of Data Mining and Deep Learning Methods in Diagnosis with Medical Image Processing, Hacettepe University Graduate School of Health Sciences Master Thesis in Biostatistics, Ankara, 2021.

Image processing is a set of methods in which different mathematical algorithms are applied in order to obtain useful information from images transferred to the computer. In this method, the primary goal is to improve the quality of images. Another positive aspect of the image processing method is the acquisition of numerical values (attributes) that describe the images. The features obtained from the images by the image processing method are used to classify the images. Although imaging methods are advanced, statistical methods and various image processing algorithms may also be important in diagnosing various diseases early. In this study, images were digitized using medical image processing techniques and its success in detecting normal-abnormal lesions and then in distinguishing malign-benign lesions was examined. The effects of five different image pre-processing algorithms, two different segmentation methods and breast tissue type on the results were investigated. In addition, classification performances of data mining and deep learning algorithms in different image processing techniques used were compared. For this, 322 mammography images, 209 normal, 61 benign and 52 malign, in the open source MIAS database were used. When five different pre-processing and two different segmentation algorithms created in image processing were compared, it was seen that the preprocessing and image segmentation techniques applied on images actually affected the classification performance measures. In general, SVM, RF and ANN, which are classical data mining methods, performed better than other algorithms in this data set. Deep learning methods also gave similar results to classical data mining methods.

Keywords: Computer aided diagnosis, image pre-processing methods, image segmentation methods, data mining methods, classification performances.

İÇİNDEKİLER

ONAY	iii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYAN	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER VE KISALTMALAR	xi
ŞEKİLLER	xii
TABLolar	xiv
1. GİRİŞ	1
2. GENEL BİLGİLER	5
2.1. Görüntü İşleme	5
2.1.1. Görüntü İşlemede Temel Kavramlar	7
2.2. Görüntü	10
2.2.1. Gri Tonlamalı Görüntü (Grayscale Image)	10
2.2.2. İkili Görüntü (Binary Image)	11
2.2.3. Renkli Görüntü (Red, Green, Blue-RGB)	11
2.2.4. Renkli Görüntünün Gri Düzeyli Görüntüye Dönüşümü	11
2.2.5. Gri Düzeyli Görüntünün İkili Görüntüye Dönüşümü	11
2.3. Gürültü (Noise)	12
2.3.1. Gürültü Çeşitleri	12
2.4. Görüntü Filtreleme Yöntemleri	14
2.4.1. Görüntü Yumuşatma Filtreleri (Bulanıklaştırma, Smoothing)	14
2.4.2. Görüntü Keskinleştirme Filtreleri	15
2.5. Görüntü İyileştirme Yöntemleri	17
2.5.1. Histogram Eşitleme (Histogram Equalization)	18
2.5.2. Kontrast Sınırlı Uyarlanabilir Histogram Eşitleme (Contrast Limited Adaptive Histogram Equalization-CLAHE)	19

2.6. Görüntüler Üzerinde Morfolojik İşlemler	20
2.6.1. Aşınma (Erosion)	21
2.6.2. Genişleme (Dilation)	21
2.6.3. Açma (Opening)	22
2.6.4. Kapama (Closing)	22
2.7. Görüntü Bölütleme (Segmentasyon) Yöntemleri	22
2.7.1. Kenar Tabanlı Bölütleme (Edge based segmentation)	23
2.7.2. Bölge Tabanlı Bölütleme (Based region segmentation)	24
2.8. Öznitelik Çıkarımı	26
2.9. Sınıflandırma	34
2.10. Sınıflandırma Yöntemleri İçin Performans Ölçüleri	38
3. GEREÇ VE YÖNTEM	41
3.1. MIAS veri seti	41
3.2. Yöntem	43
3.2.1. Ön işleme	43
3.2.2. Bölütleme ve ROI	44
3.2.3. Öznitelik çıkarımı	45
3.2.4. Sınıflandırma	46
3.3. Verinin değerlendirilmesinde kullanılan programlar ve paketler	50
4. BULGULAR	51
5. TARTIŞMA	65
6. SONUÇ VE ÖNERİLER	68
7. KAYNAKLAR	71
8. EKLER	
EK-1: Tez Çalışması Orijinallik Raporu	
EK-2: Dijital Makbuz	
9. ÖZGEÇMİŞ	77

SİMGELER VE KISALTMALAR

A1	CLAHE
A2	MF&CLAHE
A3	MF&USM
A4	CLAHE&USM
A5	MF&CLAHE&USM
ARCH	Yapısal Bozulma
ASYM	Asimetri
AUC	ROC Eğrisi Altında Kalan Alan
B	İyi huylu
BDT	Bilgisayar Destekli Tanı
BT	Bilgisayarlı Tomografi
CALC	Kalsifikasyon
CART	Sınıflama ve Regresyon Ağacı
CC	Craniocaudal view
CIRC	İyi tanımlanmış/sınırlı kitleler
CL	Kontrast Sınırlı Uyarlanabilir Histogram Eşitleme
D	Yoğun glandular
DDO	Düzeltilmiş Doğruluk Oranı
DÖ	Derin Öğrenme
DVM	Destek Vektör Makinesi
F	Yağlı
G	Yağlı glandular
GLCM	Gri Seviye Eş Oluşum Matrisi
GLDM	Gri Seviye Fark Yöntemi
GLRLM	Gri Seviye Çalışma Uzunluğu Matrisi
KA	Karar Ağaçları

k-NN	k-En Yakın Komşuluk
LBP	Yerel İkili Örüntüler
LoG	Laplacian of Gaussian
LR	Lojistik Regresyon
M	Kötü huylu
MF	Ortanca Filtre
MIAS	Mammographic Image Analysis Society
MISC	Kötü Tanımlanmış Kitleler
MLO	Medio-lateral oblique
MRG	Manyetik Rezonans Görüntüleme
mRMR	Minimum Redundancy Maximum Relevance
NB	Naive Bayes
NKD	Negatif Kestirim Değeri
NORM	Normal
PKD	Pozitif Kestirim Değeri
RGB	Renkli Görüntü
RO	Rastgele Orman
ROI	İlgili Bölge
SPIC	Spiculated Kitleler
US	Ultrasonografi
USM	Keskin Olmayan Maskeleme
YSA	Yapay Sinir Ağları

ŞEKİLLER

Şekil	Sayfa
2.1. Tam vücut kemik taraması (A), Gamma düzeltmesi ile keskinleştirilmiş görüntü (B)	6
2.2. Görüntü işlemedeki temel adımlar	7
2.3. 1000×1000 boyutunda bir dijital görüntü (A), 132×108 boyutundaki piksellerin büyütülmüş hali (B)	8
2.4. Aynı görüntünün farklı çözünürlüklerinin gösterimi	9
2.5. Orijinal görüntü (A), Parlaklığı artan görüntü (B) ve Parlaklığı azalan görüntü (C)	9
2.6. Renkli görüntü 384×512×3 (A), Gri tonlamalı görüntü 384×512 (B) ve İkili görüntü 384×512 (C)	12
2.7. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Gauss gürültüsü eklenmiş 1024×1024 piksel boyutunda mamografi görüntüsü (B)	13
2.8. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Tuz ve biber gürültüsü eklenmiş 1024×1024 piksel boyutunda mamografi görüntüsü (B)	13
2.9. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Mamografi görüntüsünün histogram grafiği (B)	18
2.10. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (B), Uyarlanabilir histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (C)	19
2.11. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Kontrast sınırlı uyarlanabilir histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (B)	19
2.12. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Canny kenar belirleme yöntemi ile oluşan mamografi görüntüsü (B)	24
2.13. 1024×1024 piksel boyutunda mamografi görüntüsü (A), K-ortalamalar Kümeleme yöntemi ile mamografi görüntüsünün bölütlenmesi (B)	26
2.14. LBP operatörü ile pikseller için etiketler elde edilmesi	33

2.15.	Verilerin doğrusal olarak ayrılabilme (A) ve ayrılamama durumu (B)	35
3.1.	MIAS veri tabanından normal (A), İyi huylu (B) ve Kötü huylu (C) mamografi örnekleri	42
3.2.	Mamografi görüntüsündeki farklı bileşenler	43
3.3.	mdb023 nolu mamografi görüntüsü (A), Ön işleme yöntemleri ve A5 algoritması ile iyileştirilmiş mamografi görüntüsü (B), k-ortalamar kümeleme yöntemi uygulanmış mamografi görüntüsü (C), Maskeleme işlemi uygulandıktan sonra gerçek zemin görüntüsü (D)	45
3.4.	mdb023 nolu mamografi görüntüsü (A), Ön işleme yöntemleri ve A5 algoritması ile iyileştirilmiş mamografi görüntüsü (B), Canny kenar algılama kümele yöntemi uygulanmış mamografi görüntüsü (C), Maskeleme işlemi uygulandıktan sonra gerçek zemin görüntüsü (D)	45
3.5.	Ön işleme yöntemlerinin performanslarının değerlendirilmesi	48
3.6.	Bölütleme yöntemlerinin performanslarının değerlendirilmesi	49
4.1.	Normal ve Anormal sınıflamasında ön işleme kombinasyonlarının sınıflama yöntemlerinde AUC değerlerinin karşılaştırılması	53
4.2.	Benign ve Malign sınıflamasında ön işleme kombinasyonlarının sınıflama yöntemlerinde AUC değerlerinin karşılaştırılması	55
4.3.	MF&CL&USM ön işleme yöntemleri uygulandıktan sonra normal ve anormal sınıflamasında sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması	57
4.4.	MF&CL&USM ön işleme yöntemleri uygulandıktan sonra benign ve malign sınıflamasında sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması	58
4.5.	Ön işleme yöntemlerinin doku tipine göre normal-anormal sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması	61

TABLOLAR

Tablo	Sayfa
2.1. Temel istatistiksel özellikler	27
2.2. Gri seviye eş oluşum matrisinden (GLCM) çıkarılan özellikler	28
2.3. Gri seviye fark matrisinden (GLDM) çıkarılan özellikler	30
2.4. Gri seviye çalışma uzunluğu matrisinden (GLRLM) çıkarılan özellikler	31
2.5. 2×2'lik Örnek Olumsuzluk Tablosu	39
3.1. MIAS veri tabanı	42
4.1. Korelasyon katsayısı kullanarak seçilen 9 özellik	51
4.2. Seçilen özelliklerin ön işleme yöntemlerinden elde edilen AUC değerleri	52
4.3. Normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansları	54
4.4. Benign-malign sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı	56
4.5. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra k-ortalamlar kümeleme ve Canny kenar algılama algoritmalarının normal-anormal sınıflama için sınıflandırma yöntemlerinin performansı	59
4.6. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra k-ortalamlar kümeleme ve Canny kenar algılama algoritmalarının benign-malign sınıflama için sınıflandırma yöntemlerinin performansı	59
4.7. Meme doku tipi D (Yoğun, Dense glandular) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı	62
4.8. Meme doku tipi F (Yağlı, Fatty) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı	63

- 4.9.** Meme doku tipi G (Yađlı glandular, Fatty glandular) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

1. GİRİŞ

Bilimsel teknolojilerdeki ilerlemeler sağlık alanına şimdiye dek görülmemiş büyüklükte karmaşık veriler getirmiştir. Bu karmaşık veri yığınının, faydalı verilerin elde edilmesi veri madenciliğidir. Başka bir deyişle, büyük veri kümelerinden, veriler arasındaki benzer eğilimleri keşfetmeyi amaçlayan, bir anlamda veri tabanlarından bilginin keşfedilmesini sağlayan ve ileriye yönelik doğru tahminlerde bulunmamıza yardımcı olan yöntemler bütünüdür.

Veri madenciliğinde birçok algoritma bulunur. Bu algoritmaların bazıları sayısal değerlerin kestirimi, bazıları da sınıflandırma amaçlı kullanılmaktadır. Ancak, bazı durumlarda büyük ölçekli verilerin analiz edilmesinde klasik veri madenciliği yöntemleri yetersiz kalabilmektedir. Bu problemin üstesinden gelebilmek için çok katmanlı ağ yapılarına sahip ve yapay sinir ağlarının uzantısı olan derin öğrenme yöntemi geliştirilmiştir. Son yıllarda veri boyutunun artması, veri setindeki değişken sayısının fazlalığı, değişkenler arasındaki ilişkilerin çözümlenmesine yönelik girişimler derin öğrenme yönteminin gelişmesine neden olmuştur. Derin öğrenme, nesnelerin tespiti ve sınıflandırılması dahil olmak üzere çeşitli makine öğrenim görevlerinde yüksek performans sergilediğinden, son yıllarda çığır açan önemli bir teknoloji olarak kabul edilir (1). Derin öğrenme kavramının kullanılmaya başlanması 2000’li yıllara dayansa da özellikle son yıllarda çok kullanılan bir çalışma alanı olmasının nedeni, veri boyutunun gün geçtikçe artması ve bu verileri işleyebilecek bilgisayar donanımlarının geliştirilmiş olmasıdır. Derin öğrenme yöntemleri, görüntü işleme, ses analizi, robotik, gen analizleri, kanser teşhisi, gen seçimi ve sınıflandırması, doku sınıflandırması vb. birçok alanda kullanılmaktadır.

Sağlık alanında, hekimlere yardımcı olmak, hastalıkların belirlenmesinde çeşitli yöntemler geliştirmek, ilaç maliyetlerini azaltmak gibi amaçlarla çok boyutlu veriler toplanmakta ve depolanmaktadır. Hastalıklara erken tanı konulma olasılığını artırmak için manyetik rezonans görüntüleme (MRG), ultrasonografi (US), bilgisayarlı tomografi (BT), mamografi, X-ray (Röntgen) olmak üzere çeşitli biyomedikal görüntüleme teknikleri geliştirilmiştir. Veri madenciliği ve derin öğrenme algoritmalarının son yıllarda sık kullanıldığı alanlardan biri de tıbbi görüntülemedir.

Çünkü, bir hastalığın ortaya çıkış sürecini aydınlatmada ve hastalığın seyrini incelemede en önemli sağlık hizmetlerinden biri radyolojidir. Bu nedenle, doktorlara tanı koymada yardımcı olmak için veri madenciliği ve derin öğrenme algoritmaları ile radyolojik görüntüler işlenebilir. Görüntü işleme, çok çeşitli bilimsel, sanatsal ve ticari uygulamalar için de kullanılır. Son yıllarda özellikle sağlık alanında görüntü işleme yönteminden sıklıkla yararlanılmaktadır. Tıbbi görüntü işlemenin amacı, tanı ve tedavi süreçlerinde kullanılan tıbbi görüntülerin daha anlaşılır olmasını sağlamaktır.

Dünyada ve ülkemizde en çok görülen ölüm nedenlerinden birisi kanserdir. Kanserde erken tanı koymak çok önem taşımaktadır. Meme kanseri, tüm dünyada kadınlarda çok sık ortaya çıkan kanser türlerinden biridir. Kansere erken evrede tanı konulduğu zaman tedavi yöntemlerine daha olumlu yanıt alınabilmektedir. Erken tanı koymada, görüntüleme yöntemlerinin gelişmiş olmasına rağmen istatistiksel yöntemler ve çeşitli görüntü işleme algoritmaları da önemli olabilir. Bu nedenle, Bilgisayar Destekli Tanı (BDT) sistemleri aracılığıyla sayısal (dijital) radyolojik görüntülerde matematiksel ve istatistiksel yöntemlerin kullanılması, radyologlar için iyi (benign) ve kötü huylu (malign) kitleleri ayırt etmede faydalı olabilir (2). Alanyazına bakıldığında, mamografi görüntülerindeki, en erken meme kanseri belirtilerinden biri olan, mikrokalsifikasyonların belirlenmesi için çeşitli algoritmalar önerilmiştir (3).

Bu tez çalışmasının amacı, tıbbi görüntüleme tekniklerinden elde edilen görüntüleri çeşitli yöntemlerle işleyerek, görüntülerden çıkarılan özneliklerin hastalık durumunu belirlemedeki başarısını araştırmak ve ayrıca veri madenciliği ve derin öğrenme algoritmalarını kullanarak bu algoritmaların sınıflama performanslarına olan katkısını ortaya koymaktır. Buna ek olarak, mamografi görüntülerinin meme dokusu alanları birbirinden farklı olduğu için meme doku tipine göre de ön işleme algoritmalarının sınıflama yöntemleri üzerindeki etkisini incelemektir.

Bu tez çalışmasında, meme kanseri tanısı ve sınıflandırması sürecinde tıbbi görüntü işleme tekniklerinden yararlanılmıştır. Çalışmada açık erişimli bir veri tabanından elde edilen mamografi görüntüleri kullanılmıştır. Çalışma kapsamında görüntü ön işleme, bölütleme (segmentasyon), ilgili bölgeleri belirleme (region of interest, ROI), öznelik çıkarımı ve sınıflandırma adımları kullanılmıştır.

Bu tez kapsamında, ilk olarak üç farklı ön işleme yönteminin farklı kombinasyonlarının sınıflandırma performansları üzerindeki etkisi karşılaştırılmıştır. Ön işleme yöntemi olarak görüntü keskinleştirme filtresi olan keskin olmayan maskeleme (un-sharp masking, USM), görüntü yumuşatma filtresi olan ortanca filtre (median filter, MF) ve görüntünün kontrastını artırmak için kontrast sınırlı uyarlanabilir histogram eşitleme (contrast limited adaptive histogram equalization, CLAHE) algoritmalarının farklı kombinasyonları kullanılmıştır. İlk olarak CLAHE (CL) algoritması tek olarak test edilmiştir. Daha sonra Ortanca filtre ve CLAHE (MF&CL), Ortanca filtre ve Keskin olmayan maskeleme (MF&USM), CLAHE ve Keskin olmayan maskeleme (CL&USM) ve Ortanca filtre, CLAHE ve Keskin olmayan maskeleme (MF&CL&USM) gibi filtrelerin olası bazı kombinasyonları oluşturulmuştur. Bu kombinasyonlardan her biri, her görüntüye uygulanmıştır. Bu çalışmanın amacı, çeşitli ön işleme yöntemlerinin farklı kombinasyonlarının kullanımının sınıflama performanslarına olan katkısını araştırmaktır. Çünkü, herhangi bir görüntü işleme algoritması uygulamadan önce elde edilen sonuçları daha iyi hale getirmek için ön işleme algoritmaları büyük önem taşımaktadır (4). Alanyazına bakıldığında, birçok çalışma farklı ön işleme yöntemleri önermektedir. Ancak adım adım bu ön işleme yöntemlerinin farklı kombinasyonlarının kullanıldığı ve sınıflama performanslarının değerlendirildiği çalışma bulunmamaktadır. Sonrasında ise farklı görüntü bölütleme (segmentasyon) yöntemlerinin sınıflama performansları da incelenmiştir. Ayrıca meme doku tipinin sınıflama performansında değişikliğe neden olup olmadığı da araştırılmıştır.

Bu amaç ile; tez çalışmasının hipotezleri aşağıda sıralanmıştır:

- i. Görüntü işleme ile elde edilen sayısal verilerin tanı koymaya katkısı vardır.
- ii. Kullanılan ön işleme yöntemlerinin farklı kombinasyonlarına göre sınıflama yöntemlerinin performansları farklıdır.
- iii. Farklı bölütleme yöntemlerinin sınıflama performanslarına etkisi vardır.
- iv. Meme dokusu yoğunluğunun sınıflama performanslarına etkisi vardır.
- v. Derin öğrenme yöntemlerinin veri madenciliği yöntemlerine göre performansı farklılık gösterir.

Bu çalışmanın birinci bölümünde meme kanseri problemi ve alanyazında bu problem için son yıllarda bilgisayar destekli tanı modellerinin geliştirilmesinde görüntü işlemenin rolü, bu çalışmanın amacı ve hipotezleri yer almaktadır. İkinci bölümünde, alanyazında bahsedilen görüntü işleme süreci ve çeşitli yöntemler ayrıntılı olarak verilmiştir. Üçüncü bölümde kapsamlı bir görüntü işleme süreci tasarlanmıştır. Çalışmada kullanılan MIAS veri setinden, verinin elde edilmesinden, kullanılan görüntü ön işleme ve bölütleme yöntemlerinden, sınıflamada kullanılan yöntemlerden ve sınıflama başarısının değerlendirilmesi için kullanılan performans ölçülerinden bahsedilmiştir. Dördüncü bölümde, görüntü ön işleme yöntemlerinin farklı kombinasyonları ve bölütleme yöntemlerinin sınıflamadaki başarılarına ilişkin sonuçlar sunulmuştur. Bulgulara ait sonuçlar, grafikler ve tablolar ile birlikte özetlenmiştir. Beşinci bölümde, sınıflama yöntemlerinin performansları incelenmiş, alanyazındaki benzer çalışmaların sonuçları ile bulgular sonuçları karşılaştırılmıştır. Son bölümde de tezin önemli genel sonuçları ve ileride yapılması planlanan araştırma konuları sunulmuştur.

2. GENEL BİLGİLER

2.1. Görüntü İşleme

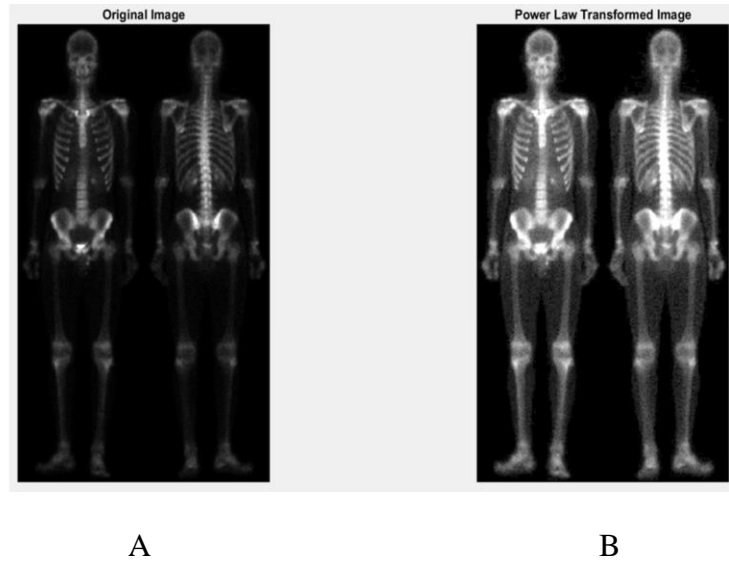
Görüntü işlemenin başlangıcı olarak Londra ve New York arasındaki denizaltı kablosu aracılığıyla gönderilen gazete resimlerinin sayısallaştırılması kabul edilmektedir. 1920'lerin başında Bartlane kablolu resim iletim sisteminin geliştirilmesiyle, resimlerin gönderilme süresi bir haftadan daha fazlayken üç saatten daha az bir süreye düşmüştür. Görüntü işleme ile ilgili çalışmalar, gelişen teknoloji ile birlikte 1960'ların başında üst düzey programlama dillerinin (COBOL ve FORTRAN) gelişimi ve yüksek kapasiteli bilgisayarların ortaya çıkmasıyla yeni bir boyut kazanmıştır. 1964 yılında Jet Propulsion Laboratuvarında uzay aracından gönderilen ay görüntüleri, televizyon kameralarından kaynaklanan görüntü bozulmalarını ortadan kaldırmak için sayısal olarak işlenmiştir (5). O günden bu yana görüntü işleme büyük bir gelişme göstermiş ve askeri, endüstri, ticaret, tıp, jeoloji, reklamcılık ve uzay gibi birçok alanda kullanılmıştır.

Görüntü işleme, bir görüntüyü bir formdan diğerine dönüştürme yöntemi olarak tanımlanabilir. Görüntü işleme ile sayısal (dijital) görüntüler üzerinde çeşitli işlemler uygulanarak yeni görüntüler elde edilir. Bu dijital görüntü verilerine, amaca uygun bilgisayar ve yazılımlar sayesinde matematiksel algoritmalar uygulanır. Görüntü işlemede girdi (input) değişkeni sayısal görüntülerdir. Çeşitli görüntü işleme algoritmaları yardımıyla, çıktı (output) değişkeni görüntülerden çıkartılan özniteliklerdir. Özetle görüntü işleme, elde edilen görüntülerden amaçlanan hedefe uygun sayısal değerler elde etme işlemidir. Görüntülerden elde edilen sayısal değerler daha sonra istatistiksel analizlerde kullanılmaktadır.

Görüntü işlemenin 2 temel amacı vardır:

1. Çeşitli aygıtlar yardımıyla elde edilen sayısal görüntülerde geometrik bozulma, gürültü ve etiket bilgisi gibi istenmeyen bilgileri kaldırmaktır. Bu hataları düzeltmek için görüntü ön işleme yöntemlerinden yararlanılmaktadır. Ön işleme yöntemi olarak görüntülere kontrast artırma, keskinleştirme ve gürültü azaltma teknikleri uygulanır. Böylece görüntü geliştirme teknikleri sayesinde görüntüler, insan

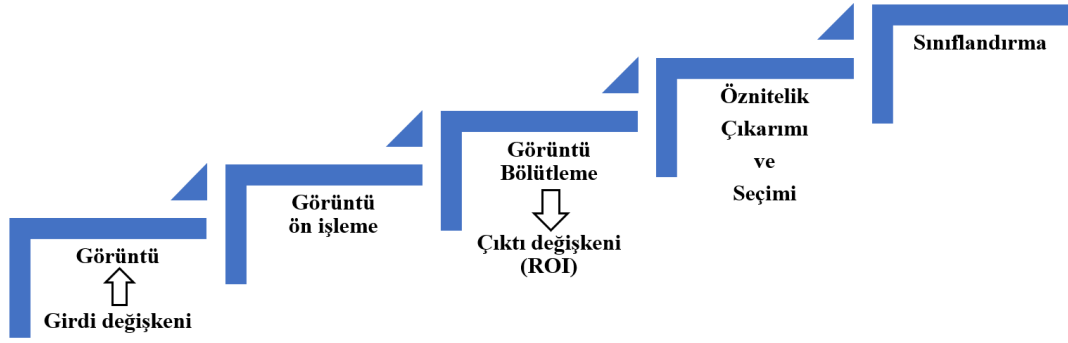
gözüne daha net ve önemli ayrıntıları daha kolay görülebilir duruma getirilmiş olur. Gonzales RC, Woods RE'nin kitabından alınan (5) tam vücut kemik taraması görüntüsünün Gamma düzeltmesi yardımıyla ince ayrıntılarının zenginleştirilmesi örnek olarak verilebilir (Şekil 2.1).



Şekil 2.1. Tam vücut kemik taraması (A), Gamma düzeltmesi ile keskinleştirilmiş görüntü (B).

2. Görüntülerden sayısal öznitelikler elde edip sonuçlar çıkarma ve öğrenmeyi sağlamaktır. Bu da bilgisayarla görme (computer vision) alanı olarak tanımlanır. Uydu görüntüleri üzerinden fayların belirlenmesi örnek olarak verilebilir.

Genel olarak görüntü işleme uygulamaları 4 aşamadan oluşmaktadır. Görüntüler üzerinde uygulanan birinci aşama görüntü ön işleme, ikinci aşama görüntü bölütleme, üçüncü aşama öznitelik çıkarımı ve dördüncü aşama ise sınıflamadır (Şekil 2.2).



Şekil 2.2. Görüntü işlemedeki temel adımlar.

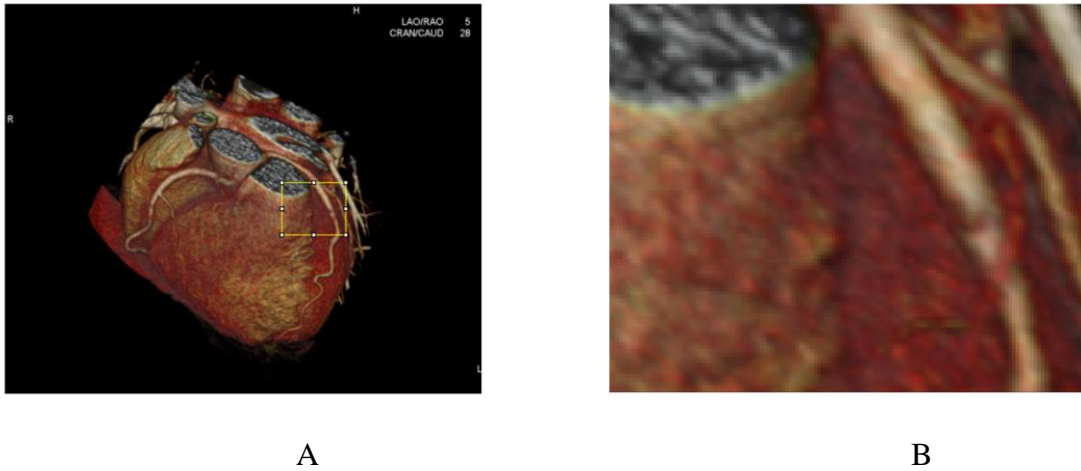
2.1.1 Görüntü İşlemede Temel Kavramlar

Görüntü işleme, görüntüden faydalı bilgi çıkarmak amacıyla görüntü üzerinde yapılan birçok farklı işlemi içerir. Bu işlemleri daha iyi anlayabilmek için görüntü işleme alanında kullanılan bazı temel kavramların bilinmesi önemlidir. Görüntü işleme ile ilgili literatürde birçok temel kavram mevcuttur. Görüntü işleme alanı ile ilgili çalışmalarda, bilinmesi gereken temel kavramlar aşağıdaki gibi sıralanabilir:

Sayısal (Dijital) Görüntü: Temel olarak piksel öğelerinden oluşan görüntüdür. Sayısal bir görüntü iki boyutlu bir matris şeklinde satır ve sütun indisleri ile matematiksel olarak ifade edilebilir (5). Sonlu sayıdaki elemanlardan oluşmaktadır.

Piksel: Sayısal bir görüntünün iki boyutlu matris şeklinde $M \times N$ 'lik gösteriminde satır ve sütunlarının kesim noktaları için en yaygın kullanılan terimdir. Piksel, bir görüntünün en küçük birimidir. Matrise ilişkin satır ve sütun indisleri görüntü içerisindeki noktaları tanımlamaktadır. Pikseller matrisinin elemanlarına ait değerler, o noktanın renk ton değerini göstermektedir. Bu matris toplam $M \times N$ tane pikselden oluşur. Sayısal görüntünün elemanları, x ve y koordinat noktalarına karşılık gelen piksellerin parlaklık değerlerinden oluşmaktadır. Başka bir deyişle görüntüler pikseller matrisidir.

Fiji-Image J açık erişimli Java tabanlı görüntü işleme programından Cardio.dcm (RGB DICOM) görüntüsü içerisinde herhangi bir kesiti tanımlayan elemanlardan meydana gelmiş pikseller matrisi Şekil 2.3.'te gösterilmiştir.



Şekil 2.3. 1000×1000 boyutunda bir dijital görüntü (A), 132×108 boyutundaki piksellerin büyütülmüş hali (B).

Dijital görüntüler, x ve y gibi iki sürekli sayısal değişkenin fonksiyonu olan $f(x,y)$ biçiminde iki boyutlu fonksiyonlar yardımıyla gösterilir (5). $f(x, y)$ fonksiyonu,

$$0 < f(x, y) < \infty, x=0, \dots, M-1 \quad (2.1)$$

$$y=0, \dots, N-1$$

Eşitlik 2.1.'de gösterildiği gibi sıfırdan farklı ve sonlu olmalıdır. Sayısal bir görüntü ve elemanlarını göstermek için matris gösterimi Eşitlik 2.2. kullanılabilir.

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix} \quad (2.2)$$

Çözünürlük: Sayısal görüntülerin çözünürlüğü, piksel sayısını ifade eder. Çözünürlük, görüntüdeki yatay ve dikey piksel sayılarının çarpımıdır (5). Çözünürlük arttıkça görüntüdeki ayrıntılar belirginleşmektedir. Başka bir deyişle, çözünürlük bir kalite göstergesidir. Çözünürlük düştükçe yani görüntüdeki piksel sayısı azaldıkça görüntüdeki bazı bilgiler yok olmaktadır. Bu olay MATLAB R2017 paket programından alınan görüntü üzerinde Şekil 2.4.'te gösterilmiştir.



720×576

180× 180

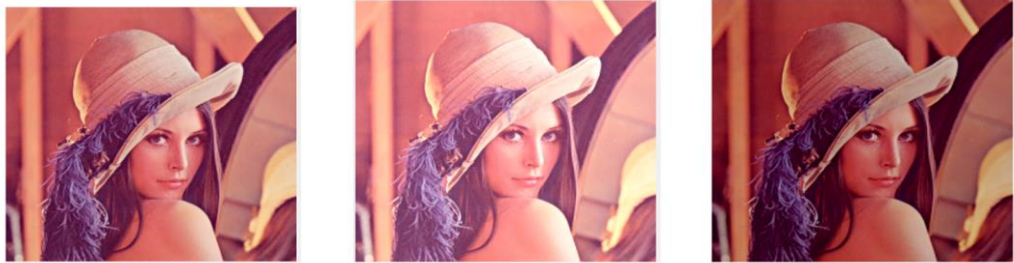
90×90

Şekil 2.4. Aynı görüntünün farklı çözünürlüklerinin gösterimi.

Parlaklık: Görüntülerin parlaklığı, 0–255 aralığındaki piksel değerlerine, pozitif veya negatif bir değer eklemekle ayarlanabilir. Eklenen sayı pozitif ise görüntünün parlaklığı artar, negatif ise azalır. Orijinal görüntünün fonksiyonu $f(x, y)$ ve parlaklığı değiştirilmiş $g(x, y)$ fonksiyonu, a sabit sayı olmak üzere,

$$g(x, y) = f(x, y) + a \quad (2.3.)$$

Eşitlik 2.3. ile bulunur. Parlaklık, görüntüdeki en karanlık bölge (siyah) ile en parlak bölgenin (beyaz) arasındaki fark olarak tanımlanabilir. Piksel değerlerine eklenen sayıya göre görüntü işleme uygulamalarında sık kullanılan MATLAB R2017 paket programından alınan lena.bmp görüntüsünün değişimi Şekil 2.5.'te gösterilmiştir.



A

B=A+40

C=A-40

Şekil 2.5. Orijinal görüntü (A), parlaklığı artan görüntü (B) ve parlaklığı azalan görüntü (C).

Karşıtlık (Kontrast): Görüntüdeki en küçük renk değerine sahip bölüm ile en büyük renk değerine sahip bölüm arasındaki farka denir. Kontrastı artırılmış görüntülerde nesnelere farklılık belirgin biçimde görünürken, düşük kontrasttaki görüntülerde nesnelere renk değerleri birbirine daha yakın olur ve fotoğrafın bulanık görünmesine (yumuşak, soft) neden olur. İnsan gözünün, birbirine yakın renk aralığındaki nesnelere ayırt etme kapasitesi sınırlıdır. Bu nedenle bir görüntüdeki düşük ve yüksek değerler arasındaki fark az olursa insan gözü görüntüdeki detayları algılayamayabilir (6).

Görüntülerin sayısal veriye dönüştürülmeden önce bir ön işleme sürecinden geçmesi gerekmektedir. Bu ön işleme adımında kullanılan çeşitli yöntemler vardır. Aşağıda ön işleme adımlarının detaylarından bahsedilmiştir. Bu yöntemlerin hepsinin görüntü sayısal veriye dönüşmeden uygulanması gerekmektedir. Bölüm 2.2'den 2.7. görüntü bölütleme (segmentasyon) yöntemlerine kadar olan tüm bölümler, görüntü ön işleme adımında kullanılan çeşitli uygulamalardır.

2.2. Görüntü

Görüntü, matematikte $f(x, y)$ fonksiyonu biçiminde tanımlanan ve iki boyutlu bir sinyal olan kavramdır. Görüntüler iki boyutlu bir sinyal olduğu için iki değişken vardır. Görüntüdeki piksel değerlerinin belirli aralıklarda olması, ortaya çıkan görüntünün niteliğini değiştirmektedir.

2.2.1. Gri Tonlamalı Görüntü (Grayscale Image)

Siyah ve beyaz renkleri arasındaki değerler kullanılarak gri seviyeli görüntü elde edilir. Bu görüntülerde 256 adet farklı renk değeri bulunabilir. Gri değerler kümesi $G=[0,1,2,\dots,255]$ biçiminde gösterilir. Sıfır sayısal değeri siyah ve 255 sayısal değeri ise beyaz renk değerini ifade eder. Bu iki renk değeri arasındaki sayılar da gri tonları belirtmektedir (7).

2.2.2. İkili Görüntü (Binary Image)

Görüntüdeki tüm pikseller için 0 ya da 1 renk değerlerinden oluşan görüntüler ikili görüntülerdir. Sembolik olarak 1 renk değeri beyazı, 0 renk değeri de siyahı göstermektedir. Elde edilen görüntü yalnızca siyah ve beyaz renkten oluşur. Bu nedenle ikili görüntü olarak da adlandırılır. İkili görüntü, bölütleme algoritmaları için çok yararlı olmasına rağmen, diğer görüntü türlerine kıyasla daha piksellere ilişkin özellik azdır (8).

2.2.3. Renkli Görüntü (Red, Green, Blue, RGB)

Renkli görüntülerde temel renk değerleri kırmızı, yeşil ve mavi olduğu için kısaca RGB (R=kırmızı, G=yeşil, B=mavi) kısaltmasıyla gösterilir. Görüntüleme bir görüntüde aynı nesneye ait 3 farklı renk değerinin görüntü yakalama cihazına ile elde edilmesi sonucu oluşur. Üç farklı renk değeri elektro-manyetik spektrumda farklı dalga boylarına sahiptir. Bu farklı dalga boylarındaki renk değerlerinin görüntü yakalama cihazında, kırmızı-yeşil-mavi sırasıyla, aynı anda yakalanırsa renkli görüntü oluşur.

2.2.4. Renkli Görüntünün Gri Düzeyli Görüntüye Dönüşümü

Birçok görüntü işleme görevi, işlemi basitleştirmek ve hızlandırmak için renkli bir görüntü, gri tonlamalı veya ikili görüntüyü dönüştürülerek gerçekleştirilebilir. Renkli görüntüden gri tonlamalı görüntü elde edilirken, görüntüler tek boyutlardan oluşmaktadır. Başka bir deyişle, renkli görüntüdeki üç farklı boyuttan oluşan tüm pikseller (0-255) arasında tek bir renk tonu ile gösterilir. Görüntüyü gri tonlamalı görüntüye dönüştürme işlemine bazen görüntü bölütleme algoritmalarında ihtiyaç duyulmaktadır.

2.2.5. Gri Düzeyli Görüntünün İkili Görüntüye Dönüşümü

Bir görüntünün, gri düzeyli bir görüntüden ikili görüntüye dönüşümü yapılırken eşikleme yöntemi kullanılır. Eşik sınır değerini belirlemek için, görüntülerin histogram grafiklerinin incelenmesi gerekmektedir. Görüntüdeki tüm piksel değerleri eşik sınır değeri ile karşılaştırılır. Eşik sınır değerinden küçük olan

piksel değerlerine 0, eşik sınır değerinden büyük olan piksel değerlerine ise 1 renk değeri verilerek siyah ve beyaz renkleri içeren iki tonlu bir görüntü elde edilir. Renkli bir görüntünün önce gri tonlamalı görüntüye, daha sonra ikili görüntüye dönüşümü MATLAB R2017 paket programından alınan görüntü üzerinde Şekil 2.7.'de verilmiştir.



A

B

C

Şekil 2.6. Renkli görüntü $384 \times 512 \times 3$ (A), Gri tonlamalı görüntü 384×512 (B) ve İkili görüntü 384×512 (C).

2.3. Gürültü (Noise)

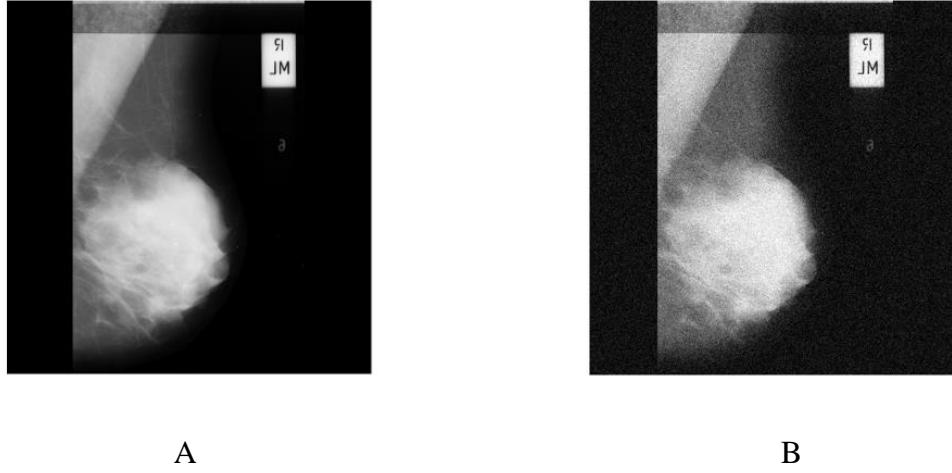
Sayısal görüntülerdeki gürültü; görüntüler üzerinde ortaya çıkan leke, bozukluk, düz çizgiler, etiket bilgisi gibi istenmeyen işaretlerdir. Gürültü, görüntünün kalitesini düşürdüğü gibi görüntüdeki anlamlı kısımların da fark edilmesine engel olabilmektedir. Gürültü, bir görüntüdeki çeşitli kaynaklardan oluşabilir. Görüntüler elde edilirken ortamda yer alan birçok faktörden (çevresel faktörler, nem, elektromanyetik dalgalar, görüntü yakalama aygıtlarından vb.) dolayı görüntüde istenmeyen işaretler oluşur. Görüntü türlerindeki gürültüler farklılık göstermektedir.

2.3.1. Gürültü Çeşitleri

Dijital görüntüler genellikle gürültü olarak adlandırılan rasgele hatalarla bozulmaktadır. Görüntüde hangi aşamada ne tür bir gürültü oluştuğunu bilmek mümkün değildir. Eğer bunun bilinmesi mümkün olsaydı, görüntülerdeki gürültüler sıfır hata ile kaldırılabilirdi. Ancak pratikte böyle bir durum mümkün değildir. Görüntülerdeki bu gürültüler temel olarak 2'ye ayrılır:

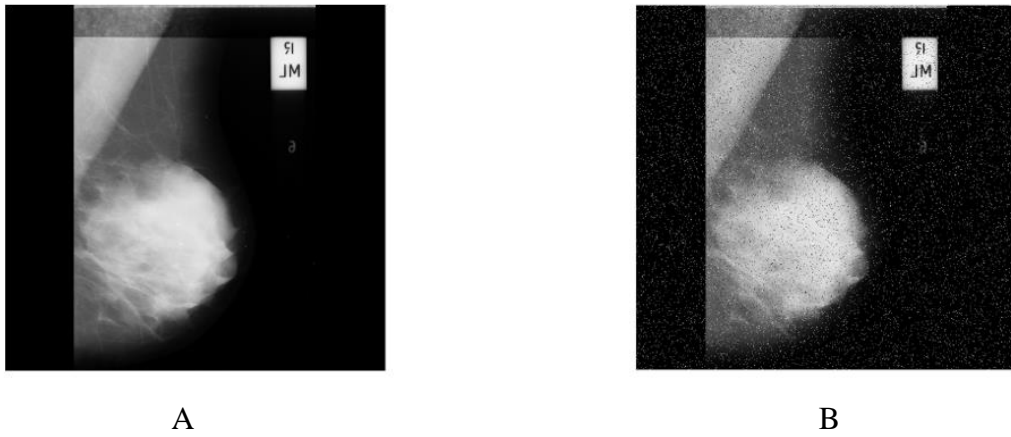
- Gauss gürültüsü, Gauss Dağılımı olarak da bilinen, normal dağılıma eşit olasılık yoğunluk işlevine sahip istatistiksel bir gürültüdür. Gauss gürültüsü olarak

sınıflandırılmasına neden olan rasgele dalgalanan sinyaldir. Kötü aydınlatmadan veya sensörün yüksek sıcaklığından kaynaklanabilir (9). Mamografi görüntüsüne Gauss gürültüsü eklenmiş hali Şekil 2.7’de gösterilmiştir.



Şekil 2.7. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Gauss gürültüsü eklenmiş 1024×1024 piksel boyutunda mamografi görüntüsü.

- Tuz ve biber gürültüsü (salt and pepper) görüntü yakalama cihazının algılayıcılarındaki piksel bozukluklarından, hatalı kapasite alanlarından ya da sayısallaştırma aşamasında zamanlama gibi hatalardan meydana gelen gürültü türleridir. Alanyazında impulse gürültüsü olarak da tanımlanır (9). Mamografi görüntüsüne tuz ve biber gürültüsü eklenmiş hali Şekil 2.8’de gösterilmiştir.



Şekil 2.8. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Tuz ve biber gürültüsü eklenmiş 1024×1024 piksel boyutunda mamografi görüntüsü.

2.4. Görüntü Filtreleme Yöntemleri

Görüntü işleme algoritmalarını uygulamadan önce görüntüler üzerindeki bu gürültülerin kaldırılması gerekir. Çünkü sayısal görüntü üzerinde yer alan bu gürültüler ve istenmeyen bilgiler, bölütleme ve öznitelik çıkarımı adımlarından elde edilecek sonuçları etkiler. Söz konusu durumda, farklı görüntü filtreleme yöntemlerinden faydalanılmaktadır. Alanyazında görüntülerdeki gürültülerin yok edilmesi ve görüntülerin iyileştirilmesi için farklı filtreleme teknikleri önerilmiştir (4, 5, 8, 9). Genel olarak uygulanan tüm bu yöntemlere görüntü ön işleme algoritmaları denir. Çeşitli görüntü ön işleme yöntemleri ile görüntü üzerinde netleştirme ve kontrastı artırma, bazı ayrıntıları belirginleştirme, görüntüyü yumuşatma, kenar keskinleştirme gibi işlemler gerçekleşir.

Bu filtreleme yöntemleri istenilen görüntüyü netleştirirken herhangi bir piksel kaybına yol açmamaktadır. Bu nedenle, görüntülere uygun filtreleme yöntemlerinin uygulanması gerekir.

2.4.1. Görüntü Yumuşatma Filtreleri (Bulanıklaştırma, Smoothing)

Görüntü yumuşatma filtreleri, görüntüdeki gürültüyü azaltmak için kullanılmaktadır. Filtreleme işlemlerinde, bu filtre matrisleri görüntüdeki tüm pikseller üzerinde uygulanır. Gürültü azaltma doğrusal ya da doğrusal olmayan filtre işlemleriyle yapılabilir.

Doğrusal yumuşatma filtresine ortalama (mean, averaging) filtre örnek olarak verilebilir. Ortalama filtre, görüntüdeki her pikselin değerini tüm piksellerin ortalama değeri ile değiştirmektedir. Ortalama filtre olarak 3×3 , 5×5 ya da 7×7 gibi tek sayıları içeren kare matrisler uygulanır. Bu yöntem özellikle Gauss gürültülerinin temizlenmesinde etkilidir. Filtrelerin boyutları büyüdükçe gürültülerin büyük bir kısmı görüntülerden kaldırılırken görüntü daha da bulanıklaşır (10). Öte yandan, daha küçük boyutlu filtreler görüntüyü bulanıklaştırmaz, ancak görüntüde önemli miktarda gürültü bırakabilirler. Bu problemi gidermek için görüntü işlemede kullanılan farklı filtreleme algoritmaları geliştirilmiştir.

Doğrusal olmayan yumuşatma filtresine ortanca filtre örnek olarak verilebilir. Ortanca filtre, pikselin değerini, o piksele komşu tüm piksellerin sıralanması ve sonra da merkezdeki piksel değerinin bu sıralama sonunda ortanca değerle değiştirilmesidir. Ortanca filtre 3×3 , 5×5 ya da 7×7 gibi tek sayıları içeren kare matrislerden oluşur. Görüntüyü temsil etmeyen piksel değerleri için, ortanca değer ile filtre uygulamak, ortalama filtrenin uygulanmasından daha etkili sonuç verir. Bu nedenle, ortanca filtre görüntüde daha az bulanıklaştırma yaptığı için keskin kenarlar daha belirgin görünür. Ortanca filtre, tuz ve biber gürültüsünün temizlenmesinde etkilidir.

Görüntü ön işleme aşamasında yaygın olarak kullanılan bir diğer yumuşatma filtresi Gauss bulanıklaştırma filtresidir. Bu filtre Gauss fonksiyonunu kullanır. Bu nedenle, gürültüleri görüntüden temizleme etkisi Gauss dağılımının standart sapmasına bağlıdır. Gauss dağılımında standart sapma değeri arttıkça Gauss bulanıklaştırma filtresinin de boyutu da büyür. Her bir pikselin değeri komşularının ağırlıklarının ortalama değerine dönüştürülür. Bu durum, ortalama filtrenin aksine görüntüde daha iyi bir iyileştirme sağlar ve kenarlar aynı boyuttaki ortalama filtreye kıyasla daha belirgin görünür. Gauss filtreleri, Gauss gürültüsünün temizlenmesinde etkilidir.

Genel olarak görüntü yumuşatma algoritmaları görüntü ayrıntılarını bulanıklaştırır. Bu nedenle, görüntüye yumuşatma filtrelerine ek olarak görüntü keskinleştirme filtreleri de uygulanmalıdır.

2.4.2. Görüntü Keskinleştirme Filtreleri

Bir görüntüdeki nesnelere kenarlarının ve ince ayrıntılarının daha net görünmesi için geliştirilen görüntü ön işleme yöntemleri keskinleştirme teknikleri olarak bilinir (11). Keskinleştirme filtreleri, görüntüler üzerinde sayısal türev alma aracılığıyla gerçekleştirilmektedir. Alanyazında farklı keskinleştirme filtreleri mevcuttur. Bunlardan bazıları Laplacian filtre, keskin olmayan maskeleme, yüksek hızlı filtreleme (high boost filter) ve gradyan maske (gradient mask)'dir.

Laplacian filtresi, görüntülerin ikinci türevleri hesaplanır ve birinci türevlerin değişme hızını ölçen görüntü keskinleştirme filtreleme yöntemidir. Laplacian filtre

pikselleri, genellikle matris ortalanmış negatif değerlerden oluşmaktadır. Köşeler sıfır ya da pozitif değerlerden oluşmaktadır. Merkez piksel değeri negatif ya da pozitif olabilir. Bu türev, kenarları belirgin hale getirdiği için görüntü bölütleme algoritmalarında, özellikle kenar tabanlı tekniklerde tercih edilmektedir (5).

Keskin olmayan maskeleme, genellikle görüntü işleme alanında kullanılan bir görüntü keskinleştirme tekniğidir. Bu işlemde, bir görüntünün maskesini oluşturmak için önce orijinal görüntü bulanıklaştırılır. Daha sonra orijinal görüntüden, bulanık görüntü çıkarılarak keskin olmayan maske elde edilir. Elde edilen maske orijinal görüntüye eklenerek keskin olmayan maskeleme yöntemi uygulanmış olur. Görüntüdeki kenarlar ve yüksek frekanslı bileşenleri keskinleştirmek amacıyla kullanılan bir çeşit filtreleme tekniğidir (12). $\bar{f}(x, y)$ 'nin bulanık bir görüntüyü ifade ettiği varsayılırsa keskin olmayan maskeleme,

$$g_{mask}(x, y) = f(x, y) - \bar{f}(x, y) \quad (2.4.)$$

Eşitlik 2.4. ile ifade edilir. Elde edilen maskeyi orijinal görüntüye ekleyerek keskin olmayan maskeleme yöntemini uygulamış oluruz (Eşitlik 2.5).

$$g(x, y) = f(x, y) + g_{mask}(x, y) \quad (2.5.)$$

Yüksek hızlı filtreleme, görüntüdeki düşük frekans bileşenlerini ortadan kaldırmadan görüntü ayrıntılarını temsil eden yüksek frekans bileşenlerini vurgulamak amacı ile kullanılan görüntü keskinleştirme tekniğidir. Bu durumda, yüksek hızlı filtreleme, görüntülerdeki düşük frekans bileşenlerini korurken yüksek frekans bileşenlerini geliştirmek için kullanılır.

Gradyanlar, çoğunlukla bir görüntüdeki nesnelere kolayca görülemeyen kenarlarını zenginleştirmek için bir ön işleme adımı olarak kullanılır (5). Alanyazında, görüntüdeki kenarları belirlemek amacıyla kullanılan yöntemler için çeşitli görüntü gradyanları mevcuttur: Canny, Sobel, Robert, Prewitt gibi kenar bulmak için kullanılan yöntemler görüntü gradyanlarını kullanır.

2.5. Görüntü İyileştirme Yöntemleri

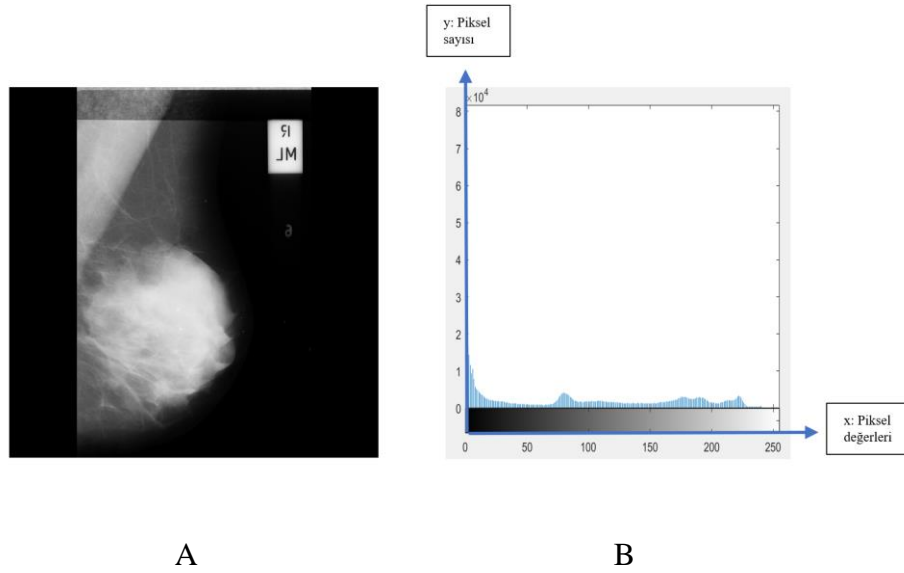
Görüntü yumuşatma ve keskinleştirme filtrelerine ek olarak, görüntü iyileştirme yöntemleri de ön işleme aşamasında uygulanan yöntemlerdir. Görüntü iyileştirme yöntemleri, görüntüyü birtakım işlemlerden geçirerek istenen amaca daha uygun hale getirmektedir. Bu yöntemler doğrudan görüntüdeki pikseller üzerinde işlem yapmaya dayanır. Temel görüntü geliştirme kontrasttaki artıştır (13). Bu nedenle, görüntü iyileştirme işlemleri, görüntülerin kontrastlarını arttırarak görünümelerini geliştirmek için kullanılır (14).

Görüntülerin kontrastını artırarak, görüntüleri iyileştirmek için önce histogram grafikleri incelenir. Her bir gri ton seviyesinin [0-255] tüm aralıktaki dağılımı histogram ile gösterilir. Histogram, o görüntüde bulunan her farklı yoğunluk değerinde, görüntüdeki piksel sayısını gösteren bir grafikdir. Görüntünün histogramı ile uygulanacak olan eşik değerleri de belirlenebilir (15). Histogram,

$$p(r_k) = n_k \quad (2.6.)$$

Eşitlik 2.6. ile ifade edilir. Bu denklemde, r_k k. gri ton seviyesini, n_k ise görüntüdeki r_k gri ton seviyesine sahip piksellerin sayısını göstermektedir.

Eğer görüntü histogram grafiğinde renk değerleri sol tarafa doğru kümelenmiş ise görüntü kontrastı koyu, histogramın sağ tarafına doğru oluşmuş ise görüntü kontrastı açık olur. Histogram grafiğinde x eksenini piksel değerlerini, y eksenini ise her bir gri değer için bulunan piksel sayısını ifade eder (Şekil 2.9).



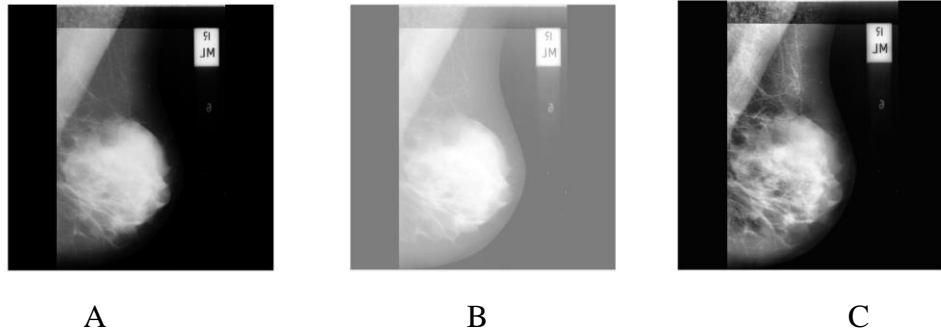
Şekil 2.9. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Mamografi görüntüsünün histogram grafięi (B)

Her bir gri ton seviyesinin daęılımı homojen olan görüntülerde görüntü iyileştirme yöntemlerinin uygulanmasına gerek yoktur. Ancak, görüntüdeki renk deęerleri belli bir yerde kümелendięinde renk daęılımı bozukluęu olabilir. Bu nedenle görüntünün belirli bazı gri seviyelerinde yoğunlaşan histogram grafięini 0-255 gri seviyesi aralıęına daęıtmak için alanyazında çeşitli yöntemler önerilmiştir (15, 16, 17, 18).

2.5.1. Histogram Eşitleme

Histogram eşitleme, histogram üzerinde gri ton seviyeleri aralıęını genişleterek eşit daęılımlı bir yapı elde etmeye çalışan kontrast zenginleştirme yöntemleri arasına girmektedir. Histogram üzerindeki bu yayma işlemi, görüntünün tamamına uygulanabileceęi gibi sadece seçili bölgeler için de uygulanabilir. Histogram eşitleme yöntemi ile görüntünün olasılık yoğunluk fonksiyonu doğrusallaştırılır ve normalleştirilir. Ancak bu yöntem uygulandıktan sonra bulanık ve soluk görüntüler oluşabilir. Bu sorunlara çözüm bulmak için uyarlanabilir (adaptive) histogram eşitleme yöntemleri geliştirilmiştir (16). Bu yöntemde de gürültü problemi ortaya çıktığı için kontrast sınırlı uyarlanabilir histogram eşitleme yöntemi geliştirilmiştir

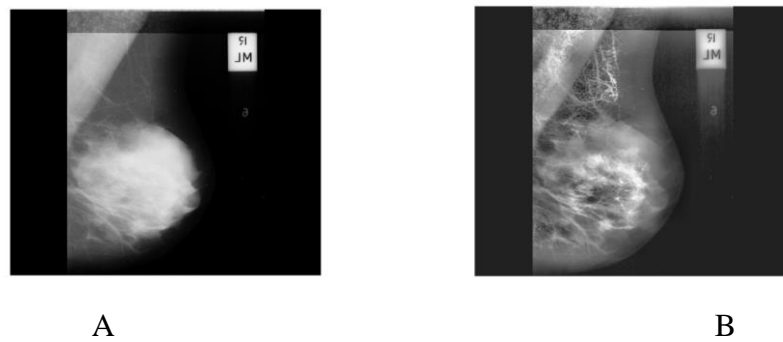
(17). Histogram eşitleme algoritması ile uyarlanabilir histogram eşitleme algoritması arasındaki farklılıklar mamografi görüntüsü üzerinde Şekil 2.10'da gösterilmiştir.



Şekil 2.10. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (B), Uyarlanabilir histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (C).

2.5.2. Kontrast Sınırlı Uyarlanabilir Histogram Eşitleme (CLAHE)

Uyarlanabilir histogram eşitleme yöntemi ile iyileştirilen görüntülerde, gürültü problemini engellemek için homojen bölgelerde kontrast iyileştirmenin sınırlandırılması gerekmektedir (16). Bu nedenle, kontrast sınırlı uyarlanabilir histogram eşitleme yöntemi geliştirilmiştir. CLAHE başlangıçta tıbbi görüntüleme için geliştirilmiş ve portal filmler gibi düşük kontrastlı görüntülerin geliştirilmesinde başarılı olduğu kanıtlanmıştır (18). Bu algoritma, kullanılan gri değerlerin dağılımını dengeler ve böylece görüntünün gizli bölgelerini daha görünür hale getirir. CLAHE yöntemi uygulandıktan sonra görüntüyü nasıl etkilediği mamografi görüntüsü üzerinde Şekil 2.11'de gösterilmiştir.



Şekil 2.11. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Kontrast sınırlı uyarlanabilir histogram eşitleme algoritması uygulandıktan sonra elde edilen görüntü (B).

2.6. Görüntüler Üzerinde Morfolojik İşlemler

Matematiksel morfolojik işlemler, görüntülerin üzerindeki yapıların biçimleri ile ilgili olan yöntemler bütünü olarak, görüntüden ayırt etmek istediğimiz nesnelere için kullanılmaktadır (19). Aynı zamanda görüntüdeki sınırlar, dışbükey ve içbükey gibi bölgesel şekillerin tanımlanmasında kullanılır. Görüntülere morfolojik operatörleri uygulamak için bir yapısal elemana (structure element) ihtiyaç vardır. Her bir piksel yapısal eleman kadar büyür ya da küçülür. Morfolojik işlemler, ön işleme ve görüntü bölütleme aşamalarında farklı amaçlar için kullanılabilir.

Ön işleme aşamasında uygulanan yöntemler sonucunda kalan alanların şekillerinde bazı bozukluklar oluşabilir. Böyle bir durumda, bazı bölgeler birbirinden ayırmak istenirse, o ince çizgileri temizlemek gerekir. Bazen görüntüler üzerindeki yapıların içindeki çok küçük alanlar kapatılıp büyük olan alana dahil edilmek istenebilir. Girinti ve boşluk gibi meydana gelen bozukluklar morfolojik işlemlerle düzeltilerek, görüntülerdeki bölgeler öznelik çıkarımı için hazırlanır. Özetle, görüntüleri temizlemek ve görüntüler üzerinde boşlukları azaltmak ve kenar belirlemek için morfolojik işlemler gerçekleştirilebilir.

Bölüm 2.7.'de detayları anlatılmış olan görüntü bölütleme aşamasında da çeşitli morfolojik işlemler kullanılabilir. Bölütleme yöntemi uygulandıktan sonra, morfolojik işlemler bölümlenmiş görüntüdeki kusurları gidermek ve görüntünün şekli ile yapısı hakkında bilgi sağlamak için kullanılabilir. Görüntü işlemedeki morfolojik operatörler görüntüye bölütleme algoritması uygulandıktan sonra, görüntüde ortaya çıkan bozuklukları gidermek için uygulanır.

Özetle, matematiksel morfolojik işlemler, görüntüde yer alan sınırlar (borders), iskelet (skeleton) vb. şekillerin belirlenmesi ve kaldırılması, görüntülerin temizlenmesi ve bölütleme aşamalarında uygulanması gereken işlemlerdir. Morfolojik işlemleri kullanmanın amacı, görüntünün yapısındaki hataları ortadan kaldırmaktır (20). Bu yöntemler, gri tonlamalı görüntülerde uygulanabileceği gibi genellikle ikili görüntüler üzerinde uygulanır (21).

Görüntü işlemede en çok kullanılan morfolojik operatörler genel olarak iki temel işlemden oluşmaktadır. Bunlar; aşınma (erosion) ve genişleme (dilation) işlemleridir. Bu iki işlemin kombinasyonlarının kullanım sırasına göre, görüntü işlemede açma (opening) ve kapama (closing) olmak üzere iki tane de matematiksel morfoloji operatörü yaygın bir şekilde kullanılmaktadır.

2.6.1. Aşınma (Erosion)

Aşınma operatörü, birbirine çizgi benzeri yapılar ile bağlanmış olan görüntü üzerindeki alanları (bölgeleri) birbirinden ayırmak için kullanılır. Aşınma, bir görüntüdeki nesnenin sınırlarındaki piksellerin alanlarını küçülterek daraltma işlemi yapar. Görüntüdeki nesnelerin küçültülmesi ve inceltilmesi için kullanılan bir yöntemdir. Böylece aşınma ile birbirine bitişik yapılar ayrılarak daha doğru bir şekilde değerlendirilmeleri sağlanır. Özetle, aşınmanın, görüntü üzerindeki yapısal elemandan daha küçük görüntü detaylarını yok eden bir morfolojik filtre olarak görev yaptığı söylenebilir. A'nın B yapısal elemanı ile aşınması $A \ominus B$ olarak gösterilir ve Eşitlik 2.7. ile ifade edilir:

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (2.7.)$$

Yukarıdaki eşitlikten, A girdi görüntüsünün B yapısal elemanı ile aşındırılması, z kadar öteleme durumunda, B'nin A içerisinde yer alabileceği bütün z noktaları kümesi olarak tanımlanır.

2.6.2. Genişleme (Dilation)

Daraltma ve inceltme işlemlerinin gerçekleştirildiği aşınmanın tersine, genişleme görüntü içindeki nesne sınırlarına pikseller ekleyerek büyütür ve kalınlaştırır. Genişleme işlemi, bir görüntü içindeki nesnelerin alanlarının ince bir çizgi ile bölünerek ayrı iki nesne gibi görünmesini engellemek için kullanılır. Bu kalınlaşmanın büyüklüğü ve durumu kullanılan yapısal elemanın biçimi ile kontrol edilir. Böylece genişleme operatörü ile, nesnelerin alanı artmış olur. A'nın B yapısal elemanı ile genişlemesi $A \oplus B$ olarak gösterilir ve Eşitlik 2.8. ile ifade edilir:

$$A \oplus B = \{z | (\widehat{B})_z \cap A \neq \emptyset\} \quad (2.8)$$

Görüntü bileşenlerini yok etmek için aşınma operatörü kullanılırken, görüntüdeki boşlukları birleştirmek için genişleme operatörü kullanılır.

2.6.3. Açma (Opening)

Açma işlemi için önce aşınma operatörü uygulanır. Böylece, görüntünün kenarlarındaki çok küçük parçalar kaybolur. Ancak, bu durumda genel alan küçülmüş olur. Küçülen alanı tekrar eski haline getirmek için genişleme işlemi uygulanır. Aşınma ile gürültü ve fazlalıklar temizlendikten sonra, genişleme operatörü uygulanarak görüntüdeki nesnelerin orijinal boyutları tekrar kazandırılmış olur. Bir kez uygulanması yeterli olmaz ise aynı işlem tekrar uygulanabilir. Açma işlemi nesnelerin keskin ayrıntılarını yumuşatır, bölgeler arasındaki dar nesnelere ayırır ve küçük girintileri kaldırır. Açma işlemi Eşitlik 2.9. ile gösterilir.

$$A \circ B = (A \ominus B) \oplus B \quad (2.9.)$$

2.6.4. Kapama (Closing)

Kapama operatörü açma operatörünün tam tersidir. Önce genişleme, ardından aşınma operatörü uygulanır. Genişleme operatörü uygulandığı için görüntüdeki nesnelerin boyutları büyür. Daha sonra aşındırma operatörü ile görüntüdeki nesnelere orijinal boyutuna geri kazandırılmış olur. Kapama işlemi sonucunda nesnelerin sınırındaki girintiler yok olur ve küçük delikler ortadan kalkar. Kapama işlemi Eşitlik 2.10. ile gösterilir.

$$A \bullet B = (A \oplus B) \ominus B \quad (2.10.)$$

2.7. Görüntü Bölütleme (Segmentasyon) Yöntemleri

Görüntü bölütleme, bir görüntüyü her bir bölgede farklı özelliklerin olduğu bölge ve nesnelere ayırarak bölmektir. Diğer bir deyişle, görüntü bölütleme bir görüntüyü anlamlı parçalara bölme işlemidir. Bölütleme işlemi sonunda elde edilen bölgelerdeki pikseller ortak özelliklere sahiptir. Görüntü işleme alanında en çok incelenen konudur. Bu durum, görüntü bölütleme adımının önemine ve farklı yapıdaki görüntüleri bölütleme için evrensel bir yöntemin bulunmamasına bağlanabilir.

Görüntü işlemede, öznitelik çıkarımı yapmadan önce ilgili bölgenin uygun bölütleme yöntemi ile belirlenmesi gerekir. Görüntü bölütlemenin, görüntü işlemedeki en zor adım olduğu söylenebilir. Çünkü, bölütleme işleminin başarısının öznitelik çıkarımı ve sınıflandırma performansı üzerinde etkisi oldukça büyüktür.

İncelenecek görüntü türlerine göre çeşitli segmentasyon yöntemleri vardır. Segmentasyon algoritmaları genellikle görüntünün yoğunluk değerlerine bağlı iki temel özelliğine dayanır. Bunlardan biri süreksizlik (discontinuity) diğeri ise benzerliktir (similarity) (5). Süreksizliğe bağlı yaklaşım, farklı bölgeler arasındaki sınır veya kenarları bularak görüntünün bölgelerini ayırt etmeye çalışan algoritmaları içerir. Bu yöntemlere kenar tabanlı teknikler (edge based techniques) denir. Benzerliğe bağlı yaklaşım ise, önceden tanımlanan kriterler kümesine göre görüntüyü benzer olan bölgelere ayıran algoritmaları içerir.

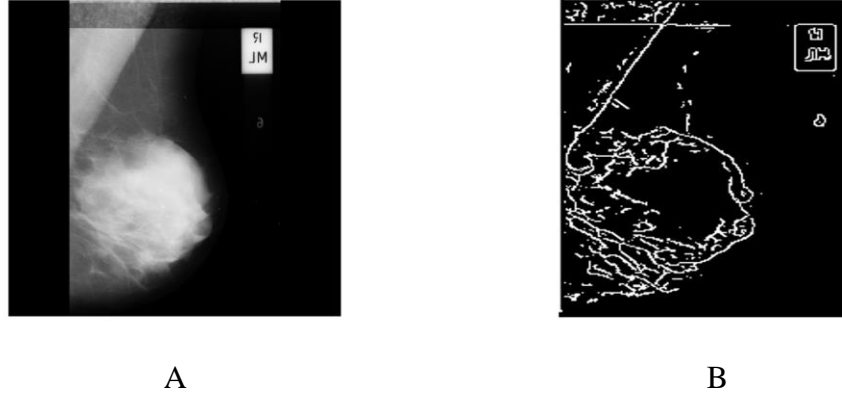
Alanyazında görüntü bölütleme yöntemleri için farklı şekillerde sınıflandırmalara rastlanılmaktadır. Bu tezde görüntü bölütleme algoritmaları kenar tabanlı teknikler ve bölge tabanlı teknikler olmak üzere 2 başlık altında açıklanmıştır.

2.7.1. Kenar Tabanlı Bölütleme (Edge based segmentation)

Gri tonlamalı görüntülerde yoğunluklardaki keskin değişiklikler kenar olarak kabul edilir (10). Sınır çizgileri veya kenarlar, çeşitli görüntü işleme uygulamalarında ihtiyaç duyulan önemli özelliklerdir. Kenar bulma teknikleri ile görüntüdeki nesnelere hem birbirinden hem de arka plandan ayırt edilir. Kenar tabanlı bölütleme yöntemleri, bölge tabanlı bölütleme yöntemlerine göre nesnelere ait bütün pikselleri belirtmek yerine, sadece sınırlara ait pikselleri belirler.

Kenar algılama (edge detection) yöntemleri, görüntüdeki nesnelere ait sınırları veya parçalarını bulmak için etkili bir yöntemdir. Kenar tabanlı görüntü bölütleme yöntemleri, kenar operatörleri ile bölgeler için sınırları oluşturan kenarları belirleyerek görüntünün bölgelerini belirlemeye çalışır. Alanyazında birçok kenar algılama yöntemi mevcuttur. Canny, Sobel, Prewitt, Roberts, Laplacian of Gaussian (LoG), Hough transform algoritmaları örnek olarak verilebilir (5,10,22).

Kenar tabanlı bölütleme yöntemlerinde amaç, görüntüde yer alan nesnelerin piksel değerlerindeki renk geçişlerini belirleyip, o bölgeleri keskinleştirmektir. Mamografi görüntüsüne Canny kenar algılama algoritması uygulandıktan sonraki görüntü Şekil 2.12’de gösterilmiştir.



Şekil 2.12. 1024×1024 piksel boyutunda mamografi görüntüsü (A), Canny kenar belirleme yöntemi ile oluşan mamografi görüntüsü (B)

2.7.2. Bölge Tabanlı Bölütleme (Based region segmentation)

Bölge tabanlı görüntü bölütleme yöntemleri, görüntüdeki her pikseli değerlendirir ve farklı grupları belirlemek için tüm pikselleri karşılaştırır.

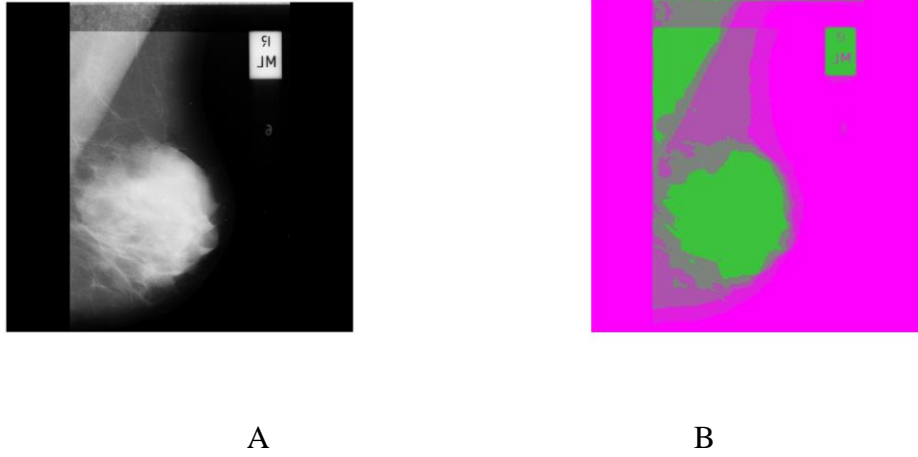
Alanyazında birçok bölge tabanlı bölütleme yöntemi mevcuttur. Eşikleme yöntemleri (Global, Manual, Otsu vb.), k-ortalamlar kümeleme, tohum bölge büyütme (seed region growing), Watershed algoritmaları bölge tabanlı tekniklere örnek olarak verilebilir (22,23).

Bölütleme yöntemi olarak eşikleme algoritmalarının kullanılmasının temel amacı, görüntülerin arka planını kaldırarak görüntüleri bölmektir. Böylece görüntüde ilgili alana odaklanarak, öznelik çıkarımı gerçekleştirilir (10). Eşikleme algoritmasını görüntüye uygulamadan önce görüntüyü gri tonlu (gray scale) formata, daha sonra da ikili (0-255 arasında değişen gri ton seviyeleri) görüntüye dönüştürülür. Bir referans piksel değeri (eşik değeri) belirlenir. Eşik değerinin üstündeki ya da altındaki tüm değerler siyah ve beyaza dönüştürülür. Böylece görüntü sadece siyah veya beyaz renkleri içerir. Görüntüdeki piksel değeri eşik değerinden büyükse beyaza (255), küçükse siyaha dönüşür (0).

K-ortalamlar kümeleme algoritması, görüntü işleme uygulamalarında yaygın olarak kullanılan görüntü bölütleme tekniklerinden biridir. Bu algoritma, iyi bilinen bir denetimsiz kümeleme yöntemidir. Bu segmentasyon yöntemi, önceden belirlenen küme sayısı aracılığıyla görüntüyü farklı bölgelere ayırmanın basit ve kolay bir yolunu sunar. İlgilenilen görüntüye ait olan bir başlangıç piksel ya da bölgesini seçtikten sonra, komşu piksellerin aynı piksele ya da bölgeye ait olup olmadığına karar vererek, piksel kümelerini oluşturduktan sonra görüntüyü farklı bölgelere ayırmayı sağlayan görüntü bölütleme algoritmasıdır. Mamografi görüntüleri, pektoral kas, meme dokusu ve arka planı içerecek şekilde 3 ana kümeyle bölünebilir. İlk olarak, her küme için bir tane olmak üzere, başlangıç noktası olan K ağırlık merkezleri tanımlanır. Daha sonra, her bir merkez için benzer pikselleri gruplayan bir özellik bölgesi belirlenmiş olur. K-ortalamlar kümeleme bölütleme yönteminin prensibi aşağıda verilmiştir:

$$J = \sum_{j=1}^k \sum_{i=1}^n \left\| x_i^{(j)} - c_j \right\|^2 \quad (2.11.)$$

Rasgele yöntemlere dayalı olarak belirlenen küme sayısına göre, kümelerin merkezleri belirlenir. Burada $\left\| x_i^{(j)} - c_j \right\|^2$, $x_i^{(j)}$ noktasından c_j grubunun merkezine olan uzaklıktır. Her pikselin en yakın kümeyle atanması, $x_i^{(j)}$ noktasından c_j grubunun merkezi arasındaki Öklit uzaklığına göre yapılır. Böylece J fonksiyonu, her bir küme için n tane pikselin (nesnenin) benzerlik ölçüsünü temsil eder. Mamografi görüntüsüne K-ortalamlar kümeleme algoritması uygulandıktan sonra farklı bölgelere ayrılması Şekil 2.13'te gösterilmiştir.



Şekil 2.13. 1024×1024 piksel boyutunda mamografi görüntüsü (A), K-ortalamlar kümeleme yöntemi ile mamografi görüntüsünün bölütlenmesi (B)

2.8. Öznitelik Çıkarımı (Feature Extraction)

Görüntü işlemenin önemli bir diğer adımı görüntülerden öznitelikleri çıkarmaktır. Görüntü işleme alanında öznitelik çıkarım yöntemleri, görüntüdeki ROI'lere ait çeşitli özniteliklerin hesaplandığı algoritmalarıdır. Başka bir deyişle öznitelik çıkarma, görüntülerin görsel içeriğini yakalama yöntemidir.

Görüntülerin doğru sınıflandırılması, en uygun özniteliklerin seçilmesine bağlıdır. Çünkü bu öznitelikler, belirli bir bölgeyi karakterize ettiği için sınıflandırma aşamasında girdi (input) değişkeni olarak kullanılır.

Görüntüdeki ilgili bölgeleri tanımlamak için en basit yaklaşımlar, bir görüntü veya bölgenin yoğunluk histogramının istatistiksel özellikleri ve görüntüdeki komşu piksellerin göreceli konumu değerli bilgiler sağlamamıza yardımcı olacaktır. Komşu piksellerin birbirlerine göre konumları, öznitelik çıkarımında doku öznitelikleri olarak adlandırılır (24). Doku analizi, bir anlamda piksel çeşitliliğini görme yöntemlerinden biridir. Gri seviye eş oluşum matrisi (gray level co-occurrence matrix, GLCM), gri seviye fark yöntemi (gray level difference method, GLDM), gri seviye çalışma uzunluğu matrisi (gray level run length matrix, GLRLM) ve yerel ikili örüntüler (local binary patterns, LBP) doku analizi için kullanılan farklı yöntemlerdir.

Temel istatistiksel öznitelikler, ROI'nin temel istatistiklerinin hesaplanmasıdır. Görüntülerden hesaplanabilecek istatistiksel özellikler Tablo 2.1.'de verilmiştir.

Aşağıdaki denklemlerde I , şüpheli bölge içinde N piksel değeri içeren vektörü temsil eder.

Tablo 2.1. Temel istatistiksel öznitelikler

Öznitelikler	Formüller
Minimum	$\min I(i)$
Maksimum	$\max I(i)$
Ortalama	$\frac{1}{N} \sum_{i=1}^N I(i)$
Standart sapma	$\sqrt{\frac{1}{N} \sum_{i=1}^N \left(I(i) - \frac{1}{N} \sum_{i=1}^N I(i) \right)^2}$
Ortanca	$\begin{cases} I\left(\frac{N+1}{2}\right) & N \text{ tek sayı ise} \\ I\left(\frac{N}{2}\right) + I\left(\frac{N+2}{2}\right) & N \text{ çift sayı ise} \end{cases}$
Dağılım aralığı	$\max I(i) - \min I(i)$
Çarpıklık	$\frac{\sum_{i=1}^N (I(i) - \frac{1}{N} \sum_{i=1}^N I(i))^3}{(N-1) \sqrt{\frac{1}{N} \sum_{i=1}^N \left(I(i) - \frac{1}{N} \sum_{i=1}^N I(i) \right)^2}^3}$
Basıklık	$\frac{\sum_{i=1}^N (I(i) - \frac{1}{N} \sum_{i=1}^N I(i))^4}{(N-1) \sqrt{\frac{1}{N} \sum_{i=1}^N \left(I(i) - \frac{1}{N} \sum_{i=1}^N I(i) \right)^2}^4}$
Boyut	N

Gri seviye eş oluşum matrisi, bir görüntüde belirli değerlere sahip ve piksel çiftlerinin ne sıklıkla meydana geldiğini hesaplayarak bir görüntünün dokusunu karakterize eden istatistiksel bir yöntemdir (25). GLCM ikinci dereceden olasılık yoğunluk fonksiyonun tahmini Eşitlik 2.12.'ye dayanır:

$$P(i, j|d, \theta) \quad (2.12.)$$

GLCM yöntemi pikseller arasındaki uzaklık d ve açı θ olduğunda gri ton seviyesinin i 'den j 'ye değişme olasılığını tanımlar. GLCM 4 farklı açıda θ (0° , 45° , 90° , 135°) ve farklı d uzaklık değerleri için hesaplanabilir. Farklı açılardan özellikler çıkarıldığı için ortalamalar alınır (26). Gri seviye eş oluşum matrisi, dokusal

özelliklerin görüntülerden çıkarılmasında yaygın olarak kullanılan bir matristir (26). GLCM yöntemi ile hesaplanabilecek özellikler Tablo 2.2.'de verilmiştir. Tablo 2.2.'deki denklemlerde N_g giriş görüntüsü için tanımlanan gri seviyelerin sayısıdır.

Tablo 2.2. Gri seviye eş oluşum matrisinden (GLCM) çıkarılan öznelikler

Öznelikler	Formüller
Otokorelasyon	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij)p(i,j)$
Kontrast	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j), i-j =n \right\}$
Korelasyon	$\sum_{i=1}^{N_g-1} \sum_{j=1}^{N_g} \frac{(i,j)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Küme önemi	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i+j - \mu_x - \mu_y)^4 p(i,j)$
Küme gölgesi	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i+j - \mu_x - \mu_y)^3 p(i,j)$
Farklılık	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i-j p(i,j)$
Enerji	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \{p(i,j)\}^2$
Entropi	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j))$
Homojenlik	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1+(i-j)^2}$
Maksimum olasılık	$\max p(i,j)$

Tablo 2.2. (Devam) Gri seviye eş oluşum matrisinden (GLCM) çıkarılan öz nitelikler

Öz nitelikler	Formüller
Karelerin toplamı (varyans)	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i, j)$
Toplam ortalama	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
Toplam varyans	$\sum_{i=2}^{2N_g} \left(i - \left(- \sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} \right) \right)^2 p_{x+y}(i)$
Toplam entropi	$- \sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\}$
Fark varyansı	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
Fark entropisi	$- \sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$
Korelasyon bilgisi ölçümü 1	$\frac{- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (\log(p(i, j))) - (- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (\log\{p_x(i) - p_y(j)\}))}{\max \left\{ - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (\log(p(i, j))) - (- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (\log\{p_x(i) - p_y(j)\})) \right\}}$
Korelasyon bilgisi ölçümü 2	$(1 - \exp \left[-2 \left(- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i) p_y(j) (\log\{p_x(i) - p_y(j)\}) - \left(- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) (\log(p(i, j))) \right) \right) \right] \right)^{\frac{1}{2}}$
Normalleştirilmiş ters fark	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 + i - j /N_g}$

Tablo 2.2. (Devam) Gri seviye eş oluşum matrisinden (GLCM) çıkarılan öznitelikler

Öznitelikler	Formüller
Normalleştirilmiş ters fark momenti	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + (i-j)^2/N_g^2}$

Gri seviye fark matrisi, gri seviyesinde belirli bir mutlak farklılığa sahip olan ve belirli bir yer değiştirme ile ayrılan iki pikselin ortaya çıkmasına dayanan istatistiksel bir yöntemdir. Yer değiştirme vektörünün dört olası biçimi mevcuttur: (0, d), (-d, d), (d, 0) ve (d, -d). Burada d, pikseller arasındaki mesafedir (27). GLDM matrisinden hesaplanabilecek özellikler Tablo 2.3.'te verilmiştir. Tablo 2.3.'teki denklemlerde δ yer değiştirme vektörü, N_g giriş görüntüsü için tanımlanan gri seviyelerin sayısı ve $D(i/\delta)$ olasılık yoğunluk fonksiyonudur.

Tablo 2.3. Gri seviye fark matrisinden (GLDM) çıkarılan öznitelikler

Öznitelikler	Formüller
Kontrast	$\sum_{i=0}^{N_g} i^2 D(i/\delta)$
Açısal ikinci moment (enerji)	$\sum_{i=0}^{N_g-1} [D(i/\delta)]^2$
Entropi	$\sum_{i=0}^{N_g-1} D(i/\delta) \log(D(i/\delta))$
Ortalama	$\sum_{i=0}^{N_g-1} i D(i/\delta)$
Ters fark momenti	$\sum_{i=0}^{N_g} \frac{D(i/\delta)}{i^2 + 1}$

Gri seviye çalışma uzunluğu matrisi, doku özelliklerinin hesaplanması için yaygın olarak kullanılan başka bir matris olan gri seviye çalışma uzunluğu matrisi (28), farklı gri seviyelerini temsil eden satırlar ve belirli bir yönde θ farklı sayıda çalışmayı temsil eden sütunlar içeren iki boyutlu bir matristir. Bu matris için de dört farklı açı yönünde (0° , 45° , 90° ve 135°) çıkarılan özelliklerin ortalaması alınır. GLRLM tekniği ile hesaplanabilecek özellikler Tablo 2.4.'te verilmiştir. Tablo 2.4.'teki denklemlerde N_g giriş görüntüsü için tanımlanan gri seviyelerin sayısı, n_r toplam çalışma sayısı, n_p görüntüdeki piksel sayısı ve M maksimum sayı uzunluğudur. Eğer GLRLM, p ile gösterilirse, $p(i,j)$ 'nin değeri, i değerinin j değeriyle bazı belirlenmiş konumsal ilişkiler içinde kaç kez birleştiğini belirtir. Bu şöyle de açıklanabilir: $P(i,j)$, i gri seviyesinin j uzunluğunda kaç defa meydana geldiğini gösterir.

Tablo 2.4. Gri seviye çalışma uzunluğu matrisinden (GLRLM) çıkarılan öznitelikler

Öznitelikler	Formüller
Short run emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M \frac{p(i,j)}{j^2}$
Long run emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M p(i,j)j^2$
Gray-level non-uniformity	$\frac{1}{n_r} \sum_{i=1}^{N_g} \left(\sum_{j=1}^M p(i,j) \right)^2$
Run length non-uniformity	$\frac{1}{n_r} \sum_{j=1}^M \left(\sum_{i=1}^{N_g} p(i,j) \right)^2$
Run percentage	$\frac{n_r}{n_p}$

Tablo 2.4.(Devam) Gri seviye çalışma uzunluğu matrisinden (GLRLM) çıkarılan öznelikler

Öznelikler	Formüller
Low gray-level run emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M \frac{p(i,j)}{i^2}$
High gray-level run emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M p(i,j) i^2$
Short run low gray-level emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M \frac{p(i,j)}{i^2 j^2}$
Short run high gray-level emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M \frac{p(i,j) i^2}{j^2}$
Long run low gray-level emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M \frac{p(i,j) j^2}{i^2}$
Long run high gray-level emphasis	$\frac{1}{n_r} \sum_{i=1}^{N_g} \sum_{j=1}^M p(i,j) i^2 j^2$

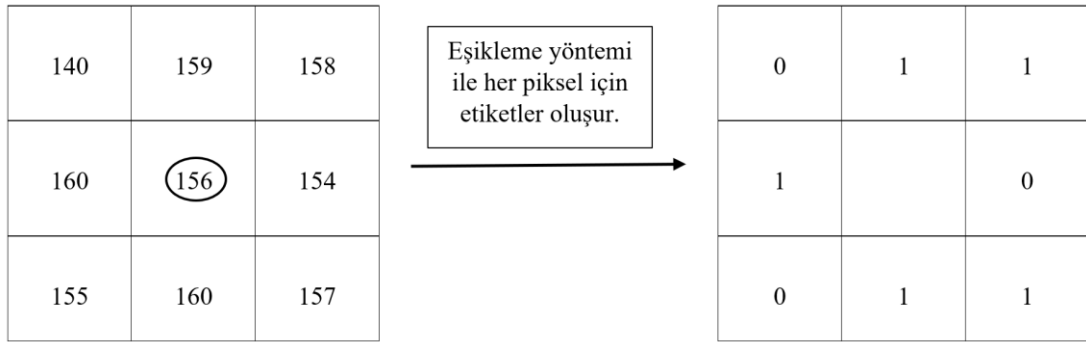
Yerel ikili örüntüler algoritması, görüntülerin gri seviyelerinden bağımsız bir doku özelliği ölçümü yöntemidir. LBP algoritması, her pikselin $R \times R$ komşuluğundaki pikselleri karşılaştırarak ve sonucu ikili bir sayı olarak değerlendirerek (eşikleme) sıfır ve birlerden oluşan etiketler oluşturur (29). Bir görüntünün her c pikseli için LBP değeri Eşitlik 2.12. çözümlenerek hesaplanır:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2.12.)$$

g_c merkezdeki pikselin gri ton seviyesi, g_p merkez pikselin komşusunun gri ton seviyesi, P merkez piksele komşu olan piksellerin sayısıdır ve s fonksiyonu Eşitlik 2.13'te olduğu gibi tanımlanır:

$$s(g_p - g_c) = \begin{cases} 1 & (g_p - g_c) \geq 0 \\ 0 & (g_p - g_c) < 0 \end{cases} \quad (2.13.)$$

LBP operatörünün en önemli özelliği parlaklık değişimlerine karşı dayanıklı olmasıdır. LBP operatörü ile piksellerin etiketlenmesi Şekil 2.14'te gösterilmiştir (30). 156'dan büyükse 1, küçükse 0 değerini alır. LBP Eşitlik 2.14.'teki gibi hesaplanabilir.



Şekil 2.14. LBP operatörü ile pikseller için etiketler elde edilmesi

Öznitelik çıkarma adımı, büyük bir veri kümesini doğru bir şekilde tanımlamak için gereken özniteliklerin boyutunun basitleştirilmesini de kapsar (31). Çok boyutlu verilerin analizi yapılırken en büyük sorunlardan biri ilgili değişkenlerin sayısından kaynaklanır. Görüntülerden çıkarılan özelliklerin çoğu yüksek korelasyon gösterir (32). Bu durum görüntüleri temsil etmeye yardımcı olmayan fazladan özelliklere yol açar. Bu nedenle, öncelikle boyut azaltma işleminin gerçekleştirilmesi gerekir. Öznitelik seçim yöntemleri, sınıflandırılma yapılacak veri setinde doğru bir ayırım yaparken, veriyle ilgili önemli değişkenleri tutarak, veri setindeki çok gerekli olmayacak değişken sayısını azaltmak için tercih edilir. Alanyazına bakıldığında, elde edilen özellikler arasından seçim yapmak için kullanılan korelasyon filtre yöntemleri (Correlation Filter Methods), temel bileşenler analizi (principal component analysis), doğrusal diskriminant analizi (lineer diskriminant analizi), ki-kare testi, kaba set yaklaşımı (Rough set approach), AdaBoost training, minimum fazlalık maksimum

ilişki (mRMR), destek vektör makinesi-özyinelemeli özellik eleme, LASSO regresyon gibi algoritmalar özellik azaltma yöntemlerindedir.

2.9. Sınıflandırma

Sınıflandırma aşaması, ROI bölgelerinin normal ve anormal bölgeler olduğunu belirleyen son adımdır (33). Alanyazında çeşitli sınıflandırma algoritmaları önerilmiştir. Sınıflandırma algoritmalarının performansları özellikle görüntü bölütleme ve özellik çıkarma işlemlerine bağlıdır. Bu algoritmalar, görüntülerden elde edilen özelliklerin daha kısa sürede incelenmesini sağlar. Klasik veri madenciliği yöntemlerinden yaygın olarak Rastgele Orman (RO), Destek Vektör Makineleri (DVM), Yapay Sinir Ağı (YSA), k-En Yakın Komşuluk (k-NN), Naive Bayes (NB), Karar Ağaçları (KA), Lojistik Regresyon (LR) vb. sınıflandırıcılar kullanılır. Derin öğrenme yöntemlerinden Derin Sinir Ağları, Evrişimli Sinir Ağları ve Tekrarlayan Sinir Ağları yaygın olarak sınıflandırma amaçlı kullanılır.

2.9.1. Rastgele Orman

Rastgele Orman (RO) algoritması, sınıflandırma adımında çeşitli karar ağacı ile sınıflandırma performansını iyileştirmeyi amaçlayan bir yöntemdir. Oluşturulan karar ağaçları birlikte karar ormanını meydana getirir. Bu algoritma 2001 yılında Brieman (34) tarafından geliştirilmiştir. RO algoritması yine Brieman'ın 1996 yılında geliştirdiği Bagging yöntemi ve Amit ve Geman tarafından 1997'de tanımlanan (35) her düğüm için en iyi ayrımın rastgele değişken seçimi yöntemlerinin birleşimidir. RO yönteminde, ağaçlar oluşturulurken Sınıflama ve Regresyon Ağacı (Classification and Regression Tree-CART) yöntemi kullanılmaktadır. Tüm veri eğitim ve test seti olmak üzere ikiye ayrılır.

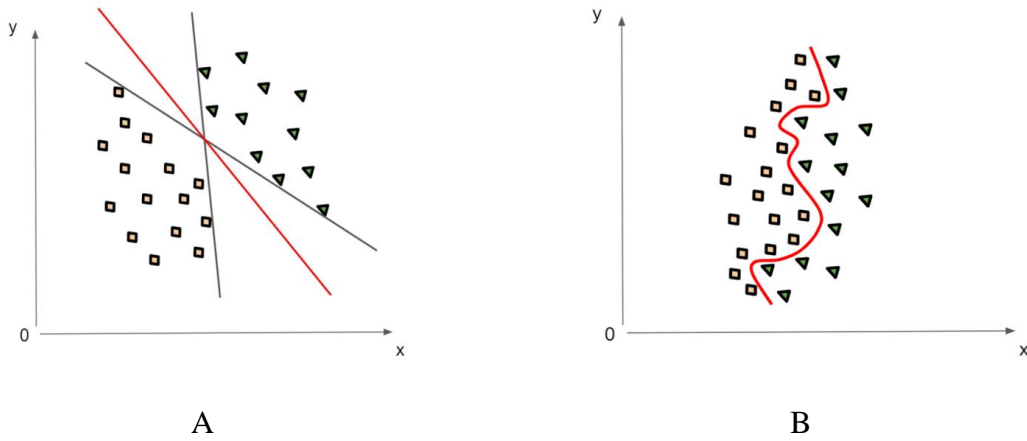
Rastgele orman algoritmasının bir avantajı veri setindeki bağımlı değişken türüne göre hem sınıflama hem de regresyon amaçlarıyla kullanılmasıdır. RO algoritmasında hiperparametre (optimize edilmesi gereken parametre) sayısı fazla olmadığı için uygulanması kolaydır. Sınıflandırma algoritmalarındaki en büyük sorun veride aşırı uyum sorunudur. RO modelinde ağaç sayısının yeterli olduğu durumlarda bu sorun çözülebilir.

2.9.2. Destek Vektör Makineleri

Destek Vektör Makineleri ilk olarak 1960'lı yılların sonunda Vapnik tarafından (36) teorik olarak ortaya konmuş, 1995 yılında yine Vapnik (37) tarafından ilk kez sınıflama başarısının yüksek olduğu gösterilmiştir. Destek Vektör Makineleri (DVM), örüntü tanıma ve sınıflandırma problemleri için Vapnik ve Chervoenkis tarafından geliştirilmiştir.

DVM ilk olarak iki sınıflı doğrusal verilerin sınıflandırılması amacıyla geliştirilmiş, daha sonra çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması için geliştirilmiştir. DVM iki sınıflı doğrusal verileri ayırmak için en uygun karar fonksiyonunu tahmin eder, iki sınıfı en uygun biçimde ayırabilen hiper düzlemi tanımlar (38). DVM çekirdek (kernel) yöntemini temel alır.

DVM'de, sınıflar birbirinden ayrılarak optimal düzlemin elde edilir. Başka bir deyişle, farklı sınıflara ait destek vektörleri arasındaki mesafeyi maksimum yapmaktır.



Şekil 2.15. Verilerin doğrusal olarak ayrılabilme (A) ve ayrılama durumu (B).

2.9.3. Yapay Sinir Ağı

Yapay Sinir Ağlarının (YSA) dayandığı ilk hesaplama modelinin temelleri 1940'ların başında araştırmalarına başlayan W.S. McCulloch ve W.A. Pitts'in, 1943 yılında yayınladıkları bir makaleyle (39) atılmıştır. 1985 yapay sinir ağlarının oldukça tanındığı, yoğun araştırmaların başladığı yıl olmuştur (40).

Yapay Sinir Ağları, biyolojik sinir sistemlerini taklit ederek, eldeki büyük ve karmaşık veri yığımından farklı algoritmalar ile yeni bilgi keşfedilmesini sağlayan bir yöntemdir. İnsan beyni yapısına benzer şekilde, yapay sinir ağları ağırlıklı bağlantılar ile birbirine bağlanan, bir anlamda, biyolojik sinir ağlarına benzer yapıda olan algoritmalarıdır. YSA, hücrelerin katmanlarla birleştirilmesiyle kurulan, basitleştirilmiş modeller ile karmaşık problemlerin çözümlenmesinde kullanılmaktadır. YSA öncelikle sınıflandırma, modelleme ve tahmin uygulamalarında ve daha pek çok alanda (sağlık, bankacılık, savunma, güvenlik vb.) kullanılmaktadır.

2.9.4. k-En Yakın Komşuluk

k-En Yakın Komşuluk (k-NN) algoritması, sınıflandırılacak veriyi başka verilerle olan yakınlık ilişkisine göre sınıflandırma yapan yöntemdir. Her yeni gözlemin sınıflandırması veya tahmini, ağırlıklı ortalamalara göre belirlenen bir uzaklığa göre hesaplanır. k-NN algoritması uzaklık ölçümü olarak Öklid, Manhattan, Minkowski vb. uzaklıkları kullanır. k bir veri kümesindeki sonuçları sınıflandırmak veya tahmin etmek için kullanılan en yakın komşuların sayısını ifade eden hiperparametredir. Bu yöntemde k için sabit bir değer yoktur. Bu yüzden farklı k değerlerini deneyip, en iyi uyumu elde etmek için bir k değerinin bulunması daha iyi sonuç verecektir.

2.9.5. Naive Bayes

Bayes Teoremi, 1812 yılında Thomas Bayes tarafından bulunan koşullu olasılık hesaplama formülüdür. Naive Bayes (NB) sınıflandırıcısının temeli Bayes teoremine dayanır. Algoritma, her eleman için tüm olası sınıfların olasılıklarını hesapladıktan sonra en yüksek olasılık değerine göre sınıflandırma

yapar. Değişkenler arasında güçlü bağımsızlık varsayımlarına sahiptir. Bu, yüksek boyutluluk problemlerinin çözülmesine yardımcı olur. Naive Bayes modeli, karmaşık hiperparametre tahmini olmadığı için basit ve kolayca uygulanabilir.

2.9.6. Karar Ağaçları

Karar ağaçları (KA), sınıflandırma ve regresyon problemleri için, yaygın olarak kullanılan algoritmalarıdır. KA algoritmaları, birçok değişken olan veri setini, birtakım kurallar ile küçük kümelerle bölen yöntemlerdir. Başka bir deyişle, karmaşık olmayan kararlar uygulanarak, çok değişkenden oluşan veri setlerini, az değişkenli gruplara bölen bir yapısı vardır. Bu yöntemin ilk hücrelerine kök olarak adlandırılır. Veri setindeki tüm gözlem değerlerini kökte yer alan duruma göre “evet” ya da “hayır” olarak sınıflandırır. Kök hücrelerinden sonra düğümler yer almaktadır. Düğümler aracılığıyla sınıflandırma yapılmaktadır. Düğüm sayısı ile model karmaşıklığı doğru orantılıdır. Karar ağacı algoritmasının en altında yapraklar bulunur. Yapraklar, sınıfı belirtir. KA, sonuçların yorumu ve anlaşılabilirliği açısından kolay olması nedeniyle tercih edilmektedir.

Aşırı uyum (overfitting) problemini çözmek için, modeldeki parametrelere kısıtlamalar ya da budama (pruning) yöntemleri uygulanabilir. Budama yöntemi, az sayıda nesne içeren yaprak düğümlerinin karar ağacı algoritmasından kaldırılması işlemidir. Karar ağaçlarında, ID3, C4.5, C5.0, CART, CHAID ve QUEST kullanılan çeşitli algoritmalarıdır.

2.9.7. Derin Öğrenme

Derin öğrenme (DÖ), yapay sinir ağlarının katman yapısının genişlemiş biçimi olarak tanımlanabilir. Başka bir deyişle DÖ, yapay sinir ağları modelleri için kullanılan gizli katman sayısının ikiden çok olması veya bu katmanlardaki nöron sayısının fazla olması seçilmesi ile oluşturulan algoritmalarıdır. DÖ modellerinin performansları; veri seti yapısına, derinliğe, seçilen optimizasyon ve aktivasyon yöntemlerinin türüne göre farklılık göstermektedir. Buna ek olarak, optimizasyon algoritmalarının performansı, parametre değerlerine ve sinir ağı yapısına göre değişiklik göstermektedir. Derin öğrenme mimarileri temel olarak doğrusal olmayan

çoklu katmanların yapay sinir ağlarıdır ve girdi özelliklerine, araştırma amaçlarına göre çeşitli türler önerilmiştir. Derin öğrenme mimarileri genel olarak üç gruba ayrılır: Derin sinir ağları, Evrişimli sinir ağları, Tekrarlayan sinir ağlarıdır (41).

Veri seti boyutu büyüdükçe öğrenme oranı da artmaktadır. Ancak veri seti büyüdükçe doğru sınıflama oranı sürekli olarak artmaz, belli bir noktadan sonra doğru sınıflama oranındaki değişim küçük olur. Bu nedenle, aynı başarı performansına boyutu küçük veri seti kullanarak ulaşabiliyorsa ve kapasite alanı önemliyse, veri seti belli bir oranda küçültülerek çalışmaya dahil edilebilir. Eğer veri setindeki gözlem sayısı ve değişken az ise derin öğrenme ile yüksek sınıflama performans sonuçlarına ulaşamayabilir. Veri setindeki gözlem sayısı artırılabilir ya da “transfer öğrenmesi” yöntemleri ile öznelik transferi yapılabilir.

2.10. Sınıflandırma Yöntemleri İçin Performans Ölçüleri

Sınıflandırma yöntemleri için kurulan modellerin performanslarını karşılaştırmak için çeşitli ölçüler vardır. Bunlardan bazıları aşağıda verilmiştir:

- Doğruluk (Accuracy)
- Duyarlılık (Sensitivity)
- Seçicilik (Specificity)
- Pozitif Kestirim Değeri (Positive Predictive Value)
- Negatif Kestirim Değeri (Negative Predictive Value)
- ROC Eğrisi Altında Kalan Alan (Area Under the ROC Curve)
- Düzeltilmiş Doğruluk Oranı (Balanced Accuracy)
- F1-Ölçüsü (F1-Measure)

Yukarıda yer alan performans ölçüleri, Tablo 2.5.’deki 2×2’lik olumsuzluk tablosu dikkate alınarak hesaplanmaktadır.

Tablo 2.5. 2×2'lik Örnek Olumsuzluk Tablosu

Kestirim	Gerçek Durum	
	Pozitif	Negatif
Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

Doğruluk (Accuracy): Doğru sınıflamaların yüzdesi olarak tanımlanır. Doğruluk 0 ile 1 arasında değişmektedir. Bu istatistiğin 1'e yakın değerleri yüksek sınıflandırma performansını göstermektedir.

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (2.14)$$

Duyarlılık (Sensitivity): Gerçek pozitiflerin içinde doğru kestirilen pozitiflerin yüzdesidir. Duyarlılık 0 ile 1 arasında değişir. 1'e yaklaştıkça performansı artar. Bu performans ölçüsü “geri çağırma (recall)” olarak da bilinir.

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (2.15)$$

Seçicilik (Specificity): Gerçek negatiflerin içinde doğru sınıflandırılan negatiflerin yüzdesidir. Seçicilik 0 ile 1 arasında değişir. 1'e yaklaştıkça performansı artar.

$$\text{Seçicilik} = \frac{DN}{YP+DN} \quad (2.16)$$

Pozitif Kestirim Değeri (Positive Predictive Value): Pozitif olarak kestirilenlerin içinde gerçek pozitiflerin yüzdesidir. Pozitif kestirim değeri (PKD) 0 ile 1 arasında değişir. Bu performans ölçüsü aynı zamanda “kesinlik (precision)” olarak da bilinir.

$$\text{Pozitif kestirim değeri} = \frac{DP}{DP+YP} \quad (2.17)$$

Negatif Kestirim Değeri (Negative Predictive Value): Negatif olarak kestirilenlerin içinde gerçek negatiflerin yüzdesidir. Negatif kestirim değeri (NKD) 0 ile 1 arasında değişir. NKD 1'e yaklaştıkça sınıflandırma performansı da artar.

$$\text{Negatif kestirim değeri} = \frac{DN}{DN+YN} \quad (2.18)$$

ROC Eğrisi Altında Kalan Alan (Area Under the ROC Curve): ROC eğrisi değişen sınıflandırma eşik değerlerine göre doğru pozitiflerin sayısının, yanlış pozitiflerin bir fonksiyonu olarak çizilmesiyle oluşur. Dikey eksen de duyarlılık, yatay eksen de 1-seçicilik değerleri bulunur. ROC eğrisi altında kalan alan (AUC), 0 ile 1 arasında değişen değerler olabilir. Bu değer 1'e yaklaşması sınıflandırma performansının da yüksek olduğunu göstermektedir.

$$\widehat{AUC} = \int_0^1 \widehat{ROC}(t) dt \quad (2.19)$$

Düzeltilmiş Doğruluk Oranı (Balanced Accuracy): Düzeltilmiş doğruluk oranı (DDO), duyarlılık ve seçicilik değerlerinin aritmetik ortalamasıdır. DDO ölçüsü, 0 ile 1 arasında değişir. Bu değer 1'e ne kadar yakınsa, sınıflandırma performansı o kadar yüksektir.

$$\text{Düzeltilmiş doğruluk oranı} = \frac{\text{Duyarlılık} + \text{Seçicilik}}{2} \quad (2.20)$$

F1 Ölçüsü (F1-Measure): F1 ölçüsü, duyarlılık ve PKD'nin harmonik ortalamasıdır. F1 ölçüsü 0 ile 1 arasında değişir. F1 ölçüsünün daha yüksek değerleri daha yüksek performansı gösterir.

$$F - \text{ölçüsü} = 2 \times \frac{PKD \times \text{Duyarlılık}}{PKD + \text{Duyarlılık}} \quad (2.21)$$

3. GEREÇ VE YÖNTEM

3.1. MIAS veri seti

Çalışmada açık erişimli Mamografik Görüntü Analiz Derneğinin (An open access Mammographic Image Analysis Society) MIAS veri tabanı kullanılmıştır (42). Bu veri seti 161 hastanın sağ ve sol meme görüntülerini içeren 322 dijitalleştirilmiş mamografi görüntüsünden oluşmaktadır. Görüntüler “pgm” formatındadır. Veri tabanında bulunan mamografi görüntüleri medio-lateral oblique (MLO) tekniğine göre çekilmiştir. Veri tabanındaki tüm görüntüler 1024 x 1024 piksel boyutunda, 200 mikron piksel kalınlığında ve 8 bitlik (256 gri seviye) yapıdadır. Verilen görüntüler 209 normal, 61 iyi huylu (benign), 52 kötü huylu (malign) lezyonların bulunduğu görüntüleri içermektedir. Görüntüler deneyimli radyologlar tarafından değerlendirilmiş, normal olmayan lezyonlar etiketlenmiştir. Etiketleme işleminde memelerin doku yapısı, normal olmayan lezyonların sınıflandırılması, lezyonların koordinatlarının belirlenmesi ve normal olmayan yapıların yaklaşık yarıçapları verilmiştir. MIAS veri tabanı yedi sütundan oluşan bir yapıda verilmiştir.

1. sütunda görüntülere mdbxxx olarak bir referans numarası verilmiştir. Burada xxx, hasta numarasını göstermektedir.

2. sütunda meme dokusunun yapısı verilmiştir. Bu yapı, F Yağlı (Fatty), G Yağlı-glandular (Fatty-glandular), D Yoğun-glandular (Dense-glandular) olarak gösterilmiştir.

3. sütunda normal olmayan kitlelerin sınıflandırılması yapılmıştır. Bu sınıflandırma aşağıdaki gibi yapılmıştır.

CALC: Kalsifikasyon (calcification)

CIRC: İyi tanımlanmış / sınırlı kitleler (well-defined/circumscribed masses)

SPIC: Spiculated kitleler (spiculated masses)

MISC: Kötü tanımlanmış kitleler (ill-defined masses)

ARCH: Yapısal bozulma (architectural distortion)

ASYM: Asimetri (Asymmetry)

NORM: Normal (Normal)

4. sütunda normal olmayan kitlelerin iyi huylu veya kötü huylu olma durumu verilmiştir.

B: İyi huylu (benign)

M: Kötü huylu (malign)

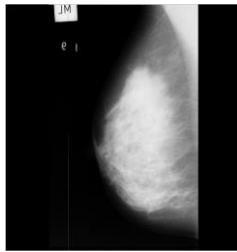
5. sütunda kitlenin x eksenini,

6. sütunda y eksenini olmak üzere piksel olarak koordinatları verilmiştir. Koordinat sisteminin başlangıç noktası sol alt köşe seçilmiştir.

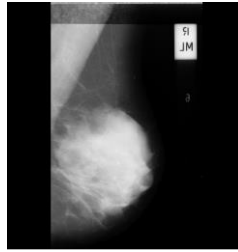
7. sütunda ise normal olmayan kitlelerin piksel olarak yarıçapı verilmiştir. MIAS veri tabanı yapısı Tablo 3.1.' de verilmiştir. MIAS veri tabanından farklı sınıfta örnek mamografi görüntüleri Şekil 3.1.' de verilmiştir.

Tablo 3.1.MIAS veri tabanı

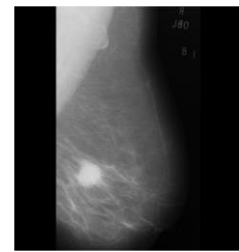
1. sütun	2. sütun	3. sütun	4. sütun	5. sütun	6. sütun	7. sütun
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM	N			
mdb004	D	NORM	N			
mdb005	F	CIRC	B	477	130	30
mdb006	F	NORM	N			
mdb007	G	NORM	N			
.
.
.
mdb322	D	NORM	N			



(A) mdb003



(B) mdb002



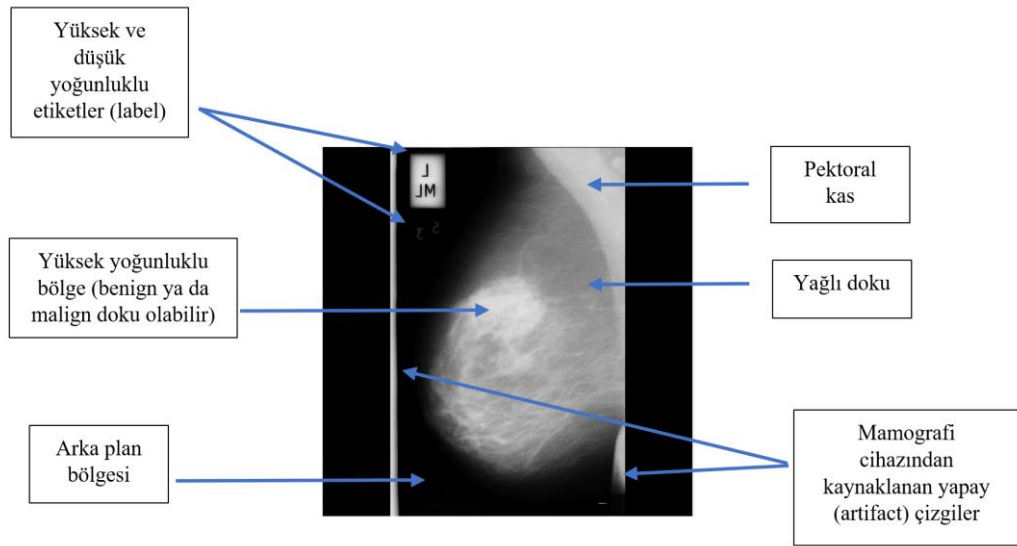
(C) mdb028

Şekil 3.1. MIAS veri tabanından normal (A), iyi huylu (B) ve kötü huylu (C) mamografi örnekleri.

3.2. Yöntem

3.2.1. Ön işleme

MIAS veri tabanında birçok mamografi görüntüsü etiket bilgisi içermektedir. Mamografi cihazından kaynaklanan bu etiketler yüksek yoğunluk değerine sahip oldukları için görüntülerden yanlış sonuçların elde edilmesine neden olabilirler. Bu nedenle, etiketlerin mamografi görüntülerinden temizlenmeleri gerekmektedir. Mamografi görüntüsündeki farklı bileşenler Şekil 3.2.'de gösterilmiştir.



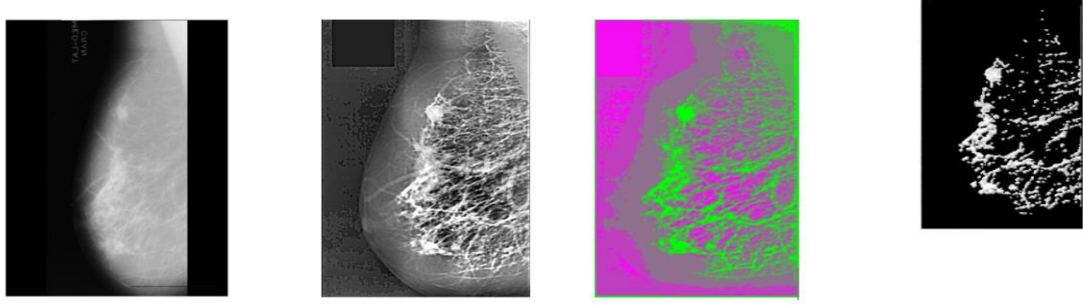
Şekil 3.2. Mamografi görüntüsündeki farklı bileşenler.

Etiket ve yapay çizgilerin görüntüden kaldırılması için önce ön işleme adımında eşikleme ve morfolojik işlemler kullanılmıştır. İlk olarak görüntüler eşikleme yöntemi ile ikili (siyah-beyaz) görüntülere dönüştürülmüştür. Eşik değerinin üzerindeki pikseller beyaz, altındaki pikseller siyah yapılarak ikili (binary) bir görüntü elde edilir. Siyah-beyaz görüntülerde etiket ve rakamları içeren bölgeleri silmek için aşınma ve genişleme morfolojik işlemleri uygulanmıştır. Daha sonra elde edilen bu siyah-beyaz görüntüler, orijinal görüntü üzerinde maske olarak kullanılmıştır. Bu işlemler ile elde edilen görüntülerdeki meme bölgesi dışındaki alanların elenmesi amaçlanmıştır.

Daha sonra mamografi görüntülerinde, görüntü kalitesini artırmak ve segmentasyon sonuçlarını daha iyi hale getirebilmek için farklı ön işleme yöntemleri kullanılmıştır. Bu çalışmada CLAHE, MF ve USM algoritmalarının farklı kombinasyonları ön işleme yöntemi olarak kullanılmıştır. İlk olarak CLAHE (A1) algoritması tek olarak test edilmiştir. Daha sonra MF & CLAHE (A2), MF & USM (A3), CLAHE & USM (A4), MF & CLAHE & USM (A5) gibi filtrelerin olası bazı kombinasyonları oluşturulmuştur. Bu kombinasyonlardan her biri, her görüntüye uygulanmıştır.

3.2.2. Bölütleme ve ROI

Ön işleme yöntemleri ile etiketler ve yapay çizgilerden temizlenen mamografi görüntüleri için bir sonraki önemli adım, uygun bölütleme yöntemi ile görüntülerden pektoral kası vb. temizleyerek ROI'leri çıkarmaktır. Bu çalışmada, ön işleme yöntemlerinin performansını incelemek için k-ortalamlar kümeleme algoritması seçilmiştir. Bu yöntem herhangi bir ön bilgi gerektirmediği ve diğer bölge büyütme tekniklerinden daha iyi olması sebebiyle tercih edilmiştir. Daha sonra, ikinci bir segmentasyon yönteminin de görüntülerden çıkarılan özellikler üzerindeki etkisini görebilmek amacıyla, kenar tabanlı görüntü bölütleme yöntemlerinden Canny kenar algılama algoritması mamografi görüntülerine uygulanmıştır. Görüntü bölütleme algoritmalarının uygulanmasından sonra elde edilen görüntüler, orijinal görüntüler üzerinde maske olarak uygulanarak sonra gerçek zemin (ground truth) görüntüleri elde edilir. Şekil 3.3 ve Şekil 3.4'te ön işleme yöntemleri, A5 algoritması ve k-ortalamlar kümeleme ve Canny kenar algılama bölütleme yöntemleri uygulandıktan sonra mamografi görüntüsünün son hali verilmiştir.



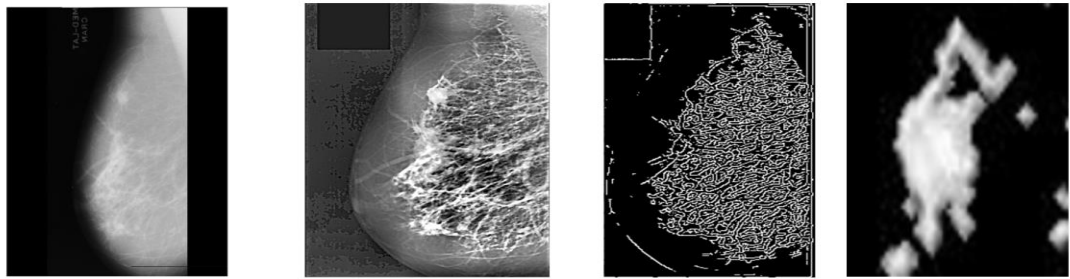
A

B

C

D

Şekil 3.3. mdb023 nolu mamografi görüntüsü (A), Ön işleme yöntemleri ve A5 algoritması ile iyileştirilmiş mamografi görüntüsü (B), k-ortalamlar kümeleme yöntemi uygulanmış mamografi görüntüsü (C), Maskeleye işlemi uygulandıktan sonra gerçek zemin görüntüsü (D).



A

B

C

D

Şekil 3.4. mdb023 nolu mamografi görüntüsü (A), Ön işleme yöntemleri ve A5 algoritması ile iyileştirilmiş mamografi görüntüsü (B), Canny kenar algılama yöntemi uygulanmış mamografi görüntüsü (C), Maskeleye işlemi uygulandıktan sonra gerçek zemin görüntüsü (D).

3.2.3. Öznitelik çıkarımı

Mamografi görüntüleri üzerinde görüntü ön işleme ve bölütleme işlemleri uygulandıktan sonra özellik çıkarımı gerçekleştirilmiştir. Bu çalışmada, özellik çıkarma tekniği olarak GLCM ve GLRLM yöntemlerinden yararlanılmıştır. GLCM yöntemi ile 22 özellik, GLRLM yöntemi ile 11 özellik çıkarılmıştır. Bu özellikler, ROI bölgelerinden, GLCM ve GLRLM matrislerinden dört farklı yönde 0° , 45° , 90° ve 135° çıkarılmıştır. Çıkarılan bu özelliklerin ortalaması alınarak özellik matrisi elde

edilmiştir. Böylece, elde edilen bu özellik matrisi yardımıyla mamografi görüntüleri, sayısal verilere dönüştürülmüştür.

Özellik seçimi için korelasyon matrisi incelenmiştir. Elde edilen 33 özellik arasından 0,90 üzerinde korelasyona sahip olan özelliklerden sadece birisi seçilerek eleyerek bir seçim yapılmıştır. Seçilen her özelliğin tek başına sınıflama performansları, ROC Analizi sonucu elde edilen eğri altında kalan alan (AUC) değerleri ile incelenmiştir.

3.2.4. Sınıflandırma

Görüntü işleme sonucu elde edilen bu sayısal veriler sınıflandırma yöntemlerinde girdi değişkeni olarak kullanılmıştır. Böylece gerçekleştirilen uygulamada, hesaplanan özellikler ile ROI bölgeleri ilk aşamada normal-anormal doku olarak, ikinci aşamada benign-malign doku olarak 7 farklı sınıflandırma yöntemi ile sınıflandırılmıştır. Sınıflandırma işlemi için kullanılan yöntemler:

- (1) Destek Vektör Makinesi (DVM)
- (2) Rastgele Orman (RO)
- (3) Yapay Sinir Ağı (YSA)
- (4) k-En Yakın Komşuluk (k-NN)
- (5) Naive Bayes (NB)
- (6) Karar Ağaçları (KA)
- (7) Derin Öğrenme (DÖ1 ve DÖ2)

Özniteliklerden oluşan veri kümesi %70 ve %30 oranında eğitim ve test seti olacak şekilde ikiye ayrılmıştır. Normal-anormal sınıflamasında her bir ön işleme algoritmasında eğitim setinde 225 ve test setinde 97 mamografi görüntüsü vardır. Benign-malign sınıflamasında her bir ön işleme algoritmasında eğitim setinde 79 ve test setinde 34 mamografi görüntüsü vardır.

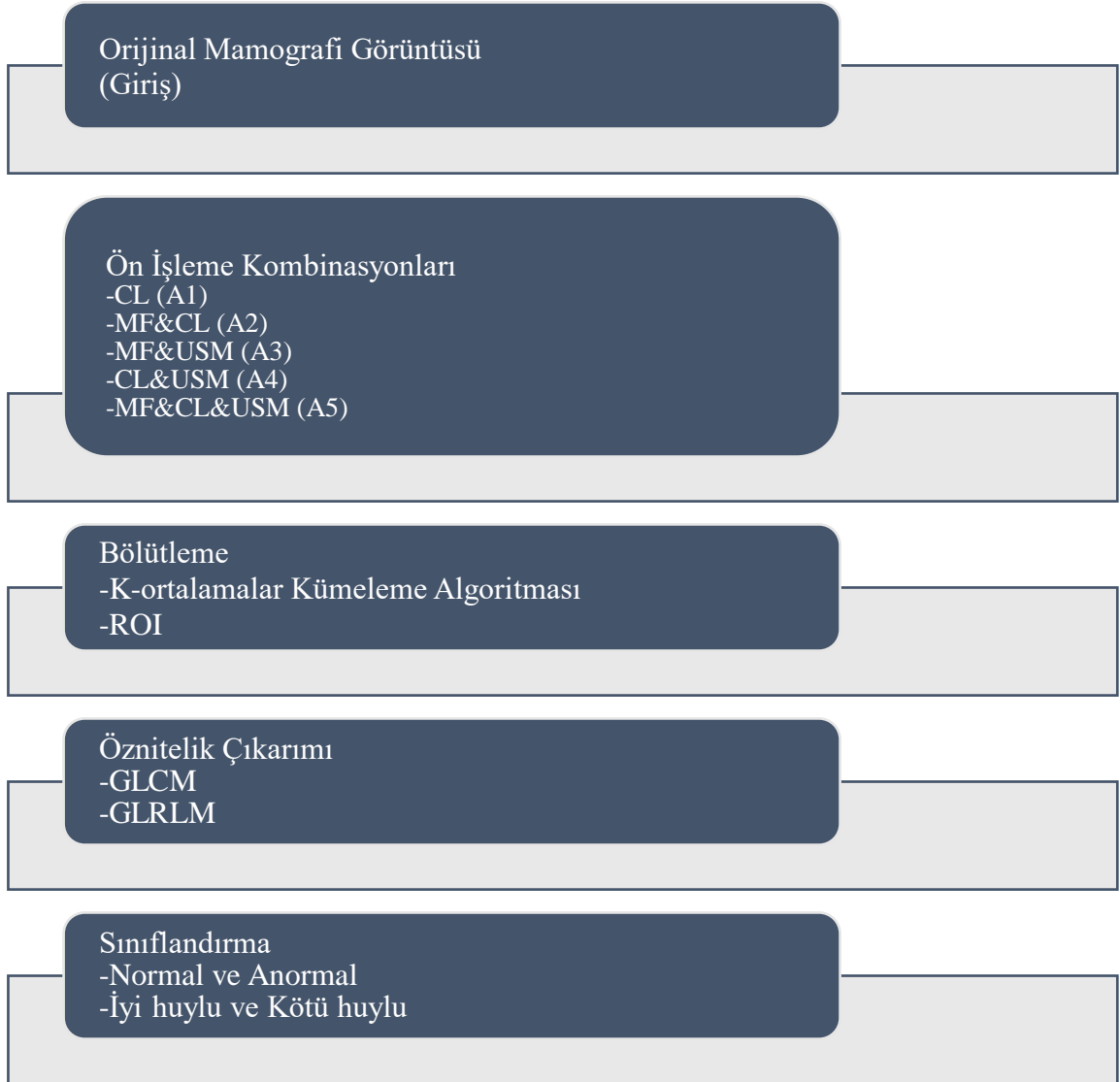
DVM yöntemi için radyal tabanlı çekirdek fonksiyonu kullanılmıştır. RO yönteminde ağaç sayısı (ntree) 500 ve özellik sayısı (mtry) 4 olarak alınarak sınıflandırma performans ölçüleri incelenmiştir. YSA'da iki katman ve her katmanda 10 nöron vardır. Karar ağacı yöntemi olan CART kullanılmıştır. k-NN yönteminde k

değeri 26 olarak alınmıştır. Klasik veri madenciliği yöntemleri ile kurulan modeller için birini dışarıda bırak (leave-one-out) çapraz doğrulama (LOOCV) yöntemi kullanılarak sonuçlar genellenmiştir.

Derin öğrenme algoritmaları için model parametreleri optimize edilerek en uygun olan 2 model seçilmiştir. Oluşturulan derin öğrenme modelleri logloss fonksiyonu incelenerek karşılaştırılmıştır. Logloss fonksiyon değerinin küçük olması istenmektedir. 1. modelde 3 gizli katman (hidden layer) ve her katmanda 20 nöron vardır. 2. modelde 5 gizli katman ve her katmanda 50 nöron vardır.

Sınıflandırma yöntemleri için kurulan bu modellerin performanslarını karşılaştırmak için genel doğruluk oranı, duyarlılık, seçicilik, pozitif kestirim değeri (PKD), negatif kestirim değeri (NKD), ROC eğrisi altında kalan alan (AUC), düzeltilmiş doğruluk oranı (DDO) ve F1 Ölçüsü hesaplanmıştır. Bu çalışmada uygulanan adımlar Şekil 3.5'te verilmiştir.

Ayrıca mamografi görüntülerini değerlendirmede meme doku tipi önemli bir faktör olabilir (32). Bu nedenle 3 farklı meme doku tipi (D, F ve G) için ön işleme algoritmalarına ve sınıflama yöntemlerine göre normal-anormal doku sınıflama sonuçları incelenmiştir.



Şekil 3.5. Ön işleme yöntemlerinin performanslarının değerlendirilmesi.



Şekil 3.6. Bölütleme yöntemlerinin performanslarının değerlendirilmesi.

3.3 Verinin deęerlendirilmesinde kullanılan programlar ve paketler

Bu alıřmada mamografi grntlerinin n iřleme adımımda MATLAB R2017 paket programında Image Processing Toolbox kullanılmıřtır. Grnt bltleme adımımda MedPic ara yz yazılımı kullanılmıřtır. Grntlerin boyutlarını kltme iřlemi iin Fiji-Image J aık eriřimli Java tabanlı grnt iřleme programı kullanılmıřtır. z nitelik ıkarma adımımda MATLAB R2017’de GLCM_Features1 ve GLRL paketlerinin iindeki GLCM.Features1.m, findmaxnum, grayrlprops.m, grayrlmatrix.m, rle_0.m, rle_45.m ve zigzag.m fonksiyonları kullanılmıřtır. Sınıflandırma adımımda eřitli sınıflama yntemlerinin incelenmesi iin R Studio paket programı kullanılmıřtır. R Studio programında “class”, “e1071”, “caret”, “caTools”, “party”, “rpart”, “neuralnet”, “ISLR” ve “h2o” paketleri kullanılmıřtır.

4. BULGULAR

Bu çalışmada 161 hastaya ait 322 mamografi görüntüsü içeren MIAS veri seti kullanılmıştır. MIAS veri tabanındaki bazı mamografi görüntülerini gürültü, etiket ve sayısallaştırma gürültüleri (bazı düz çizgiler) gibi istenmeyen bilgilerden arındırmak için çeşitli ön işleme yöntemleri kullanılmıştır. Bu çalışma dört aşamada gerçekleştirilmiş olup, birinci aşama mamografi görüntülerini iyileştirme, ikinci aşama bölütleme yöntemi ile ROI'leri belirleme, üçüncü aşama ROI bölgelerinden çeşitli istatistiksel ve doku öznitelikleri çıkarma ve son aşama bulunan bu ROI bölgelerinin normal-anormal meme dokusu olup olmadığının ayırımı yapacak 1. sınıflandırma işlemi ve bu ROI bölgelerin benign-malign ayırımı yapacak olan 2. sınıflandırma işlemidir.

Ön işleme adımı olarak, görüntü yumuşatma algoritması olan ortanca filtre, görüntü keskinleştirme filtresi olan keskin olmayan maskeleyme ve görüntülerin kontrastını artırmak için CLAHE algoritmalarının farklı kombinasyonları (A1, A2, A3, A4 ve A5) kullanılmıştır. Bir sonraki adımda, k-ortalamlar kümeleme algoritması ile ROI'ler elde edilmiştir.

Elde edilen ROI bölgelerinden, GLCM ve GLRLM teknikleri kullanılarak toplam 33 özellik elde edilmiştir. Daha sonra bu özelliklerin korelasyon matrisi incelenmiştir. 0,90 ve üzerinde korelasyon gösteren değişkenlerden biri elenerek özellik seçimi yapılmıştır. Seçilen bu özellikler Tablo 4.1.'de özetlenmiştir.

Tablo 4.1. Korelasyon katsayısı kullanılarak seçilen 9 özellik

Özniteliğin çıkarıldığı matris	Öznitelik Adı
GLCM	Otokorelasyon
GLCM	Kontrast
GLCM	Küme önemi
GLCM	Entropi
GLRLM	Short run emphasis
GLRLM	Long run emphasis
GLRLM	Gray-level non-uniformity
GLRLM	Short run low gray-level emphasis
GLRLM	Long run low gray-level emphasis

Tüm özelliklerin tek başına A1, A2, A3, A4 ve A5 algoritmalarında AUC değerlerinin ortalaması 0,746 (0,520-1,00) aralıklarında değişmektedir. Tablo 4.2.'de seçilen 9 özelliğin her bir ön işleme yönteminde AUC değeri verilmiştir. Bazı özelliklerin, bazı ön işleme algoritmalarında tek başına bile çok yüksek performans gösterdiği görülmüştür.

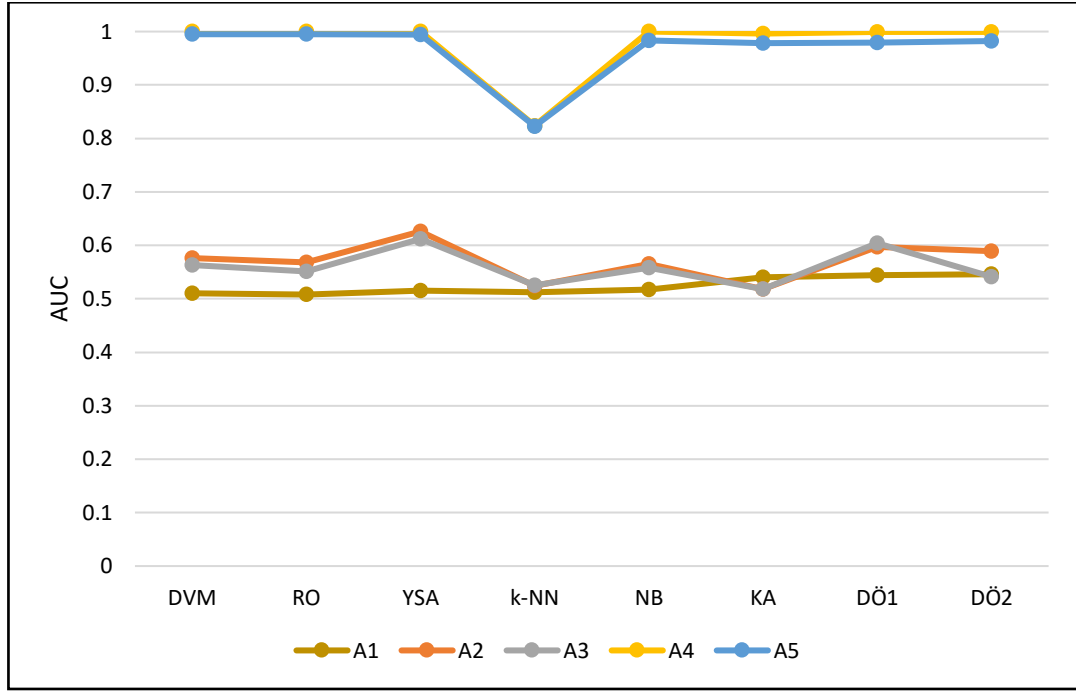
Tablo 4.2. Seçilen özellikler için ön işleme yöntemlerinden elde edilen AUC değerleri

Seçilen Özellikler	Ön işleme yöntemleri				
	CL (A1)	MF&CL (A2)	MF&USM (A3)	CL&USM (A4)	MF&CL&USM (A5)
Otokorelasyon	0,538	0,575	0,521	0,531	0,579
Kontrast	0,587	0,612	0,505	0,551	0,524
Küme önemi	0,543	0,511	0,511	0,600	0,553
Entropi	0,521	0,533	0,512	0,520	0,592
Short run emphasis	0,534	0,578	0,505	0,956	0,946
Long run emphasis	0,502	0,529	0,540	1,000	0,989
Gray leven non-uniformity	0,502	0,481	0,538	0,787	0,714
Short run low gray leven emphasis	0,514	0,559	0,516	0,766	0,732
Long run low gray level emphasis	0,501	0,559	0,603	1,000	0,984

Normal-Anormal Doku Sınıflandırması ve Elde Edilen Sonuçların Değerlendirilmesi

İncelenen veri kümesine ait özelliklerin normal-anormal sınıflama sonuçları Tablo 4.3.'te verilmiştir. CLAHE algoritması tek başına ön işleme yöntemi olarak kullanıldığında, sınıflandırma yöntemlerinin performansları üzerindeki etkisi düşük bulunmuştur. CLAHE algoritmasıyla birlikte ortanca filtre kullanıldığında (A2), (A1) algoritmasına göre sınıflama performansları yükselmiştir. MF&CL (A2) kombinasyonu kullanıldığında AUC değerleri, MF&USM (A3) kombinasyonu ile yakın sonuçlar vermiştir. CL&USM (A4) kombinasyonu ile, tekli ve diğer ikili kombinasyonlara göre daha yüksek sınıflama performansları elde edilmiştir. MF&CL&USM (A5) kombinasyonunda, yani 3 farklı ön işleme yöntemi birlikte kullanıldığında, elde edilen sınıflama performansları A4 kombinasyonu ile çok yakındır. Tablo 4.3.'teki sonuçlara göre, genel doğruluk oranı, duyarlılık, AUC, F1-Ölçüsü gibi ölçülere baktığımızda, DVM, RO, YSA, NB, KA, DÖ1 ve DÖ2 sınıflandırma yöntemlerinin performansları k-NN yöntemine göre daha başarılıdır.

5 farklı ön işleme kombinasyonu ve 6 farklı sınıflama yöntemine ilişkin AUC değerlerini gösteren grafik Şekil 4.1.'de sunulmuştur.



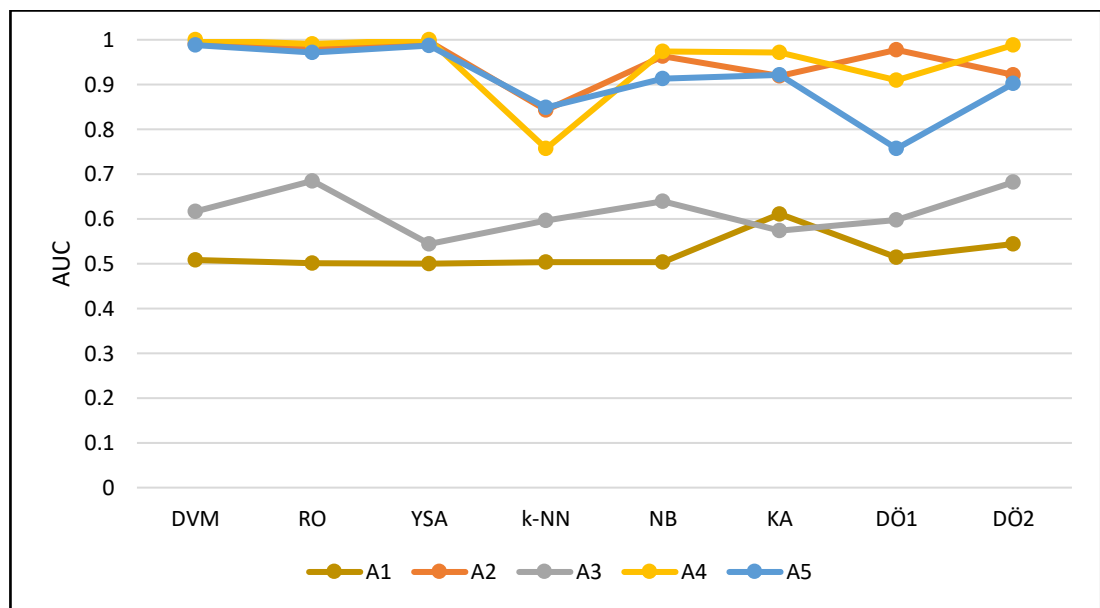
Şekil 4.1. Normal ve Anormal sınıflamasında ön işleme kombinasyonlarının sınıflama yöntemlerinde AUC değerlerinin karşılaştırılması.

Tablo 4.3. Normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Ön işleme	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	CL	0,649	0,035	0,912	0,143	0,689	0,510	0,473	0,055
	MF&CL	0,639	0,281	0,815	0,429	0,697	0,576	0,548	0,339
	MF&USM	0,659	0,069	0,912	0,250	0,697	0,563	0,490	0,108
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	0,995	1,000	1,000
RO	CL	0,608	0,181	0,828	0,353	0,662	0,508	0,505	0,240
	MF&CL	0,649	0,282	0,897	0,647	0,650	0,568	0,589	0,393
	MF&USM	0,659	0,355	0,803	0,458	0,726	0,551	0,579	0,400
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,989	1,000	0,984	0,971	1,000	0,995	0,992	0,985
YSA	CL	0,609	0,609	0,701	0,529	0,670	0,515	0,655	0,536
	MF&CL	0,658	0,658	0,754	0,631	0,650	0,626	0,706	0,624
	MF&USM	0,665	0,665	0,760	0,640	0,730	0,612	0,713	0,633
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,991	0,991	0,990	0,991	0,990	0,994	0,990	0,991
k-NN	CL	0,618	0,304	0,716	0,250	0,768	0,512	0,510	0,274
	MF&CL	0,598	0,210	0,847	0,470	0,625	0,524	0,529	0,291
	MF&USM	0,567	0,133	0,761	0,200	0,662	0,525	0,447	0,160
	CL&USM	0,701	0,552	0,796	0,636	0,734	0,824	0,675	0,591
	MF&CL&USM	0,742	0,697	0,766	0,605	0,830	0,823	0,731	0,648
NB	CL	0,649	0,187	0,877	0,429	0,687	0,517	0,532	0,260
	MF&CL	0,618	0,114	0,903	0,400	0,644	0,565	0,509	0,178
	MF&USM	0,732	0,400	0,881	0,600	0,766	0,558	0,640	0,480
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,979	1,000	0,968	0,944	1,000	0,983	0,984	0,971
KA	CL	0,526	0,054	0,817	0,154	0,583	0,540	0,435	0,080
	MF&CL	0,557	0,289	0,729	0,407	0,614	0,518	0,509	0,338
	MF&USM	0,670	0,457	0,790	0,552	0,721	0,518	0,624	0,500
	CL&USM	1,000	1,000	1,000	1,000	1,000	0,996	1,000	1,000
	MF&CL&USM	0,969	1,000	0,957	0,900	1,000	0,977	0,978	0,947
DÖ1	CL	0,652	1,000	0,651	0,008	1,000	0,544	0,825	0,016
	MF&CL	0,652	0,667	0,652	0,018	0,995	0,597	0,659	0,035
	MF&USM	0,658	0,714	0,657	0,044	0,990	0,604	0,686	0,080
	CL&USM	0,997	0,991	1,000	1,000	0,995	0,999	0,995	0,995
	MF&CL&USM	0,984	0,957	1,000	1,000	0,976	0,979	0,979	0,978
DÖ2	CL	0,658	0,636	0,659	0,062	0,981	0,546	0,647	0,115
	MF&CL	0,686	0,700	0,685	0,186	0,957	0,589	0,692	0,293
	MF&USM	0,661	1,000	0,657	0,035	1,000	0,541	0,828	0,067
	CL&USM	0,997	0,991	1,000	1,000	0,995	0,999	0,995	0,995
	MF&CL&USM	0,978	0,949	0,995	0,991	0,971	0,982	0,972	0,969

Benign-Malign Doku Sınıflandırması ve Elde Edilen Sonuçların Değerlendirilmesi (k-ortalamlar kümeleme görüntü bölütleme algoritması)

Elde edilen özelliklerin benign-malign sınıflama sonuçları Tablo 4.4.'te verilmiştir. Normal ve anormal sınıflandırma sonuçlarında olduğu gibi, CLAHE (A1) algoritmasının tek başına ön işleme olarak kullanılması sınıflandırma yöntemlerinin performanslarının düşük olmasına neden olmuştur. Ancak benign ve malign sınıflamasında MF&CLAHE (A2) kombinasyonu ön işleme yöntemi olarak kullanıldığında sınıflama yöntemlerinin performansları yükselmiştir. MF&USM (A3) kombinasyonu kullanıldığında, CLAHE (A1) algoritmasına göre daha yüksek ancak MF&CLAHE (A2) kombinasyonuna göre daha düşük sınıflama performansları elde edilmiştir. k-NN ve DÖ1 algoritmaları hariç CLAHE&USM (A4) kombinasyonu en iyi sınıflama sonuçlarını vermiştir. MF&CLAHE&USM (A5) kombinasyonu uygulandığında da sınıflama sonuçları yüksektir ama A4'e göre biraz düşüktür. Tablo 4.4.'teki sonuçlara göre, genel doğruluk oranı, duyarlılık, AUC, F1-Ölçüsü gibi ölçülere baktığımızda, DVM, RO, YSA, NB ve KA sınıflama yöntemlerinin performansları k-NN, DÖ1 ve DÖ2 yöntemine göre daha başarılıdır. Benign-malign sınıflandırmasına göre 5 farklı ön işleme kombinasyonu ve 6 farklı sınıflama yöntemine ilişkin AUC değerlerini gösteren grafik Şekil 4.2.'de sunulmuştur.



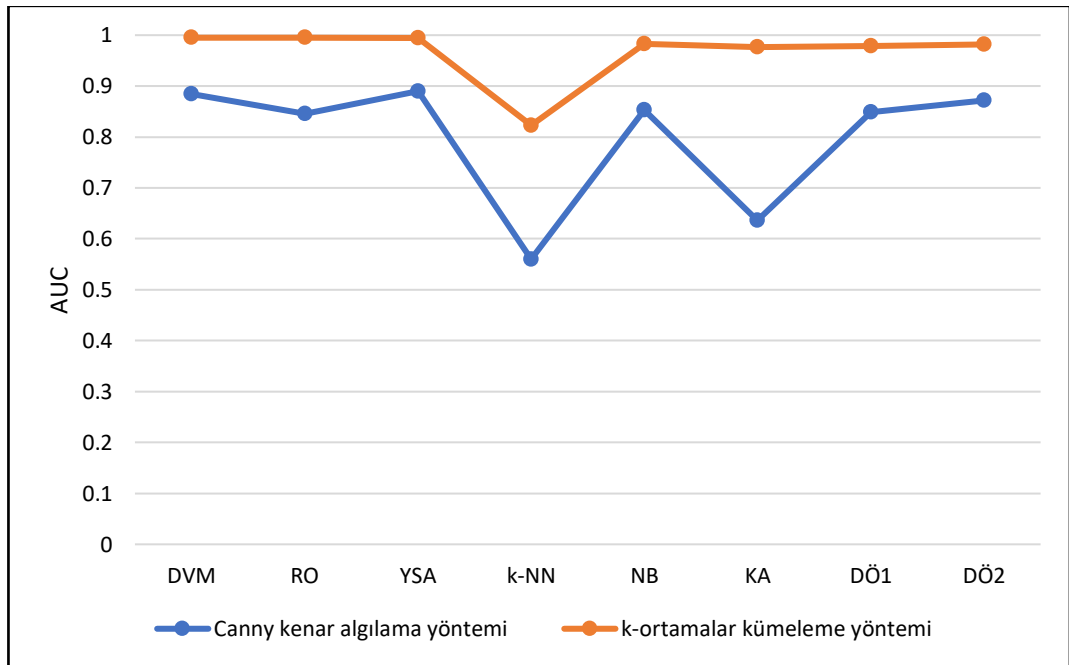
Şekil 4.2. Benign ve Malign sınıflamasında ön işleme kombinasyonlarının sınıflama yöntemlerinde AUC değerlerinin karşılaştırılması.

Tablo 4.4. Benign-malign sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Ön işleme	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	CL	0,470	0,167	0,812	0,500	0,464	0,508	0,489	0,250
	MF&CL	0,970	1,000	0,944	0,941	1,000	0,991	0,972	0,969
	MF&USM	0,647	0,389	0,937	0,875	0,577	0,617	0,663	0,538
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,941	0,937	0,937	0,937	0,944	0,988	0,941	0,937
RO	CL	0,441	0,461	0,428	0,333	0,562	0,501	0,445	0,387
	MF&CL	0,920	0,891	0,804	0,726	0,912	0,982	0,847	0,800
	MF&USM	0,647	0,565	0,722	0,643	0,650	0,685	0,642	0,600
	CL&USM	0,970	0,941	1,000	1,000	0,944	0,991	0,970	0,969
	MF&CL&USM	0,970	1,000	0,947	0,937	1,000	0,971	0,973	0,967
YSA	CL	0,531	0,531	0,430	0,530	0,520	0,500	0,480	0,531
	MF&CL	0,991	0,991	0,980	0,991	0,990	0,996	0,985	0,991
	MF&USM	0,602	0,602	0,725	0,600	0,605	0,544	0,663	0,590
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,935	0,929	0,920	0,929	0,930	0,987	0,925	0,929
k-NN	CL	0,471	0,333	0,545	0,286	0,600	0,503	0,439	0,307
	MF&CL	0,794	0,583	0,909	0,778	0,800	0,843	0,746	0,667
	MF&USM	0,676	0,538	0,762	0,583	0,727	0,596	0,650	0,560
	CL&USM	0,617	0,526	0,733	0,714	0,550	0,757	0,630	0,605
	MF&CL&USM	0,823	1,000	0,667	0,727	1,000	0,849	0,833	0,842
NB	CL	0,441	0,400	0,458	0,235	0,647	0,504	0,430	0,296
	MF&CL	0,912	1,000	0,842	0,833	1,000	0,963	0,921	0,910
	MF&USM	0,647	0,500	0,778	0,667	0,636	0,639	0,639	0,571
	CL&USM	0,971	1,000	0,941	0,944	1,000	0,974	0,971	0,971
	MF&CL&USM	0,706	1,000	0,545	0,545	1,000	0,913	0,773	0,706
KA	CL	0,471	0,636	0,391	0,333	0,692	0,611	0,514	0,437
	MF&CL	0,882	0,923	0,857	0,800	0,947	0,919	0,890	0,857
	MF&USM	0,588	0,600	0,579	0,529	0,647	0,574	0,589	0,562
	CL&USM	0,970	1,000	0,933	0,950	1,000	0,971	0,967	0,974
	MF&CL&USM	0,941	1,000	0,888	0,888	1,000	0,922	0,944	0,941
DÖ1	CL	0,549	1,000	0,544	0,019	1,000	0,514	0,772	0,037
	MF&CL	0,920	0,877	0,964	0,961	0,885	0,977	0,920	0,917
	MF&USM	0,619	0,909	0,588	0,192	0,984	0,597	0,749	0,316
	CL&USM	0,867	1,000	0,803	0,711	1,000	0,909	0,901	0,831
	MF&CL&USM	0,814	0,712	1,000	1,000	0,656	0,757	0,856	0,832
DÖ2	CL	0,566	0,667	0,558	0,115	0,951	0,544	0,612	0,197
	MF&CL	0,911	0,920	0,905	0,885	0,934	0,921	0,912	0,902
	MF&USM	0,655	0,724	0,631	0,404	0,869	0,682	0,677	0,517
	CL&USM	0,982	0,981	0,984	0,981	0,983	0,988	0,982	0,981
	MF&CL&USM	0,849	0,787	0,923	0,923	0,787	0,902	0,855	0,849

Canny Kenar Algılama ve k-Ortalamalar Görüntü Bölütleme Yöntemleri ile Elde Edilen Sonuçların Değerlendirilmesi

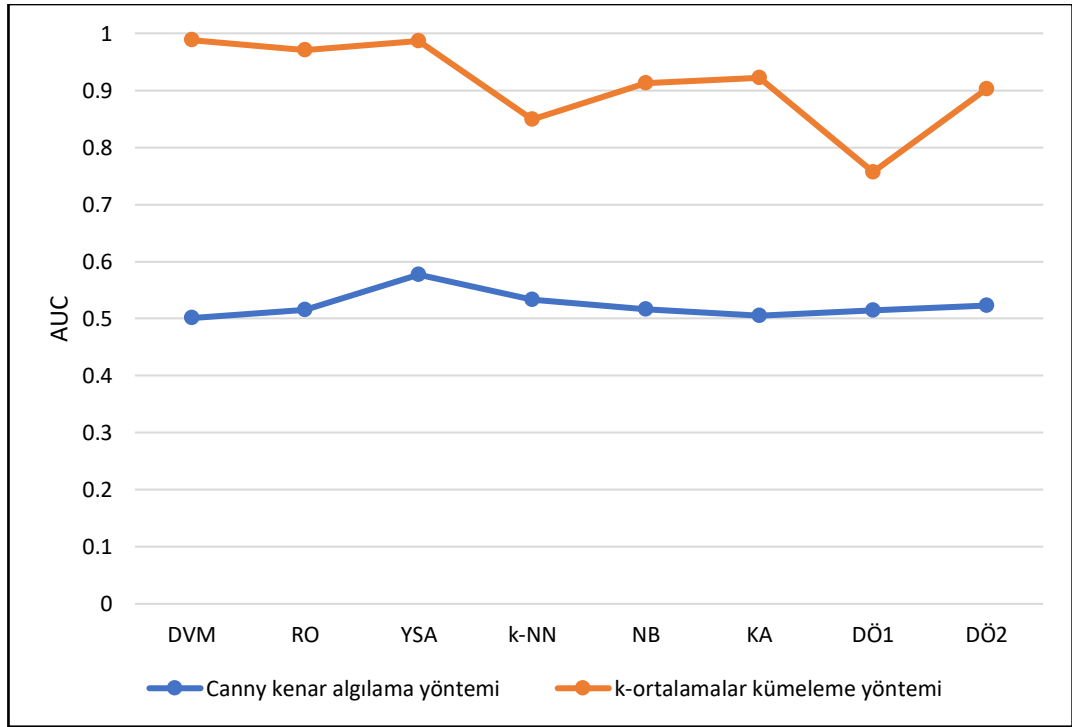
Genellikle k-ortalamlar bölütleme yönteminde en yüksek sınıflama performansını veren ön işleme kombinasyonlarından MF&CL&USM (A5) seçilerek kenar tabanlı bölütleme tekniği olan Canny kenar algılama algoritması uygulanarak ROI'ler elde edilmiştir. Canny kenar algılama bölütleme yöntemi uygulandıktan sonra normal-anormal sınıflama sonuçları Tablo 4.5.'te verilmiştir. Normal-anormal doku sınıflaması için Canny kenar algılama algoritması tüm sınıflama yöntemlerinde k-ortalamlar kümeleme algoritmasına göre düşük performans göstermiştir. Bu sonuçlara göre, normal-anormal sınıflamasında genel doğruluk oranı, duyarlılık, AUC, F1-Ölçüsü gibi ölçülere baktığımızda, DVM, RO, YSA, NB, DÖ1 ve DÖ2 sınıflama yöntemlerinin performansları k-NN ve KA yöntemine göre daha başarılıdır. Normal-anormal sınıflaması için k-ortalamlar kümeleme ve Canny kenar algılama algoritmaları kullanıldığında sınıflama yöntemlerinden elde edilen AUC değerlerini içeren grafik Şekil 4.3.'te sunulmuştur.



Şekil 4.3. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra normal ve anormal sınıflamasında sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması.

Canny kenar algılama yöntemi uygulandıktan sonra elde edilen özelliklerin normal-anormal doku sınıflama sonuçları ise Tablo 4.5.'te verilmiştir. Normal-anormal doku sınıflaması için Canny kenar algılama algoritması tüm sınıflama yöntemlerinde k-ortalamlar kümeleme algoritmasına göre düşük performans göstermiştir.

Canny kenar algılama bölütleme algoritması ile benign-malign doku sınıflamasında normal-anormal doku sınıflamasına göre daha düşük sınıflama performans sonuçları elde edilmiştir. Benign-malign sınıflaması için göre k-ortalamlar kümeleme ve Canny kenar algılama algoritmaları kullanıldığında sınıflama yöntemlerinden elde edilen AUC değerlerini içeren grafik Şekil 4.4.'te sunulmuştur.



Şekil 4.4. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra benign ve malign sınıflamasında sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması.

Tablo 4.5. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra k-ortalamlar kümeleme ve Canny kenar algılama algoritmalarının normal-anormal sınıflama için sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Bölütleme Yöntemi	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	K-ortalamlar	1,000	1,000	1,000	1,000	1,000	0,995	1,000	1,000
	Canny kenar algılama	0,773	0,667	0,828	0,667	0,828	0,884	0,747	0,667
RO	K-ortalamlar	0,989	1,000	0,984	0,971	1,000	0,995	0,992	0,985
	Canny kenar algılama	0,835	0,667	0,922	0,815	0,843	0,846	0,794	0,733
YSA	K-ortalamlar	0,991	0,991	0,990	0,991	0,990	0,994	0,990	0,991
	Canny kenar algılama	0,842	0,720	0,930	0,840	0,850	0,890	0,825	0,840
k-NN	K-ortalamlar	0,742	0,697	0,766	0,605	0,830	0,823	0,731	0,648
	Canny kenar algılama	0,567	0,108	0,850	0,308	0,607	0,560	0,479	0,159
NB	K-ortalamlar	0,979	1,000	0,968	0,944	1,000	0,983	0,984	0,971
	Canny kenar algılama	0,763	0,656	0,815	0,636	0,828	0,853	0,736	0,646
KA	K-ortalamlar	0,969	1,000	0,957	0,900	1,000	0,977	0,978	0,947
	Canny kenar algılama	0,784	0,757	0,800	0,700	0,842	0,636	0,778	0,727
DÖ1	K-ortalamlar	0,984	0,957	1,000	1,000	0,976	0,979	0,979	0,978
	Canny kenar algılama	0,773	0,772	0,778	0,923	0,495	0,849	0,775	0,841
DÖ2	K-ortalamlar	0,978	0,949	0,995	0,991	0,971	0,982	0,972	0,969
	Canny kenar algılama	0,792	0,782	0,829	0,943	0,513	0,872	0,805	0,854

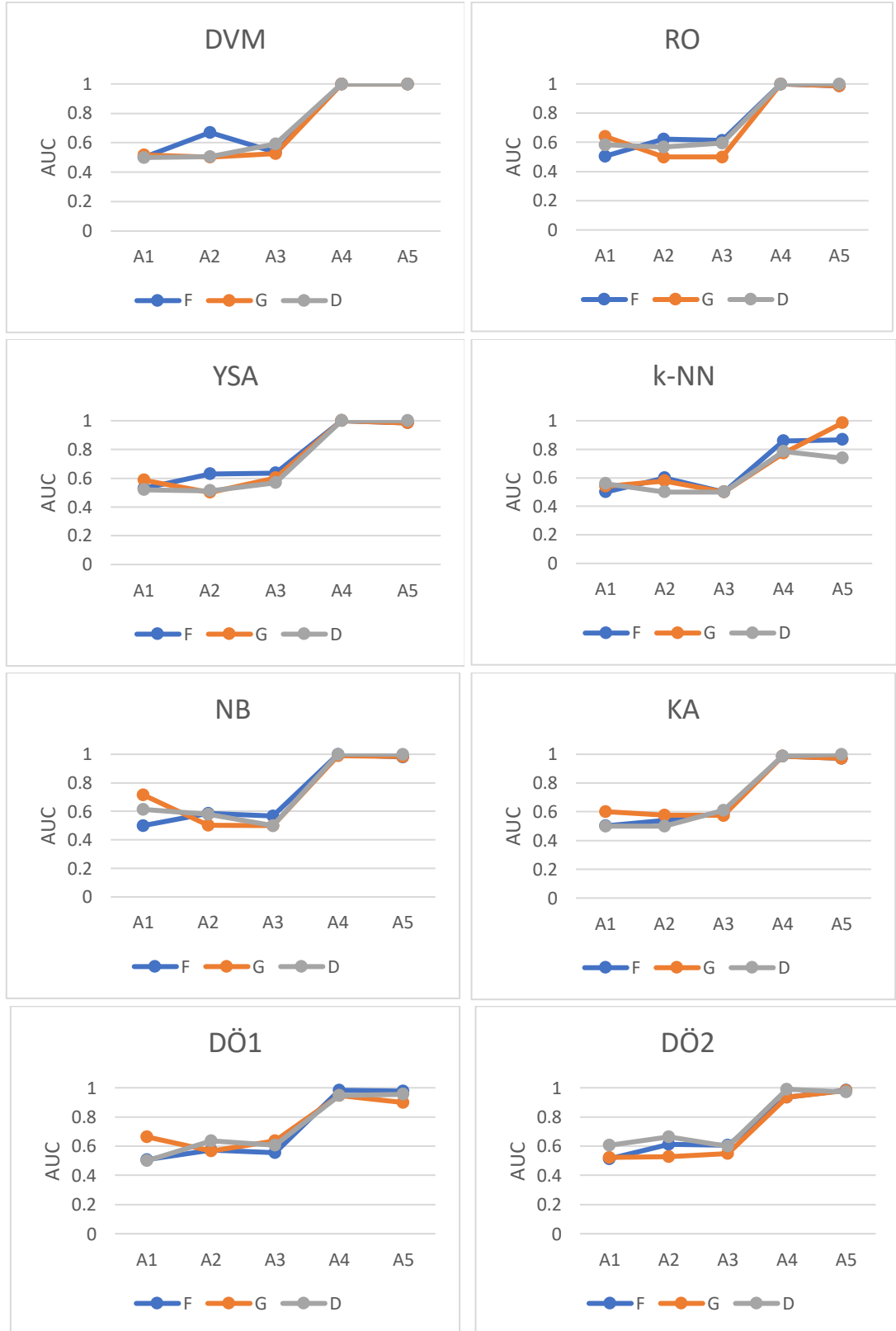
Tablo 4.6. MF&CL&USM ön işleme yöntemleri uygulandıktan sonra k-ortalamlar kümeleme ve Canny kenar algılama algoritmalarının benign-malign sınıflama için sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Bölütleme Yöntemi	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	K-ortalamlar	0,941	0,937	0,937	0,937	0,944	0,988	0,941	0,937
	Canny kenar algılama	0,412	0,357	0,450	0,312	0,500	0,501	0,404	0,333
RO	K-ortalamlar	0,970	1,000	0,947	0,937	1,000	0,971	0,973	0,967
	Canny kenar algılama	0,471	0,400	0,526	0,400	0,526	0,515	0,463	0,400
YSA	K-ortalamlar	0,935	0,929	0,920	0,929	0,930	0,987	0,925	0,929
	Canny kenar algılama	0,558	0,558	0,615	0,553	0,450	0,577	0,586	0,553
k-NN	K-ortalamlar	0,823	1,000	0,667	0,727	1,000	0,849	0,833	0,842
	Canny kenar algılama	0,530	0,545	0,522	0,353	0,706	0,533	0,534	0,428
NB	K-ortalamlar	0,706	1,000	0,545	0,545	1,000	0,913	0,773	0,706
	Canny kenar algılama	0,617	0,467	0,737	0,583	0,636	0,516	0,602	0,519
KA	K-ortalamlar	0,941	1,000	0,888	0,888	1,000	0,922	0,944	0,941
	Canny kenar algılama	0,588	0,714	0,500	0,500	0,714	0,505	0,607	0,588
DÖ1	K-ortalamlar	0,814	0,712	1,000	1,000	0,656	0,757	0,856	0,832
	Canny kenar algılama	0,557	0,551	0,667	0,967	0,077	0,514	0,609	0,702
DÖ2	K-ortalamlar	0,849	0,787	0,923	0,923	0,787	0,902	0,855	0,849
	Canny kenar algılama	0,566	0,554	1,000	1,000	0,058	0,523	0,777	0,713

Tablo 4.3., Tablo 4.4., Tablo 4.5. ve Tablo 4.6.'daki sonuçlara göre genel olarak DÖ1 ve DÖ2 veri madenciliği yöntemlerinden daha yüksek bir performans göstermemiştir.

Doku Tiplerine Göre Sonuçların Değerlendirilmesi

MIAS veri tabanındaki mamografi görüntülerinin meme doku tipine göre, ön işleme kombinasyonlarının normal-anormal sınıflandırma performansları Tablo 4.7., Tablo 4.8. ve Tablo 4.9.'da verilmiştir. Bu sonuçlara göre meme doku tipi yoğun (Dense glandular, D) olan mamografi görüntülerinin sınıflandırma performansları diğer doku tiplerine göre DVM, k-NN ve KA sınıflama yöntemlerinde A2 ön işleme yönteminde daha düşüktür. Meme doku tipi yağlı (Fatty, F) olan mamografi görüntülerinin sınıflandırma performansları diğer doku tiplerine göre DVM, YSA ve k-NN A2 ön işleme yönteminde daha yüksektir. Her bir doku tipine 5 farklı ön işleme kombinasyonu ve 6 farklı sınıflama yöntemi sonuçları, AUC değerlerini bulunduran Şekil 4.5.'te grafikler halinde sunulmuştur. CLAHE&USM (A4) ve MF&CLAHE&USM (A5) ön işleme algoritmalarında tüm yöntemlerin sınıflama performansları daha yüksek elde edilmiştir. Bu sonuçlara göre meme doku tipleri arasında çok farklılık yoktur.



Şekil 4.5. Ön işleme yöntemlerinin doku tipine göre normal-anormal sınıflamasında sınıflama yöntemlerinin AUC değerlerinin karşılaştırılması.

Tablo 4.7. Meme doku tipi D (Yoğun, Dense glandular) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Ön işleme	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	CL	0,676	0,333	0,800	0,375	0,769	0,500	0,567	0,353
	MF&CL	0,647	0,154	0,952	0,667	0,645	0,505	0,553	0,250
	MF&USM	0,764	0,250	0,923	0,500	0,800	0,592	0,586	0,333
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,412	1,000	0,913	0,846	1,000	1,000	0,956	0,917
RO	CL	0,598	0,598	0,700	0,450	0,500	0,583	0,350	0,570
	MF&CL	0,529	0,125	0,654	0,100	0,708	0,568	0,389	0,111
	MF&USM	0,558	0,062	1,000	1,000	0,545	0,596	0,531	0,117
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,991	0,991	1,000	0,991	1,000	1,000	1,000	0,991
YSA	CL	0,598	0,598	0,590	0,526	0,540	0,518	0,594	0,549
	MF&CL	0,670	0,670	0,660	0,634	0,630	0,512	0,665	0,635
	MF&USM	0,696	0,696	0,690	0,668	0,650	0,567	0,693	0,682
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
k-NN	CL	0,598	0,598	0,575	0,570	0,565	0,560	0,586	0,583
	MF&CL	0,764	0,001	0,928	0,001	0,812	0,500	0,464	0,001
	MF&USM	0,676	0,091	0,956	0,500	0,687	0,500	0,523	0,152
	CL&USM	0,765	0,500	0,462	0,500	0,846	0,785	0,673	0,500
	MF&CL&USM	0,768	0,768	0,800	0,765	0,750	0,739	0,784	0,765
NB	CL	0,588	0,333	0,680	0,272	0,739	0,613	0,506	0,300
	MF&CL	0,706	0,500	0,818	0,600	0,750	0,581	0,659	0,545
	MF&USM	0,558	0,154	0,801	0,333	0,607	0,500	0,482	0,210
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
KA	CL	0,529	0,111	0,680	0,111	0,680	0,501	0,395	0,111
	MF&CL	0,323	0,143	0,370	0,055	0,625	0,500	0,256	0,080
	MF&USM	0,676	0,375	0,769	0,333	0,800	0,610	0,572	0,353
	CL&USM	0,991	0,991	1,000	0,991	1,000	0,986	0,995	0,991
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
DÖ1	CL	0,687	1,000	0,684	0,027	1,000	0,500	0,842	0,052
	MF&CL	0,768	0,857	0,755	0,333	0,974	0,638	0,806	0,479
	MF&USM	0,705	0,800	0,700	0,111	0,987	0,605	0,750	0,193
	CL&USM	0,919	0,909	0,924	0,833	0,960	0,947	0,916	0,869
	MF&CL&USM	0,964	0,900	1,000	1,000	0,947	0,955	0,950	0,947
DÖ2	CL	0,696	0,667	0,698	0,111	0,974	0,605	0,682	0,190
	MF&CL	0,750	1,000	0,731	0,222	1,000	0,665	0,865	0,363
	MF&USM	0,687	0,667	0,688	0,055	0,987	0,599	0,677	0,099
	CL&USM	0,982	0,947	1,000	1,000	0,974	0,990	0,974	0,973
	MF&CL&USM	0,902	0,766	1,000	1,000	0,855	0,970	0,883	0,867

Tablo 4.8. Meme doku tipi F (Yağlı, Fatty) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Ön işleme	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	CL	0,594	0,077	0,947	0,500	0,947	0,501	0,512	0,133
	MF&CL	0,625	0,200	0,818	0,333	0,692	0,670	0,509	0,250
	MF&USM	0,656	0,273	0,857	0,500	0,692	0,534	0,565	0,353
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RO	CL	0,468	0,077	0,737	0,167	0,538	0,505	0,407	0,105
	MF&CL	0,594	0,333	0,750	0,444	0,652	0,621	0,542	0,381
	MF&USM	0,594	0,250	0,800	0,428	0,640	0,614	0,525	0,316
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,968	1,000	0,956	0,900	1,000	0,986	0,978	0,947
YSA	CL	0,585	0,585	0,580	0,565	0,560	0,530	0,582	0,570
	MF&CL	0,651	0,651	0,650	0,635	0,645	0,628	0,650	0,633
	MF&USM	0,651	0,636	0,640	0,643	0,650	0,636	0,638	0,639
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,981	0,981	0,980	0,982	0,981	0,985	0,980	0,981
k-NN	CL	0,562	0,363	0,667	0,363	0,667	0,500	0,515	0,091
	MF&CL	0,718	0,300	0,909	0,600	0,740	0,598	0,604	0,400
	MF&USM	0,656	0,083	1,000	1,000	0,645	0,500	0,541	0,153
	CL&USM	0,125	0,800	0,818	0,667	0,900	0,857	0,809	0,726
	MF&CL&USM	0,687	0,400	0,818	0,500	0,750	0,866	0,609	0,444
NB	CL	0,406	0,273	0,476	0,214	0,555	0,500	0,374	0,240
	MF&CL	0,687	0,363	0,857	0,571	0,720	0,585	0,610	0,444
	MF&USM	0,594	0,214	0,888	0,600	0,592	0,566	0,551	0,316
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,968	1,000	0,947	0,928	1,000	0,981	0,937	0,963
KA	CL	0,500	0,643	0,389	0,450	0,583	0,502	0,516	0,529
	MF&CL	0,531	0,444	0,643	0,615	0,473	0,539	0,543	0,516
	MF&USM	0,562	0,357	0,722	0,500	0,591	0,591	0,539	0,417
	CL&USM	0,991	0,991	1,000	0,991	1,000	0,987	0,995	0,991
	MF&CL&USM	0,968	1,000	0,952	0,916	1,000	0,973	0,976	0,956
DÖ1	CL	0,641	1,000	0,634	0,050	1,000	0,508	0,817	0,095
	MF&CL	0,651	0,636	0,652	0,175	0,939	0,572	0,644	0,274
	MF&USM	0,669	0,778	0,659	0,175	0,969	0,555	0,719	0,285
	CL&USM	0,924	0,970	0,903	0,825	0,985	0,983	0,937	0,891
	MF&CL&USM	0,962	0,928	0,984	0,975	0,954	0,976	0,956	0,951
DÖ2	CL	0,641	0,750	0,637	0,075	0,984	0,514	0,693	0,136
	MF&CL	0,632	0,667	0,631	0,050	0,985	0,613	0,648	0,092
	MF&USM	0,669	0,857	0,656	0,150	0,984	0,606	0,757	0,254
	CL&USM	0,915	0,844	0,967	0,950	0,894	0,935	0,906	0,894
	MF&CL&USM	0,972	0,930	1,000	1,000	0,954	0,984	0,965	0,964

Tablo 4.9. Meme doku tipi G (Yağlı glandular, Fatty-glandular) olan mamografi görüntüleri için normal-anormal sınıflama için ön işleme algoritmalarına göre sınıflandırma yöntemlerinin performansı

Sınıflandırıcı	Ön işleme	Performans Ölçüleri							
		Doğruluk	Duy.	Seç.	PKD	NKD	AUC	DDO	F1
DVM	CL	0,594	0,091	0,857	0,250	0,643	0,516	0,474	0,133
	MF&CL	0,625	0,091	0,905	0,333	0,655	0,502	0,498	0,143
	MF&USM	0,656	0,091	0,952	0,500	0,667	0,527	0,521	0,154
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RO	CL	0,719	0,556	0,782	0,500	0,818	0,641	0,669	0,526
	MF&CL	0,594	0,214	0,889	0,600	0,592	0,500	0,551	0,316
	MF&USM	0,718	0,400	0,863	0,571	0,760	0,500	0,632	0,470
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,968	1,000	0,952	0,917	1,000	0,988	0,976	0,956
YSA	CL	0,644	0,644	0,640	0,618	0,620	0,587	0,642	0,617
	MF&CL	0,577	0,577	0,570	0,522	0,565	0,501	0,573	0,534
	MF&USM	0,663	0,663	0,625	0,650	0,640	0,602	0,644	0,860
	CL&USM	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	MF&CL&USM	0,990	0,990	0,991	0,991	0,990	0,985	0,990	0,990
k-NN	CL	0,656	0,091	0,952	0,500	0,667	0,541	0,521	0,152
	MF&CL	0,628	0,628	0,630	0,602	0,625	0,578	0,629	0,608
	MF&USM	0,718	0,100	1,000	1,000	0,701	0,500	0,550	0,182
	CL&USM	0,812	0,714	0,889	0,833	0,800	0,771	0,801	0,769
	MF&CL&USM	0,844	0,714	0,880	0,625	0,916	0,985	0,797	0,666
NB	CL	0,750	0,750	0,750	0,500	0,900	0,714	0,750	0,600
	MF&CL	0,625	0,300	0,772	0,375	0,708	0,502	0,536	0,333
	MF&USM	0,718	0,222	0,913	0,500	0,750	0,500	0,567	0,307
	CL&USM	1,000	0,990	1,000	1,000	1,000	0,990	1,000	0,990
	MF&CL&USM	0,968	1,000	0,960	0,875	1,000	0,983	0,980	0,933
KA	CL	0,625	0,538	0,684	0,538	0,684	0,602	0,611	0,538
	MF&CL	0,437	0,111	0,565	0,091	0,619	0,578	0,338	0,100
	MF&USM	0,468	0,250	0,600	0,272	0,571	0,575	0,425	0,261
	CL&USM	0,990	0,990	1,000	1,000	1,000	0,986	0,995	0,990
	MF&CL&USM	0,968	1,000	0,950	0,923	1,000	0,972	0,975	0,960
DÖ1	CL	0,711	0,667	0,723	0,378	0,895	0,663	0,695	0,482
	MF&CL	0,673	1,000	0,663	0,081	1,000	0,567	0,832	0,149
	MF&USM	0,702	1,000	0,683	0,162	1,000	0,636	0,842	0,279
	CL&USM	0,952	0,881	1,000	1,000	0,925	0,947	0,940	0,937
	MF&CL&USM	0,856	0,867	0,851	0,703	0,940	0,899	0,859	0,776
DÖ2	CL	0,653	0,667	0,653	0,054	0,985	0,523	0,660	0,099
	MF&CL	0,673	0,800	0,667	0,108	0,985	0,528	0,733	0,189
	MF&USM	0,653	1,000	0,650	1,027	1,000	0,549	0,825	0,053
	CL&USM	0,836	0,954	0,805	0,567	0,985	0,935	0,879	0,711
	MF&CL&USM	0,981	0,949	1,000	1,000	0,970	0,981	0,974	0,974

5. TARTIŞMA

Mamografi görüntülerinde, şüpheli bölgelerin bilgisayar destekli sistemler aracılığıyla belirlenmesi son yirmi yılda birçok araştırmacı tarafından araştırılmıştır. Alanyazında yapılan çalışmalarda, şüpheli bölgeleri belirlemede ön işleme adımının, bölütleme ve öznelik çıkarımı aşamalarında çok önemli olduğu vurgulanmaktadır. Çeşitli ön işleme yöntemleri arasından yaygın olarak CLAHE, ortanca filtre, keskin olmayan maskeleme algoritmaları kullanılmaktadır (12,32,43). Nadiren Gauss filtreleme, ortalama filtreleme, sobel gradyan gibi farklı yöntemler de ön işleme algoritması olarak kullanılmıştır (44). Alanyazında çoğu çalışmada ön işleme algoritmaları, tekli ya da ikili şekilde kullanılmıştır. Ancak, ön işleme yöntemleri ile ilgili bu çalışmada olduğu kadar adım adım detaylı bir çalışma bulunmamaktadır. Bu nedenle, bu çalışma, farklı ön işleme yöntemlerine ait farklı kombinasyonların kullanımının sınıflama yöntemlerinin performanslarına etkilerini görmek için planlanmıştır.

Al-Najdawi ve diğ. (12) yaptıkları çalışmada ön işleme yöntemi olarak CLAHE, ortanca filtre ve Gauss filtre algoritmalarını kullanmışlardır. Bu 3 filtreleme yöntemini, tekli ve ikili kombinasyonlarını, MLO ve craniocaudal view (CC) tekniklerine göre çekilen mamografi görüntülerinde uygulamışlardır. Bu çalışmada CLAHE ve MF ön işleme algoritmalarına ek olarak, literatürden farklı olarak USM ön işleme yönteminin de etkisini görmek için, USM ön işleme yöntemi de uygulanarak farklı ön işleme kombinasyonları oluşturuldu.

Tiedeu ve diğ. (44) yaptıkları çalışmada kendi geliştirdikleri CEI (i,j) algoritması ile görüntülerin kontrastlarını artırmışlardır. Orijinal görüntüyü yumuşatmak için Gauss filtreleme yöntemi kullanmışlardır. Bu iki ön işleme yöntemini uyguladıktan sonra, geliştirilmiş 66 mamografi görüntüsünde adaptive eşikleme segmentasyon yöntemini uygulayarak elde edilen ROI bölgelerinden, bir dizi moment tabanlı geometrik özellik çıkarılmıştır. Elde edilen bu özellikler ile, seçicilik %87,77, duyarlılık %100 olarak bulunmuştur. Al-Najdawi ve diğ. (12) yaptıkları çalışmada ön işleme yöntemi olarak CLAHE, ortanca ve Gauss filtre algoritmalarını kullanmışlardır. Performans ölçüsü olarak duyarlılık %96,2 ve seçiciliği %94,4

bulunmuştur. Aswini Kumar Mohanty ve diğ. (45) ROI bölgelerinden GLCM ve GLRLM tekniklerine göre toplam 19 özellik çıkararak C5 karar ağacı algoritmasına göre sınıflama yapmışlardır. Doğruluk %93,6 ve eğri altında kalan alan (AUC) değerini %99,5 olarak elde etmişlerdir. Punitha ve diğ. (46) GLCM ve GLRLM teknikleri yardımıyla elde edilen 45 özellik için ileri beslemeli geri yayılım sinir ağı algoritmasına göre sınıflama yapmışlardır. Duyarlılık %98,1 ve seçicilik %97,8 olarak bulunmuştur. Bu çalışmada normal-anormal doku sınıflamasında CLAHE&USM (A4) ve MF&CLAHE&USM (A5) algoritmalarını, benign-malign doku sınıflamasında ise MF&CLAHE (A2), CLAHE&USM (A4) ve MF&CLAHE&USM (A5) algoritmalarını ve k-ortalamlar kümeleme bölütleme yöntemi kullanıldığında, duyarlılık ve seçicilik değerleri %100 olarak elde edilmiştir. Genelde çalışmalarda duyarlılık ve seçicilik üzerinden yorum yapılmıştır. Ancak bu çalışmada her iki performans ölçüsünü birleştiren AUC değerleri üzerinden yorum yapılmıştır.

Genellikle tüm ön işleme kombinasyonlarında DVM, RO ve YSA yöntemleri k-NN, NB ve KA'na göre daha iyi performans göstermişlerdir. Alanyazında yapılan çalışmalarda, DVM' nin genellikle yüksek sonuçları verdiği görülmüştür (25,47).

Bu çalışmada k-ortalamlar kümeleme bölütleme yöntemi uygulandıktan sonra normal-anormal doku sınıflamasında CLAHE (A1), MF&CLAHE (A2) ve MF&USM (A3), benign-malign doku sınıflamasında CLHAE (A1) ve MF&USM (A3) ön işleme kombinasyonlarında düşük, CLAHE&USM (A4) ve MF&CLAHE&USM (A5) ön işleme kombinasyonlarında ise her iki doku sınıflamasında performans ölçülerinin çok yüksek olduğu görülmüştür. Bu durumun sınıflama yöntemlerine ait aşırı uyum (over-fitting) sorunu olduğu düşünülmektedir. Çünkü GLRLM tekniğine göre elde edilen short run emphasis, long run emphasis ve long run low gray level özelliklerinin tek başına AUC değerleri incelendiğinde oldukça yüksek olduğu görülmüştür. Genel anlamda elde edilen özelliklerin hepsi çok iyi performanslı özelliklerdir. Yapılan araştırmalara göre, görüntü işleme sonucu aşırı yüksek sonuç elde edilmesinin nadirinde olsa mümkün olabileceği düşünülmektedir (48).

Saif ve diğ. (49) Canny kenar algılama ve Otsu eşikleme gibi iki farklı segmentasyon algoritması tekniğini sunmuşlardır. Önerilen algoritmaların etkinliği, tıbbi ve tıbbi olmayan görüntüler için değerlendirilmiştir. Endoskopik görüntülerde

nesnelerin arka plandan net bir ayrımı olmadığı için Canny segmentasyon yönteminin Otsu eşikleme yönteminden daha uygun olduğu belirtilmiştir. Ancak herhangi bir sınıflama yöntemi ile Canny kenar algılama ve Otsu eşikleme yöntemlerinin sınıflama performansları incelenmemiştir. Shrivastava ve diğ. (50) Canny kenar algılama tekniği ile şüpheli alanın yaklaşık olarak tespit edildiğini belirtmişlerdir. Gelecek çalışmalarda, çıkarılan bölgenin şekli ve özellikleri temelinde, kötü huylu veya iyi huylu olup olmadığına göre sınıflandırması için kullanılabileceğini söylemişlerdir. Bu tez çalışması kapsamında literatürden farklı olarak kenar tabanlı görüntü bölütleme tekniği olan Canny kenar algılama yöntemi ile bölge tabanlı görüntü bölütleme tekniği olan k-ortalamlar kümeleme yönteminin sınıflama performansları karşılaştırılmıştır. k-ortalamlar kümeleme algoritmasının Canny kenar belirleme algoritmasından daha yüksek sınıflama özelliğine sahip olduğu bulunmuştur.

Yapılan çalışmalar, meme doku tiplerinin meme kanserinin tespit edilmesinde üzerinde zorluklar olduğunu göstermiştir (51, 52, 53). Meme doku tipine göre kullanılabilecek farklı bölütleme yöntemleri ile ilgili bir çalışma olmakla birlikte (32), farklı meme doku tiplerinin sınıflandırma yöntemlerinin performanslarına katkısının incelendiği herhangi bir çalışma yoktur. Bu nedenle, mamografi görüntülerinin, meme doku tiplerine göre yoğunluk değişikliklerinin incelenmesi risk değerlendirmesi açısından önemli olabilir. Bu tez kapsamında normal-anormal sınıflamasında meme doku tipine göre de değerlendirme yapılmıştır. Bu değerlendirmeye göre meme doku tipi yoğun (Dense glandular, D) olan mamografi görüntülerinin sınıflandırma performansları diğer doku tiplerine göre DVM, k-NN ve KA sınıflama yöntemlerinde MF&CLAHE (A2) ön işleme yönteminde daha düşüktür. Tüm sınıflama yöntemlerinde CLAHE&USM (A4) ve MF&CLAHE&USM (A5) ön işleme kombinasyonları üç farklı meme doku tipi için yüksek sınıflama performans sonuçları elde edilmiştir.

6. SONUÇ VE ÖNERİLER

Mamografi görüntülerinde şüpheli bölgelerin belirlenmesi için görüntü işleme algoritmaları uygulanmadan önce, görüntülerin kalitesinin iyileştirilmesi gerekmektedir. Bu çalışmada, mamografi görüntüleri üzerindeki gürültülerin ve etiket bilgisi gibi istenmeyen nesnelere kaldırılması için farklı ön işleme kombinasyonları uygulanmıştır. Aynı zamanda farklı kombinasyonlarda veri madenciliği yöntemlerinin sınıflama performansları karşılaştırılmıştır.

- Çalışmada yapılan tüm olası sınıflamalarda ön işleme yöntemi olarak CLAHE algoritması tek başına kullanıldığında, diğer ön işleme kombinasyonlarına kıyasla düşük sınıflama performans değerleri elde edilmiştir.
- Normal-anormal doku sınıflamasında, CL&USM (A4) ve MF&CL&USM (A5) algoritmalarının sınıflama performansları ile diğer ön işleme kombinasyonlarına göre daha yüksek elde edilmiştir.
- Benign-malign doku sınıflamasında, MF&CL (A2), CL&USM (A4) ve MF&CL&USM (A5) algoritmalarının sınıflama performansları diğer ön işleme kombinasyonlarına göre daha yüksek elde edilmiştir.
- Sınıflama yöntemleri açısından bakıldığında, performans ölçülerine göre DVM, RO, YSA yöntemlerinin k-NN, NB, KA, DÖ1 ve DÖ2 yöntemlerine göre daha başarılı olduğu söylenebilir.
- Mamografi görüntülerinde görüntü bölütleme yöntemi olarak canny kenar algılama yöntemi uygulandığında k-ortalamlar kümeleme yöntemine göre hem normal-anormal hem de benign-malign doku sınıflaması için daha düşük sınıflama performans sonuçları elde edilmiştir. Bu nedenle, mamografi görüntülerindeki nesnelere arka plandan ayırımı olduğu için kenar tabanlı bir görüntü bölütleme tekniği olan Canny yöntemi yerine, bölge tabanlı bir görüntü bölütleme yöntemi olan k-ortalamlar kümeleme yöntemi, tüm farklı ön işleme ve veri madenciliği algoritmaları açısından değerlendirildiğinde, tercih edilebilir.

- Bu veri setinde veri madenciliği yöntemlerine ek olarak kullanılan iki farklı derin öğrenme modellerinin DVM ve RO algoritmalarına bir üstünlük sağlamadığı söylenebilir. Bunun nedeni olarak çıkarılan öznelik ve mamografi görüntü sayısının az olması gösterilebilir.

- Meme doku tipine göre farklı ön işleme algoritmalarında, veri madenciliği ve derin öğrenme algoritmalarının performansları karşılaştırılmıştır. Meme doku tipine göre sınıflama yöntemlerinin performansları incelendiğinde MF&CLAHE (A2) ön işleme kombinasyonunda ve DVM, k-NN ve KA sınıflama yöntemleri içerisinde yoğun meme dokusunun sınıflama performansının daha düşük olduğu görülmüştür. Meme doku tipi yağlı (Fatty, F) olan mamografi görüntülerinin sınıflandırma performansları diğer doku tiplerine göre DVM, YSA ve k-NN MF&CLAHE (A2) ön işleme yönteminde daha yüksektir. Bunun birinci nedeni olarak, meme doku tipinin yoğun olması durumunda lezyonları belirlemenin zor ve meme doku tipinin yağlı olması durumunda ise lezyonları belirlemenin daha kolay olmasıdır.

Alanyazında, ortanca filtreleme yönteminin, tuz ve biber gürültüsüyle aşırı derecede bozulmuş bir görüntünün, gürültü azaltma başarısının daha yüksek olduğu görülmüştür (12). Bu nedenle, ileride yapılacak çalışmalarda, normal görüntülere tuz ve biber gürültüsü eklenerek ön işleme yöntemleri uygulandıktan sonra normal-anormal doku olarak sınıflama performansları incelenebilir. Çünkü görüntülerdeki gürültü türlerine göre kullanılması gereken ön işleme teknikleri arasında da farklılıklar olabilir (9).

Çalışmada açık erişimli veri tabanından elde edilen sadece 161 kişiye ait 322 mamografi görüntüsü üzerinde görüntü işleme adımları gerçekleştirilmiştir. Kullanılan ön işleme kombinasyonları daha büyük veri setlerinde de uygulanabilir. Daha büyük veri setlerinde, ön işleme yöntemleri uygulandıktan sonra elde edilecek özneliklerin derin öğrenme algoritmaları ile sınıflama başarıları tekrar incelenebilir.

Çalışmada oluşturulan farklı katman ve nöron sayılarına ek olarak, farklı hiperparametreler ile derin öğrenme modelleri oluşturularak normal-anormal ya da benign-malign sınıflama performansları incelenebilir.

MIAS veri tabanındaki mamografi görüntüleri MLO (mediolateral oblique view) tekniğine göre çekilmiştir. Mamografi görüntüleri CC (craniocaudal view) tekniğine göre de çekilebilmektedir. CC tekniğine göre çekilen mamografi görüntülerinde pektoral kas yer almamaktadır. Gelecek çalışmalarda CC tekniğine göre çekilmiş mamografi görüntülerinden elde edilen özellikler ile MLO tekniğine göre elde edilen özelliklerin sınıflandırma yöntemlerinin performansı üzerindeki etkileri incelenebilir.

7. KAYNAKLAR

1. Tsochatzidis L, Costaridou L, Pratikakis L. Deep Learning for Breast Cancer Diagnosis from Mammograms- A Comparative Study. *Journal of Imaging*. 2019;5(3):37-48
2. Mehdy MM, Ng PY, Shair EF, Md Saleh NI, Gomes C. Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. *Open Access Hindawi Computational and Mathematical Methods in Medicine*. 2017;1-15
3. Besl PJ, Jain RC. Segmentation through variable-order surface fitting. *IEEE*. 1988; 10(2):167-192
4. Bandyopadhyay S. Pre-processing of Mammograms Images. *International Journal of Engineering Science and Technology*. 2010;2(11):6753-58
5. Gonzales RC, Woods RE. *Digital Image Processing*. 3th ed. USA: Pearson Education; 2008.
6. Bellanger M. *Digital Signal Processing of Signals: Theory and Practice*. 3th ed: Wiley; 2000
7. Johnson S. *On Digital Photography*. 1st ed. USA: O'Reilly Media; 2006
8. Russ JC, Woods RP. *The Image Processing Handbook*. USA: CRC press; 1995
9. Kovalevsky V. *Modern Algorithms for Image Processing: Computer Imagery by Example Using C#*. New York: Springer Science+Business Media; 2018
10. Singh H. *Practical Machine Learning and Image Processing*. New York: Springer Science+Business Media; 2019
11. Tyagı V. *Understanding Digital Image Processing*. USA: Taylor&Francis Group;2018
12. Al-Najdawi N, Biltawi M, Tedmori S. Mammogram image visual enhancement, mass segmentation and classification. *Aplied Soft Computing*. 2015;35(2015):175-85
13. Patel VK, Uvaıd S, Suthar AC. Mammogram of Breast Cancer detection Based using Image Enhancement Algorithm. *International Journal of Emerging Technology and Advanced Engineering*. 2012;2(8):143-47
14. Maitra IK, Bandyopadhyay S. CAD Based Method for Detection of Breast Cancer. *Oriental Journal of Computer Science and Technology*. 2018;11(3):154-68
15. Qidwai U, Chen CH. *Digital Image Processing: An Algorithmic Approach with MATLAB*. 1st ed. USA: Chapman and Hall/CRC; 2010
16. Kurt B, Nabiyev V. Dijital Mamografi Görüntülerinin Kontrast Sınırlı Adaptif Histogram Eşitleme ile İyileştirilmesi. VII. Ulusal Tıp Bilişimi Kongresi Bildirileri; 14-17 Ekim 2010; Gazimagusa, s.67-78.
17. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ve ark. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*. 1987;39(3):355-68.

18. Suganya R, Rajaram S, Abdullah AS. Big Data in Medical Image Processing. Taylor&Francis Group.
19. Serra J. Image Analysis and Mathematical Morphology. USA. Academic Press, Inc., 6277 Sea Harbor Drive Orlando, FL;1994
20. Shrivastava N, Baharti J. Breast Tumor Detecetion in Digital Mammogram Based on Efficient Seed Region Growing Segmentation. IETE Journal of Research. 2020. <https://doi.org/10.1080/03772063.2019.1710583>
21. Naresh S, Kumari SV. Breast Cancer Detection using Local Binary Patterns. International Journal of Computer Applications. 2015;123(16):6-9
22. Muthuselvi S, Prabhu P. Digital Image Processing Techniques-A Survey. Golden Research Thoughts Journal. 2016;5(11):1-7
23. Preim B, Botha C. Image Analysis for Medical Visualization. Visual Computing for Medicine. 2nded. 2014. Chapter 04, 111-175
24. Rizal A, Hidayat R, Nugroho HA. Modification of Grey Level Difference Matrix (GLDM) for Lung Sound Classification. 4th International Conference on Science and Technology (ICST); 2018; Indonesia.
25. Radovic M, Milosevic M, Ninkovic S, Filipovic N, Peulic A. Parameter optimization of a computer-aided diagnosis system for detection of masses on digitized mammograms. Technol Health Care. 2015;23(6):757-74
26. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics. 1973;3(6): 610–21.
27. Weszka J, Dyer C, Rosenfeld A. A comparative Study of Texture Measures for Terrain Classification. IEEE Transactions on Systems, Man, Cybern. 1976;6(4):269-85.
28. Galloway MM. Texture Analysis Using Gray Level Run Lengths. Computer Graphics and Image Processing. 1975;4:172–179.
29. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. Pattern Recognition. 1996; 29(1): 51–59.
30. Yuan L, Cai-ming WU, Yi Z. Facial expression feature extraction using hybrid PCA and LBP. The Journal of China Universities of Posts and Telecommunications. 2013;20(2):120-124.
31. Mohanaiah P, Sathyanarayana P, GuruKumar L. Image Texture Feature Exraction Using GLCM Approach. International Journal of Scientific and Research Publications. 2013;3(5):1-5
32. Elmoufidi A, Fahssi KE, Andaloussi SJ, Sekkaki A. Automatically Density Based Breast Segmentation for Mamograms by using Dynamic K-means Algorithm and Seed Based Region Growing. Instrumentation and Measurement Technology Conference; July 2015:p.533-538
33. Gardezi SJS, Elazab A, Lei B , Wang T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. Journal of Medical Internet Research. 2019;21(7):1-23
34. Brieman L. Random Forests. Machine Learning. 2001;45(1):5-32

35. Amit Y, Geman D. Shape Quantization and Recognition with Randomized Trees. *Neural Computation*. 1997;9:1545-1588
36. Vapnik VN, Chervonenkis A. A Note on One Class of Perceptrons. *Automation and Remote Control*. 1964;25
37. Cortes C, Vapnik VN. Support vector networks *Machine Learning*. 1995;20:273-297
38. Vapnik VN. *The nature of statistical learning theory*, New York, USA. Springer-Verlag New York, Inc. 1995.
39. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*. 1990;52(1/2):99-115
40. Mehra P, Wah BW. *Artificial neural networks: Concepts and theory*. IEEE Computer Society Press;1992
41. Gillespy T, Rowberg AH. Displaying Radiologic Images on Personal Computers: Image Storage and Compression: Part 1. *Journal of Digital Imaging*. 1993;6(4):197-204
42. The Mini-MIAS Database of Mammograms (Internet). [Erişim tarihi 27 Eylül 2019]. Erişim adresi: <http://peipa.essex.ac.uk/info/mias.html>
43. Ramani R, Vanitha S, Valarmathy S. The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images. *International Journal of Image, Graphics and Signal Processing*. 2013;5(5):47-54
44. Tiedeu A, Daul C, Kentsop A, Graebing A, Wolf D. Texture-based analysis of clustered microcalcifications detected on mammograms. *Digital Signal Processing*. 2012;22(1):124-132
45. Mohanty AK, Senapati MR, Bebartha Swapnasikta, Lenka SK. Texture-based features for classification of mammograms using decision tree. *Neural Computing and Applications*. 2012;23:1011-1017
46. Punitha S, Amuthan A, Joseph KS. Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal*. 2018;3(2):348-358
47. Meenalosini S, Janet J. Computer aided diagnosis of malignancy in mammograms. *European Journal of Scientific Research*. 2012;72(3):360-368
48. Improve OCR Accuracy With Advanced Image Preprocessing (Internet). [Erişim tarihi 20 Mayıs 2021]. Erişim adresi: <https://docparser.com/blog/improve-ocr-accuracy/>
49. Saif JAM, Al-Kubati AAM, Haza AS, Al-Moraish M. Image Segmentation using Edge Detection and Thresholding. *International Arab Conference on Information Technology*. 2013. pp: 473-476.
50. Shrivastava P, Kirar KG. Detection of Tumor in Mammograms Using Canny Edge Detection Technique. *International Journal of Engineering Trends and Technology*. 2014;14(5):213-216.

51. Saidin N, Ngah UK, Amylia H, Shuaib IL. Density Based Breast Segmentation for Mammograms Using Graph Cut and Seed Based Region Growing Techniques. 2nd International Conference on Computer Research and Development. May 2010. pp. 246-250.
52. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal Pattern. *Cancer*. 1976;37(5):2486-92.
53. Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at Mammography. *Radiology*. 1988;167(2):353-8.

8. EKLER

EK-1: Tez Çalışması Orijinallik Raporu

Tıbbi görüntü işleme ile tanı koymada veri madenciliği ve derin öğrenme yöntemlerinin performanslarının incelenmesi

ORJİNALLİK RAPORU

% 7	% 6	% 3	% 3
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	archive.org İnternet Kaynağı	% 1
2	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	% 1
3	Submitted to Thapar University, Patiala Öğrenci Ödevi	<% 1
4	ceng.gazi.edu.tr İnternet Kaynağı	<% 1
5	perweb.firat.edu.tr İnternet Kaynağı	<% 1
6	mafiadoc.com İnternet Kaynağı	<% 1
7	www.researchgate.net İnternet Kaynağı	<% 1
8	S. Kolhe, S. Sawarkar. "Detection of mass and circumscribed mass in mammograms employing radial-basis-function neural networks", Proceedings of the International	<% 1

EK-2: Dijital Makbuz



Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Hanife Avcı
 Assignment title: Tıbbi görüntü işleme ile tanı koymada veri madenciliği ve de...
 Submission title: Tıbbi görüntü işleme ile tanı koymada veri madenciliği ve de...
 File name: tez.docx
 File size: 6M
 Page count: 81
 Word count: 15,179
 Character count: 104,314
 Submission date: 02-Aug-2021 11:33AM (UTC+0300)
 Submission ID: 1626901909



9. ÖZGEÇMİŞ