



**T.C.  
DÜZCE ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**GÖRÜNTÜ İŞLEME TEKNİKLERİYLE MEME KANSERİNİN  
TEŞHİSİ**

**GÜLİZ TOZ**

**DOKTORA TEZİ  
ELEKTRİK-ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ  
ANABİLİM DALI**

**DANIŞMAN  
DOÇ. DR. PAKİZE ERDOĞMUŞ**

**DÜZCE, 2018**

**T.C.**  
**DÜZCE ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**GÖRÜNTÜ İŞLEME TEKNİKLERİYLE MEME KANSERİNİN**  
**TEŞHİSİ**

Güliz TOZ tarafından hazırlanan tez çalışması aşağıdaki jüri tarafından Düzce Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı'nda **DOKTORA TEZİ** olarak kabul edilmiştir.

**Tez Danışmanı**

Doç. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

**Jüri Üyeleri**

Doç. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

Prof. Dr. İbrahim YÜCEDAĞ

Düzce Üniversitesi

Doç. Dr. Cihan KARAKUZU

Bilecik Şeyh Edebali Üniversitesi

Doç. Dr. Devrim AKGÜN

Sakarya Üniversitesi

Dr. Öğr. Üyesi Bahar TERCAN

Düzce Üniversitesi

Tez Savunma Tarihi: 15/05/2018

## BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

15 Mayıs 2018

Güliz TOZ

## TEŐEKKÜR

Doktora öğrenimimde ve bu tezin hazırlanmasında gösterdiği her türlü destek ve yardımlarından dolayı çok değerli hocam Doç. Dr. Pakize Erdoğan'a en içten dileklerle teşekkür ederim. Ayrıca bu çalışma süresince çalışmaya katkıda bulunan değerli hocalarım Prof. Dr. İbrahim Yücedağ ve Doç. Dr. Devrim Akgün'e teşekkür ederim.

Tez çalışmam boyunca değerli katkılarını ve desteğini esirgemeyen sevgili eşim Dr. Öğr. Üyesi Metin Toz'a şükranlarımı sunarım. Ayrıca kızım Berra'ya ve oğlum Muhammed Kayra'ya da sabırlarından ve sevgilerinden dolayı teşekkür ederim.

Bana hayatı ve doğruları öğreten, en güzel şekilde beni yetiştiren ve her türlü fedakârlığı gösteren değerli anne ve babama da teşekkürü bir borç bilirim.

**15 Mayıs 2018**

**Güliz TOZ**

# İÇİNDEKİLER

	<u>Sayfa No</u>
ŞEKİL LİSTESİ.....	VIII
ÇİZELGE LİSTESİ.....	XII
KISALTMALAR.....	XIII
SİMGELER .....	XV
ÖZET .....	XVI
ABSTRACT .....	XVIII
EXTENDED ABSTRACT.....	XX
<b>1. GİRİŞ.....</b>	<b>1</b>
1.1 GENEL BİLGİLER.....	1
1.2 MEME ANATOMİSİ .....	1
1.3 MEME KANSERİ VE İSTATİSTİKLER .....	2
1.4 MAMOGRAFİ .....	5
1.5 BİLGİSAYAR DESTEKLİ TEŞHİS SİSTEMLERİ .....	6
1.6 LİTERATÜR ÖZETİ.....	7
1.7 TEZ ÇALIŞMASI KONUSU VE LİTERATÜRE KATKILAR.....	14
<b>2. MATERYAL VE YÖNTEM.....</b>	<b>17</b>
2.1 KULLANILAN VERİTABANLARI .....	17
2.1.1 MIAS Veri Tabanı .....	17
2.1.2 INBREAST Veri Tabanı .....	18
2.2 ÖNİŞLEMLER .....	18
2.2.1 Meme Bölgesinin Bölütlenmesi.....	19
2.2.2 Mamogramlarda Gürültü Giderme .....	20
2.2.2.1 Uzaysal Alan .....	20
2.2.2.2 Dönüşüm Alanı .....	21
2.3 ŞÜPHELİ BÖLGELERİN (ROI) BÖLÜTLENMESİ .....	39
2.3.1 w-BSAFCM Algoritması .....	40
2.3.1.1 FCM (Fuzzy C-Means) Algoritması .....	41

2.3.1.2 BSA (Backtracking search optimization algorithm) Algoritması .....	43
2.3.1.3 FCM ile Bir Optimizasyon Algoritmasının Birleştirilmesi .....	45
2.3.1.4 BSA ve FCM Algoritmalarının Görüntü Bölütleme İşlemi için Birleştirilmesi.....	46
2.3.1.5 BSAFCM Algoritması için İyileştirme .....	47
<b>2.3.2 Entropi Yöntemleri.....</b>	<b>49</b>
<b>2.3.3 Otsu N Eşikleme.....</b>	<b>53</b>
<b>2.4 ÖZNETLİK ÇIKARIM ALGORİTMALARI .....</b>	<b>59</b>
<b>2.4.1 Şekilsel Öznitelikler:.....</b>	<b>60</b>
<b>2.4.2 Dokusal Öznitelikler:.....</b>	<b>60</b>
2.4.2.1 Gri Seviye Eş-Oluşum Matrisi (GLCM) .....	60
<b>2.4.3 Frekansal Öznitelikler .....</b>	<b>65</b>
<i>İki Boyutlu Dalgacık Dönüşümü ile Öznitelik Çıkarımı .....</i>	<i>66</i>
2.4.3.1 Eğricik Dönüşümü ile Öznitelik Çıkarımı.....	68
<b>2.5 ÖZELLİK AZALTMA ALGORİTMALARI .....</b>	<b>72</b>
<b>2.5.1 Lineer Ayırıcı Analiz(LAA) .....</b>	<b>73</b>
<b>2.6 SINIFLANDIRMA .....</b>	<b>75</b>
<b>2.6.1 Destek Vektör Makineleri (DVM).....</b>	<b>75</b>
2.6.1.1 Doğrusal Ayrılamayan Veriler için Destek Vektör Makineleri .....	76
<b>2.6.2 Yapay Sinir Ağları (YSA) .....</b>	<b>77</b>
2.6.2.1 Yapay Sinir Hücresi (İşlemci Eleman).....	78
2.6.2.2 Yapay Sinir Ağlarında Öğrenme.....	79
2.6.2.3 Yapay Sinir Ağının Yapısı .....	80
2.6.2.4 Yapılarına Göre Yapay Sinir Ağları .....	81
<b>2.6.3 K-En Yakın Komşuluk (K-EK) Sınıflandırıcı .....</b>	<b>81</b>
<b>2.6.4 Sınıflandırma Performansının Değerlendirmesi .....</b>	<b>83</b>
2.6.4.1 Sınıflandırma Başarı Oranı .....	83
2.6.4.2 Duyarlılık ve Özgüllük.....	83
2.6.4.3 ROC(Receiver Operating Characteric) Analizi.....	83
2.6.4.4 k-Kat Çapraz Doğrulama .....	84
<b>3. DENEYSEL ÇALIŞMALAR.....</b>	<b>86</b>
<b>4. BULGULAR VE SONUÇ.....</b>	<b>108</b>
<b>5. KAYNAKLAR.....</b>	<b>111</b>



## ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 1.1. Memenin anatomik yapısı [2].	2
Şekil 1.2. Memenin dış görünümü ve kadranları [3].	2
Şekil 1.3. Kitle örnekleri (a) Kalsifikasyon (b) İyi tanımlanmış kitle (c) Mimari bozulma (d) Spiküle kitle [41].	4
Şekil 1.4. Dijital mamografi cihazı ve görüntüleme birimi [8].	5
Şekil 1.5. Dijital mamogram örnekleri: (a) Sağ memenin kranio kaudal (CC) görünümü (b) Sol memenin CC görünümü (c) Sağ memenin medio lateral eğik görünümü (MLO) (d) Sol memenin MLO görünümü [41].	6
Şekil 1.6. Önerilen BDT sisteminin akış diyagramı.	14
Şekil 2.1. Meme bölgesinin bölütlenmesine ait görüntüler (a) Orijinal mamogram (Mdb012) (b) $\delta = 0.07$ eşik değeri ile ikilik resme dönüştürülmüş görüntü (c) En büyük alana sahip bölgenin filtrelendiği görüntü (d) Art efektlerden temizlenmiş sonuç görüntüsü.	19
Şekil 2.2. 3x3'lük medyan filtre örneği.	21
Şekil 2.3. Medyan filtre uygulama sonucu mamogram görüntüsü (a) Orijinal mamogram (b) 3x3 medyan filtre ile filtrelenmiş mamogram.	21
Şekil 2.4. Ortalama filtre ve gauss filtre uygulama sonucu mamogram görüntüsü	21
Şekil 2.5. Üçüncü seviyeye kadar ayırık dalgacık dönüşümü katsayıları (a) 1. Seviye ADD (b) 2. Seviye ADD (c) 3. Seviye ADD.	22
Şekil 2.6. Bir sinyalin bir seviyeli bileşenlerine ayrıştırılması [45].	23
Şekil 2.7. Tek boyutlu sinyal için eşikleme örneği.	24
Şekil 2.8. Sinyalin yeniden elde edilmesi [45].	24
Şekil 2.9. Biortogonal dalgacık 3.5 için ölçekleme fonksiyonu ve dalgacık.	25
Şekil 2.10. Biortogonal 3.5 ayrışma ve yeniden yapılanma filtreleri.	25
Şekil 2.11. Biortogonal dalgacık ile gürültü giderme (a) Orijinal mamogram (b) Biortogonal dalgacık 3.5 ile gürültü giderilmiş mamogram görüntüsü.	26
Şekil 2.12. 64x64 boyutundaki görüntünün 16 eşit kareye bölünmüş görüntüsü [55].	27
Şekil 2.13. Histogram eşitleme yönteminin örnek bir mamogram üzerinde uygulanması (a) Orijinal mamogram (b) Histogram eşitleme sonucu mamogram.	28
Şekil 2.14. CLAHE uygulaması ve kırpma sınır değerlerinin etkisi (a) Orijinal mamogram (b) 0.01 kırpma sınırlı CLAHE sonucu mamogram (c) 0.1 kırpma sınırlı CLAHE sonucu mamogram (d) 0.05 kırpma sınırlı CLAHE sonucu mamogram.	28
Şekil 2.15. Tüm yönlerde bölünen bölge sayısının CLAHE yöntemine etkisi (a) Bölünen bölge sayısı=8 (b) Bölünen bölge sayısı=16 (c) Bölünen bölge sayısı=32 (d) Bölünen bölge sayısı=64.	29
Şekil 2.16. Farklı difüzyon zamanları ve $\Delta = 0.25$ için anizotropik difüzyon örneği (a) Difüzyon zamanı=2 (b) Difüzyon zamanı=6 (c) Difüzyon zamanı=10 (d) Difüzyon zamanı=16.	30
Şekil 2.17. Farklı difüzyon zamanları ve $\Delta = 0.5$ için anizotropik difüzyon	

	örneği (a) Difüzyon zamanı=2 (b) Difüzyon zamanı=6 (c) Difüzyon zamanı=10 (d) Difüzyon zamanı=16... ..	31
Şekil 2.18.	Pektoral kas çıkartılması amacıyla gerçekleştirilen ön işlemler (a) Orijinal mamogram (b) 512x512 boyutlarına ve 256 renk skalasına indirgenmiş, ayna görüntüsü alınmış mamogram (c) Meme bölgesi bölütlenmiş mamogram (d) Meme kaba sınırı çıkartılmış mamogram. ....	33
Şekil 2.19.	SSEM yönteminin grafiksel gösterimi ve örnek bir uygulama (a) SSEM yöntemi ile sınır tespiti için grafiksel gösterim (b) 6x18 boyutlu kas ve meme bölgesine ait örnek bir gri seviye matrisi (c) SSEM yöntemi sonucunda pektoral kas sınırı işaretlenmiş yeni mamogram görüntüsü. ....	34
Şekil 2.20.	Pektoral kas için <i>Alanmax</i> ve <i>Alanmin</i> örnekleri (a),(b) Orijinal mamogram (c) Pektoral kas bölgesi için <i>Alanmax</i> örneği (d) Pektoral kas bölgesi için <i>Alanmin</i> örneği. ....	35
Şekil 2.21.	SSEM yöntemiyle kaba pektoral kas alanının belirlenmesi (a) Orijinal mamogram (b) SSEM sonucu elde edilen görüntünün ikilik formu (c) Morfolojik işlemler sonrası elde edilen mamogram (d) En büyük iki alanı seçilmiş mamogram (e) Kaba pektoral kas alanı belirlenmiş mamogram. ....	36
Şekil 2.22.	Doğrusal interpolasyon yöntemi ile pektoral kas sınırının düzeltilmesi (a) Orijinal mamogram (b) Kaba pektoral kas sınırları elde edilmiş mamogram görüntüsü (c) Doğrusal enterpolasyon yöntemi ile nihai sınırların elde edildiği mamogram görüntüsü. ....	37
Şekil 2.23.	Önerilen yöntemin örnek bir mamograma uygulanması sonucu elde edilen görüntü (uzmanlar tarafından verilen ground truth değerleri kullanılarak çizilen kas sınırı (kırmızı) ve önerilen yöntem ile bulunan kas sınırı(yeşil)). ....	37
Şekil 2.24.	Aday pektoral kas bölgesi ve sınırının tespitine ait akış diyagramı. ....	38
Şekil 2.25.	Önerilen yöntemin farklı mamogram görüntüleri üzerindeki performansı (a-d) Orijinal mamogram görüntüleri (e-h) Uzmanlar tarafından verilen ground truth değerleri kullanılarak çizilen kas sınırları (kırmızı) ve önerilen yöntem ile bulunan kas sınırları(yeşil). ....	39
Şekil 2.26.	Klasik FCM algoritmasına ait akış diyagramı. ....	42
Şekil 2.27.	BSA algoritmasına ait akış diyagramı. ....	45
Şekil 2.28.	BSAFCM algoritmasının akış diyagramı. ....	47
Şekil 2.29.	w-BSAFCM ile örnek mamogram görüntülerinin bölütlenmesi (a,c,e) Orijinal mamogramlar (b,d,e) w-BSAFCM ile 5 Kümeye bölütlenmiş mamogramlar. ....	49
Şekil 2.30.	Mdb001 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram (b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $toptK=151$ , $Oweşik=112$ , $\alpha=0.4118$ , $\beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $toptHC=198$ , $Oweşik=112$ , $\alpha=0.4118$ ).....	56
Şekil 2.31.	Mdb005 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram (b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $toptK=151$ , $Oweşik=111$ , $\alpha=0.4353$ , $\beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $toptHC=198$ , $Oweşik=111$ , $\alpha=0.4353$ ) .....	57
Şekil 2.32.	Mdb184 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram	

	(b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $toptK=161$ , $Oweşik=120$ , $\alpha=0.3725$ , $\beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $toptHC=208$ , $Oweşik=120$ , $\alpha=0.3725$ ) .....	58
Şekil 2.33.	GLCM için gri seviye giriş matris görüntüsü ve referans piksel açılarının gösterimi(a) GLCM için 4x4'lük bir gri seviye giriş matris görüntüsü (b) Referans piksel için açılarının gösterimi .....	62
Şekil 2.34.	Şekil 2.33 (a)'da verilen giriş görüntüsü için $d=1$ uzaklıkta (a) $\theta=0^\circ$ (b) $\theta=45^\circ$ (c) $\theta=90^\circ$ (d) $\theta=135^\circ$ yönlerinde elde edilmiş GLCM matrisleri .....	62
Şekil 2.35.	Farklı Daubechies dalgacıklarının gösterimi (a) Orijinal sinyal (b) db1 (c) db2 (d) db3 (e) db4 (f) db5(g) db6 (h) db7 (i) db8 .....	67
Şekil 2.36.	Takoz ve eğricik gösterimi (a) Fourier düzleminde takoz gösterimi .....	70
Şekil 2.37.	Farklı ölçeklerde takozların frekans düzleminde ve uzaysal düzlemde gösterimi [106].....	70
Şekil 2.38.	Verinin orijin etrafında sarmalanması .....	71
Şekil 2.39.	Lineer ayırıcı analiz örneği .....	73
Şekil 2.40.	Doğrusal olarak ayrılabilen veri setleri için hiper- düzlemin belirlenmesi [116].....	76
Şekil 2.41.	Doğrusal olarak bölünemeyen veri setleri için hiper düzlem tespiti (a) Doğrusal olarak bölünemeyen veri seti , (b) Doğrusal olarak bölünemeyen veri setleri için hiper-düzlemin tespiti .....	77
Şekil 2.42.	Yapay sinir hücresinin yapısı .....	79
Şekil 2.43.	K-EK için K parametresinin seçimi .....	82
Şekil 2.44.	Performanslarına göre ROC eğrileri .....	84
Şekil 2.45.	Veri setinin 5 kat çapraz doğrulama için test ve eğitim verilerine ayrılması.....	85
Şekil 3.1.	Geliştirilen BDT sistemi ve kullanılan yöntemler .....	88
Şekil 3.2.	Önişlem aşamasına ait mamogram görüntüleri (a) Orijinal mamogram (b) Meme bölgesi bölütlenmiş ve medyan filtre uygulanmış mamogram (c) Pektoral kas çıkartılmış mamogram (d) Biortogonal analiz, anizotropik difüzyon ve kontrast sınırlı adaptif histogram eşitleme işlemleri uygulanmış mamogram .....	89
Şekil 3.3.	MIAS veri tabanından alınan 128x128 kırılmış mamogram görüntüleri anormal(üstte) ve normal(alтта).....	90
Şekil 3.4.	INBREAST veri tabanından alınan 500x500 kırılmış mamogram görüntüleri anormal(üstte) ve normal(alтта) .....	90
Şekil 3.5.	MIAS veri tabanından Mdb010 isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önişlemden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 128x128 elde edilmiş yeni görüntüler .....	91
Şekil 3.6.	MIAS veri tabanından Mdb184 isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önişlemden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 128x128 elde edilmiş yeni görüntüler .....	92
Şekil 3.7.	Inbreast veri tabanından 22613822_45c7f44839fd9e68_MG_R_ML_ANON isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önişlemden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 500x500 elde edilmiş	

	yeni görüntüler. ....	93
Şekil 3.8.	Inbreast veri tabanından Şekil 3.8 Inbreast veri tabanından 50996406_6aba0b402889a16f_MG_R_CC_ANON isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önışlemeden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 500x500 elde edilmiş yeni görüntüler. ....	94
Şekil 3.9.	MIAS 140 mamogram için normal-anormal sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü +LAA. ....	103
Şekil 3.10.	MIAS 70 mamogram için iyi huylu-kötü huylu sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA. ....	103
Şekil 3.11.	INBREAST 106 mamogram için normal-anormal sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA. ....	104
Şekil 3.12.	INBREAST 53 mamogram için iyi huylu-kötü huylu sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA. ....	104

## ÇİZELGE LİSTESİ

	<b><u>Sayfa No</u></b>
Çizelge 2.1. Kitle Teşhisi için öznitelik çıkarım yöntemleri [93].	59
Çizelge 3.1. MIAS veri tabanından alınan mamogramlara ait özellikler.	86
Çizelge 3.2. INBREAST veri tabanından alınan mamogramlara ait özellikler.	87
Çizelge 3.3. MIAS veri tabanı 140 (normal/anormal) mamogram için sınıflama başarısı (%).	95
Çizelge 3.4. Lineer ayırıcı analiz kullanılarak MIAS veri tabanı 140 (normal/anormal) mamogram için sınıflama başarısı (%).	96
Çizelge 3.5. MIAS veri tabanı 70 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).	97
Çizelge 3.6. Lineer ayırıcı analiz kullanılarak MIAS veri tabanı 70 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).	98
Çizelge 3.7. INBREAST veri tabanı 106 (normal/anormal) mamogram için sınıflama başarısı (%).	99
Çizelge 3.8. Lineer ayırıcı analiz kullanılarak INBREAST veri tabanı 106 (normal/anormal) mamogram için sınıflama başarısı (%).	100
Çizelge 3.9. INBREAST veri tabanı 53 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).	101
Çizelge 3.10. Lineer ayırıcı analiz kullanılarak INBREAST veri tabanı 53 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).	102
Çizelge 3.11. Kullanılan tüm mamogramlar için duyarlılık ve özgüllük analizi.	105

## KISALTMALAR

ABD	Amerika Birleşik Devletleri
ACR	Amerikan Radyoloji Koleji
ADC	Anormallik Saptama Sınıflandırıcı
ADD	Ayrık Dalgacık Dönüşümü
AIM	Açısal İkinci Moment
ARC	Mimari bozulma
ASYM	Asimetri
BBHE	İki Boyutlu Histogram Eşitleme
BDT	Bilgisayar Destekli Teşhis
BIRADS	Breast Imaging Reporting and Data Systems
BSA	Back Tracking Search
BSAFCM	Back Tracking Search&Fuzzy C-Means
CAD	Computer Aided Diagnosis
CC	Kraniokaudal Görünüm
CFS	Korelasyona Dayalı Özellik Seçimi
CHSJ	Centro Hospitalar de S. João
CII	Kontrast Artış İndeksi
CIRC	İyi tanımlanmış/sınırlandırılmış kitleler
CLAHE	Kontrast Sınırlı Adaptif Histogram Eşitleme
CNR	Kontrast-Gürültü Oranı
DDSM	Digital Database for Screening Mammography
DICOM	Digital Imaging and Communications in Medicine
DVM	Destek Vektör Makinesi
DWT	Ayrık Dalgacık Dönüşümü
EPI	Kenar Koruma İndeksi
FCM	Fuzzy C-Means
FDCT	Hızlı Ayrık Eğricik Dönüşümü
FFT	Hızlı Fourier Dönüşümü
FOS	Birinci Seviye İstatistiksel Özellikler
FPR	Yanlış Pozitif Oranı(Özgüllük)
GA	Genetik Algoritma
GLCM	Gri Seviye Eş-Oluşum Matrisi
GLDM	Gri Seviye Fark Matrisi
GLRLM	Gray Level Run Length Matrices
GN	Gerçek Negatif
GP	Gerçek Pozitif
HMF	Hibrid Medyan Filtresi
HOHM	Yüksek Mertebeden Hibrit Ortalama
IARC	International Agency for Research on Cancer
IQI	Görüntü Kalitesi İndeksi
K-EK	K-En Yakın Komşuluk
LAA	Lineer Ayrıcı Analizi
LMMSE	Doğrusal Minimum Ortalama-Kare Hata

MAE	Ortalama Mutlak Hata
MD	Maksimum Fark
MIAS	Mammographic Image Analysis Society
MISC	Diğer, kötü tanımlanmış kitleler
MLO	Mediolateral Eğik Görünüm
MMBEBHE	Minimum Ortalama Parlaklık Hatası İki Boyutlu Histogram Eşitleme
MSE	Ortalama Kare Hata
MRMR	Minimum-Redundancy Maximum-Relevancy
MRG	Manyetik Rezonans Görüntüleme
NAE	Normalleştirilmiş mutlak hata
NCI	National Cancer Institute
NLM	Yerel Olmayan Ortalama Seviye
NORM	Normal
NSS	Yapısal Komşuluk Benzerliği
ORNRAD	Yönlü Rician Gürültü Azaltma Anizotropik Difüzyon
PSNR	En Yüksek Sinyal Gürültü Oranı
PSO	Parçacık Sürü Optimizasyonu
QDA	Karesel Ayrımcılık Analizi
RMSHE	Özyineli Ortalama-Ayrı Histogram Eşitleme
ROC	Receiver Operating Characteric
ROI	Region of Interest
SC	Yapısal İçerik
SPIC	Spiküle kitleler
SRN	Sinyal Gürültü Oranı
SSEM	Single Sided Edge Marking(Tek Yönlü Kenar İşaretleme)
SVM	Support Vector Machine
SVM-RFE	Support vector machine recursive feature elimination
TBA	Temel Bileşen Analizi
TFM	Ters Fark Momenti
TPR	Doğru Pozitif Oranı(Duyarlılık)
UCI	University of California,Irvine
USFFT	Eşit Olmayan Aralıklı Hızlı Fourier Dönüşümü
w-BSAFCM	Weighted- Back Tracking Search&Fuzzy C-Means
WHO	World Health Organization
XML	Genişletilebilir İşaretleme Dili
YN	Yanlış Negatif
YP	Yanlış Pozitif
YSA	Yapay Sinir Ağları

## SİMGELER

%	Yüzde
$\Phi$	Meme bölgesini arka plandan ayırmak için kullanılan eşik değeri
$\infty$	Sonsuz
$\partial$	Türev
$\nabla$	Gradyan
$\Delta$	Laplace
$\delta$	Meme bölgesini bölütleme için kullanılan eşik değeri
$\alpha$	Piksellerin yoğunluk benzerlik değeri
$\emptyset$	Pektoral kas ve meme bölgesi arasındaki yoğunluk farkı için eşik değeri
$\varrho$	Derece
$\Sigma$	Toplam sembolü
$\varepsilon$	Üyelik matrisleri arasındaki farklar için belirlenen eşik değeridir.
$\sigma$	Standart Sapma
$\mu$	Ortalama
$\theta$	Açı
$\varphi$	Eğricik

## ÖZET

### GÖRÜNTÜ İŞLEME TEKNİKLERİYLE MEME KANSERİNİN TEŞHİSİ

Güliz TOZ

Düzce Üniversitesi

Fen Bilimleri Enstitüsü, Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. Pakize ERDOĞMUŞ

Mayıs 2018, 120 sayfa

Meme kanseri hem dünyada hem de ülkemizde kadınlar arasında en sık görülen ve yüksek oranlarda ölüme neden olan kanser türüdür. Ülkemizde kanser istatistiklerine bakıldığında meme kanseri tüm kadın kanserlerinin yaklaşık % 25,6'sını oluşturmaktadır. Türkiye'de her 8 kadından birisinde meme kanseri görülmektedir. Meme kanseri yüzünden gerçekleşen ölümlerin azaltılmasında en etkin yöntem erken teşhis ve tedavidir. Erken teşhis için kullanılan en yaygın yöntem mamografidir. Mamografi meme kanseri için tasarlanan özel bir röntgen tekniğidir. Mamografi ile elde edilen görüntüler sayesinde memede elle muayene ile fark edilemeyecek kadar küçük değişiklikler bile belirlenebilmektedir. Mamografilerin yorumlanması uzman radyologlar tarafından yapılmaktadır. Fakat yorgunluk, iş yoğunluğu, insan gözünün kısıtlılığı vb. gibi etkenler radyologların mamografileri yanlış yorumlamasına neden olabilmektedir. Günümüzde, meme kanserinin teşhisinde yanlış ya da eksik yorumlamaya neden olabilecek etkileri azaltmak amacıyla radyoloğa yardımcı olacak otomatik algılama sistemleri geliştirilmektedir. Bu sistemlere Bilgisayar Destekli Teşhis (BDT) sistemleri denilmektedir. BDT yazılımları, memedeki anormallikleri tespit etmek için çeşitli görüntü işleme algoritmaları kullanılmaktadır ve meme kanserinin tanısı ve teşhisinde radyoloji uzmanlarına ikinci bir görüş olarak yardımcı olmaktadır. Nihai karar radyolog tarafından verilmektedir. Bu çalışmada mamogramlarda görülen anormalliklerin tespiti ve sınıflandırılması amacıyla bir BDT sistemi geliştirilmiştir. Bu sistem başlıca beş bölümden oluşmaktadır. Bunlar sırasıyla; ön işleme, bölütleme, öznitelik çıkarımı, öznitelik seçimi ve sınıflandırmadır. Ön işleme aşamasında görüntüyü iyileştirmek ve gürültüyü azaltmak amacıyla medyan filtre, biortogonal dalgacık analizi, anizotropik difüzyon yöntemi, adaptif histogram eşitleme yöntemi kullanılmıştır. Ayrıca ön işleme aşamasının önemli bir kısmını teşkil eden pektoral kasın çıkarılması için bu tez kapsamında geliştirilen Tek Yönlü Kenar İşaretleme (Single Sided Edge Marking-SSEM) isimli yeni bir algoritma kullanılmıştır. Şüpheli bölgelerin tespiti amacıyla Otsu N eşikleme, Havrda & Charvat entropi yöntemleri ile birlikte yine bu tez kapsamında geliştirilen bir görüntü kümeleme algoritması olan w-BSAFCM, kullanılmıştır. Öznitelik çıkarımı için gri seviye eş oluşum matrisi, dalgacık analizi ve eğricik dönüşümü yöntemleri kullanılmıştır. Öznitelik azaltımı aşamasında Lineer Ayırıcı Analiz yöntemi kullanılmıştır. Son olarak, sınıflandırma aşamasında ise yapay sinir ağları, destek vektör makinesi (DVM) ve K- En yakın komşuluk sınıflandırıcıları kullanılmıştır. Tez çalışması kapsamında önerilen yöntem ve

geliştirilen algoritmalar literatürde sıkça kullanılan MIAS veri tabanına ek olarak yeni bir mamogram veritabanı olan INBREAST veri tabanındaki görüntüler üzerinde de test edilmiş ve elde edilen sonuçlar literatürdeki diğer çalışmalarla karşılaştırılmıştır. Yapılan deneysel çalışmalarda mamogram görüntüleri, geliştirilen yöntemle ilk olarak normal ve anormal olarak sınıflandırılmış ardından anormal mamogramlar da kendi aralarında iyi huylu ve kötü huylu olarak sınıflandırılmışlardır. Yapılan tez çalışması için toplam 246 adet mamogram kullanılmıştır. MIAS veritabanından seçilen 70 normal ve 70 anormal mamogram ile INBREAST veri tabanından seçilen 53 normal ve 53 anormal mamogram üzerinde önerilen yöntemlerle yapılan sınıflandırma sonucunda en iyi sonuçlar eğricik analizi ile çıkarılan özneliklerle sağlanmıştır. Eğricik analizi ile çıkarılan özneliklerin lineer ayırıcı analiz ile boyutu azaltılmış ve eğricik analizi ile lineer ayırıcı analizin birlikte kullanıldığı durumda sınıflama başarısı % 100 olarak elde edilmiştir. Ayrıca mamogramların sınıflandırılması amacıyla gri seviye eş oluşum matrisi ve dalgacık analizi ile elde edilen özneliklere ait sonuçlar karşılaştırmalı olarak sunulmuştur.

**Anahtar sözcükler:** Bilgisayar destekli teşhis, Eğricik dönüşümü, Lineer ayırıcı analiz, Mamogram, Yapay sinir ağları.

# ABSTRACT

## DIAGNOSIS OF BREAST CANCER WITH IMAGE PROCESSING TECHNIQUES

Güliz TOZ

Düzce University

Graduate School of Natural and Applied Sciences, Department of Electrical-Electronics  
and Computer Engineering

Doctoral Thesis

Supervisor: Assoc. Prof. Dr. Pakize ERDOĞMUŞ

May 2018, 120 pages

Breast cancer is the most frequent type of cancer that causes death in women both in the world and in our country. When we look at cancer statistics in our country, breast cancer constitutes about 25.6 % of all female cancers. Breast cancer is seen in one out of every 8 women in Turkey. The most effective method for reducing deaths due to breast cancer is early diagnosis and treatment. The most common method used for early diagnosis is mammography. Mammography is a special X-ray technique designed for breast cancer. Images obtained with mammography can be used to determine even small changes that cannot be noticed by manual examination. Interpretation of mammography is done by expert radiologists. But fatigue, workload, lack of human vision, etc. may cause radiologists to misinterpret mammograms. Nowadays, automatic detection systems are being developed to help radiologists in order to reduce the effects that may lead to incorrect or incomplete interpretation in the diagnosis of breast cancer. These systems are called Computer Aided Diagnosis (CAD) system. CAD software uses a variety of image processing algorithms to detect abnormalities in the breast, and as a second opinion to radiology specialists in the diagnosis and diagnosis of breast cancer. The final decision is made by the radiologist. In this study, a CAD system was developed to detect and classify abnormalities in mammograms. This system mainly consists of five sections. These are respectively; pre-processing, segmentation, feature extraction, feature selection and classification. In the pre-processing phase, are used median filter, biorthogonal wavelet analysis, anisotropic diffusion method, adaptive histogram equalization method for image enhancement and noise reduction. In addition, a new method developed within the scope of this thesis has been used to remove the pectoral muscle. In order to detect the suspicious areas, w-BSAFCM which is an image clustering algorithm developed within the scope of this thesis, Otsu N threshold, Havrda & Charvat entropy methods are used together. Gray level co-occurrence matrix, wavelet transformation and curvelet transformation methods are used for feature extraction. Linear Discriminant Analysis method is used in the feature reduction phase. In the classification phase, artificial neural networks, support vector machine (SVM) and K-nearest neighborhood classifiers are used. In addition to the MIAS database frequently used in the literature, the proposed method and developed algorithms for the thesis

study were also tested on images of the INBREAST database and the results obtained were compared with other studies in the literature. In experimental studies, mammogram images were first classified as normal and abnormal by the developed method and then abnormal mammograms were classified as benign and malignant among themselves. A total of 246 mammograms were used for the thesis work. The best results were obtained with the proposed methods and the curvelet transform for feature extraction by using 70 normal and 70 abnormal mammograms selected from the MIAS database, and 53 normal and 53 abnormal mammograms selected from the INBREAST database. The dimensionality was reduced by linear discriminant analysis of the extracted features by the curvelet transform and the classification performance was obtained as 100 % when the linear analysis and curvelet transform were used together. In addition, gray level co-occurrence matrices and wavelet analysis results for the classification of mammograms are presented comparatively.

**Keywords:** Artificial neural networks, Computer assisted diagnosis, Curvelet transform, Linear discriminant analysis, Mammogram.



## **EXTENDED ABSTRACT**

### **DIAGNOSIS OF BREAST CANCER WITH IMAGE PROCESSING TECHNIQUES**

Güliz TOZ

Düzce University

Graduate School of Natural and Applied Sciences, Department of Electrical-Electronics  
and Computer Engineering

Doctoral Thesis

Supervisor: Assoc. Prof. Dr. Pakize ERDOĞMUŞ

May 2018, 120 pages

#### **1. INTRODUCTION**

Breast cancer is one of the most common types of cancer among the women both in the world and in our country. Although high mortality rates in breast cancer, death rates can be reduced by early detection and treatment of the disease. Mammography is one of the most common methods used for early diagnosis. With mammography, even small changes that cannot be noticed by hand can be determined. The interpretation of the mammograms is performed by expert radiologists. Although mammography is an effective tool for fighting breast cancer, accurate reading of mammography is not always possible. Even though approximately 85-90 % of breast cancers are viewed in mammography examinations; because of limitation of the human eye, workload and fatigue etc. approximately 10-15 % of the tumors in the mammograms are missed by radiologists. Nowadays, automatic detection systems are being developed to help radiologists. These systems reduce the cases which can lead to incorrect or incomplete interpretation of breast cancer diagnosis. These systems are called Computer Aided Diagnosis (CAD) system. CAD are systems that use various image processing algorithms to detect abnormalities in the breast and serve as a second opinion to radiologists in the diagnosis of breast cancer. The ultimate decision in breast cancer is given by the radiologist. In this thesis, a CAD system was developed to classify and identify abnormalities in mammograms. This system mainly consists of five sections. These are respectively; pre-processing, segmentation, feature extraction, feature

selection and classification. Many methods have been used to image enhancement and reduce noise. These methods can be listed as, median filter, biorthogonal wavelet analysis, anisotropic diffusion method, adaptive histogram equalization method and Single Sided Edge Marking (SSEM) method. Single Sided Edge Marking (SSEM) method is an important part of the preprocessing stage and it developed in scope of this thesis for the purpose of removal the pectoral muscle from mammograms. In order to detect suspicious areas, w-BSAFCM that is new image clustering algorithm developed in this thesis, Otsu N thresholding and Havrda & Charvat entropy methods were used. Gray level co-occurrence matrix, wavelet transform and curvelet transform are used for feature extraction. Linear Discriminant Analysis method is used in the feature reduction phase. Finally, in the classification phase, artificial neural networks, support vector machine (DVM) and K- nearest neighborhood classifiers have been used.

## **2. MATERIAL AND METHODS**

Two different data bases have been used to test the results of the proposed methods in different databases. These are the INBREAST database which is a current database and not accessible. In order to obtain data from the INBREAST database, necessary correspondences were made to access the database. The second is the MIAS database, which is frequently used in the literature and is open to access. The methods developed within the scope of this thesis study were tested using mammograms from both databases.

## **3. RESULTS AND DISCUSSIONS**

Since pectoral muscle shows similar features with abnormal structures such as mass, it is confused with suspicious regions in CAD systems and makes accurate diagnosis difficult. Since the removal of pectoral muscle from mammograms is important for accurate diagnosis, a new method called SSEM has been developed with this aim in the scope of this thesis. Besides in this thesis, a new image clustering algorithm called w-BSAFCM has been developed. Suspicious region detection has been performed by using Otsu N algorithm, Havrda & Charvat Entropy and w-BSAFCM clustering algorithm in a hybrid way on mammograms. By using these three methods together, the number of ROIs is reduced. So, the workload has also been reduced. Features have been

obtained with three different methods, GLCM, Wavelet and Curvelet analysis and the effect of these features on classification of mammograms both before and after feature reduction were examined and the results are presented comparatively. As far as we research, a study, which comparing of these three methods was not found in the literature. In the literature, coefficients are often used for the features obtained for the curvelet transform and this leads to the dimension problem. In order to reduce the dimension problem, researchers generally use the first hundreds of the largest coefficients for each level. When the feature is mostly obtained by the curvelet analysis, level 4 is selected, which means that the feature matrix size is  $N \times 400$  ( $N$ : the number of mammogram). In this study, instead of using coefficients directly in curvelet transform, some statistical formulation is used for all coefficients for each level of curvelet transform. These are as standard deviation and skewness. A total of  $N \times 162$  ( $N$ : the number of mammograms) features was obtained for the 4th level curvelet analysis so the dimension was further reduced and all the coefficients included in the calculation of the data. In the normal-abnormal classification of mammograms, curvelet transform and linear discriminant analysis were used together and the best results have been achieved by using this combination. For 140 mammograms taken from MIAS database and 106 mammograms taken from INBREAST databases, 100 % successful in the normal-abnormal classification were obtained. Similarly, in the benign-malignant classification of mammograms, 100 % success was achieved as a result of the use of curvelet transform and linear discriminant analysis for the 53 abnormal mammograms taken from INBREAST databases and 70 abnormal mammograms taken from the MIAS database.

#### **4. CONCLUSION AND OUTLOOK**

CAD systems are aiming attracting radiologist attention to the suspicious region by highlighting abnormalities. These systems are intended to serve as a "second reader" to radiologists. This is also the main goal of the developed CAD systems. In the proposed system, a new pectoral muscle removal method and a new image clustering algorithm are developed. The proposed CAD system has been tested on mammograms taken from the INBREAST database and the MIAS database. According to the results obtained, it was seen that the best results in both normal-abnormal and benign-malignant classifications in both databases were obtained as a result of the curvelet transformation

and LAA combination with 100 % success rate. Considering the other studies in the literature and the results obtained in this thesis study, it is seen that curvelet transformation gives better results than GLCM and wavelet analysis in the diagnosis of breast cancer. Finally, two articles and three conference papers were published with the studies realized within the scope of this thesis.



# 1. GİRİŞ

## 1.1. GENEL BİLGİLER

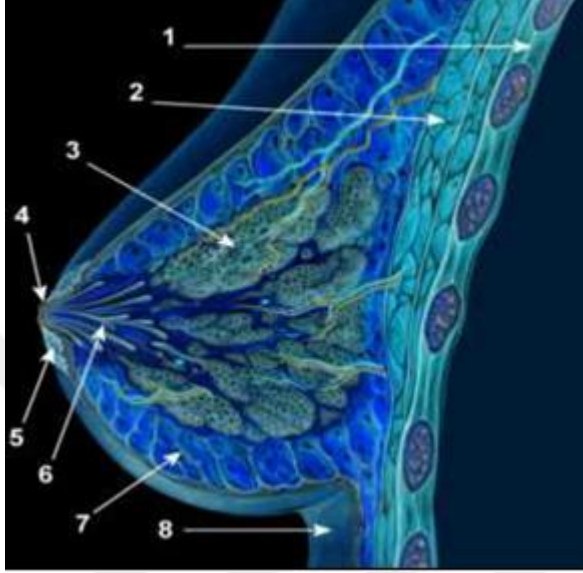
Meme kanseri kadınlarda kanserden kaynaklanan ölüm nedenleri arasında ön sıralarda yer almaktadır. Meme kanserinin belirtileri arasında mikro kalsifikasyonlar, kitleler, mimari bozulma ve ikili asimetri sayılabilmektedir. Bu kanser mamogram, ultrason, MRI ve biyopsinin de aralarında bulunduğu farklı yöntemlerle teşhis edilebilmektedir. Hastalığın erken teşhisi sayesinde ölüm oranı büyük oranda azaltılabilmektedir. Günümüzde makine öğrenmesi ve yapay zekâ tekniklerindeki gelişmeler sayesinde meme kanserinin teşhisine yardımcı olacak farklı bilgisayar destekli sistemler kullanılabilmektedir. Bilgisayar Destekli Tanı (BDT) sistemi, ikinci bir göz olarak kullanılacak uygun maliyetli alternatif bir yöntemdir. Bu bölümde ilk olarak meme anatomisi tanıtılmış daha sonra meme kanseri ile ilgili bazı istatistiksel bilgilere yer verilmiş ardından meme görüntüleme yöntemlerinden en çok kullanılanlardan biri olan mamografi tanıtılmış ve son olarak bilgisayar destekli teşhis sistemleri hakkında bilgi verilmiştir.

## 1.2. MEME ANATOMİSİ

Erişkin kadın memesi, göğüs ön duvarındaki göğüs kaslarının üstünde bulunan ve orta hatta göğüs kemiğinin dış kenarına, yan hatta koltuk altının ön sınırından aşağı doğru uzanan, süt üreten bezlerden oluşan göğüs duvarına bağlı damla şeklinde bir çift organdır [1]. Meme yapısının özünde kas dokusu yoktur fakat göğüs duvarının en büyük kasları olan pektoral kasların üzerine inşa edilmiştir [1]. Her memede her birinde birçok lobül bulunan 15- 20 lop bulunmaktadır ve bu lopların etrafı yağ dokusuyla çevrilidir [1]. Her lopun içinde süt üretimi yapan süt bezleri vardır [1]. Süt bezlerinden çıkan süt kanalları birleşerek daha büyük kanalları oluşturur [1]. Meme başının çevresinde yer alan pigmentli alana areola denir [1]. Memenin anatomik yapısını gösteren bir resim Şekil 1.1'de verilmiştir.

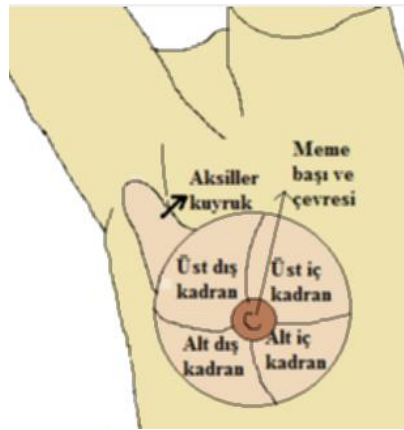
Meme hastalıklarının yerini belirtmek için meme bölgesi kadranslara ayrılmıştır. Her bir

kadranın yer aldığı görüntü Şekil 1.2’de verilmiştir. Bazı durumlarda bir miktar meme dokusu koltuk altına kadar uzanabilir. Buna memenin koltuk altı kuyruğu (aksiller kuyruk) adı verilir. Şekil 1.2’de bu kuyruk görülebilmektedir. Memenin şekli ve büyüklüğü kişiden kişiye hatta aynı kişide her iki memede bile değişiklik gösterebilmektedir [3].



1. Göğüs duvarı
2. Pektoral kas
3. Lobüller
4. Meme başı
5. Areola
6. Laktiferöz kanal
7. Yağ dokusu
8. Cilt

Şekil 1.1. Memenin anatomik yapısı [2].



Şekil 1.2. Memenin dış görünümü ve kadranları [3].

### 1.3. MEME KANSERİ VE İSTATİSTİKLER

Kanser, bir organ veya dokudaki hücrelerin kontrol dışı bölünüp çoğalmasıyla oluşan kötü urlara denir [4]. Normalde vücudun sağlıklı ve düzgün çalışması için hücrelerin büyümesi, bölünmesi ve daha çok hücre üretmesine gereksinim vardır [4]. Bu ölen hücrelerin yenilenmesi ve yaralanan dokuların onarılması için gereklidir [4]. Fakat

bazen bu süreç yanlış işler, yeni hücrelere gerek olmadan hücreler kontrolsüz bölünmeye başlar ve çoğalırlar [4]. Kontrolsüz çoğalan bu hücreler tümörü oluştururlar [4]. Tümörler iyi huylu veya kötü huylu olarak sınıflandırılırlar [4]. İyi huylu tümörler kanser değildir [4]. Bunlar sıklıkla alınır ve çoğu zaman tekrarlamazlar. Kötü huylu tümörlerdeki hücreler anormaldirler, kontrolsüz ve düzensiz bölünürler [4]. Bu tümörler normal dokuların içine sızabilirler ya da dokuları tahrip edebilirler [4]. Tedavi edilmemesi durumunda ciddi rahatsızlıklara yol açabilirler, hatta ölüme neden olabilirler [4].

Meme kanseri de meme dokusundaki hücrelerde meydana gelen kanserlerdir. En çok görülen türü “duktal” kanser olarak anılan meme kanallarından kaynaklanan kanser türüdür [4]. İkinci sırada ise “lobüler” kanserler yer almaktadır ve süt üreten bezlerden kaynaklanır. Ayrıca daha az rastlanan medüller, tübüler, müsinöz gibi türleri de vardır [4]. Mamografik değerlendirmede meme lezyonlarının doğru ve standart bir şekilde tanımlanması, malign patoloji olasılığının belirlenmesi, uzmana mamografi sonucu olarak net bir mesaj verilmesi amacıyla Amerikan Radyoloji Koleji (ACR) tarafından 1993 yılında ‘Meme Görüntüleme Raporlama ve Veri Sistemi’ (Breast Imaging Reporting and Data System -BIRADS) adı altında bir raporlama sistemi geliştirilmiştir [5]. Bu sisteme göre memede karşılaşılan sorunların sınıflandırılması ve raporlama sistemi standart hale getirilmiştir. BIRADS sınıflandırmayı 6 kategoride yapmaktadır. Bunlar;

Kategori 0: Yetersiz tetkik.

Kategori I: Negatif. Memeler simetrik, kitle yok, şüpheli kalsifikasyon yok.

Kategori II: İyi huylu bulgular.

Kategori III: Muhtemelen iyi huylu, kısa aralıkla kontrol önerilir.

Kategori IV: Kötü huylu olma şüphesi. Biyopsi önerilir.

Kategori IV a: Kötü huylu olma şüphesi düşük

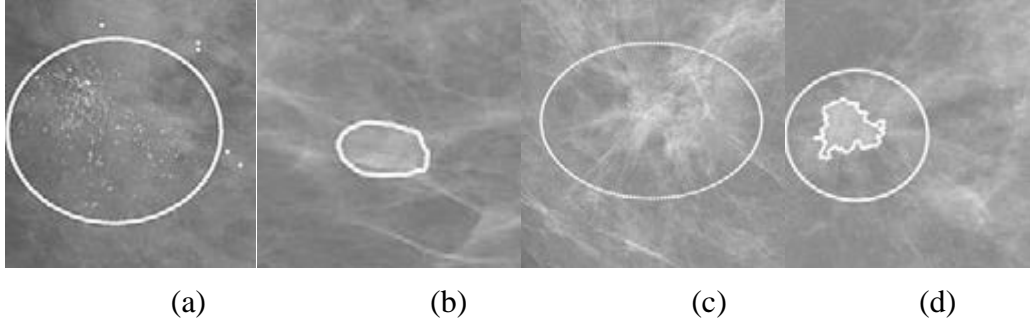
Kategori IV b: Kötü huylu olma için ara şüphe

Kategori IV c: Kötü huylu olma şüphesi orta

Kategori V: Kötü huylu olma şüphesi kuvvetli. Gerekli işlemler yapılmalıdır.

Kategori VI: Bilinen (biyopside kanıtlanmış) kötü huylu tümör [5].

Memede karşılaşılan kitle türlerine, kalsifikasyon, iyi tanımlanmış kitle, spiküle kitle ve mimari bozulma örnek verilebilir. Bu örneklere ait görüntüler Şekil 1.3’te verilmiştir.



Şekil 1.3. Kitle örnekleri (a) Kalsifikasyon (b) İyi tanımlanmış kitle (c) Mimari bozulma (d) Spiküle kitle [41].

Yakın akrabada meme kanseri görülmüş olması, yaşlanma ve yoğun meme dokusuna sahip olunması meme kanserine yakalanmada yüksek derecede risk oluştururken ilk adet yaşı, doğum kontrol hapları, hiç doğum yapmamış olmak, ilk doğum yaşı, emzirme, menopoz yaşı, menopoz sonrası hormon tedavisi, boy ve kilo, hareketsizlik, proliferatif meme hastalıkları, beslenme ve alkol kullanımı da meme kanseri için risk faktörleri olarak görülmektedir [6].

Meme kanserinin erken evrede tespit edilmesi tedavi süreci ve hayatta kalma olasılığı açısından oldukça önemlidir. Erken tanı için kendi kendine yapılan meme kontrolleri, doktor tarafından yılda bir yapılan meme muayeneleri ve meme görüntüleme yöntemleri kullanılmaktadır [4].

Meme kanseri Amerikan Kanser Derneği'nin (American Cancer Society) 2012 verilerine göre kadınlarda görülen tüm kanserlerin yaklaşık % 30'unu oluşturmaktadır [2]. Ülkemizde ise bu oran % 25,6'dır. Türkiye'de kadınlarda kanserden en sık ölüm nedeni de yine meme kanseridir (% 17,6) [8]. IARC (International Agency for Research on Cancer) Globocan 2008 verilerinde Türkiye'de meme kanseri sıklığı % 28,3 olarak hesaplanmıştır [8]. Dünya Sağlık Örgütü'nün (WHO) bildirdiğine göre her yıl dünya genelinde 1,2 milyon kişi meme kanseri tanısı almakta ve 500 000 kadarı bu hastalık nedeniyle hayatını kaybetmektedir [2]. ABD genelinde 2008 yılı için hesaplanan yeni meme kanseri vaka sayısını kadınlarda 182.460, erkeklerde 1990 vaka; beklenen ölümleri ise kadınlarda 40.480, erkeklerde 450 vaka olarak National Cancer Institute (NCI), tarafından rapor edilmiştir. NCI 2000-2004 verilerine göre her sekiz kadından birinde hayatı boyunca meme kanseri gelişebileceği ve her otuz kadından birinin meme kanseri nedeniyle öleceği tahmin edilmektedir [2].

#### 1.4. MAMOGRAFİ

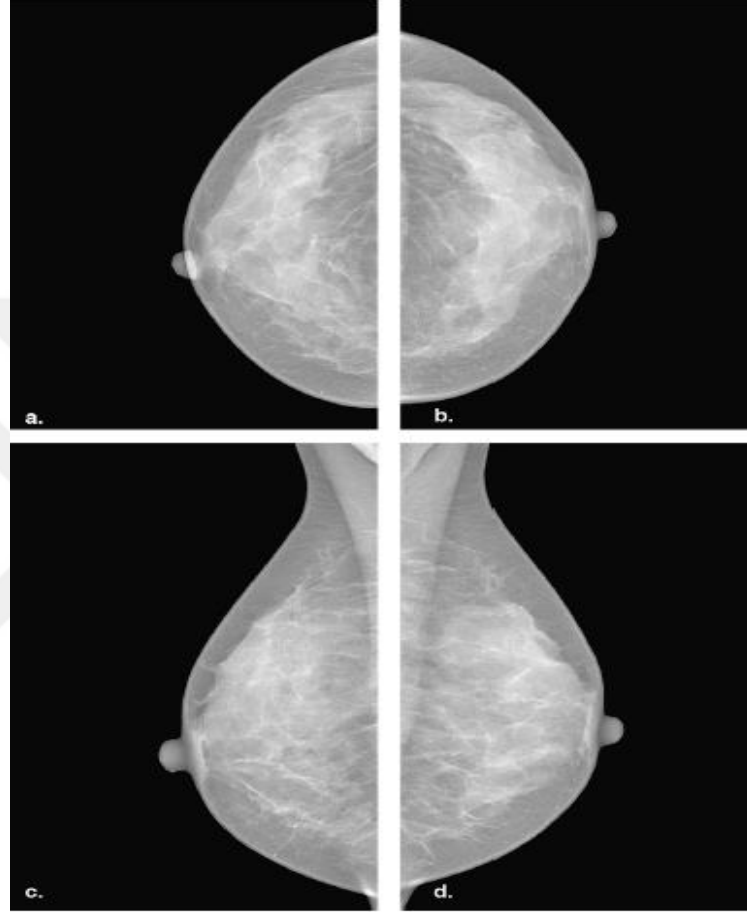
Meme hastalıklarının tanısında ve meme kanserinin erken saptanması amacıyla radyologlar tarafından farklı görüntüleme yöntemleri kullanılmaktadır. Bunlar; mamografi, ultrasonografi, bilgisayarlı tomografi ve manyetik rezonans görüntüleme (MRI) olarak sıralanabilir. Bunlardan mamografi günümüzde yaygın olarak kullanılan etkinliği kanıtlanmış temel meme görüntüleme yöntemidir. Mamografinin bu kadar yaygın olmasının nedeni düşük maliyet ve düşük radyasyon dozunda kaliteli görüntüleme performansdır. Görüntüleme işlemi esnasında, meme sıkıştırılarak kalınlığı homojen bir şekilde azaltılmış olur böylece saçılan radyasyonun azaltılması, görüntü keskinliğinin artırılması sağlanır [9]. Ayrıca, meme sabit olduğundan harekete bağlı görüntü bulanıklığı en aza indirgenir [9]. Görüntülenen meme dokusu miktarı artar, nesne ve reseptör mesafesi kısaltılarak geometrik bulanıklığın azaltılmasını sağlar [9]. Sıkıştırma sistemi ince sert plastikten yapılan kompresyon plağı ve kontrol mekanizmasından oluşur [9]. Mamografinin meme kanserini saptamadaki duyarlılığı % 63–98 oranındadır fakat bu oran doku yoğunluğu fazla olan memelerde % 30-48'lere düşmektedir [2]. Özellikle doku yoğunluğu fazla olan (dens) memelerde mamografik duyarlılığın az olması nedeniyle ultrasonografi ve manyetik rezonans görüntüleme (MRG) gibi ek tanı yöntemlerinin de kullanılması gerektiği belirtilmektedir [2]. Dijital mamografi cihazı ve görüntüleme birimi şekil 1.4'te verilmiştir.



Şekil 1.4. Dijital mamografi cihazı ve görüntüleme birimi [8].

Tanı ve tarama amaçlı olmak üzere iki çeşit mamografi incelemesi vardır. Tarama amaçlı mamografi, 40 yaş ve üzeri risk grubunda olan kadınlara hastalık ortaya

çıkmadan önce erken teşhis amacıyla yapılır [10]. Erken teşhis tedavinin etkili olması ve kanser nedeniyle gerçekleşen ölümlerin azaltılmasında oldukça önemlidir [8]. Tarama amaçlı görüntüleme meme içinde anormal yapıların fark edilebilmesi amacıyla iki farklı yönde görüntü elde edilir. Bunlar her iki meme için medio lateral eğik görünümü (MLO) ve kranio kaudal (CC) görünümüdür (Şekil 1.5) [9].



Şekil 1.5. Dijital mamogram örnekleri: (a) Sağ memenin kranio kaudal (CC) görünümü (b) Sol memenin CC görünümü (c) Sağ memenin medio lateral eğik görünümü (MLO) (d) Sol memenin MLO görünümü [41].

### 1.5. BİLGİSAYAR DESTEKLİ TEŞHİS SİSTEMLERİ

Mamografi meme kanseri ile mücadelede etkin bir araç olmasına rağmen mamografilerin doğru okunması her zaman mümkün olamamaktadır. Meme kanserlerinin yaklaşık % 85-90'ı mamografi incelemeleri sonucu elde edilirken, görüntülenen mamografilerdeki tümörlerin yaklaşık % 10-15'i çeşitli sebeplerden radyologlar tarafından gözden kaçırılmaktadır [11]. Özellikle genç ve/veya meme

dokusu yoğun kadınlarda kitlelerin görülebilmesi zor olmaktadır [12]. Ayrıca radyolojik anormalliklerin yaygınlığı, dikkati kaybetme ve yüksek miktarda mamografinin okunduğu taramalarda radyoloğun yorgunluğu, algısını bozabilmekte ve tümörlerin gözden kaçmasına neden olabilmektedir [13]. Bilgisayar Destekli Teşhis Sistemleri, anormal yapıları ön plana çıkararak radyoloğun dikkatini şüpheli bölgelere çeken, radyologlara "ikinci bir okuyucu" olarak hizmet etmesi amaçlanan, sistemler olarak tanımlanabilir [12]. BDT sistemleri ile elde edilen şüpheli bölgelerin görsel olarak kontrolü radyolog tarafından yapılır, radyolog eğitim ve tecrübelerine dayanarak nihai kararı verir [12].

Bu tez çalışmasında da meme kanserinin mamografi görüntüleri üzerinden, görüntü işleme teknikleri ile tespiti için bir BDT çalışması gerçekleştirilmiştir. Bu amaçla yapılan literatür çalışması bir sonraki bölümde sunulmuştur.

## 1.6. LİTERATÜR ÖZETİ

Meme kanserinin tespiti ve teşhisine yönelik çalışmalara bakıldığında genel olarak beş aşamadan söz edilebilir. Bunlar; önışleme, bölütleme, öznitelik çıkarımı, öznitelik seçimi ve sınıflandırmadır. Önışleme aşaması ise kendi içinde gürültünün kaldırılması, yumuşatma, görüntü iyileştirme, filtreleme ve histogram eşitleme gibi amaca en uygun başlangıç görüntüsüne ulaşmayı sağlayan işlemlerden oluşan aşamadır [14]. Gürültü görüntü üzerinde farklı nedenlerden dolayı oluşan istenmeyen sinyaller olarak tanımlanabilmektedir. Gürültü farklı nedenlerden kaynaklandığı için gürültünün giderilmesi amacıyla farklı algoritmalar geliştirilmiştir. Filtreleme bu amaçla kullanılan temel olarak alçak geçiren ve yüksek geçiren olmak üzere ikiye ayrılan işlemlerdir. Filtreler doğrusal olan veya olmayan, zamanla değişen veya değişmeyen, analog veya dijital, ayrık veya sürekli, pasif veya aktif şekilde olabilmektedir [14]. Yumuşatma işlemi de yine gürültünün azaltılması, daha az bulanıklık ve görüntü içindeki gereksiz piksellerin kaldırılmasıyla ilgili işlemlerdir [14]. Bu amaçla yapılan çalışmalara bakıldığında Vijikala ve Dhas [15] mamogram görüntülerinde gürültü giderilmesi amacıyla Hibrit Medyan Filtresi (HMF), Yönlü Rician Gürültü Azaltma Anizotropik Difüzyon (ORNRAD), Yüksek Mertebeden Hibrit Ortalama (HOHM) ve Yerel Olmayan Ortalama Seviye (NLM) filtrelerini kullanmışlar ve filtrelerin performans analizi için Ortalama Kare Hata (MSE), Doğrusal Minimum Ortalama-Kare Hata (LMMSE), En Yüksek Sinyal Gürültü Oranı (PSNR), Görüntü Kalitesi İndeksi (IQI),

Ortalama Mutlak Hata (MAE) ve Kontrast-Gürültü Oranı (CNR) parametrelerini kullanmışlardır. Bir diğer çalışma [16], uzaysal ve frekans alanında gürültü giderme amacıyla kullanılan uyarlamalı histogram eşitlemesi, medyan filtre, butterworth filtre, frost filtre ve dalgacık filtre olmak üzere beş algoritmadan hangisinin mamografi görüntülerine daha uygun olduğunu bulmak için yapılmıştır. Bu gürültü giderme tekniklerini 5 iyi huylu ve 5 kötü huylu vakadan oluşan bir dizi 10 görüntü üzerinde test etmişler ve sonuçlar uzaysal alan için medyan filtrenin ve frekans alanı için dalgacık ile gürültü gidermenin daha yüksek PSNR (sinyal gürültü oranı) ve düşük MSE (ortalama karekök hata), MD (maksimum fark), NAE (normalleştirilmiş mutlak hata) ve SC Yapısal içerik) değerlerine sahip olduğunu göstermiştir. Bir diğer çalışmada Mencattini ve arkadaşları [17] mamogramlarda gürültü giderilmesi amacıyla diydik wavelet dönüşümüne dayanan yeni bir algoritma sunmuşlardır. Gürültü giderme aşaması yerel yinelemeli gürültü varyans tahminine dayanmaktadır. Ayrıca, mikro kalsifikasyonlar için, farklı dalgacık ölçeklerinde adaptasyonlu bir uyum derecesi ayarlaması önermişlerdir ; buna karşın kitle tespiti durumunda, diydik dalgacık bilgilerini matematiksel morfoloji ile birleştiren yeni bir segmentasyon yöntemi geliştirmişlerdir. Önerilen algoritma, hem analitik indeksler hem de radyologların görüşleri ile literatürde önerilen birkaç algoritma ile elde edilen sonuçlarla karşılaştırarak, çok sayıda klinik görüntü üzerinde test edilmiştir. Ön testlerden elde edilen sonuçlar, yöntemin diğer yaklaşımlara oranla meme kanserinin erken teşhisinde anlamlı derecede etkili olduğunu göstermektedir. Görüntü iyileştirme ile ilgili yapılan çalışmalara bakıldığında yine pek çok çalışma karşımıza çıkmaktadır. Örneğin Kim [18] ve arkadaşları yaptıkları çalışmada birinci türev ve yerel istatistiklere dayanan mamografi görüntüleri için uyarlanabilir bir görüntü iyileştirme yöntemi önermişlerdir. Bu yöntem üç aşamadan oluşmaktadır. Birinci aşamada mikro kalsifikasyon zannedilen art efektler kaldırılmıştır. İkinci aşamada, birinci türev operatörleri kullanılarak görüntü değişiklikleri hesaplanmaktadır. Son aşamada ise uyarlamalı ağırlıklı gradyan görüntüleri ile mamografi imgesinin önemli özellikleri geliştirilmektedir. Uyarlamalı olarak görüntü detaylarının güçlendirilebilmesi ve görüntü gürültülerinin bastırılabilmesi için görüntünün yerel istatistiğinden faydalanmışlardır. Amutha [19] ve arkadaşları çalışmalarında, meme kanserinin erken tanısı ve teşhisi için kontrast artırma ve gürültü giderme amacıyla bir algoritma sunmuşlardır. Bu çalışma modifiye matematiksel morfoloji ve biortogonal dalgacık analizine dayanmaktadır. Önerilen algoritmanın performans değerlendirmesi için Kontrast Artış İndeksi (CII) ve Kenar Koruma İndeksi

(EPI) kullanılmıştır. Sonuçlar, önerilen algoritmanın diğer iyi bilinen algoritmalara kıyasla çok daha iyi görüntü kalitesi sağladığını göstermektedir. Moradmand ve arkadaşları [20] dijital mamogramlarda görüntü iyileştirme amacıyla dört farklı algoritmanın karşılaştırıldığı bir çalışma sunmuşlardır. Bu çalışmada dalgacık tabanlı iyileştirme (8. Seviye Asimetrik Daubechies), Kontrast Sınırlı Adaptif Histogram Eşitleme (CLAHE), morfolojik operatörler ve keskin olmayan maskeleme karşılaştırılmış ve 114 klinik dijital mamografi üzerinde test edilmiştir. Deneysel sonuçlar, dalgacık dönüşümün daha etkili olabileceğini ve özellikle yoğun meme yapıları için BDT sisteminin genel olarak iyileştirilebileceğini düşündürmektedir. Akila [21] ve arkadaşları dolaylı kontrast geliştirme tekniklerinden histogram eşitleme (HE), kontrast sınırlı adaptif histogram eşitleme (CLAHE), parlaklığı koruyan iki boyutlu histogram eşitleme (BBHE), minimum ortalama parlaklık hatası iki boyutlu histogram eşitleme (MMBEBHE), özyineli ortalama-ayrı histogram eşitleme (RMSHE) yöntemlerini düşük kontrastlı düşük kontrastlı mamografi görüntülerine uygulayarak karşılaştırılmıştır. Hem CLAHE hem de RMSHE tekniklerinin, test görüntülerinde bulunan kitlelerde ve mikrokalsifikasyonlarda daha iyi iyileştirilme sağladığı görülmüştür. Ayrıca, RMSHE tekniğinin yapılan performans değerlendirme çalışmalarına göre düşük kontrastlı mamografi görüntüleri için en iyi kontrast arttırmayı sağladığı ve aynı zamanda en iyi parlaklık koruma özelliğini sunduğu görülmüştür.

Meme kanserinin tanısı ve teşhisinde bölütleme denildiğinde , genel olarak şüpheli bölgelerin diğer bölgelerden ayrılması düşünülebilir. Bu amaçla yapılan ilk çalışmalardan biri Strauss [22] ve arkadaşları tarafından sayısal mamogramlarda mikro kalsifikasyonun tespiti ve bölütlenmesi için bilgisayar tabanlı yöntemlerin kullanılmasının incelendiği çalışmadır. Bilgisayar algılama sistemi otomatik olarak elde edilen  $r$  ve  $h$  parametrelerinin, iki filtrelenmiş görüntüye iki kez uygulandığı *rhmaxima* algoritmasına dayanmaktadır. Bölütleme işlemi, arka plan ve mikro kalsifikasyonlar için farklı işaretlerin kullanıldığı watershed dönüşümünden türetilen bir algoritma ile gerçekleştirilmiştir. Bir diğer çalışma Li [23] ve diğerlerinin gerçekleştirdiği çalışmadır. Bu çalışmada tümörün veya diğer şüpheli bölgelerin tanımlanması amacıyla dijital mamogramların bölütlenmesi için bir yöntem önerilmiştir. Bu yöntem bir görüntünün modellenmesi için gauss işlemi ile Markov Rasgele Alanını birleştirmektedir. Önerilen yöntem 13 mamografi üzerinde test edilmiş ve yeni tekniğin dijital mamogramlardaki tümörlerin teşhisinde önemli bir potansiyele sahip olduğu bildirilmiştir. Sahakyan ve

Sarukhanyan [24] çalışmalarında arka plan bölgesinden göğüs profil bölgesini ayırmak ve sayısallaştırma gürültüsünü gidermek için ileri kenar algılama ve bölge bölütlemesi amacıyla morfolojik önileme algoritmasını kullanmışlardır. Önerilen algoritma farklı meme yoğunluğuna sahip mamogramlarda test edilmiş ve genel olarak meme bölgesi tespiti amacıyla kalite ölçeği ortalama değeri 0.95 olarak elde edilmiştir, bu değer algoritmanın farklı yoğunluklarda bile son derece sağlam olduğunu göstermektedir. Ashidi ve Siong [25] mamogramlarda kitle segmentasyonu ve tespiti ile ilgili çalışmalarında ilk olarak mamogramı, meme bölgesi ve meme olmayan bölge olarak ikiye ayırmışlar daha sonra görüntü iyileştirme amacıyla mamogramlardaki doku yapısının kontrastını iyileştirmişlerdir. Son olarak mamogramlarda kitle tespiti ve bölünmesi için lokal istatistiksel doku analizine dayanan sınırlı bölge büyütme uygulamışlardır. Önerdikleri sistemi MIAS veri tabanında bulunan 322 görüntü üzerinde test etmişlerdir. Elde ettikleri veriler önerilen tekniğin % 94.59 hassasiyetle ve resim başına 3.9 yanlış pozitif sayısı ile kitleleri tespit ettiğini göstermektedir. Cheng [27] ve arkadaşları yaptıkları kapsamlı çalışmada kitle tespiti ve sınıflandırma yöntemlerini incelemişler ve bu yöntemlerin avantaj ve dezavantajlarını karşılaştırmışlardır. Gulrud [27] ve diğerleri dijital mamografide tespit edilen kitlenin bölütlenmesi için gri ölçekli matematiksel morfolojiye dayanan bir yöntem sunmuşlardır. Önileme aşamasında, lokal histogram tekniğine dayalı görüntü iyileştirmesi sonrasında morfolojik düzleştirme işlemi uygulamışlardır. Bölütlenmiş bölgelerdeki yumuşatılmış görüntünün gradyanına watershed dönüşümü uygulamışlardır. Önerilen yöntem MIAS veritabanında bulunan mamogramlar üzerinde test edilmiştir. Deneysel sonuçlar, yöntemin kitle yapılarının konturunu çıkarmada faydalı olacağını göstermektedir.

Meme kanserinin tespitine yönelik bir diğer aşama öznitelik çıkarımıdır. Öznitelik çıkarımı tüm görüntünün kullanılması yerine yalnızca görüntünün belirli özelliklerinin işleme tabi tutulmasıdır. Doku analizi ile ilgili öznitelik çıkarımına dair yaklaşımlar dört başlıkta toplanabilir. Bunlar; yapısal, istatistiksel, model tabanlı ve dönüşüm yöntemlerine dayanan yaklaşımlardır. Örneğin Haralick gri ton uzaysal bağımlılıklarına dayanan bazı kolay hesaplanabilir dokusal özellikler ve üç farklı görüntü verisinin tanımlanmasına ait uygulamaları gerçekleştirdiği bir çalışma yapmıştır. Elde ettiği sonuçlar bu yöntemin, kolayca hesaplanabilen dokusal özniteliklerin çok çeşitli görüntü sınıflandırma uygulamaları için genel bir uygulanabilirliğe sahip olduğunu

göstermektedir [28]. Lahmiri ve Boukadoum [29] kanser görüntülerini sınıflandırmak için ayrı dalgacık dönüşümü ve Gabor filtresi kombinasyonundan elde edilen istatistiksel özellikleri kullanan bir sistem önermişlerdir. Bu sistem sınıflandırıcı olarak destek vektör makinelerini kullanmaktadır. Deneysel sonuçlar, böyle karma bir işleme modelinin, ayırık dalgacık dönüşümü ve Gabor filtre bankalarının tek başına kullanımıyla karşılaştırıldığında daha düşük standart sapma ve daha yüksek doğruluk sağladığını göstermektedir. Al-Shamlan ve El-Zaart [30] çalışmalarında, mamografi görüntüsünde daha fazla anlam ifade eden 23 önemli öznitelik çıkartmışlardır. Öznitelikler seçilirken mamografi görüntü alanındaki uzman kişilerle yapılan birçok röportaj ve bu alanda yayınlanmış birçok makaleyi temel almışlardır. Kullanılan bu öznitelikleri 80 mamografiye uygulamışlar ve çok iyi sonuçlar elde ettiklerini bildirmişlerdir. Saraswathi [31] ve diğerleri dijital mamografinin analizi için verimli bir öznitelik çıkarma yöntemi uygulamışlardır. Başlangıçta, her mamogram görüntüsü ayrı ayrı dalgacık ve eğricik dönüşümleri kullanılarak çok seviyeli ayrışma ve sarmalama işlemine tabi tutulmuştur. Ardından, her ayrışma seviyesine ait katsayılar elde edilmiş ve eğricik katsayıları destek vektör makinesine verilmiştir. Performans sınıflandırma doğruluk oranı kullanılarak ölçülmüştür. Deneysel sonuçlar, eğriciğin dalgacıktan daha iyi bir performans sağladığını göstermektedir. Ayrıca elde edilen eğricik katsayıları sınıflandırılmayacak kadar fazla olduğu için zaman karmaşıklığını azaltmak ve belirgin özellikleri seçmek amacıyla bir Parçacık Sürü Optimizasyon (PSO) algoritması kullanmışlardır. Öznitelik çıkarımı ile elde özniteliklerin boyutu bazen çok büyük olabilmektedir. Bu gibi durumlarda özniteliklerin azaltılması ve seçimi amacıyla farklı algoritmalar ve yöntemler geliştirilmiştir. Örneğin Liu ve Tang [32] meme kitlelerinin sınıflandırılmasını 408 kötü huylu ve 418 iyi huylu kitlesi olan 826 ilgi bölgesi elde edilmiş veri setiyle araştırmış ve sunmuşlardır. Bölütleme amacıyla seviye belirleme yöntemiyle mekansal FCM kümeleme yöntemini birleştirmişlerdir. Bölütlemeyen sonra, her kitle için 12 geometrik öznitelik ve 19 doku özelliği olmak üzere 31 öznitelik elde etmişlerdir. F-skoru, Relief, SVM-RFE, SVM-RFE (mRMR) ve önerilen SRN de dâhil olmak üzere birçok öznitelik seçimi yöntemini araştırmışlardır. Önerdikleri öznitelik seçim yönteminin, daha az seçilmiş özniteliklere sahip diğer seçim yöntemlerinden daha iyi performans gösterdiğini belirtmişlerdir. Nandi [33] ve arkadaşları 57 kitle içeren mamogram görüntüsüne ait kenar keskinliği, şekli ve dokusu ile ilgili 22 özelliğin çıkarıldığı bir veri seti kullanmışlardır. Çalışmalarındaki yenilik, genetik programlamanın (GP) uyarlanması ve uygulanmasıdır. GP sınıflandırıcısının kullandığı

öznitelik havuzunu hassaslaştırmak için Student'ın t-testi, Kolmogorov-Smirnov Testi ve Kullback-Leibler ayırıcısı olmak üzere üç istatistiksel ölçüt içeren beş öznitelik seçimi yöntemi kullanmışlardır. Elde edilen doğruluk sonuçlarının, genel olarak eğitim için % 99.5 ve üzeri ve test için % 98 ve üzeri olduğunu belirtmişlerdir. Nugroho [34] ve arkadaşları yaptıkları çalışmada elde ettikleri mamogramlar için histogram ve Gri Seviye Eş-oluşum Matrisini (GLCM) ile ilgili toplam 12 öznitelik çıkarmışlardır. Çalışmalarında özniteliklerin % 50'sini azaltan Korelasyona Dayalı Özellik Seçimi (CFS) kullanılmışlardır. Seçilmiş bu özellikler çok katmanlı algılayıcı algoritması ile sınıflandırılmıştır. Özel Onkoloji Kliniği Kotabaru Yogyakarta'dan alınan 40 dijital mamografi verisi için elde edilen deneysel sonuçlar % 91.66 doğruluk oranına ulaşıldığını göstermektedir.

Son aşama sınıflandırma ve karar aşamasıdır. Meme kanserinin sınıflandırılmasında genel olarak kanserli veya kanserli değil ve iyi huylu veya kötü huylu şeklinde iki sonuç elde edildiği gözlemlenmektedir. Sınıflandırıcı olarak pek çok yöntem kullanılmaktadır. Örneğin Rabidas [34] ve arkadaşları mamografi kitlelerinin iyi huylu ve kötü huylu olarak sınıflandırılmasında Yapısal Komşuluk Benzerliğini (NSS) kullanan iki yeni öznitelik çıkarım yöntemi önermişlerdir. İyi huylu ve kötü huylu kitlelerdeki piksellerin gri seviye dağılımı arasındaki farktan yararlanan yöntem kitlelerin komşu bölgelerle arasındaki benzerliği kullanmaktadır. NSS-I ve NSS-II, olarak adlandırılan bu iki yeni yöntem farklı ölçeklerde global benzerlik yakalamaktadır. Yöntemin performansı, mini-MIAS ve DDSM (Digital Database for Screening Mammography) veri tabanlarından alınan görüntülerle Fisher lineer ayırıcı analizi ile 10 kat çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. Yazarlar Mini-MIAS veritabanı için % 94.57, DDSM veritabanı için % 85.42 doğruluk elde ettiklerini belirtmişlerdir. Elman ve Kadah kitleler ile normal göğüs dokusu arasındaki farkı belirlemek için bir CAD sistemi önermişlerdir. İlk önce DDSM veri tabanından alınan görüntüler ön işlemden geçirilmiş ve mamogram görüntüleri iyileştirilmiştir, ardından kitleler ve normal göğüs dokusu içeren 100 ilgi bölgesi (ROI) çıkartılmıştır. Elde edilen ilgi bölgeleri için 59 dokusal ve istatistiksel öznitelik çıkartmışlardır. Ardından öznitelik seçimi için Ardışık İleri Seçme ve Ardışık Dalgalı İleri Seçme yöntemlerini kullanmışlardır. Son olarak, sınıflandırma için K-En Yakın Komşuluk (K-EK) algoritması, Lineer Ayırıcı Analizi (LAA), Karesel Ayırıcı Analizi (QDA) ve Destek Vektör Makinalarını (DVM) kullanmışlardır. [36]'da, mamografi görüntülerinden kitle saptama ve sınıflandırma için yeni bir bilgisayar

destekli tanı sistemi önerilmiştir. Önerilen sistem tamamen otonom bir sistemdir. Dijital mamogramlardan ilgi alanlarının çıkartılması için görüntü önışleme ve Otsu N eşikleme algoritması kullanılmıştır. Daha sonra ilgi alanlarından farklı düzen ve iterasyonlarda bir grup 32 Zernike momenti çıkarılmıştır. Bu momentler esnek geri yayılım nöral ağ sınıflandırıcıya uygulanmıştır. Deneysel sonuçlar, önerilen tekniğin, dijital mamografide kötü huylu ve iyi huylu kitle tipindeki anormalliklerin etkili bir şekilde sınıflandırılabilirdiğini göstermektedir. Jen ve Yu [37] çalışmalarında birinci dereceden istatistiksel yoğunluklar ve gradyanlara dayalı öznitelikler çıkararak yeni anormallik saptama sınıflandırıcısına (ADC) gönderen ve anormal mamografileri tespit eden bir yöntem önermişlerdir. Çalışmalarında, öncelikle daha doğru göğüs bölütleme için küresel eşitleme dönüşümü, gürültü giderme, ikilik resme dönüştürme, meme oryantasyonu tayini ve pektoral kası gizleme gibi önışleme adımlarını araştırmışlardır. Öznitelik ağırlıklarının belirlenmesi için Temel Bileşen Analizini (TBA) kullanmışlardır. Deneysel sonuçlar, anormal mamografi tespiti için ADC algoritmasının öznitelik ağırlık ayarlamaları ile uygulanmasının, ilgili iki veri kümesinde % 88 ve % 86'lık belirgin hassaslık verdiğini göstermektedir. Nithya ve Santi [38] mamogramların normal ve kanserli olarak sınıflandırıldığı bir çalışma yürütmüşlerdir. Çalışmalarında mamografi görüntülerinden GLCM özniteliklerinin çıkartılması ve bu özniteliklerin sınıflandırılmasını araştırmışlardır. Normal ve kanser sınıflandırması için maksimum doğruluk oranını % 96 olarak elde etmişlerdir. Eltoukhy [39] ve diğerleri, dijital mamografi görüntülerinden meme kanseri tanısı için bir yöntem önermişlerdir. Dalgacık ve Eğriciğin çok çözünürlüklü gösterimini mamogram görüntülerinin uzun bir katsayı vektörüne dönüştürülmesi için kullanılmışlardır. Dalgacık ve Eğricik katsayılarından oluşan satırların görüntü sayısını sütunlarında katsayı sayısını belirttiği bir matris oluşturmuşlardır. İstatistiksel t-testi yöntemine dayalı olarak bir öznitelik çıkarma yöntemi geliştirmişlerdir. Geliştirilen yöntem, sınıfları ayırt etme kabiliyetine göre özniteliklerin (sütunların) sıralamasını yapmaktır. Sınıflandırma yapmak için Destek vektör makinesi (SVM) kullanmışlardır. Sınıflandırma hem normal ve anormal dokular arasında hem de iyi huylu-kötü huylu tümörleri ayırmak için kullanılmıştır. Elde ettikleri sınıflandırma doğruluk oranları, önerilen yöntemin meme kanserinin başarılı bir şekilde saptanmasına katkıda bulunabileceğini göstermektedir.

## 1.7. TEZ ÇALIŞMASI KONU VE LİTERATÜRE KATKILAR

Sunulan tez çalışmasında da meme kanserinin bilgisayar destekli teşhisi için 5 aşamadan oluşan bir BDT sistemi önerilmiştir. Bu sistem sırasıyla önişleme, bölütleme, öznetelik çıkarımı, öznetelik seçimi ve sınıflandırma aşamalarından oluşmaktadır. Geliştirilen bu sisteme ait genel şema şekil 1.6'da sunulmuştur.



Şekil 1.6. Önerilen BDT sisteminin akış diyagramı.

Önişleme aşamasında mamogram görüntülerinde gürültü giderme, iyileştirme, artefektlerin kaldırılması gibi görüntü kalitesini arttıracak işlemler araştırılmış ve daha iyi sonuç veren yöntemlerden medyan filtre, morfolojik işlemler, biortogonal dalgacık analizi, anizotropik difüzyon yöntemi, kontrast sınırlı adaptif histogram eşitleme (CLAHE) yöntemleri bu amaçlar doğrultusunda kullanılmıştır. Ayrıca mamogram görüntülerinin otomatik sınıflandırılmasında yanlış pozitif neden olabilen pektoral kasın çıkartılması amacıyla bu tez çalışması kapsamında SSEM isimli yeni bir algoritma geliştirilmiş ve kullanılmıştır.

Bölütleme kısmında mamogramlardaki şüpheli bölgelerinin bölütlenmesi işlemi gerçekleştirilmektedir. Bu amaçla Otsu N eşikleme, Havrda & Charvat Entropi ve bu tez kapsamında görüntü kümeleme amacıyla geliştirilen w-BSAFCM yöntemleri kullanılmıştır.

Öznetelik çıkarımı aşamasında yapısal, istatistiksel ve dönüşüm modellenmiş yöntemler kullanılmıştır. Öznetelik seçimi için ise Lineer Ayırıcı Analiz kullanılmıştır.

Mamogramlar Yapay Sinir Ağları, Destek Vektör Makineleri ve K-En yakın Komşuluk yöntemleri olmak üzere üç farklı yöntem kullanılarak sınıflandırılmışlardır. Mamogramlar sınıflandırılırken ilk olarak normal-anormal sınıflandırma yapılmıştır. Anormal mamogramlar içinde iyi huylu-kötü huylu sınıflandırılması gerçekleştirilmiştir. Bu tez çalışmasında önerilen yöntemlerin farklı veri tabanlarındaki sonuçlarını incelemek amacıyla iki farklı veri tabanı kullanılmıştır. Bunlardan ilki literatürde sıklıkla kullanılan ve erişime açık olarak bulunan MIAS veri tabanıdır. İkincisi ise güncel bir veri tabanı olan ve erişime açık olmayan INBREAST veri tabanıdır. INBREAST veri tabanından verilerin elde edilmesi amacıyla gerekli yazışmalar yapılarak veri tabanına erişim sağlanmıştır. Bu tez çalışması kapsamında geliştirilen yöntemler her iki veri tabanından alınan mamogramlar kullanılarak test edilmiştir.

Önerilen yeni BDT sisteminin literatüre katkıları şu şekilde özetlenebilir.

1. SSEM adını vermiş olduğumuz pektoral kasın çıkartılmasına yönelik yeni bir yöntem geliştirilmiştir. Bu yöntem sayesinde BDT Sistemlerinde sıklıkla kitle vb. anormal yapılarla karıştırılmakta olan ve doğru teşhisi zorlaştıran pektoral kasın çıkartılması mümkün olmaktadır.
2. Bu tez kapsamında w-BSAFCM adını verdiğimiz yeni bir görüntü kümeleme algoritması geliştirilmiştir. Mamogram görüntüleri üzerinde Otsu N algoritması, Havrda & Charvat entropi ve w-BSAFCM kümeleme algoritması yöntemleri hibrit bir şekilde kullanılarak şüpheli bölge tespiti gerçekleştirilmektedir. Bu üç yöntemin birlikte kullanılmasıyla elde edilen ROI sayısı azaltılmış böylece işlem yükü de hafifletilmiştir.
3. Özniteliklerin elde edilmesi aşamasında üç farklı yöntem olan GLCM, dalgacık ve eğricik analizi kullanılmış ve bu yöntemlerin hem öznitelik azaltılmadan önce hem de öznitelik azaltıldıktan sonra mamogramların sınıflandırılması üzerindeki etkileri incelenmiş ve sonuçlar karşılaştırmalı olarak sunulmuştur. Bu üç yöntemin karşılaştırmalı olarak sunulduğu bir çalışmaya literatürde rastlanmamıştır.
4. Literatürde sıklıkla eğricik dönüşümü için elde edilen özniteliklerde katsayılar kullanılmış ve bu boyut problemine neden olmuştur. Boyut problemini azaltmak amacıyla araştırmacılar genellikle her seviye için en büyük katsayıların ilk yüzünü almışlardır. Çoğunlukla eğricik analizi ile öznitelik elde edilirken 4. Seviye seçilmekte bu da öznitelik matris boyutununun 400 olması anlamına gelmektedir.

Bu çalışmada ise eğricik dönüşümlerinde katsayıların kullanılması yerine her bir eğricik dönüşümü seviyesi için tüm katsayılar için bazı istatistiksel hesaplamalar kullanılmaktadır. Bunlar standart sapma ve çarpıklık olarak belirlenmiştir. Toplamda 4. Seviye eğricik analizi için 162 adet öznitelik elde edilerek boyut daha da indirgenmiş ve tüm katsayıların içerdiği veriler hesaplamalara dâhil edilmiştir.

5. Yapılan tüm bu çalışmalar doğrultusunda mamogramların normal-anormal sınıflandırmasında en iyi sonuçlara eğricik dönüşümü ve lineer ayırıcı analizin birlikte kullanılması sonucunda ulaşılmıştır. MIAS veri tabanından alınan 140 ve INBREAST veri tabanından alınan 106 mamogram için normal-anormal sınıflandırma başarısı % 100 olarak görülmüştür. Benzer şekilde mamogramların iyi huylu-kötü huylu sınıflandırmasında MIAS veri tabanından alınan 70 anormal ve INBREAST veri tabanından alınan 53 anormal mamogram için eğricik dönüşümü ve lineer ayırıcı analizin birlikte kullanılması sonucunda % 100 başarı elde edilmiştir.

Yapılan tüm bu çalışmalarla ilgili ayrıntılı bilgiler tezin ilerleyen bölümlerinde verilmektedir.

## 2. MATERYAL VE YÖNTEM

Tez çalışması kapsamında geliştirilen BDT sisteminin eğitilmesi ve performansının değerlendirilmesi amacıyla, önceden konu uzmanları tarafından değerlendirilmiş ve kanserli olup olmadıklarına ve eğer kanserli iseler kanser çeşidinin tespitinin yapılmış olduğu örnek mamogramlara ihtiyaç vardır. Bu nedenle bu tez çalışmasında hem literatürde sıkça kullanılan ve herkese açık bir veri tabanı olan MIAS [40] veri tabanı kullanılmış hem de daha farklı örnekler elde etmek amacıyla başka veri tabanları araştırılmıştır. Araştırma sonucunda Portekiz Ulusal Bilgi Koruma ve Hastane Etik Kurulunun izniyle Centro Hospitalar de S. João [CHSJ], Göğüs Merkezi, Porto'dan INBREAST isimli veri tabanına ulaşılmış ve gerekli sözleşme yapılarak bu veri tabanı da elde edilmiştir. Bu bölümde ilk olarak bu veri tabanları tanıtılmış ardından da geliştirilen BDT sisteminin her bir aşaması ve bu aşamalarda kullanılan yöntemler detaylı olarak sırasıyla açıklanmıştır.

### 2.1. KULLANILAN VERİTABANLARI

Bu tez çalışmasında MIAS ve INBREAST olmak üzere iki farklı veri tabanından alınan mamogram görüntüleri kullanılmıştır. Bu veri tabanlarına ait bilgiler aşağıda ayrıntılı olarak açıklanmıştır.

#### 2.1.1. MIAS Veri Tabanı

MIAS [40] veri tabanı MLO pozisyonunda elde edilmiş 1024x1024 piksel boyutlarında, 322 adet sayısallaştırılmış görüntüden oluşmuştur. Bu mamogramlardan 207'si normal, 51'i kötü huylu, 64'ü iyi huyludur. Bu veri tabanında 25 adet mikro kalsifikasyon, 23 adet sınırlı kitle, 19 adet mimari bozulma, 19 adet speküle kitle ve 15 adet çeşitli bozukluk vakası bulunmaktadır. Görüntülerin neredeyse % 50'si farklı gürültü türlerine sahip arka planlardan oluşur. Bütün görüntüler uzmanlar tarafından çeşitli açılardan sınıflandırılmış ve çalışmalar için araştırmacıların kullanımına sunulmuştur. MIAS veri tabanının yer alan mamogramlar farklı kategorilere ayrılmaktadır. Bu kategoriler aşağıda sunulmuştur.

Sınıflandırma kategorileri:

Memenin dokusuna göre; yağlı, yağlı-bezel, yoğun bezel,

Anormallik durumuna göre;

**CALC** Kireçlenme,

**CIRC** İyi tanımlanmış / sınırlı kitleler,

**SPIC** spiküle kitleler,

**MISC** Diğer kötü tanımlanmış kitleler,

**ARCH** Mimari bozulma,

**ASYM** Asimetri,

**NORM** Normal,

Anormalliğin şiddetine göre;

**B** İyi Huylu,

**M** Kötü Huylu olarak sınıflandırılmıştır.

Ayrıca anormalilerin koordinatları ve büyüklükleri de veri tabanında bildirilmiştir.

### **2.1.2. INBREAST Veri Tabanı**

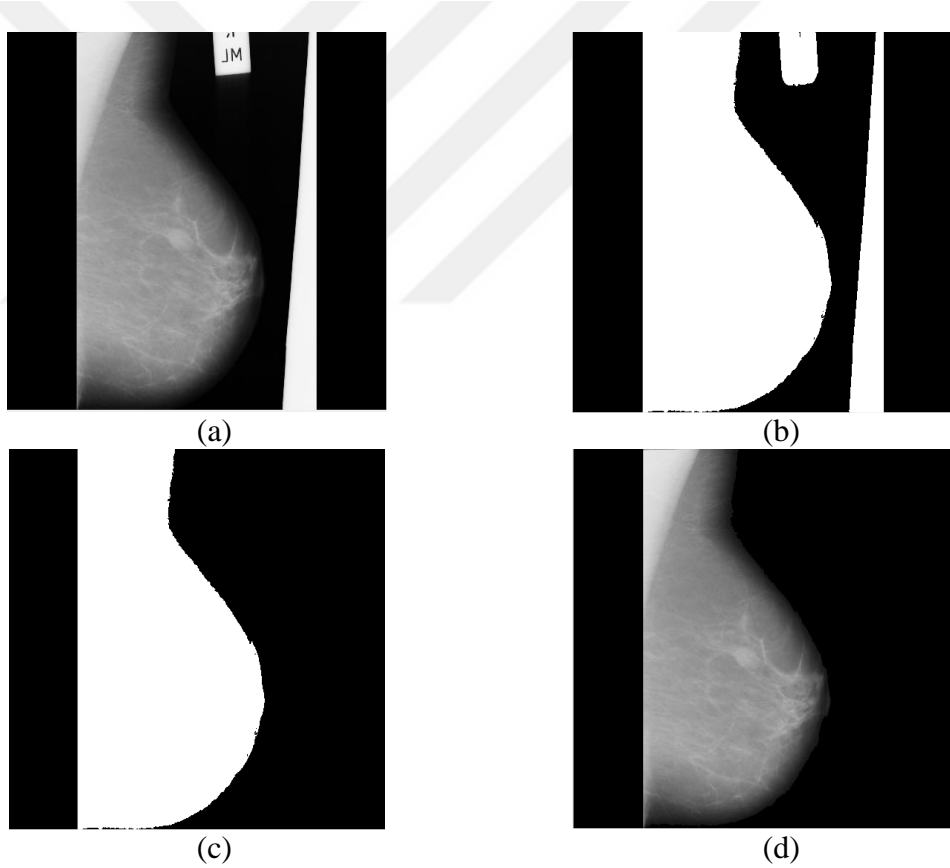
INBREAST, DICOM formatında mamografi görüntülerinin bulunduğu bir veri tabanıdır. Kullanılan mamogramlar, Portekiz Ulusal Bilgi Koruma ve Hastane Etik Kurulunun izniyle Centro Hospitalar de S. João [CHSJ], Göğüs Merkezi, Porto'dan edinilmiştir. INBREAST veri tabanında değişik anormalliklere sahip 115 vaka (410 görüntü) bulunmaktadır. Bu görüntülerde 108 kitle, 308 mikro kalsifikasyon, 3 mimari distorsiyon ve 14 asimetri vakası mevcut olup, doğru konturlar radyoloji uzmanları tarafından yapılmış ve XML formatında sunulmaktadır [41].

## **2.2. ÖNİŞLEMLER**

Önişlemler görüntü üzerindeki art efektlerin kaldırılması ve görüntünün iyileştirilmesi amacıyla yapılan işlemlerdir. Geliştirilen BDT kapsamında yapılan önişlemler bu bölümde sunulmuştur.

### 2.2.1. Meme Bölgesinin Bölütlenmesi

Önişlemler esnasında ilk olarak işlem alanının belirlenmesi ve yalnızca ilgili meme bölgesinin kullanılması amacıyla meme bölgesi bölütleme işlemi gerçekleştirilmiştir. Bu işlem aynı zamanda mamogram üzerinde bulunan art efektlerin birçoğundan kurtulmayı da sağlamaktadır. Bunun için mamogram görüntüsünde arka plan koyu renk ve meme bölgesi daha açık renk olduğu için mamogram deneysel olarak elde edilen  $\delta=0.07$  eşik değeri ile ikili görüntüye dönüştürülmüştür (Şekil 2.1-b). Daha sonra elde edilen yeni görüntüde en büyük alana sahip bölge kalacak şekilde bir filtre uygulaması yapılmıştır (Şekil 2.1-c). Son olarak orijinal mamogram görüntüsü ile bölütlenmiş mamogram görüntüsünün birleştirilmesi sonucunda sadece meme bölgesinin bulunduğu yeni görüntü elde edilmiştir (Şekil 2.1-d). Orijinal mamogram ve elde edilen görüntüler Şekil 2.1’de verilmiştir.



Şekil 2.1. Meme bölgesinin bölütlenmesine ait görüntüler (a) Orijinal mamogram (Mdb012) (b)  $\delta =0.07$  eşik değeri ile ikilik resme dönüştürülmüş görüntü (c) En büyük alana sahip bölgenin filtrelendiği görüntü (d) Art efektlerden temizlenmiş sonuç görüntüsü

## 2.2.2. Mamogramlarda Gürültü Giderme

Gürültü, görüntü üzerinde oluşan ve görüntüyü bozan etmenlerdir. Gürültü, sayısallaştırma kaynaklı, hareket kaynaklı, görüntü sıkıştırma kaynaklı vb. olabilmektedir. Tuz ve biber gürültüsü, parlaklık gürültüsü, foton gürültüsü ve gauss gürültüsü gibi farklı gürültü türleri vardır [42]. İmgenin özelliklerine, gürültünün frekans ve istatistiksel dağılımına göre gürültünün bozucu etkilerini ortadan kaldırmak amacıyla literatürde farklı yöntemler yer almaktadır. Bu yöntemler uzaysal alan ve dönüşüm alanı olmak üzere iki ana başlık altında toplanabilir. Uzaysal alanda yapılan işlemler, asıl imge üzerinde gerçekleştirilen işlemlerdir. Bu veriler, imgenin gri seviye değerleri ile ilgilidir ve genellikle piksel komşuluk ilişkisine dayanmaktadır. Dönüşüm alanındaki işlemler ise imgeye bir dönüşümün uygulandığı ve oluşan katsayıların işlenmesiyle gerçekleştirilen işlemlerdir. İşlenmiş katsayıların ters dönüşümünün alınmasıyla da gürültüden temizlenmiş imge oluşturulur. Bunların dışında bu iki alanı birleştiren hibrit çalışmalar da mevcuttur.

### 2.2.2.1. Uzaysal Alan

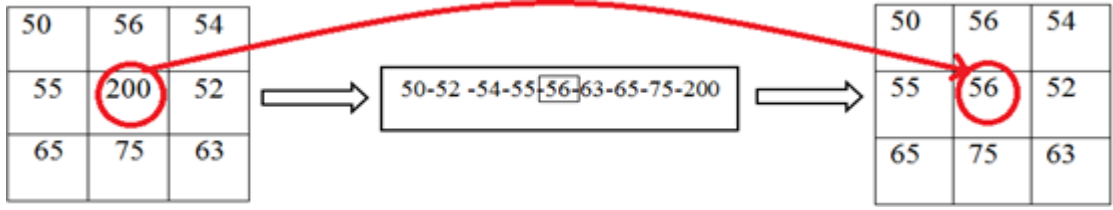
Uzaysal alandaki çalışmalar lineer ve lineer olmayan filtreler olarak iki gruba ayrılabilir. Lineer filtrelere ortalama filtre ve wiener filtre, lineer olmayan filtrelere ise medyan filtre ve türevleri örnek gösterilebilir [42]. Bu tez çalışmasında uzaysal alanda yapılacak gürültü giderme işlemleri için medyan filtre kullanılmıştır.

#### *Medyan Filtre*

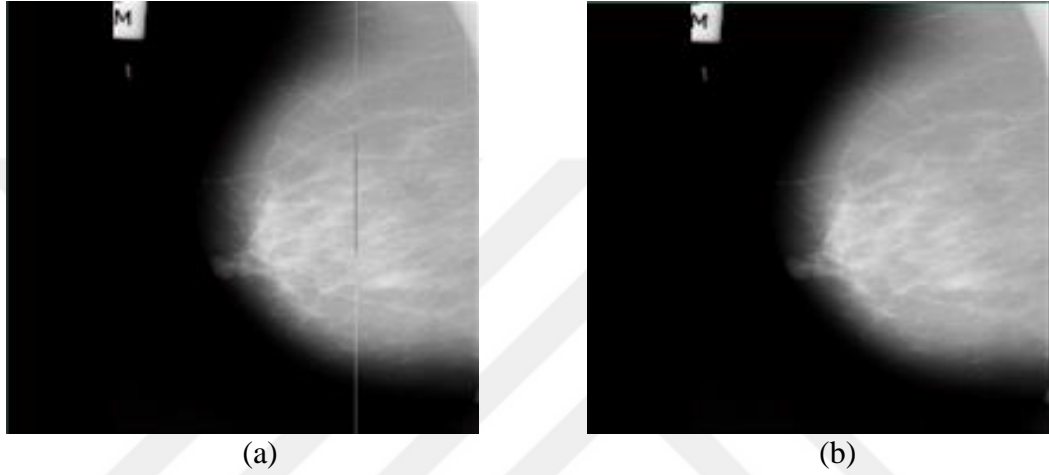
Medyan filtre gürültüyü giderirken kenarları korumada oldukça etkili bir filtredir. Temel prensibi maskeyi oluşturan boyuttaki resim piksel değerlerinin küçükten büyüğe sıralanmasına ve ortadaki değer merkez piksele atanmasına dayanmaktadır. Örneğin; 3x3'lük bir filtremiz olsun. Resim üzerinde bu filtreye denk gelen piksel numaraları küçükten büyüğe sıralanır. Daha sonra referans filtredeki merkez piksel değeri (3x3 filtre için dokuz değerden beşincisi) yeni görüntünün merkez pikseline atanır.

Tez çalışması kapsamında uzaysal alanda gürültüyü gidermek amacıyla medyan filtrenin yanı sıra ortalama filtre ve gauss filtrede kullanılmış ve elde edilen sonuçlara göre medyan filtrenin mevcut veri tabanlarındaki uzaysal alanda gürültü gidermede diğerlerinden daha etkin bir yöntem olduğu görülmüştür. Bu nedenle meme bölgesi bölütlenmiş her görüntü ilk olarak 3x3'lük medyan filtre ile filtrelenmiştir. Şekil 2.2'de 3x3'lük bir medyan filtre gösterimi, Şekil 2.3'de medyan filtre uygulaması ve Şekil

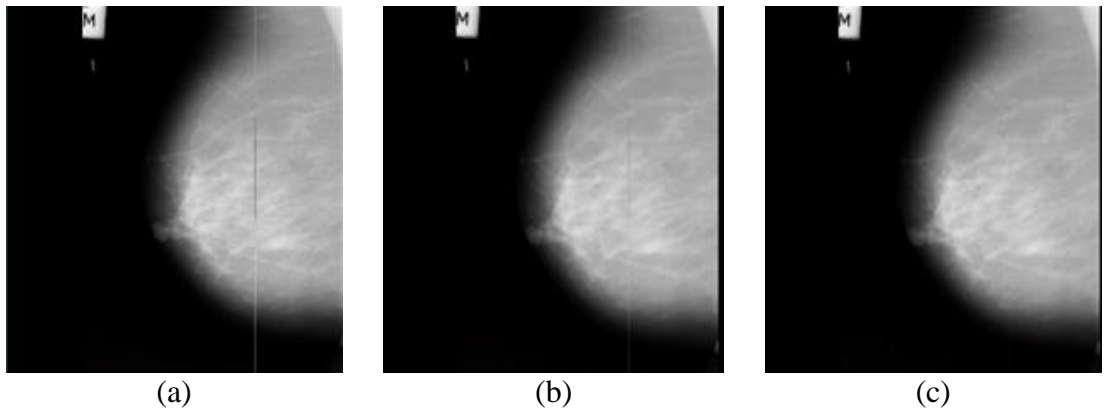
2.4'te ise deneysel çalışmalarda kullanılan diğer gürültü giderme yöntemlere ait uygulama görüntüleri yer almaktadır.



Şekil 2.2. 3x3'lük medyan filtre örneği.



Şekil 2.3. Medyan filtre uygulama sonucu mamogram görüntüsü (a) Orijinal mamogram (b) 3x3 medyan filtre ile filtrelenmiş mamogram.



Şekil 2.4. Ortalama filtre ve gauss filtre uygulama sonucu mamogram görüntüsü (a) Orijinal mamogram (b) Gauss filtre (c) Ortalama filtre.

#### 2.2.2.2. Dönüşüm Alanı

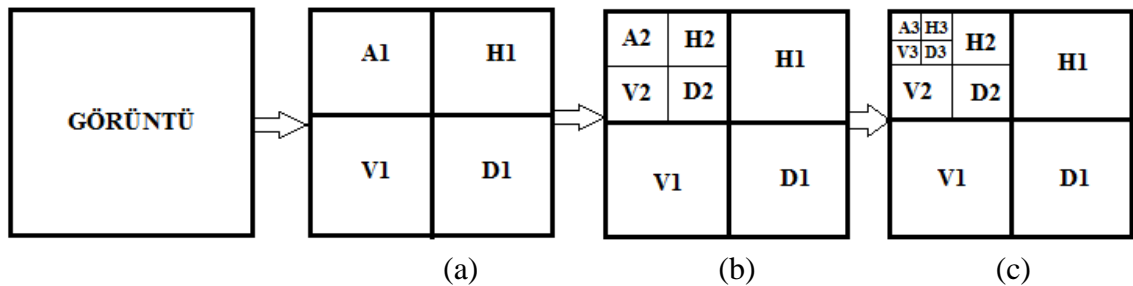
Dönüşüm alanındaki gürültü giderme tekniklerinde gürültü giderme işlemi yapılacak imgeye ilk olarak bir dönüşüm uygulanır ve bu dönüşüm sonucunda bazı katsayılar elde

edilir. Elde edilen bu katsayıların farklı yöntemlerle işlenmesi sonucunda da gürültüden temizlenmiş imge oluşturulur. Dönüşüm alanındaki gürültü giderme tekniklerine Fourier dönüşümü ve dalgacık dönüşümü örnek olarak gösterilebilir. Dalgacık dönüşümü tabanlı yöntemlerde amaç gürültüyü azaltırken imgenin kenarlarının ve ayrıntı bilgilerinin korunmasıdır. Bu yöntemlerde görüntü birçok alt banda ve yüksek frekans bileşenlerine ayrıştırılır. Oluşan katsayılar üzerinde bir eşikleme işlemi yapılır ve son olarak ters dalgacık dönüşümü uygulanarak gürültü giderilmiş görüntü elde edilir. Bu tez çalışmasında dönüşüm alanı gürültü giderme işlemleri için dalgacık analizi kullanılmıştır.

### *Dalgacık Analizi*

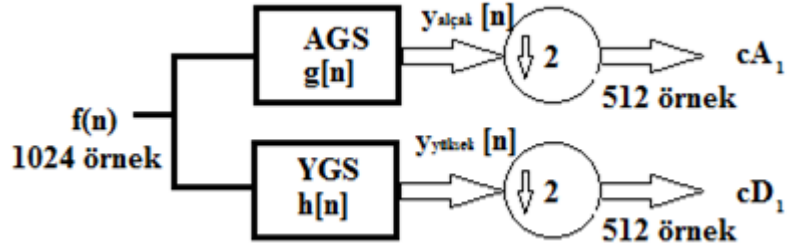
Dalgacık analizi çok çözünürlüklü bir dönüştürme olduğundan hem zaman hem de frekans bilgisini içermektedir ayrıca görüntü temsil edilirken görüntüdeki bilgi içeriği değişmez. Uzaysal ve spektral bilgileri de sağladığından ayrık dalgacık dönüşümü (ADD) , gürültü gidermede tercih edilir [42]. ADD' de, sinyal yaklaşık ve ayrıntı katsayılarının bulunduğu iki tamamlayıcı filtreden geçirilir. Yaklaşık katsayısı, yüksek frekans katsayılarına, ayrıntı katsayısı, düşük frekans katsayılarına denir. Ölçek kavramı ise sinyalin yukarı veya aşağı örnekleme ile ilgilidir. Aşağı örneklemede orijinal sinyalden bazı örnekler çıkartılırken yukarı örneklemede sinyaldeki örnek sayısı artırılır [44]. Bu işlem, ayrışma veya analiz, bu işlemin tersi ise yeniden yapılandırma veya sentez olarak bilinir.

ADD uygulanarak orijinal görüntü dört bantta ifade edilir. Bu dört parça, Şekil 2.5'te A1, H1, V1 ve D1 olarak belirtilmiştir. A1, yaklaşık olarak adlandırılır ve tekrar dört alt-banda ayrılabilir. Kalan bantlar detay bileşenleridir. Bir sonraki ayrışma seviyesini elde etmek için A1 alt bandı kullanılır.



Şekil 2.5. Üçüncü seviyeye kadar ayrık dalgacık dönüşümü katsayıları (a) 1. Seviye ADD (b) 2. Seviye ADD (c) 3. Seviye ADD.

Şekil 2.6’da bir sinyalin bir seviyeli bileşenlerine ayrıştırılması gösterilmektedir.



Şekil 2.6. Bir sinyalin bir seviyeli bileşenlerine ayrıştırılması [45].

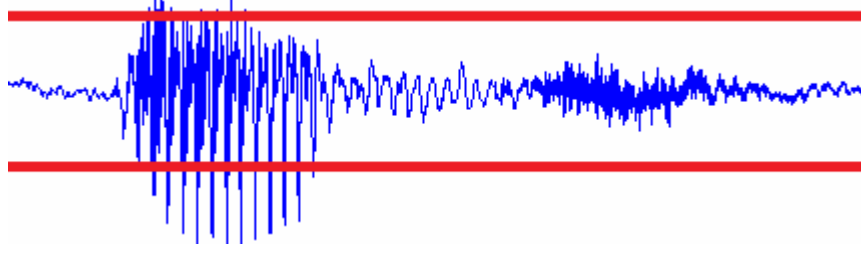
Denklem (2.1) ve (2.2)’de gösterildiği gibi  $f(n)$  sinyalinin alçak geçiren (AGS) ve yüksek geçiren (YGS) süzgeçlerle katlaması alınarak çıkışlar elde edilir [45].

$$cA_1 = \sum_{k=-\infty}^{\infty} f[k]g[2n - k] \quad (2.1)$$

$$cD_1 = \sum_{k=-\infty}^{\infty} f[k]h[2n - k] \quad (2.2)$$

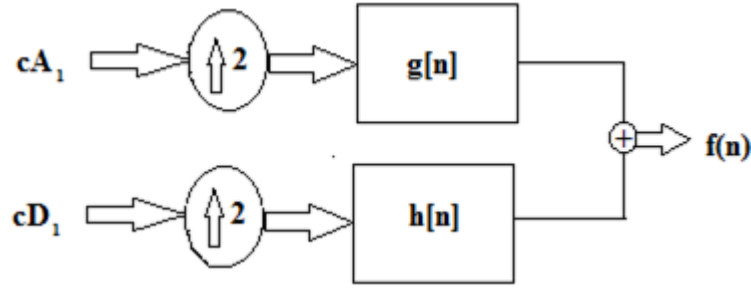
$cA_1$  yaklaşım katsayıları, sinyalin yüksek ölçekli alçak frekans bileşenlerini  $cD_1$  ayrıntı katsayıları ise sinyalin düşük ölçekli yüksek frekans bileşenlerini göstermektedir. Şekil 2.6 da sunulan sinyalde yer alan 1024 örnek için katlama alındığında 1024 yaklaşım bileşeni ve 1024 ayrıntı bileşenleri elde edilir fakat bu durumda 1024 değeri yerine 2048 değeri elde edilmiş olur. Bu durumu engellemek amacıyla iki ve ikinin katları ile aşağı örnekleme yapılır. Şekildeki örnek sinyal için bir kez katlama alındığı için iki ile aşağı örnekleme yapıldığında 512 örnekli  $cA_1$  ve  $cD_1$  katsayıları elde edilmiş olur. Aşağı örnekleme sonucunda çözünürlük yarıya iner ve ölçek iki katına çıkar [45].

İkinci aşamada, dalgacık katsayılarına eşikleme algoritmaları uygulanır. Eşikleme işlemi kabaca, işlenen katsayıların belirli bir değer altında olanlarını gürültü kabul edilmesi ve bunların elenmesi amacıyla gerçekleştirilen işlem olarak ifade edilebilir. Eğer katsayılar, belirlenen eşik değerinden küçük ise sıfırlanır, büyük ise ya olduğu gibi tutulur veya değiştirilir. Bu işleme uygulanma şekline göre sert veya yumuşak eşikleme denilir [43]. Dalgacıklarda gürültü giderilmesinde eşikleme için örnek bir görüntü Şekil 2.7’de verilmiştir.



Şekil 2.7. Tek boyutlu sinyal için eşikleme örneği.

Son aşamada ise ters dalgacık dönüşümü uygulanarak gürültü giderilmiş görüntü elde edilir. Bunun için örneğimizde  $cA_1$  yaklaşım katsayılarında ve  $cD_1$  ayrıtı katsayılarında tek indisli elemanlar arasına sıfır yerleştirilir ve iki ile yukarı örnekleme yapılır. Yukarı örnekleme alınan sinyalin süzgeçlerle katlaması alınarak sinyal yeniden elde edilmiş olur(Şekil 2.8) [45].



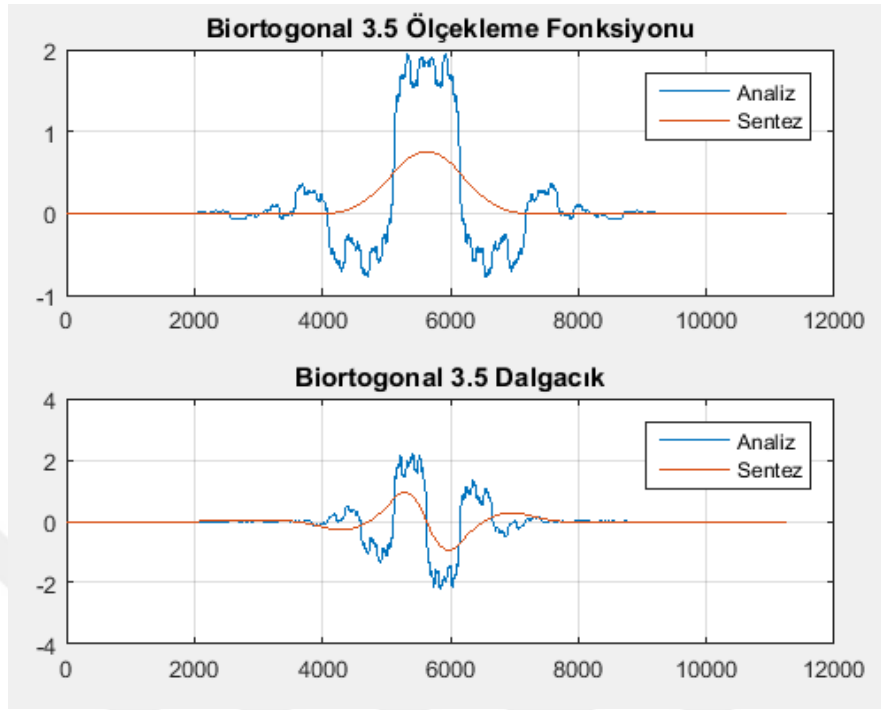
Şekil 2.8. Sinyalin yeniden elde edilmesi [45].

### *Biortogonal Dalgacık Analizi*

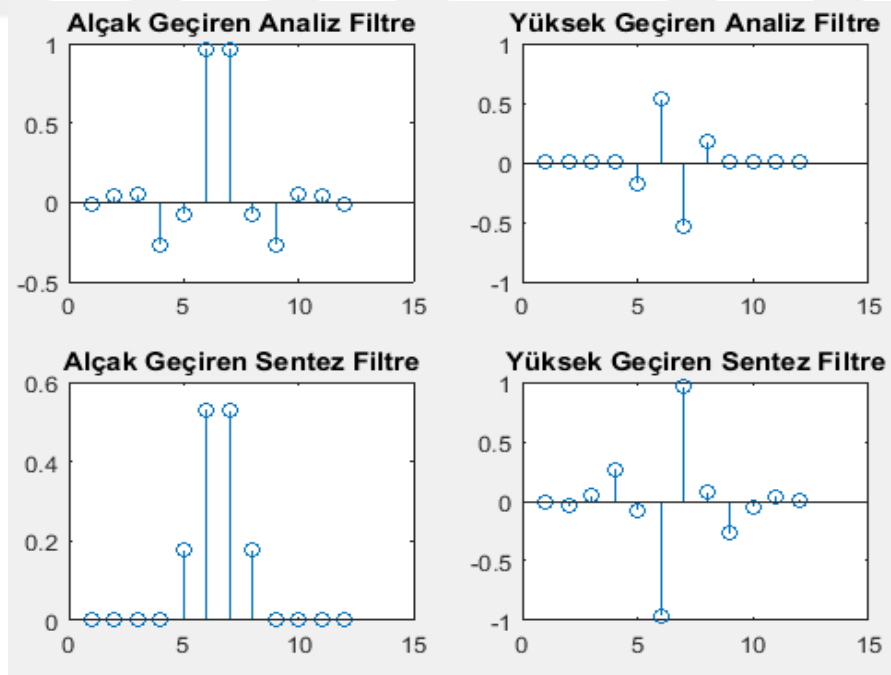
Biortogonal dalgacık, Mallat ve Zhong tarafından önerilmiş dalgacık dönüşümünün farklı ayrıştırma ve yeniden yapılandırma filtreleri ile düzenlenmesi sonucu oluşturulmuş bir türüdür [12]. Ortogonal dalgacık aileleri yerine, biorthogonal dalgacığı oluşturmak için analizde kullanılan dalgacık, sentezde kullanılan farklı oluşturulur [46]. Biortogonal dalgacıklar, ortogonal dalgacıklara oranla daha fazla serbestlik derecesine sahiptir [47].

Biortogonal dalgacıklar sinyal ve görüntünün sentezi için gerekli olan lineer faz özelliklerini gösterirler. İlginç özellikleri türetmek için tek bir sinyal yerine analiz için bir tane ve sentez için de bir tane olmak üzere iki dalgacık kullanırlar [48]. Bu tez çalışmasında görüntüdeki gürültüleri elimine etmek amacıyla Biortogonal 3.5 dalgacığı kullanılmıştır. Kullanılan dalgacığa ait ölçekleme fonksiyonu ve dalgacığa ait

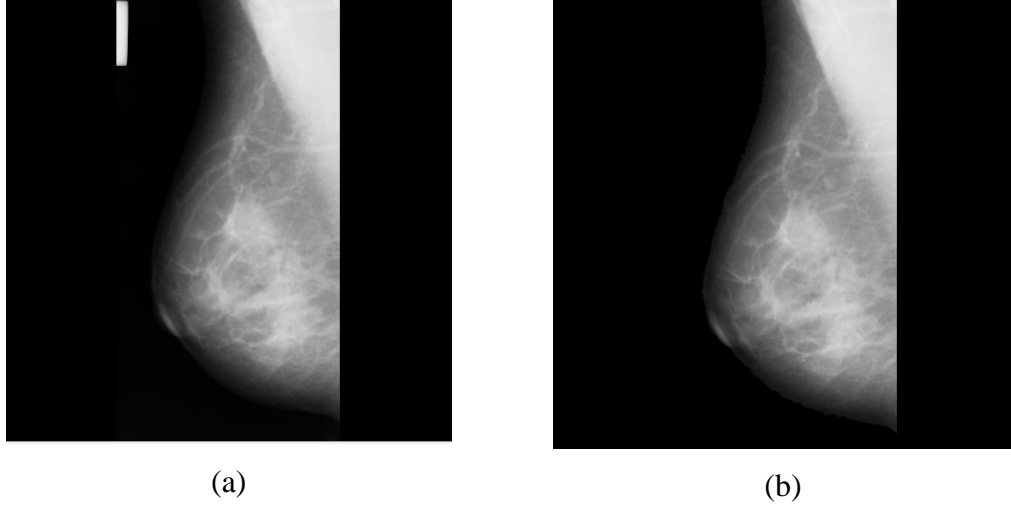
görüntüler Şekil 2.9’ da, analiz ve sentez için kullanılan filtreler Şekil 2.10’da, örnek bir mamogram görüntüsü üzerinde uygulaması Şekil 2.11’de sunulmuştur.



Şekil 2.9. Biortogonal dalgacık 3.5 için ölçekleme fonksiyonu ve dalgacık.



Şekil 2.10. Biortogonal 3.5 ayrışma ve yeniden yapılanma filtreleri.



Şekil 2.11. Biortogonal dalgacık ile gürültü giderme (a) Orijinal mamogram  
(b) Biortogonal dalgacık 3.5 ile gürültü giderilmiş mamogram görüntüsü.

#### *Kontrast Sınırlı Adaptif Histogram Eşitleme(CLAHE)*

Medikal görüntüleme, insan vücudundaki anormallikleri tanımlamak amacıyla yapılan manyetik rezonans görüntüleme, bilgisayarlı tomografi, ultrason görüntüleme gibi farklı şekillerde elde edilebilen bir tekniktir [49]. Fakat bu görüntüler elde edilirken farklı faktörlerden etkilenir ve analizinin daha iyi yapılabilmesi için görüntünün iyileştirilmesine ihtiyaç duyulur [49]. Medikal görüntülerde görüntü iyileştirme amacıyla farklı yöntemler kullanılmaktadır. Bunlardan bir tanesi de histogram eşitlemedir. Fakat mamogram görüntülerinde arka plan bölgesinin büyük olması, histogram eşitleme yönteminin görüntüyü iyileştirmede yetersiz kalmasına neden olmaktadır. Bu nedenle görüntü iyileştirme amacıyla Kontrast Sınırlı Adaptif Histogram Eşitleme(CLAHE) yöntemi geliştirilmiştir. CLAHE yöntemi ile ilgili literatürde pek çok çalışma mevcuttur [50-54] ve bu yöntem özellikle medikal görüntülerde iyi sonuçlar üretmektedir [55].

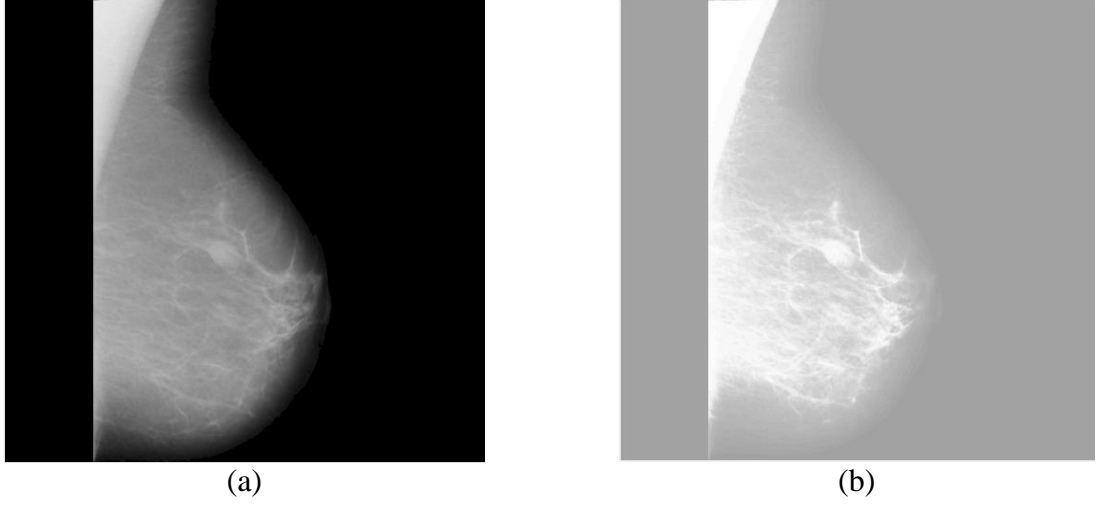
CLAHE yöntemi temelde görüntüyü yaklaşık olarak birbirine eşit ve birbiriyle örtüşmeyen bölgelere böler. Örneğin 64x64 boyutundaki bir görüntüde iyi istatistiksel bir tahminde bulunmak için görüntü her yönde eşit olarak 4'e bölünerek 16'ya eşit olacak şekilde seçilir. Bölümlenmeye ait bir görüntü şekil 2.12'de verilmiştir.

KB	SB	SB	KB
SB	IB	IB	SB
SB	IB	IB	SB
KB	SB	SB	KB

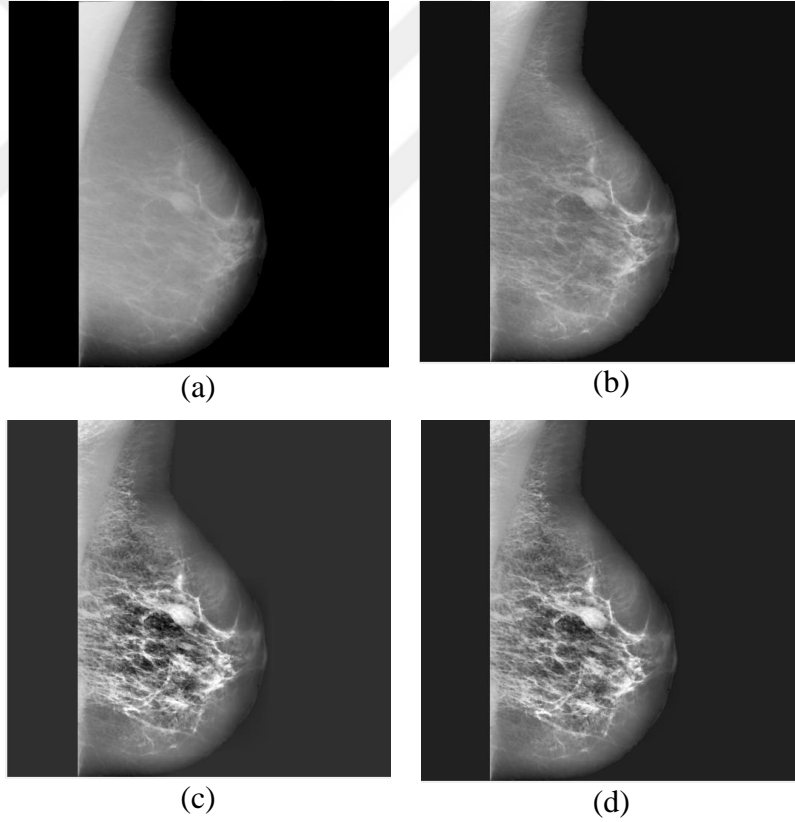
Şekil 2.12. 64x64 boyutundaki görüntünün 16 eşit kareye bölünmüş görüntüsü [55].

Bu bölümlenme sonucunda 3 grup elde edilir. Bunlar; köşe bölgelerinin oluşturduğu köşe (KB) grubu, 8 bölgeden oluşan sınır bölge grubu (SB) ve yine dört bölgeden oluşan diğer bölgelerin içinde yer alan iç bölgeler sınıfı (IB)' dir [55]. Bu yaklaşımda öncelikle her bölgenin histogramı hesaplanır. Ardından, kontrast genişlemesi için bir kırpma sınırı elde edilir. Daha sonra, her histogram, yüksekliği klip sınırının ötesine geçmeyecek şekilde yeniden dağıtılır. Son olarak, gri seviye haritaları için kontrast sınırlı histogram sonucunda elde edilen kümülatif dağılım fonksiyonları belirlenir [55]. CLAHE tekniğinde pikseller, dört en yakın komşuları kullanılarak haritalanır ve bilineer interpolasyon yöntemi ile birleştirilir.

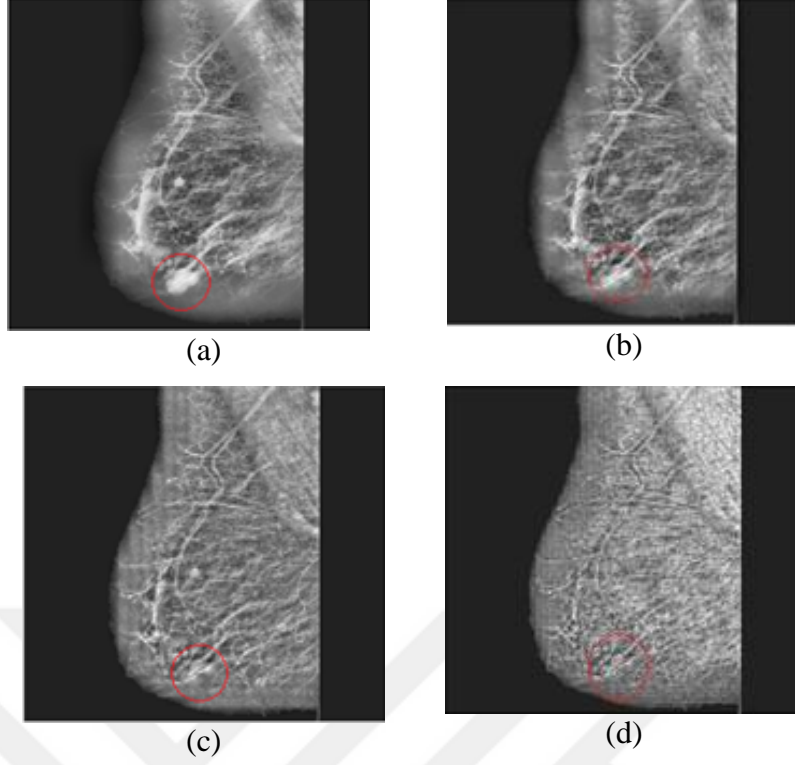
CLAHE yönteminin klasik histogram eşitleme ile farklılıklarını gösterebilmek için ilk olarak örnek bir mamogram üzerinde klasik histogram eşitleme sonucu Şekil 2.13 de gösterilmiştir. CLAHE yönteminde kırpma değerinin sonucu nasıl etkilediğini gösteren bir uygulama şekil 2.14'de, her yönde eşit bölge sayısının CLAHE sonucunu nasıl etkilediğini gösteren örnek bir uygulama ise Şekil 2.15'de verilmiştir.



Şekil 2.13. Histogram eşitleme yönteminin örnek bir mamogram üzerinde uygulanması  
(a) Orijinal mamogram (b) Histogram eşitleme sonucu mamogram.



Şekil 2.14. CLAHE uygulaması ve kırpma sınır değerlerinin etkisi (a) Orijinal mamogram (b) 0.01 kırpma sınırlı CLAHE sonucu mamogram (c) 0.1 kırpma sınırlı CLAHE sonucu mamogram (d) 0.05 kırpma sınırlı CLAHE sonucu mamogram.



Şekil 2.15. Tüm yönlerde bölünen bölge sayısının CLAHE yöntemine etkisi (a) Bölünen bölge sayısı=8 (b) Bölünen bölge sayısı=16 (c) Bölünen bölge sayısı=32 (d) Bölünen bölge sayısı=64.

Şekil 2.14'de de görülebileceği gibi CLAHE yönteminde kırpma sınır değeri mamogram görüntüsünün iyileştirilmesinde belirli bir fark oluşturmaktadır. Şekil 2.14-b'de düşük bir kırpma değeri belirlendiğinde iyileştirme daha az olurken yüksek bir kırpma değeri belirlendiğinde (Şekil 2.14-c) kenarlar bulanıklaşmaktadır. Bu nedenle bu tez çalışmasında CLAHE yöntemi için kırpma değeri=0.05 olarak belirlenmiştir. Aynı şekilde Şekil 2.15 incelendiğinde tüm yönlerde belirlenen bölge sayısının etkilerini bakılırsa, bölge sayısının artmasının kitle teşhisinde olumsuz bir etkiye neden olduğu söylenebilir. Şekil 2.15'de kırmızı çember ile işaretlenen kitle görüntüsünün, bölümlenen bölge sayısı arttıkça kaybolduğu görülmektedir. Bu nedenle CLAHE yöntemi için tüm yönlerde bölünen bölge sayısı=8 olarak belirlenmiştir.

#### *Anizotropik Difüzyon Yöntemi*

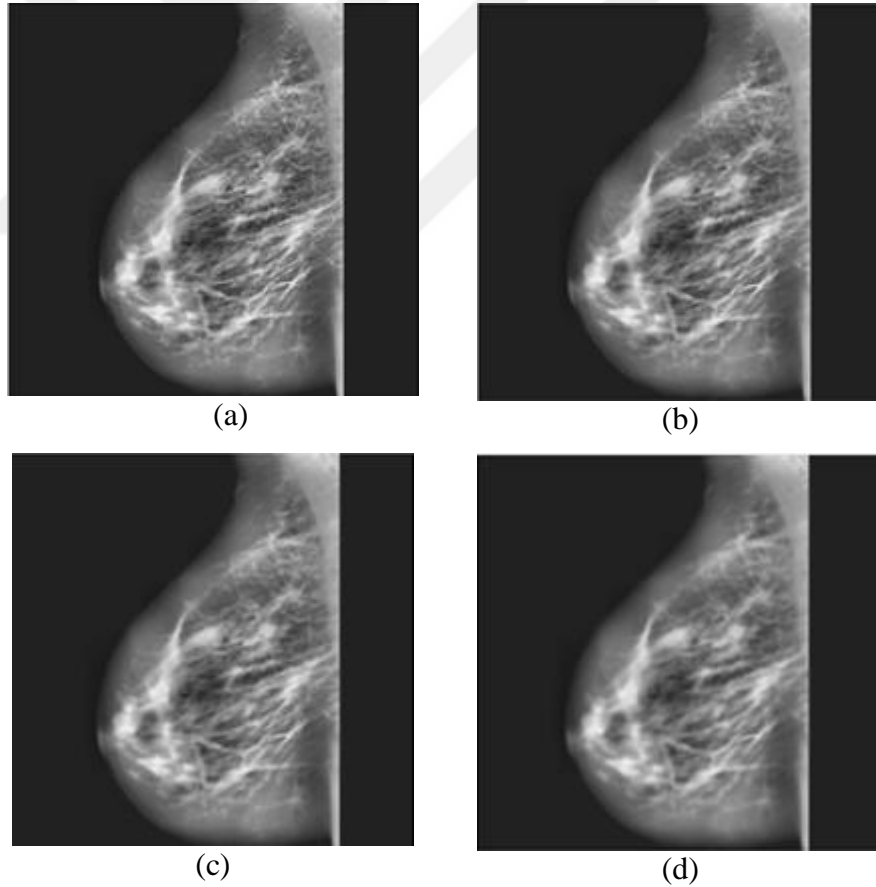
Önişleme amacıyla geliştirilen BDT sisteminde kullanılan bir diğer yöntem anizotropik difüzyon yöntemidir. Bu yöntemin bir diğer ismi Perona-Malik difüzyonudur. Bu yöntemde görüntünün önemli kısımları örneğin çizgileri, kenarları veya görüntünün yorumlanması için önemli olan diğer ayrıntıları korunurken görüntü gürültüsü azaltılır.

Anizotropik difüzyon filtreleri, tutarlı devamlılık gösteren tek boyutlu yapılardaki kesintiye uğramış kısımları düzeltmede ve yüksek oranda bozulan kenarların düzeltilmesi gibi bazı uygulamalarda izotropik olanlardan daha iyi performans gösterebilir [56].

Difüzyon denklemi aşağıdaki gibi formüle edilebilir [57].

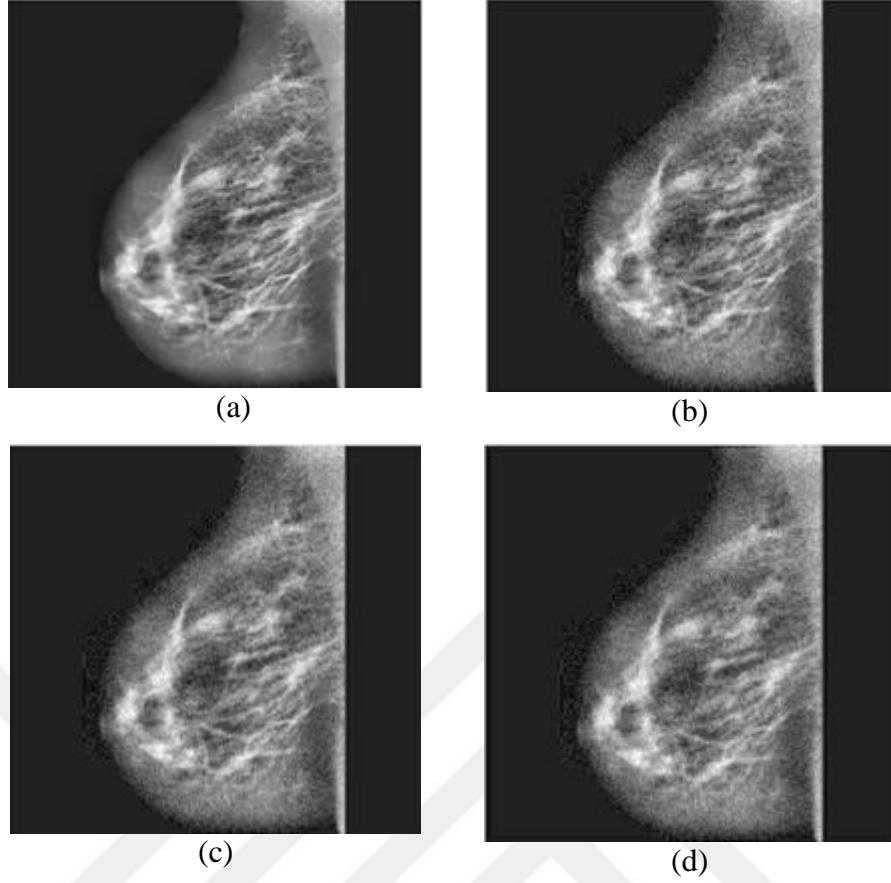
$$\frac{\partial I}{\partial t} = \text{div}(c(x, y, t)\nabla I) = \nabla c \cdot \nabla I + c(x, y, t)\Delta I \quad (2.3)$$

Burada  $\text{div}(\cdot)$  sapma operatörünü,  $\nabla$  gradyanı,  $\Delta$  Laplası ve  $c(x,y,t)$  ise difüzyon katsayısını simgeler.  $c(x, y, t)$  görüntünün kenarlarını korur ve genellikle difüzyon hızını kontrol eden görüntü gradyanının bir fonksiyonu olarak seçilir. Seçimin şekline göre Perona Malik Tip I ve Perona Malik Tip II olarak isimlendirilir. Anizotropik Difüzyon yönteminin örnek bir mamogram üzerindeki sonuçları Şekil 2.16'da ve Şekil 2.17'de verilmiştir.



Şekil 2.16. Farklı difüzyon zamanları ve  $\Delta=0.25$  için anizotropik difüzyon örneği

(a) Difüzyon zamanı=2 (b) Difüzyon zamanı=6 (c) Difüzyon zamanı=10 (d)  
Difüzyon zamanı=16.



Şekil 2.17. Farklı difüzyon zamanları ve delta( $\Delta$ )=0.5 için anizotropik difüzyon örneği (a) Difüzyon zamanı=2 (b) Difüzyon zamanı=6 (c) Difüzyon zamanı=10 (d) Difüzyon zamanı=16.

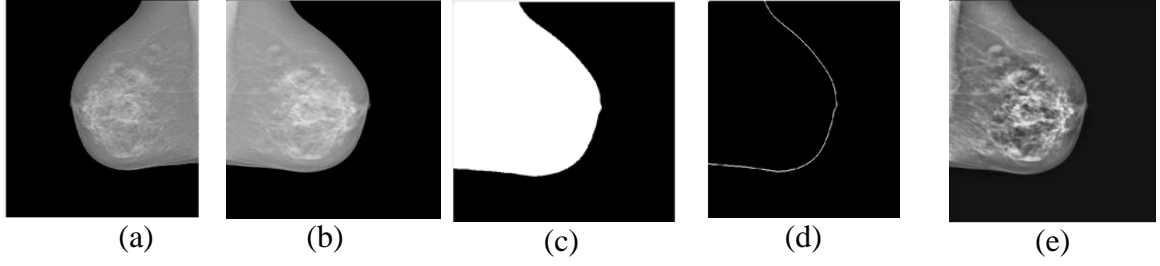
Şekil 2.16 ve Şekil 2.17’de de görülebileceği üzere difüzyon zamanı arttıkça bulanıklaşma artmakta ve delta( $\Delta$ ) değeri büyüdükçe bozunma artmaktadır. Bu nedenlerden dolayı bu tez çalışmasında anizotropik difüzyon yöntemi için Perona Malik Difüzyon Tip II,  $\Delta = 1/4$ ;  $\nabla = 30$ ; difüzyon zamanı ise 2 olarak seçilmiştir.

#### *Dijital Mamogramlardan Pektoral Kasın Çıkarılması*

BDT sistemleri, radyoloji uzmanlarına medikal görüntülerdeki anormalliklerin tespitinde yardımcı olmaktadır. MLO görünümlü mamogramlarda sıklıkla pektoral kas da yer almaktadır. Pektoral kas ise kitle vb. anormal yapılarla benzer özellikler gösterdiğinden bilgisayar destekli teşhis çalışmalarında şüpheli bölgelerle karıştırılmakta ve doğru teşhisi zorlaştırmaktadır [58]. Bu nedenle pektoral kasın mamogramlardan çıkarılması doğru teşhis için önemlidir. Pektoral kasın MLO görünümlü mamogramlardan otomatik olarak çıkarılması gerekli bir adım olmasına karşın mamogramlardaki şekil, büyüklük yoğunluk gibi değerlerin mamogramdan

mamograma farklılık göstermesi bu işlemi zorlaştırmaktadır. Literatüre bakıldığında bu alanda yapılmış çalışmaların dört ana grupta toplandığı görülmektedir. Bunlar yoğunluğa bağlı yaklaşımlar [59], [60], çizgiye dayalı tespit yaklaşımları [61], [62], dalgacık temelli bölütleme [63], [64] ve istatistiksel [63]-[65] yöntemlerdir [66] .

Bu tez çalışmasında mamogram görüntülerinden pektoral kas bölgesi sınır tespiti yapılması amacıyla SSEM adında yeni bir yöntem önerilmiştir. Bu yöntemde ilk olarak görüntü üzerinde şu ön işlemler gerçekleştirilmiştir. Birinci aşamada işlem karmaşıklığının azaltılması amacıyla görüntü 512x512 boyutuna, renk skalası da 256'ya indirgenmiştir (Şekil 2.18-a). Daha sonra bütün görüntüler pektoral kas bölgesi sol üst köşede olacak şekilde düzenlenmiştir. Bunun için pektoral kas bölgesi sağ üst köşede bulunan mamogramların ayna görüntüsü alınarak pektoral kas sol üst köşeye gelecek şekilde döndürülmüştür. Pektoral kas sol üst köşede bulunan mamogramlarda ek bir işlem yapılmamıştır (Şekil 2.18-b). Mamogram görüntüsü deneysel olarak elde edilen bir eşik değeri ( $\delta=0.09$ ) ile ikilik görüntüye dönüştürülmüştür (Şekil 2.18-c). Bu işlem sonucunda ikilik görüntüde yer alan en büyük bölge mamogram bölgesidir. Bu nedenle en büyük alana sahip bölge görüntüde kalacak şekilde bir filtre uygulanmıştır. Elde edilen görüntü iki piksel büyüklüğünde yapı elemanı kullanılarak eritilmiş ve bir önceki aşamada bulunan mamogram bölgesinden çıkartılmıştır (Şekil 2.18-d). Böylece mamogram kaba sınırı elde edilmiştir. Mamogram bölgesinin elde edildiği görüntü ile orijinal görüntü çarpılarak sadece mamogram görüntüsünün olduğu yeni görüntü elde edilmiştir. Daha sonra gürültü giderilmesi ve görüntünün iyileştirilmesi amacıyla 3x3'lük medyan filtre ve biortogonal dalgacık kullanılmıştır. Son olarak her bir mamogram adaptif histogram eşitleme ve anizotropik difüzyon yöntemleriyle iyileştirilmiş ve gürültülerden temizlenmiştir (Şekil 2.18-e).



Şekil 2.18. Pektoral kas çıkartılması amacıyla gerçekleştirilen ön işlemler  
 (a) Orijinal mamogram (b) 512x512 boyutlarına ve 256 renk skalasına indirgenmiş, ayna görüntüsü alınmış mamogram (c) Meme bölgesi bölütlenmiş mamogram  
 (d) Meme kaba sınırı çıkartılmış mamogram.

#### *Tek Yönlü Kenar İşaretleme Yöntemi (SSEM)*

Bu çalışmada önerilen yöntem için pektoral kasın bazı belli başlı karakteristik özelliklerinden faydalanılmıştır. Bu özellikler şu şekilde sıralanabilir:

- ✓ Pektoral kas bölgesi kabaca üçgensel bölgedir.
- ✓ Pektoral kas sınırı yaklaşık düz bir çizgidir.
- ✓ Pektoral kas ile göğüs bölgesi arasında belli bir yoğunluk değişimi vardır.
- ✓ Pektoral kas bölgesi kabaca homojendir.

SSEM yöntemi mamogramların geometrik özellikleri ile kas ve meme dokusu arasındaki yoğunluk farkını esas almaktadır. Mamogramların geometrik özelliklerine bakıldığında yukarıdan aşağıya daralan kabaca üçgensel bir bölge olduğu görülmektedir. Ön işlem aşamasında pektoral kas bölgesi sol üst köşeye getirildiği için kenar bulma işlemi de bu yöntemde sağdan sola ve yukarıdan aşağıya yaklaşık  $30^{\circ}$ - $45^{\circ}$  lik bir açıyla gerçekleştirilir. SSEM yönteminin detayları şu şekilde ifade edilebilir.

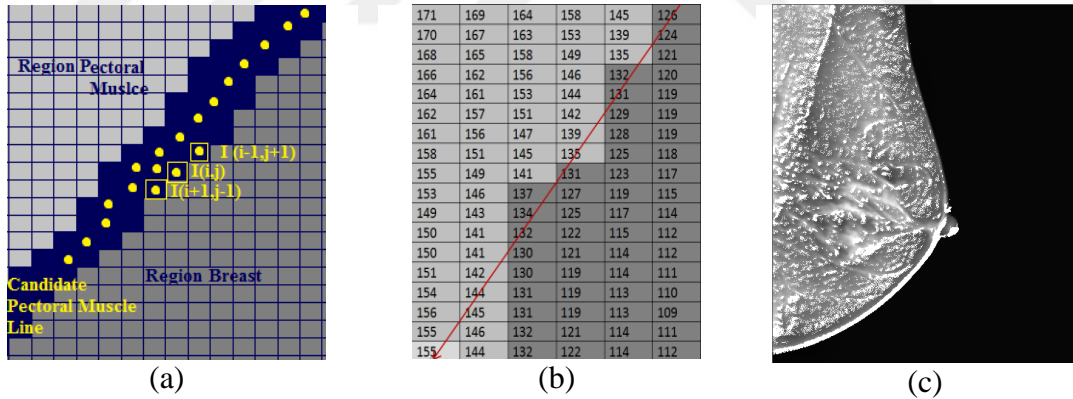
- İlk olarak kenar tespiti işlemlerinde kullanılmak üzere üç farklı eşik değeri belirlenmiştir. Bunlardan birincisi görüntülerdeki mamogram olmayan bölgelerin hesaplamaya dâhil olmaması için kullanılan eşik değeridir ( $\phi$ ), bu değer 5 olarak alınmıştır. İkincisi ise piksellerin yoğunluk değerinin birbirine ne kadar benzediğini belirleyen ( $\alpha$ ) değeridir ve bu çalışmada 1 olarak alınmıştır. Üçüncü eşik değeri ise pektoral kas ve meme bölgesi arasındaki yoğunluk farkı için kullanılan eşik değeridir ( $\theta$ ). Bu değer de bu çalışmada 1 olarak

belirlenmiştir.  $\phi$  değerinin küçük tutulmasının nedeni pektoral kas sınırının olabildiğince çok noktayla işaretlenmesini sağlamaktır.

- Ardından mamogram görüntüsü ( $I$ ) piksel piksel taranarak her bir piksel ( $I_{i,j}$  ( $i = 2,3,\dots,M ; j = 2,3, \dots, N - 1$ ), burada  $M$ ,  $I$  matrisinin satır sayısı,  $N$  ise sütun sayısıdır) sağ, üst ve alt yönlerde iki piksel uzaklığında bulunan komşuları ile birlikte değerlendirmeye alınır. Bu değerlendirmede aşağıdaki koşullar kullanılır. Bu koşullara göre yapılan değerlendirme sonucunda üzerinde pektoral kas sınırı işaretlenmiş yeni mamogram görüntüsü  $I_{new}$  elde edilir.

$$\left. \begin{array}{l} I(i,j) > \phi \\ \text{and} \\ \left( |I_{i,j} - I_{i+1,j-1}| \leq \alpha \text{ or } |I_{i,j} - I_{i-1,j+1}| \leq \alpha \right) \\ \text{or} \\ \left( |I_{i,j} - I_{i-1,j}| \leq \alpha \text{ or } |I_{i,j} - I_{i,j-1}| \leq \alpha \right) \\ \text{and} \\ \left( |I_{i,j} - I_{i,j+1}| \geq \phi \text{ and } |I_{i,j} - I_{i+1,j+1}| \geq \phi \right) \\ \text{and} \\ |I_{i,j} - I_{i+1,j}| \geq \phi \end{array} \right\} \text{then} \left. \begin{array}{l} I_{new,i,j} = 255 \\ \text{and} \\ I_{new,i+1,j-1} = 255 \\ \text{and} \\ \text{if } (j-2) \neq 0 \text{ then } I_{new,i+2,j-2} = 255 \end{array} \right\} \text{else } I_{new,i,j} = I_{i,j}$$

Pektoral kas sınırının işaretlenmesi için renk yoğunluğunun 255 olarak seçilmesinin nedeni mamogram üzerinde daha az rastlanan ve ikilik görüntüde ayırt edilmesi kolay olan bir değer olmasıdır. SSEM yönteminin grafiksel gösterimi ve örnek bir uygulama Şekil 2.19’da sunulmuştur.

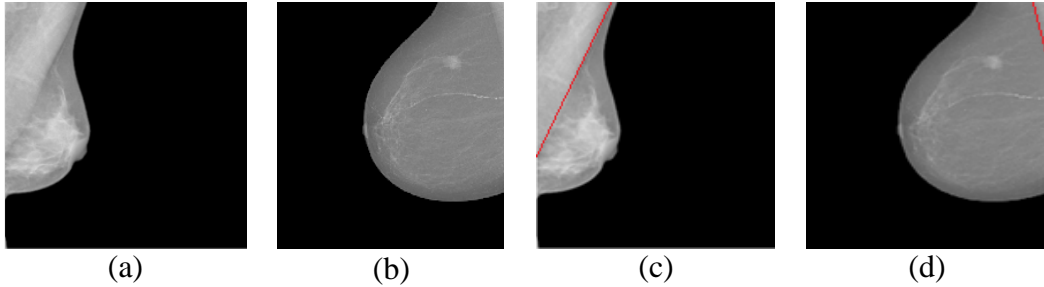


Şekil 2.19. SSEM yönteminin grafiksel gösterimi ve örnek bir uygulama (a) SSEM yöntemi ile sınır tespiti için grafiksel gösterim (b) 6x18 boyutlu kas ve meme bölgesine ait örnek bir gri seviye matrisi (c) SSEM yöntemi sonucunda pektoral kas sınırı işaretlenmiş yeni mamogram görüntüsü.

#### Aday Pektoral Kas Bölgesi ve Sınırının Tespiti

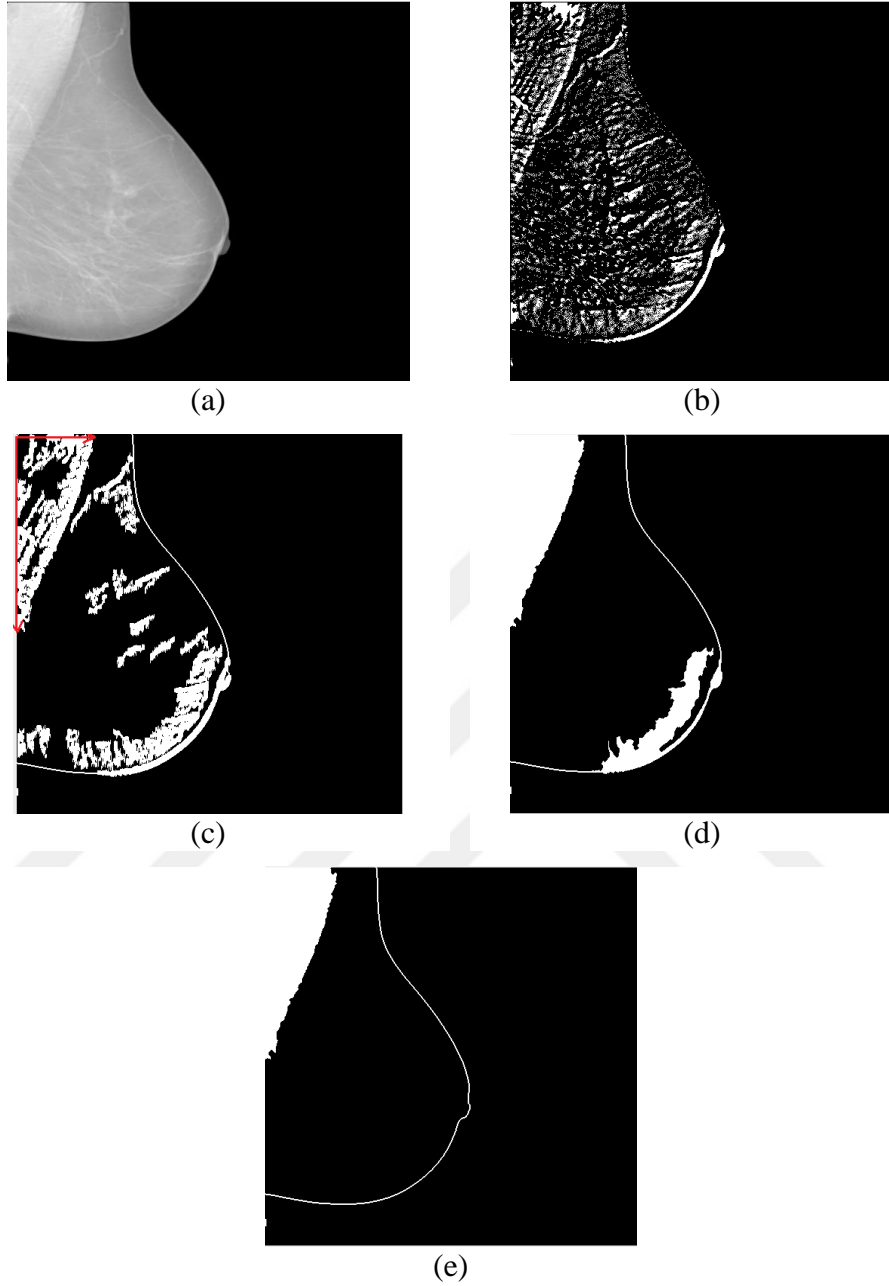
SSEM yöntemi sonucunda pektoral kas sınırı işaretlenmiş yeni mamogram görüntüsünden aday pektoral kas bölgesini bulabilmek amacıyla, bu çalışmada, deneysel olarak elde edilen ve  $Alan_{min}$  ve  $Alan_{max}$  olarak isimlendirilen iki farklı

değişken kullanılmıştır. Bunlardan  $Alan_{min}$  muhtemel en küçük pektoral kas alanını  $Alan_{max}$  ise yine muhtemel en büyük mamogram bölge alanının yarısını ifade etmektedir. Şekil 2.20’de  $Alan_{max}$  ve  $Alan_{min}$  değerlerine ait örnek mamogram görüntüleri yer almaktadır.



Şekil 2.20. Pektoral kas için  $Alan_{max}$  ve  $Alan_{min}$  örnekleri (a),(b) Orijinal mamogram (c) Pektoral kas bölgesi için  $Alan_{max}$  örneği (d) Pektoral kas bölgesi için  $Alan_{min}$  örneği.

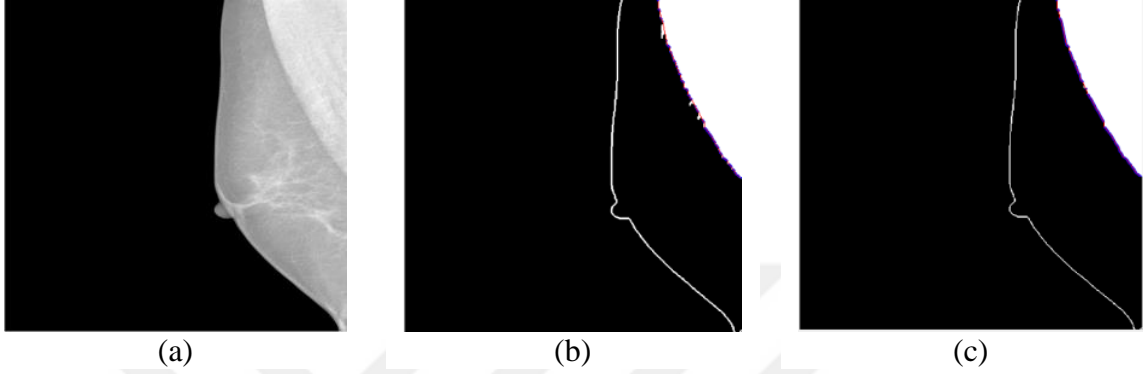
SSEM yöntemi kullanılırken aday pektoral kas sınırındaki pikseller gri seviye renk değeri 255 ile işaretlenmiştir. Bu nedenle SSEM sonucu elde edilen görüntü 0.99 eşik değeri ile ikilik resme dönüştürülür (Şekil 2.21-b) ve morfolojik operatörlerle pektoral kas sınırı güçlendirilir (Şekil 2.21-c). İkilik resme dönüştürülmüş ve pektoral kas sınırı güçlendirilmiş görüntü filtreler yardımıyla  $Alan_{min}$ 'dan daha küçük görüntülerden temizlenir. Temizlenmiş görüntü ile daha önceden elde edilen kaba meme bölgesi birleştirilir. Sonra pektoral kas alanını oluşturmak için sol üst köşeden başlayarak yukarıdan aşağıya ve soldan sağa pektoral kas köşesini kapsayan bir çizgi çizilir (Şekil 2.21-c). Bu çizgiler ile aday pektoral kas bölgesi sınırları arasında kalan boşluklar doldurulur. Pektoral kas bölgesini tespit etmek için en büyük alana sahip iki bölge seçilir (Şekil 2.21-d). Bu bölgelerin alanları ve ağırlık merkezleri kullanılarak konum olarak diğer bölgeye göre sol üstte bulunan ve alanı  $Alan_{min}$ 'dan büyük ve  $Alan_{max}$ 'dan daha küçük olan bölge pektoral kas bölgesi olarak seçilir (Şekil 2.21-e).



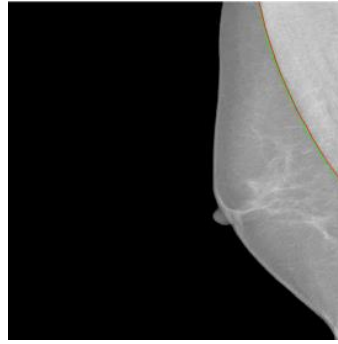
Şekil 2.21. SSEM yöntemiyle kaba pektoral kas alanının belirlenmesi (a) Orijinal mamogram (b) SSEM sonucu elde edilen görüntünün ikilik formu (c) Morfolojik işlemler sonrası elde edilen mamogram (d) En büyük iki alanı seçilmiş mamogram (e) Kaba pektoral kas alanı belirlenmiş mamogram.

Aday pektoral kas bölgesi sınırında bazen morfolojik işlemler sonrasında kalan bozukluklar olabilmektedir (Şekil 2.22-b). Bu bozuklukları gidermek amacıyla pektoral kas bölgesine ait sınırlar çıkartılırken SSEM yöntemine benzer şekilde yukarıdan aşağıya ve soldan sağa yaklaşık 30°-45° 'lik açı ile pektoral kasın konum bilgileri

alınmış ve kas sınırı oluşturulmuştur. Son olarak, elde edilen sınırlar üzerindeki eksik noktaların tamamlanması amacıyla doğrusal interpolasyon yöntemi kullanılmış ve bu işlem sonucunda elde edilen sınır nihai kas sınırı olarak kabul edilmiştir (Şekil 2.22-c). Sonuç olarak örnek bir mamogram için elde edilen nihai kas sınırı ile uzmanlar tarafından aynı mamogram için verilen ground truth değerleri kullanılarak çizilen kas sınırı Şekil 2.23’de birlikte sunulmuştur.

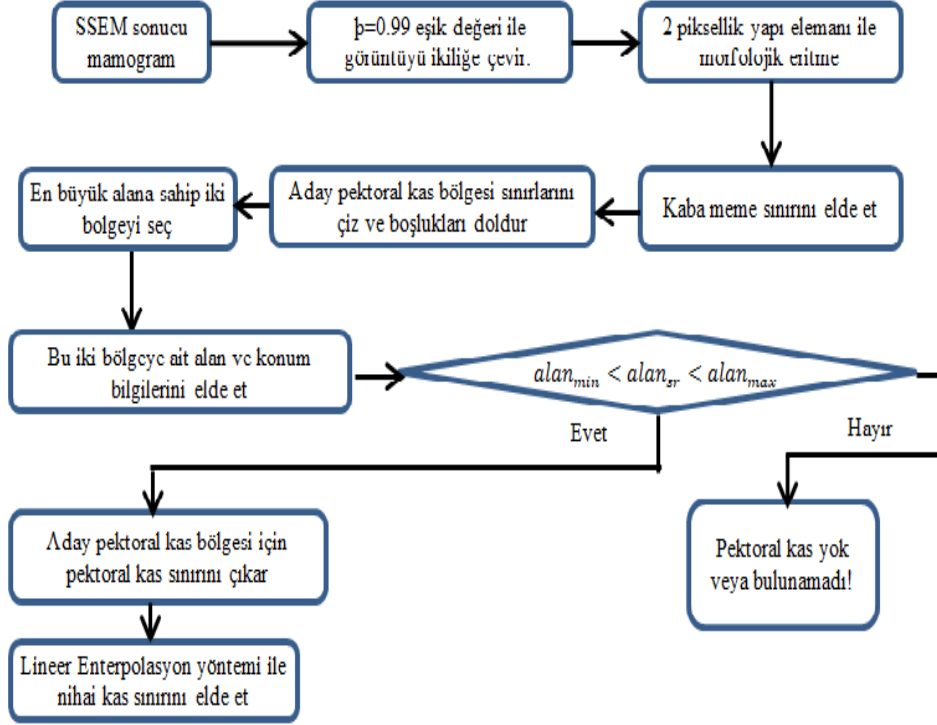


Şekil 2.22. Doğrusal interpolasyon yöntemi ile pektoral kas sınırının düzeltilmesi  
(a) Orijinal mamogram (b) Kaba pektoral kas sınırları elde edilmiş mamogram görüntüsü (c) Doğrusal enterpolasyon yöntemi ile nihai sınırların elde edildiği mamogram görüntüsü.



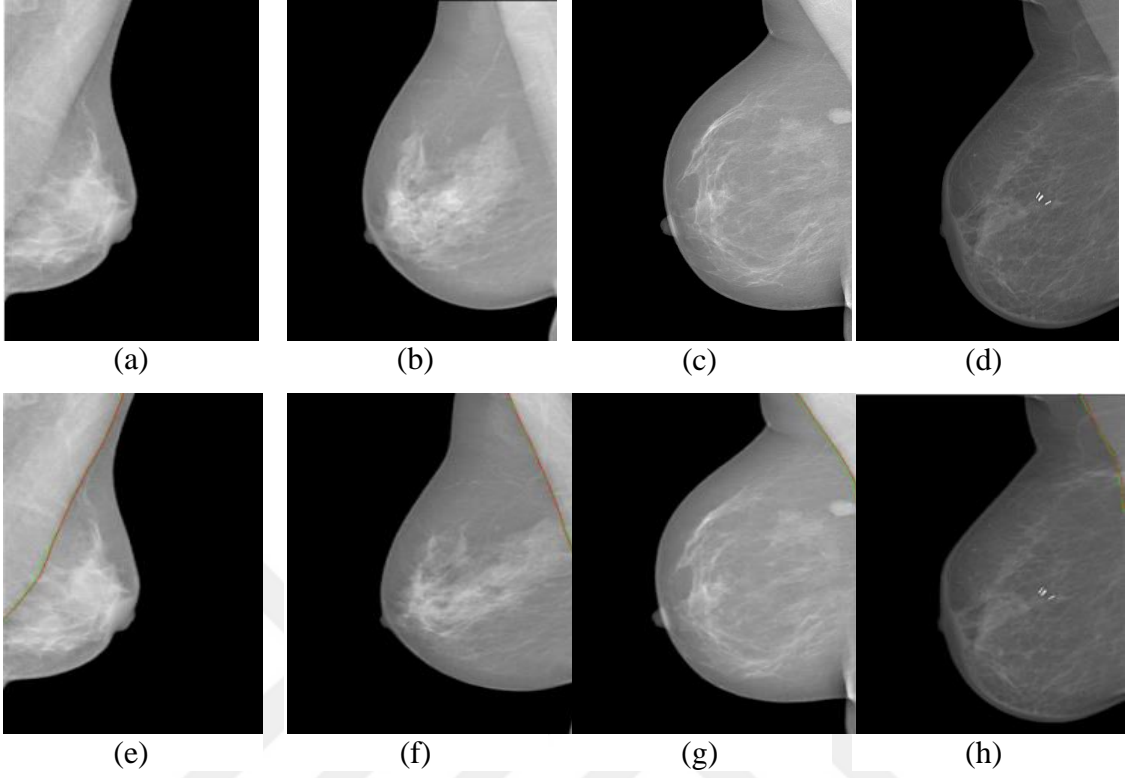
Şekil 2.23. Önerilen yöntemin örnek bir mamograma uygulanması sonucu elde edilen görüntü (uzmanlar tarafından verilen ground truth değerleri kullanılarak çizilen kas sınırı (kırmızı) ve önerilen yöntem ile bulunan kas sınırı(yeşil)).

Son olarak aday pektoral kas bölgesi ve sınır tespiti için kullanılan tüm işlemlere ait akış diyagramı, Şekil 2.24’de sunulmuştur.



Şekil 2.24. Aday pektoral kas bölgesi ve sınırının tespitine ait akış diyagramı.

Ayrıca önerilen yöntemin farklı mamogram görüntüleri üzerindeki performansını gösterebilmek amacıyla dört ayrı mamogram görüntüsü için yapılan pektoral kas sınır tespiti uygulamalarına ait görüntüler Şekil 2.25’de sunulmuştur.



Şekil 2.25. Önerilen yöntemin farklı mamogram görüntüleri üzerindeki performansı (a-d) Orijinal mamogram görüntüleri (e-h) Uzmanlar tarafından verilen ground truth değerleri kullanılarak çizilen kas sınırları (kırmızı) ve önerilen yöntem ile bulunan kas sınırları(yeşil).

Yapılan tüm bu çalışmalar sonunda önışlem aşaması tamamlanmıştır. Bir sonraki aşamada önışlem aşamasında elde edilen gürültü giderilmiş ve iyileştirilmiş mamogram görüntüsünden, şüpheli bölgelerin belirlenmesi amacıyla yapılan çalışmalar sunulacaktır.

### 2.3. ŞÜPHELİ BÖLGELERİN (ROI) BÖLÜTLENMESİ

BDT sistemindeki her aşama doğru teşhis için oldukça önemlidir. Fakat ilgili anormalliklerin tespiti için kullanılacak olan şüpheli bölgelerin belirlenmesi daha da önemlidir. Çünkü bu bölgeler üzerinde kitle ve kalsifikasyon aday bölgeler belirlenir ve sonraki aşamalarda belirlenen bu bölgeler üzerinden ilgili anormallikler tespit edilir. Bu nedenle, bu aşama ileri tespit ve teşhis aşamaları için oldukça önemlidir. ROI ile bölütleme için temelde iki yaklaşım kullanılır. Bunlardan ilki konu uzmanları tarafından

tespit edilen bölgenin manuel olarak elde edildiği yaklaşımdır. Bu yöntemi kullanan sistemler yarı otomatik olarak adlandırılırlar. Diğeri ise ROI bölgelerinin otomatik olarak tespit edildiği yaklaşımdır. ROI bölgelerinin otomatik olarak elde edilmesi meme dokusunun özellikleri, kalsifikasyonlar ve pektoral kas gibi tespiti bozucu etkenler nedeniyle oldukça zordur. Dolayısıyla, günümüzde dijital mamografide otomatik, doğru, sağlam ve verimli bir ROI bölütleme tekniğini bulmak hala bir zorluk olarak karşımıza çıkmaktadır [67]. Literatüre bakıldığında meme kanserinde ROI tespiti ile ilgili pek çok çalışma bulunduğu görülmektedir [67]-[72]. Bu tez çalışmasında ROI bölütleme için Havrda & Charvat Entropi, Otsu N Eşikleme ve bu tez çalışması kapsamında geliştirilen bir görüntü kümeleme algoritması olan w-BSAFCM algoritması kullanılmıştır. w-BSAFCM algoritması Bulanık Kümeleme Algoritması (FCM) ile yeni bir optimizasyon algoritması olan Back Tracking Search (BSA) optimizasyon algoritmasının hibrit bir şekilde kullanılması ile geliştirilmiştir. FCM algoritması literatürde oldukça sık kullanılan ve birçok farklı versiyonları geliştirilmiş bir kümeleme algoritmasıdır [73]-[76]. Ancak bu algoritmanın özellikle başlangıç değerlerinin seçimi ve lokal minimuma takılma şeklinde iki önemli problemi vardır ve yine literatüre bakıldığında bu problemlerin çözümü için FCM ile farklı optimizasyon algoritmalarının birleştirildiği görülmektedir [77]-[81]. Bu çalışmaların temel amacı FCM'nin en önemli dezavantajları olan lokal minimuma takılma ve başlangıç değerlerinden etkilenme sorunlarını çözmektir. Diğer taraftan BSA henüz yeni sayılabilecek, 2013 yılında geliştirilmiş, bir algoritmadır. Yapılan literatür taramasına göre de tez çalışmasının yapıldığı ana kadar FCM ile BSA'nın birlikte kullanıldığı bir çalışmanın henüz yayınlanmamış olduğu görülmüştür. Dolayısıyla bu tez çalışmasında bu iki yöntem birleştirilerek, şüpheli bölgelerin segmentasyonu amacıyla kullanılmıştır.

Tez çalışmasında ROI segmentasyon için kullanılan w-BSAFCM, Havrda & Charvat Entropi ve Otsu N Eşikleme yöntemleri ile ilgili temel bilgiler ilerleyen bölümlerde sunulmuştur.

### **2.3.1. w-BSAFCM Algoritması**

w-BSAFCM algoritması segmentasyon amacıyla geliştirilen ve bulanık kümeleme algoritması (FCM) ile Pınar Çivicioğlu [82] tarafından geliştirilen BSA optimizasyon algoritmasının birleştirilmesiyle oluşturulan yeni hibrit bir algoritmadır. Bu algoritmanın lokal arama yeteneklerini geliştirmek için bu tez çalışmasında ayrıca bir

ağırlık parametresi ( $w$ ) tanımlanmış ve yeni geliştirilen algoritmaya  $w$ -BSAFCM adı verilmiştir. Geliştirilen algoritma için seçilen FCM ve BSA algoritmaların temel özellikleri bir sonraki bölümde sunulmuştur.

### 2.3.1.1. FCM (Fuzzy C-Means) Algoritması

FCM algoritması bir veri kümeleme algoritmasıdır. Bu algoritma Bezdek [83] tarafından 1981 yılında geliştirilmiştir. Görüntü işleme açısından FCM görüntünün benzer özelliklere sahip kümeler ayrılmasını sağlayan bir algoritmadır. Bu algoritma kümeleme işlemi için, her bir pikselin tüm kümelerde bir üyelik değeri bulunması temeline dayanır ve bu üyelik değerlerini kullanarak kendi amaç fonksiyonunu en küçük yapmaya çalışır. Kümeleme problemi, verilen bir veri setinde yer alan elemanların kendi aralarındaki ilişkilere göre  $c$  tane kümeye ayrılması olarak tanımlanabilir. Buna göre veri seti  $H = (h_1, h_2, \dots, h_m)$  şeklinde tanımlanabilir. Bu veri seti  $m$  tane üyeye sahiptir. Her bir üye,  $h_j$ , her bir küme için bir üyelik değerine sahiptir. Örneğin  $h_j$ 'nin  $i$ 'nci küme deki üyelik değeri  $u_{i,j}$ 'dir [84]. Tüm üyelerin tüm kümeler üzerinde sahip oldukları üyelik değerlerinin bir arada tutulduğu matrise üyelik matrisi denir ve  $U = [u_{i,j}] \in [0,1]_{c \times m}$  şeklinde gösterilir[84]. Bu matris aşağıdaki kriterleri sağlamalıdır [84].

$$\sum_{i=1}^c u_{i,j} = 1, 1 \leq j \leq m \quad (2.4)$$

$$0 \leq u_{i,j} \leq 1, 1 \leq i \leq c \quad (2.5)$$

$$0 \leq \sum_{j=1}^m u_{i,j} < m \quad (2.6)$$

Denklemlere göre bir verinin tüm kümelerdeki üyelik değerlerinin toplamı 1'e eşit olmalıdır ve her bir üyelik değeri 0 ile 1 arasında pozitif değerler almalıdır. Son olarak bir kümedeki tüm elemanların toplamı eleman sayısından küçük olmalıdır. FCM algoritması FCM, Denklemler (2.4)-(2.6)' da verilen kriterlere göre aşağıda verilen amaç fonksiyonunu en küçük yapmaya çalışır [84].

$$J = \sum_{i=1}^c \sum_{j=1}^m u_{i,j}^k D_{i,j}^2 \quad (2.7)$$

Denklem (2.7)'de  $k$ ,  $J$  ve  $D_{i,j}$  sırasıyla bulanık mantık sabiti, amaç fonksiyon ve  $j$  inci elemanın  $i$ 'nci kümenin merkezine olan uzaklığını ifade etmektedir. Denklem (2.7)' de yer alan  $D_{i,j}$  denklem aşağıdaki gibi hesaplanmaktadır [84].

$$D_{i,j} = \|v_i - h_j\| \quad (2.8)$$

Denklem (2.8)'de  $\| \cdot \|$  ifadesi Öklid uzaklığını,  $v_i$  ise  $i$ 'nci kümenin merkezini ifade etmektedir. İki veri arasında öklid uzaklığı hesaplanırken tek boyutlu veride (örneğin gri seviye resimlerde) iki verinin farkı alınır. Eğer veri iki veya üç boyutlu ise bu durumda iki veya üç boyutlu uzayda iki veri arasındaki uzaklık hesaplanır. Bu çalışmada gri seviye resimler kullanıldığı için bir veri ile küme merkezi arasındaki uzaklık bu iki verinin farkının mutlak değeri alınarak hesaplanmıştır. Her bir küme merkezi ise aşağıdaki gibi hesaplanmaktadır [84].

$$v_i = \frac{\sum_{j=1}^m u_{i,j}^k h_j}{\sum_{j=1}^m u_{i,j}^k} \quad (2.9)$$

Denklem (2.9)' da yer alan üyelik değerleri de Denklem (2.10) kullanılarak hesaplanmaktadır [84].

$$u_{i,j} = \frac{1}{\sum_{r=1}^c \left( \frac{D_{i,j}}{D_{r,j}} \right)^{2/(k-1)}} \quad (2.10)$$

Son olarak FCM algoritmasına ait akış diyagramı Şekil 2.26'da sunulmuştur.



Şekil 2.26. Klasik FCM algoritmasına ait akış diyagramı.

FCM algoritması için iki farklı durdurma kriteri kullanılabilir bunlardan ilki art arda üretilen iki üyelik matrisi arasında hesaplanan ve bu matrislerin elemanları arasındaki en büyük farkın belli bir sınır değerinden küçük olması şartıdır. Bu şart Denklem (2.11)'de verilmiştir. Diğeri ise belirli bir iterasyon sayısının belirlenmesi ve bu sayıda döngülerin sonlandırılmasıdır [84]. Biz bu çalışmada belli bir iterasyon sayısını durdurma kriteri olarak seçtik.

$$\max(U^{(l+1)} - U^{(l)}) < \varepsilon \quad (2.11)$$

Denklem (2.11)'de  $\max(U^{(l+1)} - U^{(l)})$  ifadesi  $l$  ve  $l + 1$  inci üyelik matrisleri arasındaki farkların en büyüğü,  $\varepsilon$  ise belirlenen eşik değeridir.

### 2.3.1.2. BSA (Backtracking search optimization algorithm) Algoritması

BSA algoritması Civicioğlu [82] tarafından 2013 yılında geliştirilmiş bir evrimsel optimizasyon algoritmasıdır. Bu algoritma iki tane çaprazlama ve mutasyon yöntemi kullanmaktadır [82]. Bu algoritmada ayrıca arama-yönü matrisi şeklinde bir matris kullanılmaktadır. Bu matris sayesinde BSA için bir hafıza tutulmaktadır ve bu hafıza rastgele belirlenecek bir seçimle değişene kadar aynı kalmaktadır. BSA temel olarak beş bölümden oluşmaktadır ve bu bölümler aşağıdaki gibi sıralanabilir [82].

*Başlangıç:* Bu aşamada optimizasyon için başlangıç nüfusu belirlenir. Başlangıç nüfusunun elemanlarının her bir çözülecek problem için birer aday çözümdürler. BSA için başlangıç nüfusu aşağıdaki gibi belirlenir [82].

$$S_{i,j} \sim R(\min_j, \max_j) \quad (2.12)$$

Denklemden,  $S_{i,j}$  ( $i = 1,2,3, \dots, n$  ve  $j = 1,2,3, \dots, d$ ) başlangıç nüfusunun ( $S$ 'in)  $i$ 'nci elemanının  $j$ 'inci boyuttaki elemanını ifade etmektedir (örneğin bir problemin çözümü için tanımlanan başlangıç nüfusu 30 elemanlı ve bu nüfusun her bir elemanı da iki boyutlu ise denklemden belirtilen  $j = 1$  ve ya  $2$  olabilir). Denklemden geçen  $n$  ve  $d$  sayıları, sırasıyla başlangıç nüfusunun eleman sayısı ve her bir elemanın boyut sayısıdır. Son olarak  $R$  ise normal dağılımı ifade etmektedir. Başlangıç aşamasında BSA ayrıca  $S$  için amaç fonksiyon değerlerini de belirlemektedir.

$$fitness = ObjectFunc(S) \quad (2.13)$$

Denklemden (2.13)'de  $fitness$   $n \times 1$  lik bir vektördür ve nüfusun her bir elemanı için hesaplanan amaç fonksiyon değerlerini içerir.  $ObjectFunc$  ise seçilen problemin çözümü için kullanılacak amaç fonksiyondur.

*Seçim I:* Bu bölümde BSA nüfusun farklı bir formu olan ve arama-yönü matrisinin hesaplanmasında kullanılacak olan  $oldP$  matrisini üretir. İlk aşamada  $oldP$  matrisi başlangıç nüfusu gibi rastgele üretilir [82].

$$oldP_{i,j} \sim R(\min_j, \max_i) \quad (2.14)$$

$oldP$ , ilk aşamada rastgele seçildikten sonra diğer döngülerde rastgele seçilmez. Diğer döngülerde  $oldP$  belirlenirken Denklemden (2.14)'de belirtilen kurallara göre seçim yapılır. Bu sayede BSA'nın bir hafızasının olması ve dolayısıyla önceki aramalarda elde edilen sonuçların diğer aramalara aktarılması sağlanır [82].

$$\text{Eğer } r_1 < r_2 \text{ ise } oldP := S \quad ; \quad r_1, r_2 \sim R(0,1) \quad (2.15)$$

Denkleme göre rastgele belirlenen iki sayıdan  $r_1$  sayısı  $r_2$  den küçükse  $oldP$  matrisi yerine  $S$  matrisi alınmaktadır. Seçilen  $oldP$  matrisi diğer döngülerde değişene kadar (tekrar  $r_1 < r_2$  olana kadar) aynı kalmaktadır.  $oldP$  belirlendikten sonra bu matrisin elemanları kendi aralarında rastgele yer değiştirilmektedir [82].

$$oldP := permutting(oldP) \quad (2.16)$$

Denklemden  $permutting$  matris elemanlarını rastgele yer değiştiren bir fonksiyondur.

*Mutasyon:* BSA' nın mutasyon işleminde  $T$  matrisi adı verilen bir matris elde edilmektedir. Bu matris o anki  $S$  ile  $oldP$  arasındaki farkın alınması ve bu farkın bir katsayı ile çarpılması ile elde edilen sonucun yine  $S$  ile toplanması şeklinde elde edilmektedir. Burada,  $S$  ile  $oldP$  arasındaki fark BSA' nın arama yönü matrisi ismini verdiği matristir. Bu matrisin büyüklüğü ise farklı şekillerde belirlenebilecek olan bir katsayı ( $F$ ) ile belirlenmektedir.

$$T = S + F(oldP - S) \quad (2.17)$$

*Çaprazlama:* BSA çaprazlama yaparken  $T$  matrisini,  $mixrate$  adı verilen bir parametreyi ve  $n$  ve  $d$  sayılarını kullanmaktadır. Bu işlemin sonucunda elde edilen matrise ise *Mutant* matris adı verilmektedir. Buna göre çaprazlama yapılırken ilk olarak  $n \times d$  boyutunda bir birler matrisi tanımlanmaktadır. Bu birler matrisine  $map$  matrisi adı verilmektedir. Ardından Denklem (2.18)'de bulunan iki farklı seçim stratejisi kullanılarak  $T$  matrisinin bazı elemanları ile  $S$  in aynı sıradaki elemanları yer değiştirmektedirler [82].

$$\text{Eğer } r_1 < r_2 \text{ ise } map_{i,u(1:[mixrate.rand.d])} = 0; \text{ değilse } map_{i,randi(d)} = 0 \quad (2.18)$$

Denklem (2.18)'de  $u = permutting(1,2,3 \dots, d)$ ,  $rand [0, 1]$  aralığında  $randi$  ise  $[0, d]$  aralığında rastgele sayı üreten fonksiyonlardır,  $mixrate$  ise kullanıcının belirleyeceği bir parametredir. Bu parametre, bu çalışmada 1 olarak seçilmiştir. Denkleme göre birlerden oluşan  $map$  matrisinin bazı elemanları iki farklı yöntemle seçilmekte ve seçilen elemanlar yerine 0 atanmaktadır. Dikkat edilirse eğer  $r_1 < r_2$  ise birden fazla eleman seçilirken tersi durumda sadece bir eleman seçilmektedir.  $map$  matrisi oluşturulduktan sonra bu matrisin elemanlarından değeri 1 olanlara karşılık gelen  $T$  matrisi elemanlarının yerine  $S$  in aynı sıradaki elemanları gelmektedir [82].

$$\text{Eğer } map_{i,j} = 1 \text{ ise } T_{i,j} = S_{i,j} \quad (i = 1,2,3, \dots, n ; j = 1,2,3, \dots, d) \quad (2.19)$$

$$\text{Mutant} = T \quad (2.20)$$

*Mutant* matrisinin bazı elemanları arama uzayı sınırlarını aşabilir bu durumda sınırı aşanlar için rastgele seçilen elemanlar atanmaktadır (Denklem (2.12)'deki gibi).

*Seçim II*: Bu bölümde BSA elde edilen *Mutant* matrisi için amaç fonksiyon değerini hesaplamakta ve buna göre *fitness* vektörü ve *S* matrisi güncellenmektedir [82].

$$\text{fitnessM} = \text{ObjectFunc}(\text{Mutant}) \quad (2.21)$$

$$\text{Eğer } \text{fitnessM}_{i,j} < \text{fitness}_{i,j} \text{ ise } \text{fitness}_{i,j} = \text{fitnessM}_{i,j} \text{ ve } S_{i,j} = \text{Mutant}_{i,j} \quad (2.22)$$

Başlangıç kısmı hariç son dört aşama BSA'nın durdurma kriteri sağlanana kadar bir döngü içerisinde devam etmektedir. Döngülerin sonunda BSA *fitness* vektörünün en küçük değerini *global minimum* ve bu değerini bulmasını sağlayan *S* matrisi elemanını da nihai çözüm olarak seçmektedir. BSA 'ya ait akış şeması Şekil 2.27'de sunulmuştur.



Şekil 2.27. BSA algoritmasına ait akış diyagramı.

### 2.3.1.3. FCM ile Bir Optimizasyon Algoritmasının Birleştirilmesi

FCM algoritması birçok kümeleme probleminin çözümünde kullanılabilen hızlı ve verimli bir algoritmadır. Ancak bu algoritma özellikle küme merkezlerinin başlangıç aşamasındaki değerlerinin seçilmesine karşı hassastır [85]. Bu algoritmanın diğer bir önemli dezavantajı da problemin çözümü sırasında lokal minimum noktalarına kolayca takılabilesidir [85]. Bu iki sorunu ortadan kaldırmak için birçok araştırmacı FCM algoritması ile diğer optimizasyon algoritmalarını birleştirme yoluna gitmiştir. İki algoritmanın birleştirilmesi sırasında genellikle iki yöntem kullanılmaktadır. Bunlardan ilki küme merkezleri için başlangıç değerlerinin FCM dışında bir optimizasyon algoritması ile belirlenmesi ve ardından FCM ile kümeleme işlemlerinin yapılması şeklindedir. Diğer yöntemde ise FCM algoritmasının amaç fonksiyonu kullanılarak global minimumu bulma işlevi diğer optimizasyon algoritması ile

gerçekleştirilmektedir. Bu çalışmada ikinci yöntem tercih edilmiştir. FCM ile diğer bir algoritmayı bu yöntemle birleştirebilmek için gerekli olan işlem tüm algoritmalar için kullanılabilir genel bir nüfus yapısının belirlenmesidir. BSA nüfus temelli bir algoritma olduğu için tanımlanacak genel nüfus yapısı ve FCM' nin Denklem (2.7)'de verilen amaç fonksiyonu kullanılarak birleştirme işlemi yapılabilir. Belirlenen genel nüfus yapısı Denklem (2.23)'de gösterildiği şekildedir.

$$S = \begin{bmatrix} S_{1,1} & \cdots & S_{1,c} \\ \vdots & \cdots & \vdots \\ S_{n,1} & \cdots & S_{n,c} \end{bmatrix} \quad (2.23)$$

Denklemden  $S$  matrisinin her bir satırı bir aday çözümü içermektedir. Dolayısıyla nüfus  $n$  tane aday çözümden oluşmaktadır. Her bir aday çözüm de  $c$  tane elemandan oluşmaktadır. Bu elemanların her biri bir küme merkezidir. Denklemler (2.7), (2.8), (2.10) ve (2.23)'ün kullanılması ile FCM ve diğer bir optimizasyon algoritmasının birleştirilmesi işlemi mümkün olabilmektedir. FCM ile BSA'nın birleştirilmesi için yapılan işlemler bir sonraki aşamada adım adım sunulmuştur.

#### 2.3.1.4. BSA ve FCM Algoritmalarının Görüntü Bölütleme İşlemi için Birleştirilmesi

Klasik FCM ile BSA algoritmaları görüntü bölütleme işlemi için birleştirilirken sırası ile aşağıdaki işlemler gerçekleştirilmektedir.

1. Adım: Bölütlenecek resimlerin gri seviye formlarının elde edilmesi (bu çalışmada direkt gri seviye resimler kullanılmıştır. Herhangi bir dönüştürme işlemi yapılmamıştır) ve başlangıç parametrelerinin tanımlanması. Bu parametreler BSA için nüfus büyüklüğü ( $n$ ), durdurma kriteri (40 döngü sayısı), *mixrate* parametresi ( $mixrate = 1$ ) ve  $F$  (arama-yönü matrisi büyüklüğü, bu çalışmada iki farklı yöntemle belirlenmiştir. Bir sonraki bölümde verilecektir). FCM için ise bulanıklaştırma katsayısı olan  $m$  (bu çalışmada  $m = 2$  olarak alınmıştır.) parametreleridir.

2. Adım: Denklem (2.23) kullanılarak başlangıç nüfusunun ( $S$ ) belirlenmesi.

3. Adım:  $S$  de verilen her bir aday çözüm için *fitness* vektörünün hesaplanması. Bu işlem için Denklemler (2.7), (2.8) ve (2.10) kullanılmaktadır.

4. Adım: BSA döngüsünün başlaması ve her bir döngüde *fitnessM* vektörünün 3. adımda verildiği gibi elde edilmesi ve ardından *fitness* vektörü ile  $S$ 'in güncellenmesi.

5. Adım: Eğer durdurma kriteri sağlanıyorsa BSA döngüsünün sonlandırılması ve

global minimum, nihai çözüm ( $S$ 'in global minimum un bulunmasını sağlayan elemanı) ve nihai üyelik matrisinin elde edilmesi.

6. Adım: Elde edilen üyelik matrisi kullanılarak bölütlenmiş resimlerin elde edilmesi.

Yukarda sunulan adımlar bir akış şeması içerisinde Şekil 2.28'deki gibi gösterilebilir. FCM nin klasik hali ile BSA nin birleştirilmesi ile elde edilen bu yeni algoritmaya BSAFCM adı verilmiştir.



Şekil 2.28. BSAFCM algoritmasının akış diyagramı.

Şekil 2.28'de mavi kutucuklar BSA ile ilgili süreçleri ifade ederken kırmızı kutucuklar klasik FCM'nin süreçleridir.

### 2.3.1.5. BSAFCM Algoritması için İyileştirme

BSA çok iyi lokal ve global arama yeteneklerine sahip bir optimizasyon algoritmasıdır[82] ayrıca FCM de eğer küme merkezleri için başlangıçta uygun değerler verilmişse iyi bir lokal arama algoritmasıdır. Bu iki algoritmanın yukarıda da anlatıldığı gibi birleştirilmesi ile FCM'nin lokal minimuma takılma ve başlangıç parametrelerinin seçimine karşı olan hassasiyet problemi çözülebilmektedir. Ancak geliştirilen algoritmanın literatürde yer alan diğer algoritmalarla karşılaştırılabilecek düzeye gelmesi için bu çalışmada BSAFCM algoritması için yeni bir ağırlık parametresi ( $w$ ) tanımlanmıştır. Tanımlanan ağırlık parametresi bir dizi denemeden sonra aşağıdaki gibi elde edilmiştir.

$$w_{t+1} = w_{min} + \exp\left(-\exp\left(-\left(t_{max} - t\right) * \frac{\left(w_{max} - w_{min}\right)}{t_{max}} * w_t\right)\right) \quad (2.24)$$

Denklemden,  $t$  ve  $t_{max}$  sırasıyla anlık ve maksimum döngü sayılarını,  $w_{min}$  ve  $w_{max}$  ise  $w$  için seçilen minimum ve maksimum değerlerdir. Son olarak  $w_t$ ,  $t$ . iterasyonda elde edilen  $w$  değeridir. Denklemden de görüldüğü gibi,  $w$  değeri hem  $w_{min}$  ve  $w_{max}$  değerlerinden hem de bir önceki döngüden elde edilen  $w_t$  değerinden etkilenmektedir. Dolayısıyla bir başlangıç  $w$  değerine ihtiyaç vardır. Bu başlangıç  $w$  değeri 0-1

aralığında normal dağılımla rastgele seçilmektedir.

Detaylı olarak ifade etmek gerekirse Denklem (2.15)'e bakıldığında arama-yönü matrisinin üyelerinin tamamen rastgelelikle seçildiği açıkça görülmektedir. Arama-yönü matrisi bir hafıza şeklinde çalışarak eski nesilden elde edilen bilgileri bir sonraki nesile aktardığı için bu çalışmada lokal arama açısından bu bilgilerin tamamen rastgele değil de en düşük değeri  $w_{min}$  olacak şekilde ve gittikçe azalan bir ihtimalle değişmesi sağlanmıştır. Ayrıca bu ihtimal döngünün başlangıcında çok yüksek olmasına rağmen ilerleyen aşamalarda lokal minimuma takılma ihtimalini gidermek için gittikçe azalmaktadır ancak hiçbir zaman  $w_{min}$  den daha küçük olmamaktadır. Bu değerler kullanılarak oldP'nin elemanları öncelikle tamamen rastgele olan ilk süreçten daha yüksek ihtimalli olarak S in elemanlarından seçilmektedir. Ancak bu oldP'nin tamamen S'in elemanlarından seçildiği anlamına gelmemektedir. Çok yüksek olmamakla birlikte oldP'nin elemanları rastgele de seçilebilmektedir. Bu her iki ihtimali sağlamak için Denklem (2.15) aşağıdaki gibi değiştirilmiştir.

$$\text{Eğer } r_1 < w \text{ ise oldP} := S \mid r_1 \sim R(0,1) \quad (2.25)$$

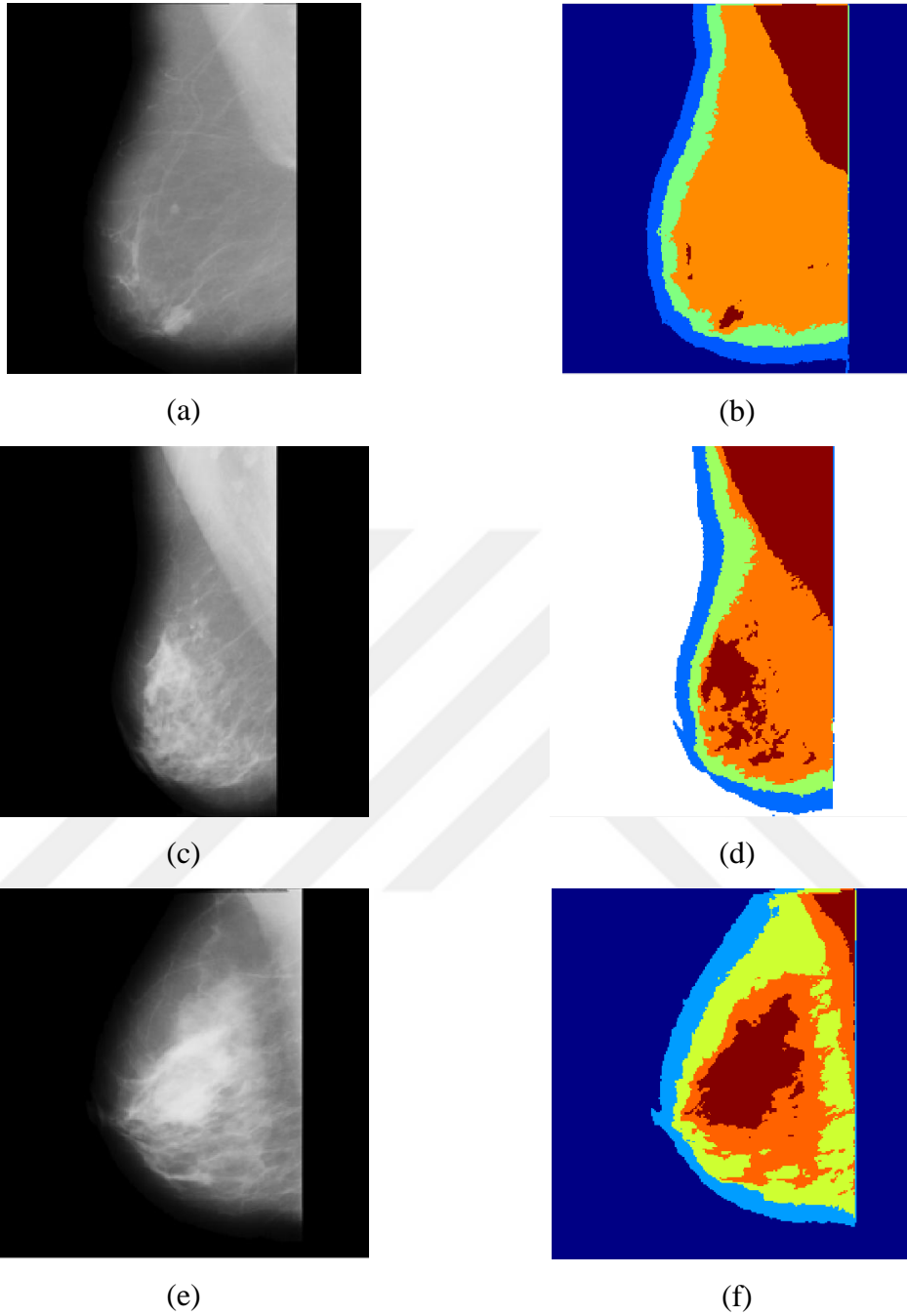
Denkleme göre rastgele seçilecek  $r_1$  sayısı  $w$ 'den küçükse oldP yerine S'in elemanları alınacak yani hafıza güncellenecek eğer değilse oldP rastgele seçilecek yani hafıza silinip yeniden rastgele doldurulacak. Şekilde verilen yüksek  $w$  değerleri sayesinde hafızanın yüksek ihtimalle güncellenmesi sağlanmaktadır.

Bu çalışmada  $w$  parametresi ayrıca arama-yönü matrisinin büyüklüğünün seçilmesinde de kullanılmıştır. Bu seçimle Denklem (2.16)'da mutasyon amacıyla elde edilen T matrisinin oluşturulmasında kullanılan F ölçeğinin yerine yine  $w$  parametresi aşağıdaki gibi tanımlanmıştır.

$$\text{Eğer } r_1 < r_2 \text{ ise } F := w \text{ değilse } F = 3r_3 \mid r_1, r_2, r_3 \sim R(0,1) \quad (2.26)$$

Denkleme göre rastgele belirlenecek bir ihtimalle,  $r_1 < r_2$  ise, F yerine  $w$  değeri alınmakta karşıt durumda ise F yerine [0-3] aralığında rastgele belirlenecek bir sayı seçilmektedir. Bu denklemde de  $w$  nin çoğunlukla [0-1] aralığında olması nedeniyle mutasyon ihtimali kısmen azaltılmakta ve dolayısıyla algoritmanın lokal arama yeteneği arttırılmaktadır. Bu parametre ile tüm optimizasyon algoritmalarının hedefi olan algoritmanın başlarında çözüm uzayının iyi taranması, sonlarına doğru ise optimum çözüme yoğunlaşılması hedefi gerçekleştirilmeye çalışılmıştır. Geliştirilen bu yeni algoritmaya da w-BSAFCM algoritması adı verilmiştir. Geliştirilen algoritma ile elde

edilen bölütleme görüntüleri Şekil 2.29'da sunulmuştur.



Şekil 2.29. w-BSAFCM ile örnek mamogram görüntülerinin bölütlenmesi (a,c,e) Orijinal mamogramlar (b,d,e) w-BSAFCM ile 5 Kümeye bölütlenmiş mamogramlar.

### 2.3.2. Entropi Yöntemleri

Entropi düzensizlik ölçüsü olarak tanımlanabilir. Bir sistem ne kadar düzenli olursa bilgi entropisi o kadar düşük olur; Aksine, sistem ne kadar karışıkça, bilgi entropisi o kadar yüksektir. Rastgele bir  $G$  olayının  $P(G)$  olasılığı ile olması için Denklem

(2.27)'de verilen bilgileri içermesi gerekmektedir [87].

$$I(G) = \log \left[ \frac{1}{P(G)} \right] = -\log[P(G)] \quad (2.27)$$

Buradaki  $I(G)$  olayın öz bilgisi olup, olayın olasılığı ile öz bilgi miktarı arasında ters orantı vardır.

$p_1, p_2, p_3, \dots, p_i, \dots, p_k$  'nin ayrık kaynağın olasılığı olduğunu farz edelim. Bu durumda  $0 \leq p_i \leq 1$  ve  $\sum_{i=1}^k p_i = 1$ 'dir. Burada  $k$  durum sayısıdır. Ayrık bir kaynağın entropisini bulmak için genellikle olasılık dağılımı kullanılır.

Çeşitli entropi ölçütleri vardır. Bunlar Shannon ve Non-Shannon olmak üzere temelde iki grupta incelenebilir. Bu tez çalışmasında 4 entropi ölçütü kullanılmıştır. Bunlardan ilki Shannon entropi ölçütüdür ve Denklem (2.28)'deki gibi tanımlanabilir.

$$S = -\sum_{i=0}^k p_i \log_e(p_i) \quad (2.28)$$

Non-Shannon Entropi ölçütlerinden;

a) Renyi Entropi:

$$R = \frac{1}{1-\alpha} \log_e \left( \sum_{i=0}^k p_i^\alpha \right) \quad \alpha \neq 1, \alpha > 0 \quad (2.29)$$

b) Havrda & Charvat Entropi:

$$HC = \frac{1}{1-\alpha} \left( \sum_{i=0}^k p_i^\alpha - 1 \right) \quad \alpha \neq 1, \alpha > 0 \quad (2.30)$$

c) Kapur Entropi :

$$K_{\alpha, \beta} = \frac{1}{\beta - \alpha} \log_e \frac{\sum_{i=0}^k p_i^\alpha}{\sum_{i=0}^k p_i^\beta} \quad \alpha \neq \beta, \alpha < 0, \beta > 0 \quad (2.31)$$

şeklinde tanımlanır [87].

Mamogramları meme bölgesi dışında kalan arka plan, meme dokusu bölgesi ve şüpheli bölge olmak üzere üç bölmede inceleyebiliriz. Arka plan olarak adlandırdığımız bölge meme bölgesi dışında olduğundan teşhis ile ilgili bilgi içermez dolayısıyla bu bölgeyi hesaplama dışında bırakabiliriz. Bunun için Denklem (2.32)'de yer alan formülü kullanabiliriz [87].

$$k = \frac{1}{n} \sum_{i,j \in R}^M \sum_{i,j \in R}^N f(i, j) \quad (2.32)$$

Burada  $M$  ve  $N$  mamogramın boyutlarıdır.  $R$  ise gri seviye değeri  $Otsu_{esik}$ 'den büyük olan bölgeyi temsil etmektedir,  $n$  değeri de bu bölgedeki piksel sayısıdır,  $f(i, j)$  ise,  $i$ . ve

$j$ . koordinatlarında bulunan gri seviye değeridir [87]. Buradaki  $Otsu_{esik}$  değeri arka plan bölgesini ve doku bölgesini ayırmak amacıyla kullanılan bir eşik değeridir ve ileriki aşamalarda nasıl hesaplandığı anlatılmıştır.

$p_1, p_2, \dots, p_{k-1}, p_k, p_{k+1}, \dots, p_L$  mamogramın gri seviye olasılık dağılımı olsun. Burada  $p_k$  mamogramın normalize histogramıdır ve Denklem (2.33)'deki gibi hesaplanır [87].

$$p_k = h_k / (M \times N) \quad (2.33)$$

Burada  $h_k$  mamogramın gri seviye histogramını belirtir. Bu dağılımdan meme arka planı, meme dokusu bölgesi ve şüpheli bölge olmak üzere üç bölgeye ait olasılık dağılımı elde edilebilir [87].

Meme arka planına ait olasılık dağılımı:

$$\frac{p_1}{P_{k-1}}, \frac{p_2}{P_{k-1}}, \dots, \frac{p_{k-1}}{P_{k-1}}$$

Meme dokusuna ait olasılık dağılımı:

$$\frac{p_k}{P_t - P_{k-1}}, \frac{p_{k+1}}{P_t - P_{k-1}}, \dots, \frac{p_t}{P_t - P_{k-1}}$$

Şüpheli bölgeye ait olasılık dağılımı:

$$\frac{p_{t+1}}{1 - P_t}, \frac{p_{t+2}}{1 - P_t}, \dots, \frac{p_L}{1 - P_t}$$

Burada  $P_{k-1} = \sum_{i=1}^{k-1} p_i$ ,  $P_t = \sum_{j=1}^t p_j$ ,  $t$  eşik değeri ve  $L$  'de maksimum gri seviye değerini belirtir [87].

Shannon entropisine dayalı eşikleme yaklaşımında doku bölge piksellerinin entropisi ve şüpheli bölge piksellerinin entropisi aşağıdaki gibi tanımlanabilir:

Doku bölge pikselleri için Shannon Entropi:

$$S_{DB}(t) = - \sum_{i=k}^t \frac{p_i}{P_t - P_{k-1}} \log_e \left( \frac{p_i}{P_t - P_{k-1}} \right) \quad (2.34)$$

Şüpheli bölge pikselleri için Shannon Entropi:

$$S_{ROI}(t) = - \sum_{i=t+1}^L \frac{p_i}{1 - P_t} \log_e \left( \frac{p_i}{1 - P_t} \right) \quad (2.35)$$

Eşikleme yaklaşımına dayalı Shannon entropi  $S_{DB}(t) + S_{ROI}(t)$  toplamını maksimize ederek optimum eşik değerini  $t_{opt}^S$  bulmaya çalışır [87].

$$t_{opt}^S = Arg_{t=k}^L max[S_{DB}(t) + S_{ROI}(t)] \quad (2.36)$$

Bu optimum eşik değeri ile şüpheli bölge ve doku bölgesi ayrılır.

Renyi entropi için ise doku bölge piksellerinin entropisi ve şüpheli bölge piksellerinin entropisi aşağıdaki gibi hesaplanabilir.

Doku bölgesi pikselleri için Renyi entropi:

$$R_{DB}(t) = \frac{1}{1-\alpha} \log_e \left[ \sum_{i=k}^t \left( \frac{p_i}{P_t - P_{k-1}} \right)^\alpha \right] \quad (2.37)$$

Şüpheli bölge pikselleri için Renyi entropi:

$$R_{ROI}(t) = \frac{1}{1-\alpha} \log_e \left[ \sum_{i=t+1}^L \left( \frac{p_i}{1-P_t} \right)^\alpha \right] \quad (2.38)$$

burada  $\alpha \neq 1$  ve  $\alpha > 0$ 'dır.

Eşikleme yaklaşımına dayalı Renyi entropi  $R_{DB}(t) + R_{ROI}(t)$  toplamını maksimize ederek optimum eşik değerini  $t_{opt}^R$  bulmaya çalışır [87].

$$t_{opt}^R = Arg_{t=k}^L max[R_{DB}(t) + R_{ROI}(t)] \quad (2.39)$$

Benzer şekilde doku bölge pikselleri için Havrda & Charvat entropi:

$$HC_{DB}(t) = \frac{1}{1-\alpha} \left[ \sum_{i=k}^t \left( \frac{p_i}{P_t - P_{k-1}} \right)^\alpha - 1 \right] \quad (2.40)$$

$$HC_{ROI}(t) = \frac{1}{1-\alpha} \left[ \sum_{i=t+1}^L \left( \frac{p_i}{1-P_t} \right)^\alpha - 1 \right] \quad (2.41)$$

olarak hesaplanır.

Havrda & Charvat entropi de  $HC_{DB}(t) + HC_{ROI}(t) + (1 - \alpha) HC_{DB}(t)HC_{ROI}(t)$  toplamını maksimize ederek optimum eşik değerini  $t_{opt}^{HC}$  bulmaya çalışır [87].

$$t_{opt}^{HC} = Arg_{t=k}^L max[HC_{DB}(t) + HC_{ROI}(t) + (1 - \alpha) HC_{DB}(t)HC_{ROI}(t)] \quad (2.42)$$

Son olarak doku bölgesi pikselleri için Kapur entropi:

$$K_{DB}(t) = \frac{1}{\beta-\alpha} \log_e \left[ \frac{\sum_{i=k}^t \left( \frac{p_i}{P_t - P_{k-1}} \right)^\alpha}{\sum_{i=k}^t \left( \frac{p_i}{P_t - P_{k-1}} \right)^\beta} \right] \quad (2.43)$$

$$K_{ROI}(t) = \frac{1}{\beta-\alpha} \log_e \left[ \frac{\sum_{i=t+1}^L \left( \frac{p_i}{1-P_t} \right)^\alpha}{\sum_{i=t+1}^L \left( \frac{p_i}{1-P_t} \right)^\beta} \right] \quad (2.44)$$

Burada  $\alpha \neq \beta, \alpha < 0, \beta > 0$ 'dır.

Benzer şekilde eşikleme yaklaşımına dayalı Kapur entropi  $K_{DB}(t) + K_{ROI}(t)$  toplamını maksimize ederek optimum eşik değerini  $t_{opt}^K$  bulmaya çalışır[87].

$$t_{opt}^K = \text{Arg}_{t=k}^L \max[K_{DB}(t) + K_{ROI}(t)] \quad (2.45)$$

Optimum eşik  $t_{opt}^*$ ,  $* \in \{S, R, HC, K\}$  ile doku bölgesinden şüpheli bölge ayrılır [87]. Literatürde yapılan çalışmalara bakıldığında Havrda& Charvat entropinin diğerlerinden daha başarılı sonuçlar ürettiği görülmektedir[88]-[90]. Dolayısıyla bu tez çalışmasında Havrda & Charvat entropi [91]'de yer alan ve  $\alpha$  değeri için yapılan geliştirilmiş hali ile kullanılmıştır.

Mamogramlar üzerinde şüpheli bölgelerin tespiti için ilk olarak mamogram arka planını ayırmak için eşikleme işlemi yapılmaktadır. Bu tez çalışmasında bu amaçla eşiklemeye dayalı entropi yaklaşımlarından Otsu N eşikleme yöntemi kullanılmıştır.

### 2.3.3. Otsu N Eşikleme

Otsu eşikleme yöntemi, Nobuyuki Otsu [92] tarafından önerilen optimum global eşikleme yöntemidir. Görüntülerin bölütlenmesi amacıyla otomatik eşik seçimi yapar ve bu işlemi de yalnızca gri-seviyeli histogramın sıfır ve birinci derece kümülatif momentlerini kullanarak gerçekleştirir [67]. Buna göre;  $M \times N$  boyutunda bir giriş görüntüsünün gri seviyeleri,  $[0, 1, 2, \dots, L - 1]$  cinsinden gösterilsin. Gri seviye değeri  $i$  olan piksel sayısı,  $n_i$  ve görüntüdeki toplam piksellerin sayısı  $MN = n_0 + n_1 + \dots + n_{L-1}$  ile gösterilsin. Görüntüyü  $C_1$  ve  $C_2$  olacak şekilde iki bölgeye ayırmak için  $0 < k < L - 1$  olacak şekilde bir  $k$  eşik değeri seçtiğimizi varsayalım, bu durumda  $C_1$  sınıfı  $[0, k]$  aralığında yoğunluk değerlerine ve  $C_2$  sınıfı ise  $[k + 1, L - 1]$  aralığındaki yoğunluk değerlerine sahip piksellerden oluşur ve bir pikselin bir sınıfa ( $C_1$ ) ait olma olasılığı aşağıdaki şekilde hesaplanabilir [67].

$$P_1(k) = \sum_{i=0}^k p_i \quad (2.46)$$

$C_2$  sınıfına ait olma olasılığı ise;

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - P_1(k) \quad (2.47)$$

şeklinde hesaplanabilir.

$C_1$  sınıfına atanan piksellerin ortalama yoğunluk değerleri:

$$m_1(k) = \frac{1}{P_1(k)} \sum_{i=0}^k i p_i \quad (2.48)$$

Benzer şekilde  $C_2$  sınıfına atanan piksellerin ortalama yoğunluk değerleri:

$$m_2(k) = \frac{1}{P_2(k)} \sum_{i=k+1}^{L-1} i p_i \quad (2.49)$$

Global ortalama ise:

$$m_G = \sum_{i=0}^{L-1} i p_i \quad (2.50)$$

şeklinde bulunabilir. Optimum  $k$  değerini maksimum yapacak bir kriter bulmak için Denklem (2.51) kullanılır.

$$y(k) = \frac{\sigma_B^2(k)}{\sigma_G^2} \quad (2.51)$$

Denklem (2.52)'de  $\sigma_B^2(k)$  sınıflar arasındaki varyanstır ve aşağıdaki gibi tanımlanır.

$$\sigma_B^2(k) = P_1(m_1 - m_G)^2 + P_2(m_2 - m_G)^2 \quad (2.52)$$

ve global varyansta Denklem (2.53)'deki gibi tanımlanır.

$$\sigma_G^2 = \sum_{i=0}^{L-1} (i - m_G)^2 P_i \quad (2.53)$$

Sonuç olarak,  $\sigma_B^2(k)$  değerini maksimize eden optimum eşik değeri  $k$  elde edilmiş olur.

Otsu eşikleme yöntemini üç sınıfa ayırmak için aşağıdaki adımlar uygulanır.

*Adım 1:* İlk olarak giriş görüntüsünün normalize histogramı hesaplanır. Burada  $p_i$ , ( $i=0,1,2,\dots,L-1$ ) histogram elemanlarını  $L$  ise gri seviye değerlerini belirtsin.

*Adım 2:*  $P_1, P_2, P_3$  ve  $P_4$  kümülatif toplamlar hesaplanır. Bunun için  $k_1, k_2$  ve  $k_3$

şeklinde üç adet eşik değeri kullanılır.

$$P_1 = \sum_{i=0}^{k_1} p_i \quad (2.54)$$

$$P_2 = \sum_{i=k_1+1}^{k_2} p_i \quad (2.55)$$

$$P_3 = \sum_{i=k_2+1}^{k_3} p_i \quad (2.56)$$

$$P_4 = \sum_{i=k_3+1}^{L-1} p_i \quad (2.57)$$

*Adım 3:* Denklem (2.50) kullanılarak global ortalama hesaplanır.

*Adım 4:* Denklem (2.58)-(2.61) kullanılarak ortalama yoğunluk değerleri hesaplanır.

$$m_1 = \frac{1}{P_1} \sum_{i=0}^{k_1} ip_i \quad (2.58)$$

$$m_2 = \frac{1}{P_2} \sum_{i=k_1+1}^{k_2} ip_i \quad (2.59)$$

$$m_3 = \frac{1}{P_3} \sum_{i=k_2+1}^{k_3} ip_i \quad (2.60)$$

$$m_4 = \frac{1}{P_4} \sum_{i=k_3+1}^{L-1} ip_i \quad (2.61)$$

*Adım 5:*  $k_1, k_2, k_3$  eşik değerleri için sınıflar arasındaki varyans  $\sigma_B^2(k_1, k_2, k_3)$  Denklem (2.62) kullanılarak hesaplanır.

$$\left\{ \begin{array}{l} k_1 = 1, \dots, k_2 - 1 \\ k_2 = k_1 + 1, \dots, k_3 - 1 \\ k_3 = k_2 + 1, \dots, L - 2 \end{array} \right\} \text{ olmak üzere}$$

$$\sigma_B^2(k_1, k_2, k_3) = P_1(m_1 - m_G)^2 + P_2(m_2 - m_G)^2 + P_3(m_3 - m_G)^2 + P_4(m_4 - m_G)^2 \quad (2.62)$$

*Adım 6:* Sınıflar arası varyansı maksimum yapan değer optimum eşik değeridir.

Eğer birden fazla maksimum değer varsa tespit edilen çeşitli maksimumlara karşılık gelen  $k$  değerlerin ortalaması alınarak  $k^*$  değeri elde edilir.

$$\sigma_B^2(k_1^*, k_2^*, k_3^*) = \max \sigma_B^2(k_1, k_2, k_3), 0 < k_1 < k_2 < k_3 < L - 1 \quad (2.63)$$

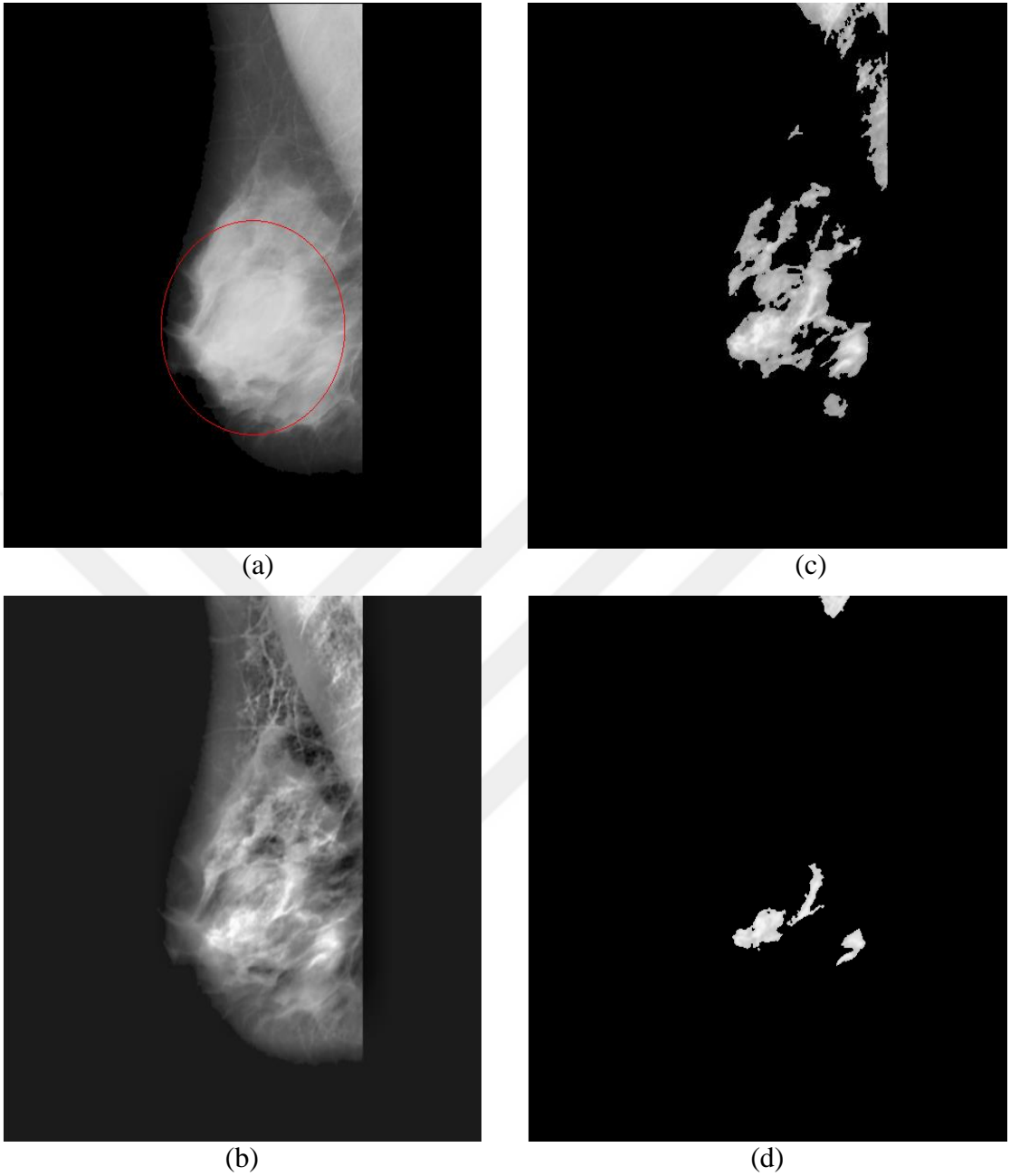
Arka plan ve meme bölgesini birbirinden ayırmak için bu tez çalışmasında, sabit bir değer kullanmak yerine w-BSAFCM yöntemi ile Otsu N eşikleme yöntemlerinin birlikte kullanılması ile elde edilen ve  $Ow_{esik}$  adı verilen bir eşik değeri kullanılmıştır. Bu uygulama sayesinde mamogramdan mamograma değişiklik gösteren gri seviye değerleri için adaptif bir eşik değeri elde edilmiştir.

$Ow_{esik}$  değeri hesaplanırken ilk olarak iyileştirilmiş mamogram görüntüsünden w-BSAFCM yöntemiyle elde edilen küme merkezleri eşik değeri olarak belirlenmiştir. Bu yöntemde küme sayısı beş olarak belirlenmiş ve  $we_1, we_2, we_3, we_4$  ve  $we_5$  olmak üzere beş adet eşik değeri elde edilmiştir. Otsu çoklu eşikleme ile de ayrıca  $k_1, k_2, k_3, k_4$  ve 5 olmak üzere beş eşik değeri elde edilmiştir. Elde edilen bu eşik değerlerinden  $Ow_{esik}$  değeri şu şekilde hesaplanmıştır.

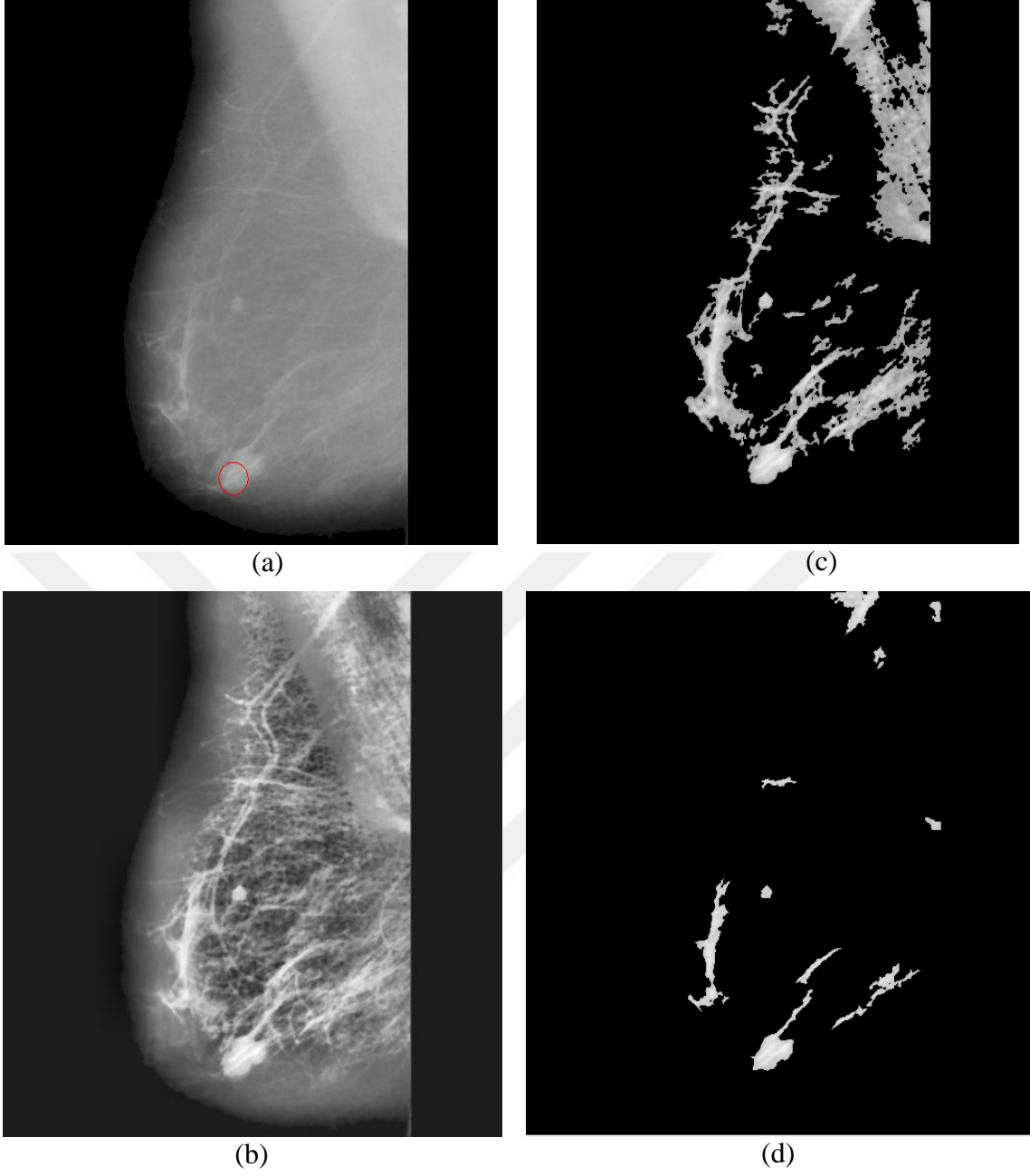
$$Ow_{esik} = (k_4 + we_4)/2 \quad (2.64)$$

Şüpheli bölgelerin tespiti için ise Havrda & Charvat ve Kapur entropiye dayalı eşik yöntemleri test edilmiş ve en iyi sonuçları veren Havrda & Charvat entropiye dayalı eşikleme yöntemi bu amaçla kullanılmıştır. Bu deneysel çalışmalara ait uygulama

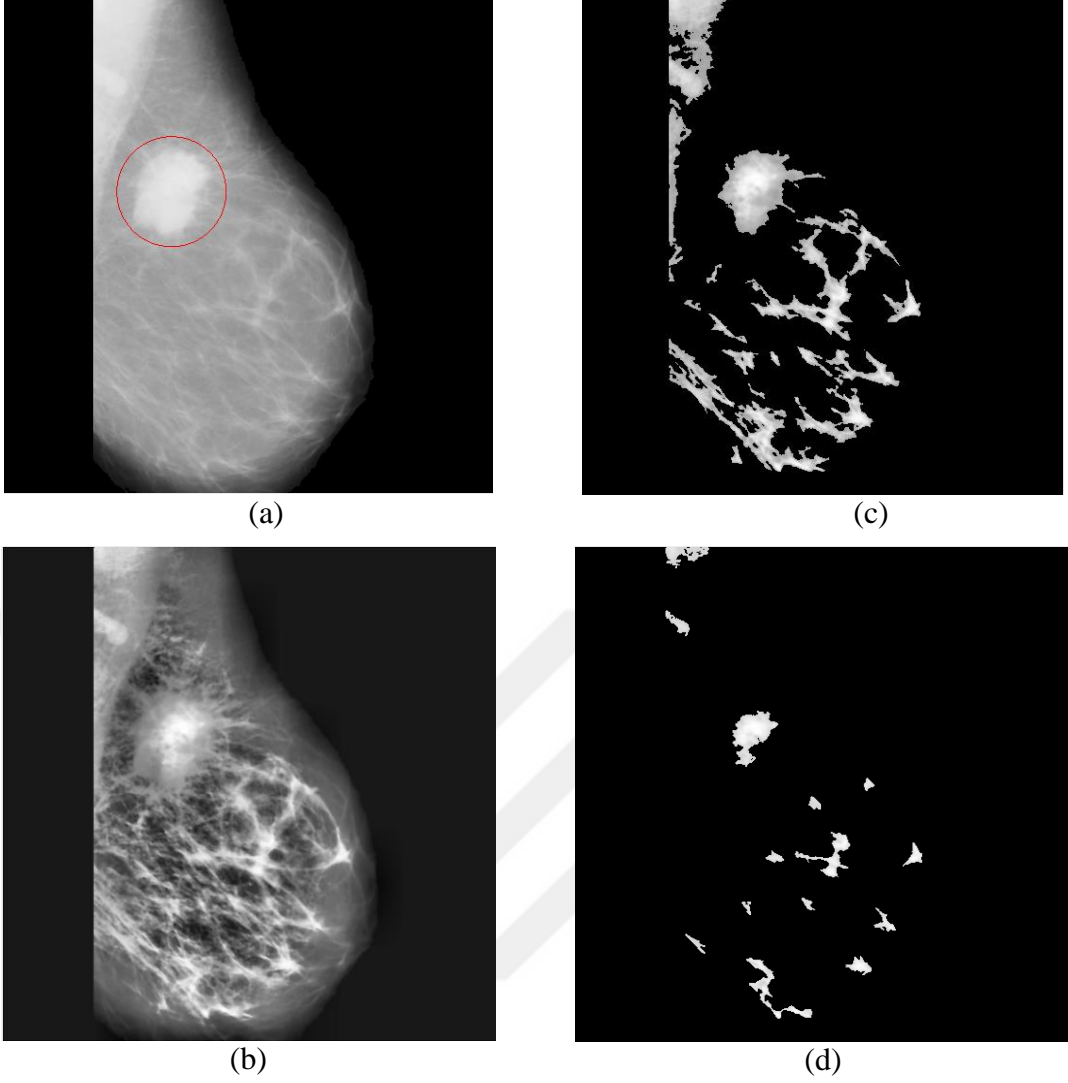
görüntüleri şekil 2.30,2.31,2.32 ve 2.33'de sunulmuştur.



Şekil 2.30. Mdb001 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram (b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $t_{opt}^K=151, Ow_{eşik}=112, \alpha=0.4118, \beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $t_{opt}^{HC} = 198, Ow_{eşik} = 112, \alpha = 0.4118$ ).



Şekil 2.31. Mdb005 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram (b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $t_{opt}^K=151, Ow_{eşik}=111, \alpha=0.4353, \beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $t_{opt}^{HC} = 198, Ow_{eşik} = 111, \alpha = 0.4353$ ).



Şekil 2.32. Mdb184 için kullanılan eşiklemeye dayalı entropi yaklaşımlarına ait örnek görüntüler (a) Kitle bölgesi çizilmiş orijinal mamogram (b) İyileştirilmiş mamogram (c) Kapur entropi ile bulunan şüpheli bölge ( $t_{opt}^K=161, Ow_{eşik}=120, \alpha=0.3725, \beta=2$ ) (d) Havrda & Charvat entropi ile bulunan şüpheli bölge ( $t_{opt}^{HC} = 208, Ow_{eşik} = 120, \alpha = 0.3725$ ).

Şekillerden de görüldüğü gibi Havrda & Charvat entropi ile bulunan şüpheli bölge daha iyi sonuç vermektedir. Çünkü bu yöntemle doğru pozitif olarak nitelendirdiğimiz gerçek kitle tespit edilirken, bulunan yanlış pozitif (kitle olmadığı halde kitle sanılan) sayısı daha az çıkmaktadır. Bir sonraki aşamada elde edilen bu ROI'ler için öznelilik çıkarımı algoritmaları üzerinde durulmuştur.

## 2.4. ÖZİNİTELİK ÇIKARIM ALGORİTMALARI

Görüntü işleme açısından öznelik çıkarım algoritmalarını, görüntünün tamamının değil de sadece gerekli kısımların işlendiği ve işlenen bu kısımlara ait çeşitli özneliklerin elde edildiği algoritmalar olarak tanımlayabiliriz. Meme kanserinin teşhisi açısından bakıldığında şüpheli bölgeler üzerinde bulunan kitleler için öznelik çıkarımı bilgisayar destekli teşhis açısından oldukça önemlidir. Literatürde kitle teşhisi için çok farklı yapılarda öznelikler sunulmuştur [93]. İyi huylu ve kötü huylu kitleleri birbirinden ayırmak için kullanılan temel öznelikler genel olarak doku özellikleri ve şekil özellikleri olmak üzere iki grupta incelenebilir. Bu iki gruba ait alt özellikler ayrıntılı olarak Çizelge 2.1’de listelenmiştir.

Çizelge 2.1. Kitle Teşhisi için öznelik çıkarım yöntemleri [93].

Şekil özellikleri	Doku özellikleri	
Hu'nun değişmez momentleri	İstatistiksel	Birinci Seviye İstatistiksel Özellikler (FOS) (Ortalama gri seviye değeri, ortalama gri seviye kare değeri, standart sapma, varyans, çarpıklık ve basıklık).
Diğer öznelikler (Alan, Çevre, Yoğunluk, En / Boy oranı, Büyük Eksen Uzunluğu, Küçük Eksen Uzunluğu, Dış merkezlik, Yönlendirme, Dışbükey Çoklu Alan, Euler sayısı, Eşdeğer çap, Katılık, Kapsam ve Ortalama Yoğunluk)		Gri Seviye Eş-Oluşum Matrisi (GLCM) (Ortalama, Varyans, Farklılık, Korelasyon, Çapraz Moment, Enerji, Homojenlik, Entropi, Kontrast, Toplam Ortalama, Toplam Entropi, Toplam Varyans, Korelasyon bilgisi 1, Korelasyon bilgisi 2, Entropi Farkı, Ortalama Farkı, Varyans Farkı, Gölge Önemi. Dört yöne karşılık gelen her özellik için dört değer elde edilebilir. ( $\theta=0^\circ, 45^\circ, 90^\circ$ ve $135^\circ$ ))
		Gri Seviye Fark Matrisi (GLDM)( Ortalama, Kontrast, Açısız İkinci Moment, Entropi ve Ters Fark Momenti)
		Gray Level Run Length Matrices (GLRLM)
	Frekansal	Tamura özellikleri(Kabalık, kontrast, yön, doğrusallık, düzenlilik ve pürüzlülük)
		Gabor Dönüşümü
		İki Boyutlu Dalgacık Dönüşümü
		Eğricik Dönüşümü

Bu tez çalışmasında daha iyi sonuç elde edebilmek amacıyla hem şekilsel hem de dokusal bazı öznelikler kullanılmıştır. Bu öznelikler bir sonraki bölümde açıklanmıştır.

### 2.4.1. Şekilsel Öznitelikler:

Alan: Seçili bölgedeki (şüpheli kitle) piksellerin sayısı olarak tanımlanabilir.

$$A = K_{pikselsayısı} \quad (2.65)$$

Çevre: Seçili bölgenin(şüpheli kitle) kontür çizgisindeki toplam piksel sayısıdır.

$$\zeta = K_{kontürpikselsayısı} \quad (2.66)$$

Yükseklik : Seçili bölgenin(şüpheli kitle) satır sayısı olarak tanımlanabilir.

$$Y = K_{satirsayısı} \quad (2.67)$$

Genişlik : Seçili bölgenin(şüpheli kitle) sütun sayısı olarak tanımlanabilir.

$$Y = K_{sutunsayısı} \quad (2.68)$$

Merkez: MxN boyutunda seçili bölgenin(şüpheli kitle) merkez noktası olup ,  $M_{satir}$  ve  $M_{sutun}$  bu noktanın koordinatlarını ifade eder.  $K_{n.sutun}$  kitlenin n. sütunundaki dolu piksel sayısını,  $K_{m. satir}$  ise kitlenin m. satırındaki dolu piksel sayısını temsil eder.

$$M_{satir} = \frac{\sum_{m=1}^M m.K_{m.satir}}{A}, M_{sutun} = \frac{\sum_{n=1}^N n.K_{n.sutun}}{A} \quad (2.69)$$

Çevreleyen Kutu: Seçili bölgeyi(şüpheli kitle) çevreleyen en küçük boyutlu dikdörtgendir.  $K_y$  kitlenin yüksekliği,  $K_g$  kitlenin genişliğidir,  $\zeta K_y$  çevreleyen kutunun yüksekliği,  $\zeta K_g$  ise çevreleyen kutunun genişliğidir.

$$\zeta K_y = K_y, \zeta K_g = K_g \quad (2.70)$$

### 2.4.2. Dokusal Öznitelikler:

Ortalama Gri Seviye Değeri: Seçili bölgenin(şüpheli kitle) gri seviye piksellerinin ortalaması olarak tanımlanabilir.

$$O_{GS} = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N K_{m,n} \quad (2.71)$$

Standart Sapma: Seçili bölgedeki(şüpheli kitle) gri seviye değerlerinin standart sapmasıdır.

$$SS_K = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (K_{m,n} - O)^2} \quad (2.72)$$

#### 2.4.2.1. Gri Seviye Eş-Oluşum Matrisi (GLCM)

Gri seviye eş-oluşum matrisi dokuyu tanımlamak amacıyla bir görüntünün veya

bölgenin yoğunluk histogramının istatistiksel momentleri kullanılarak hesaplanır [94]. Doku öznelikleri, görüntüdeki birbirine göre belirli konumlarda gözlemlenen yoğunluk kombinasyonlarının istatistiksel dağılımından hesaplanır. Her bir kombinasyonda yoğunluk puanlarının (piksel) sayısına göre istatistikler birinci, ikinci ve yüksek sıralı istatistikler olarak sınıflandırılır. Hesaplama yalnızca histogramlar kullanılırsa yoğunluk dağılımı hakkında bilgi elde edilebilir fakat bu dokudaki piksellerin birbirlerine göre görel konumu hakkında bilgi edinilemez. GLCM, bir görüntüdeki komşu piksellerin görel konumu hakkında önemli bilgiler sağlayabilmektedir. GLCM, satır ve sütunlarının sayısının resimdeki gri seviyelerin sayısına eşit olduğu bir matristir. GLCM matrisi hesaplanırken dikkat edilmesi gereken önemli konulardan bir tanesi gri seviye değeridir. Gri seviye değerinin büyük olması hesaplama yükünü arttıracığından en uygun gri seviye değerinin seçilmesi önem arz etmektedir. Diğer önemli seçimler ise mesafe( $d$ ) ve açının( $\theta$ ) seçimidir. Mesafe için  $d=1,2,\dots,64$  olabilecek şekilde seçimler yapılabilmektedir. Fakat mesafe için uygun değer seçilmesi incelenen görüntüdeki dokuların kabalığı veya inceliği ile ilişkili olduğundan önem taşımaktadır. Bu tez çalışmasında ince, detaylı dokuları tespit edebilmek amacıyla uzaklık parametresi 1 olarak seçilmiştir [94].

Açıların seçimi için aşağıdaki eşitlikleri geçerlidir.

$$P(d, 0^\circ) = P^T(d, 180^\circ),$$

$$P(d, 45^\circ) = P^T(d, 225^\circ),$$

$$P(d, 90^\circ) = P^T(d, 270^\circ),$$

$$P(d, 135^\circ) = P^T(d, 315^\circ)$$

Dolayısıyla, GLCM matrisi için açı( $\theta$ ) hesaplanırken  $\theta=0^\circ,45^\circ,90^\circ,135^\circ$  yönleri dikkate alınır. Haralick bu hesaplamalarda dört yön kullanılıyorsa elde edilen değerlerin ortalamasının alınmasını önermiştir [94].

Matris elemanı  $P(i, j | \Delta x, \Delta y)$ , uzaklığı ( $\Delta x, \Delta y$ ) olan referans piksel ve komşu pikselin göreceli frekansıdır. Ayrıca, matris elemanı  $P(i, j | d, \theta)$ , belirli bir uzaklık mesafesi ( $d$ ) ve belirli bir açıda ( $\theta$ )  $i$  ve  $j$  gri seviye değerleri arasındaki değişikliklerin ikinci dereceden istatistiksel olasılık değerlerini içerir [95].

$M \times N$  boyutunda  $G$  gri seviyeden oluşan bir  $f(m,n)$  giriş resmi verilsin. Bu durumda

$$P(i, j | \Delta x, \Delta y) = WQ(i, j | \Delta x, \Delta y) \quad (2.73)$$

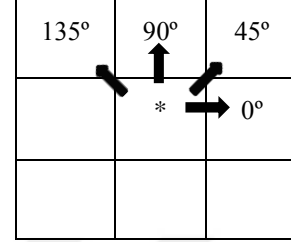
$$W = \frac{1}{(M-\Delta x)(N-\Delta y)}, Q(i, j | \Delta x, \Delta y) \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \text{ ve}$$

$$A = \begin{cases} 1 & \text{eğer } f(m, n) = i \text{ ve } f(m + \Delta x, n + \Delta y) = j \\ & \text{değilse } 0 \end{cases} \text{ olur.}$$

4x4 boyutunda 4 gri seviyeye sahip küçük bir alt görüntümüz olsun.

	0	1	2	3
0	0	0	1	2
1	0	1	3	2
2	3	3	3	1
3	2	2	1	1

(a)



(b)

Şekil 2.33. GLCM için gri seviye giriş matris görüntüsü ve referans piksel açılarının gösterimi (a) GLCM için 4x4'lük bir gri seviye giriş matris görüntüsü (b) Referans piksel için açılarının gösterimi.

Bu görüntüye ait  $d=1$  uzaklıkta ve  $\theta=0^\circ, 45^\circ, 90^\circ, 135^\circ$  yönlerinde elde edilmiş GLCM matrisleri Şekil 2.34'de verilmiştir.

	0	1	2	3
0	1	2	0	0
1	0	1	1	1
2	0	1	1	0
3	0	1	1	2

(a)

	0	1	2	3
0	1	0	0	0
1	0	2	0	0
2	0	0	0	2
3	0	1	2	1

(b)

	0	1	2	3
0	1	0	0	0
1	1	1	1	1
2	0	0	1	2
3	1	2	0	1

(c)

	0	1	2	3
0	0	0	0	0
1	1	0	0	2
2	0	1	0	1
3	2	1	0	0

(d)

Şekil 2.34. Şekil 2.33 (a)'da verilen giriş görüntüsü için  $d=1$  uzaklıkta (a)  $\theta=0^\circ$  (b)  $\theta=45^\circ$  (c)  $\theta=90^\circ$  (d)  $\theta=135^\circ$  yönlerinde elde edilmiş GLCM matrisleri.

*Haralick Özellikleri:* Haralick ve arkadaşları tarafından; açısız ikinci moment, kontrast,

korelasyon, kareler toplamı, ters fark momenti, toplam ortalama, toplam varyans, entropi, toplam entropi, fark varyansı, fark entropisi, birinci korelasyon ölçüsü, ikinci korelasyon ölçüsü, maksimum korelasyon katsayısı olmak üzere on dört öznelik belirlenmiştir [28]. Görüntüler için bu doku özellikleri GLCM vasıtasıyla hesaplanarak öznelik vektörleri oluşturulmaktadır.

Bu on dört öznelikten kontrast, fark entropisi, , toplam varyans, entropi, toplam entropi, toplam ortalama, birinci korelasyon ölçüsü, ikinci korelasyon ölçüsü aşağıdaki eşitlikler kullanılarak hesaplanabilmektedir. Burada;

$G$ : Gri seviye değerini,

$p(i,j)$ = normalize GLCM görüntünün  $(i,j)$ 'inci elemanını,

$$p_x(i) = \sum_{j=0}^{G-1} p(i, j) ,$$

$$p_y(j) = \sum_{i=0}^{G-1} p(i, j),$$

$\sigma_x, \sigma_y$  sırasıyla  $p_x$  ve  $p_y$  'nin standart sapmasını

$\mu$  ,  $p(i,j)$ 'nin ortalamasını,

$\mu_x$  ve  $\mu_y$  sırasıyla  $p_x$  ve  $p_y$  'nin ortalama değerlerini

$$\mu_x = \sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} p(i, j) = \sum_{i=0}^{G-1} i p_x(i),$$

$$\mu_y = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} j p(i, j) = \sum_{j=0}^{G-1} j p_y(j),$$

$$\sigma_x^2 = \sum_{i=0}^{G-1} (i - \mu_x)^2 \sum_{j=0}^{G-1} p(i, j) = \sum_{i=0}^{G-1} (p_x(i) - \mu_x(i))^2,$$

$$\sigma_y^2 = \sum_{j=0}^{G-1} (j - \mu_y)^2 \sum_{i=0}^{G-1} p(i, j) = \sum_{j=0}^{G-1} (p_y(j) - \mu_y(j))^2,$$

$$p_{x+y}(k) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) \quad i + j = k \quad (k = 2, 3, \dots, \dots, 2G)$$

$$p_{x-y}(k) = \sum_{i=1}^G \sum_{j=1}^G p(i, j) \quad |i - j| = k \quad (k = 0, 1, 2, \dots, \dots, (G - 1))$$

$HX$  ve  $HY$  ,  $p_x$  ve  $p_y$  değerlerine ait entropi değerlerini,

$$HXY = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) \log(p(i, j)),$$

$$HXY1 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) \log(p_x(i)p_y(j)),$$

$$HXY2 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p_x(i)p_y(j) \log(p_x(i)p_y(j)) \text{ ifade eder [28].}$$

**Kontrast:** Bir piksel ile komşusu arasındaki yoğunluk farkını ifade eder. GLCM de kontrast Denklem (2.74) ile hesaplanmaktadır [28].

$$Kontrast = \sum_{n=0}^{G-1} n^2 \{ \sum_{i=1}^G \sum_{j=1}^G p(i,j) \}, \quad |i - j| = n \quad (2.74)$$

*Entropi:* Entropi bir sistemdeki düzensizlik olarak tanımlanabilir. Entropi Denklem (2.75) ile hesaplanmaktadır [28].

$$Entropi = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \log(p(i,j)) \quad (2.75)$$

*Kareler Toplamı:* Kareler toplamı hesaplanırken Denklem (2.76) kullanılmaktadır.

$$KT = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 p(i,j) \quad (2.76)$$

*Toplam Ortalama:* Toplam ortalama hesaplanırken Denklem (2.77) kullanılmaktadır [28].

$$ToplamOrtalama = \sum_{i=2}^{2G} i p_{x+y}(i) \quad (2.77)$$

*Toplam Entropi:* Toplam entropi hesaplanırken Denklem 2.78 kullanılmaktadır [28].

$$ToplamEntropi = - \sum_{i=0}^{2G-2} p_{x+y}(i) \log(p_{x+y}(i)) \quad (2.78)$$

*Toplam Varyans:* Toplam varyans Denklem (2.79) ile hesaplanmaktadır [28].

$$ToplamVaryans = \sum_{i=2}^{2G} (i - ToplamEntropi)^2 p_{x+y}(i) \quad (2.79)$$

*Fark Entropisi:* Diferansiyel entropi hesaplanırken Denklem (2.80) kullanılır [28].

$$Fark Entropisi = - \sum_{i=0}^{G-1} p_{x-y}(i) \log(p_{x-y}(i)) \quad (2.80)$$

*Birinci Korelasyon Ölçüsü:* Birinci korelasyon ölçüsü Denklem (2.81) ile hesaplanmaktadır [28].

$$BKO = \frac{HXY - HXY1}{\max(HX, HY)} \quad (2.81)$$

*İkinci Korelasyon Ölçüsü:* İkinci korelasyon ölçüsü Denklem (2.82) ile hesaplanmaktadır [28].

$$IKO = (1 - \exp[-2(HXY2 - HXY)])^{1/2} \quad (2.82)$$

*Homojenlik:* Homojenlik Denklem (2.83) ile hesaplanmaktadır [96].

$$Homojenlik = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+|i-j|} p(i,j) \quad (2.83)$$

*Ters Fark Momenti (TFM):* Ters fark momenti homojenliği pikseller arasındaki ton farklılıklarını büyütür bulmaya çalışır. Görüntüdeki tüm elemanlar aynı değerdeyse homojenlik en üst seviyededir. Ters fark momenti Denklem (2.84) ile hesaplanmaktadır

[29].

$$TFM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+|i-j|^2} p(i,j) \quad (2.84)$$

Enerji: Enerji Denklem (2.85) ile hesaplanmaktadır [96].

$$Enerji = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j)^2 \quad (2.85)$$

Otokorelasyon: Oto korelasyon Denklem (2.86) ile hesaplanmaktadır [96].

$$OK = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{ij\} p(i,j) \quad (2.86)$$

Farklılık: Farklılık Denklem (2.87) ile hesaplanmıştır [96].

$$Frk = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i-j| p(i,j) \quad (2.87)$$

Küme Gölge: Küme gölge Denklem (2.88) ile hesaplanmaktadır [96].

$$KG = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j - \mu_x - \mu_y)^3 p(i,j) \quad (2.88)$$

Küme Önemi: Küme önemi Denklem (2.89) ile hesaplanmaktadır [96].

$$KO = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j - \mu_x - \mu_y)^4 p(i,j) \quad (2.89)$$

Maksimum Olasılık: Maksimum olasılık Denklem (2.90) ile hesaplanmaktadır [96].

$$MO = \max_{i,j} p(i,j) \quad (2.90)$$

Bu tez çalışmasında, eş oluşum matrisleri elde edilirken kullanılan parametrelerden gri seviye değeri  $G=32$ , mesafe  $d=1$  ve her bir açı değeri  $\theta=0^\circ, 45^\circ, 90^\circ, 135^\circ$  için ayrı ayrı hesaplanarak bu matrislerin ortalaması alınarak toplamda 17 öznitelik elde edilmiştir.

### 2.4.3. Frekansal Öznitelikler

Frekans uzayı sinyallerin zamana değil de frekansa bağlı olarak tanımlanıp analiz edildiği alandır. Zaman uzayı sinyalin zamana göre değişimini gösterirken frekans uzayı sinyalin bir frekans bandı içindeki değişimini gösterir. Bir sinyal bazı matematiksel denklemler ile zaman uzayından frekans uzayına veya frekans uzayından zaman uzayına geçirilebilir. Bu işlemlere dönüşüm adı verilir. Bu tez çalışmasında mamogramlardan elde edilen ROI'lere hem dalgacık hem de eğricik alanında dönüşüm uygulanarak öznitelikler elde edilmiştir.

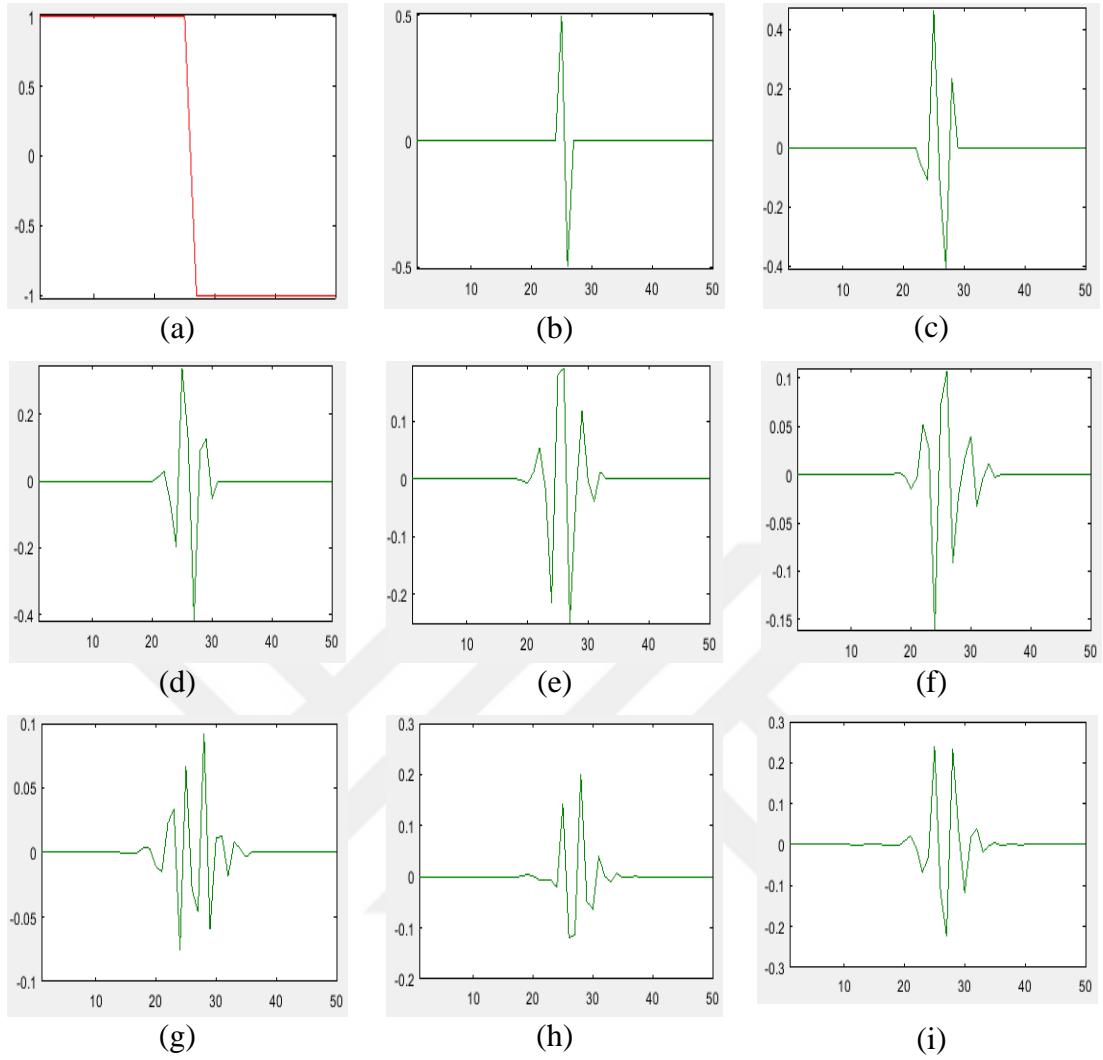
### *İki Boyutlu Dalgacık Dönüşümü ile Öznitelik Çıkarımı*

Dalgacık analizi bölüm 2.2’de ayrıntılı anlatıldığından bu bölümde sadece öznitelik çıkarımı için kullanılan daubechies dalgacı ve kullanılan istatistiksel formüller sunulmuştur.

Fourier Dönüşümü ile dalgacık dönüşümü karşılaştırıldığında aralarındaki önemli farklardan birisi dalgacık dönüşümünde dalgacık türünü seçebilme olanağının olmasıdır[97]. Fakat dalgacıkları seçmek için belirli bir algoritma yoktur bu nedenle seçilen dalgacık, üzerinde çalışılacak sinyale zaman düzleminde görsel olarak ne kadar çok benzerse o kadar çok verim elde edilir. Fakat bu şekilde bir seçim yapmak oldukça güçtür. Dolayısıyla üzerinde çalışılacak sinyale tüm dalgacıklar denenerek en yüksek verimi sağlayan dalgacık seçilmelidir.

*Daubechies Dalgacıkları:* Daubechies dalgacıkları dbM şeklinde kısaca isimlendirilebilmektedir. M daubechies dalgacıklarının türünü göstermektedir. Dalgacıklar simetrik değildir ve analizi ortogonaldır. Hem sürekli hem ayrık dalgacık dönüşümü yapabilmektedir [97]. Şekil 2.35 te farklı Daubechies dalgacıklarının gösterimi sunulmuştur.





Şekil 2.35. Farklı Daubechies dalgacıklarının gösterimi (a) Orijinal sinyal (b) db1 (c) db2 (d) db3 (e) db4 (f) db5 (g) db6 (h) db7 (i) db8.

Dalgacık türlerinden daubechies1, symlet1, coiflet1 ve biorspline1.1 haar dalgacığı ile aynıdır [98]. Dalgacıklar ile öznitelik çıkarımı için, hem yapılan deneysel çalışmalar hem de literatür taraması sonucunda daubechies 8 dalgacığı ile diğerlerinden daha yüksek verim elde edildiğı görülmüş dolayısıyla bu tez çalışmasında daubechies 8 dalgacığı öznitelik çıkarımı için tercih edilmiştir. Buna göre daha önce elde edilen şüpheli bölgelerdeki normal dokuları ve kanserli dokuları analiz etmek amacıyla db8 dalgacığı kullanılarak ayrıntı ve detay katsayıları dördüncü seviyeye kadar elde edilmiştir. Elde edilen bu katsayılardan ayrıştırma vektörü  $C$  kullanılarak Denklem (2.91)'deki formül ile enerji hesaplanmış, ayrıntı katsayıları kullanılarak ortalama ve standart sapma değerleri, Denklem (2.92)-(2.97)'deki formüller ile elde edilmiştir [99].

$$\text{Enerji } (E_2) = \frac{1}{NM} \sum_i^N \sum_j^M |C_2(i, j)|^2 \quad (2.91)$$

$$\text{Ortalama } (\bar{\mu}_{h2}) = \frac{1}{NM} (\sum_i^N \sum_j^M D_{h2}(i, j)) \quad (2.92)$$

$$\text{Standart Sapma } (\sigma_{h2}) = \sqrt{\frac{1}{NM-1} \sum_i^N \sum_j^M (D_{h2}(i, j) - \bar{\mu}_{h2})^2} \quad (2.93)$$

$$\text{Ortalama } (\bar{\mu}_{v2}) = \frac{1}{NM} (\sum_i^N \sum_j^M D_{v2}(i, j)) \quad (2.94)$$

$$\text{Standart Sapma } (\sigma_{v2}) = \sqrt{\frac{1}{NM-1} \sum_i^N \sum_j^M (D_{v2}(i, j) - \bar{\mu}_{v2})^2} \quad (2.95)$$

$$\text{Ortalama } (\bar{\mu}_{d2}) = \frac{1}{NM} (\sum_i^N \sum_j^M D_{d2}(i, j)) \quad (2.96)$$

$$\text{Standart Sapma } (\sigma_{d2}) = \sqrt{\frac{1}{NM-1} \sum_i^N \sum_j^M (D_{d2}(i, j) - \bar{\mu}_{d2})^2} \quad (2.97)$$

Tez çalışmasında dördüncü seviyeye kadar katsayılar hesaplanmıştır ve 2. , 3. ve 4. dereceden elde edilen katsayılar ile enerji için  $E_2, E_3, E_4$  olmak üzere 3 adet öznelik elde edilmiştir. Ayrıca her ayrışma seviyesine ait detay katsayıları için ortalama ve standart sapma değerleri  $(\bar{\mu}_{h2}, \bar{\mu}_{h3}, \bar{\mu}_{h4}, \bar{\mu}_{v2}, \bar{\mu}_{v3}, \bar{\mu}_{v4}, \bar{\mu}_{d2}, \bar{\mu}_{d3}, \bar{\mu}_{d4}, \sigma_{h2}, \sigma_{h3}, \sigma_{h4}, \sigma_{v2}, \sigma_{v3}, \sigma_{v4}, \sigma_{d2}, \sigma_{d3}, \sigma_{d4})$  olmak üzere 18 adet öznelik elde edilmiştir. Toplamda dalgacık analizi için 21 adet öznelik elde edilmiştir.

#### 2.4.3.1. Eğricik Dönüşümü ile Öznelik Çıkarımı

Eğricik dönüşümü Candès ve Donoho tarafından 2000 yılında ortaya konmuş bir yaklaşımdır [100]. Eğricik dönüşümü, birçok yön ve pozisyonda, her bir ölçekte analiz yeteneğine sahip, çok ölçekli standart olmayan bir yapıdır [101]. Bu dönüşümünün, temel yeniliği, anizotropik ölçekleme ilkesine dayanmasıdır [102]. Eğricik dönüşümündeki amaç genel bir eğri boyunca süreksizlikle pürüzsüz nesnelere temsil etmektir [102]. Eğricik dönüşümünde ilk olarak birinci nesil olarak adlandırılan yaklaşım sunulmuş ve bu yöntemin eksileri daha sonra aynı araştırmacılar tarafından ikinci nesil olarak adlandırılan ve farklı bir frekans ayırma tekniğinin kullanılması ile geliştirilmiş eğricik dönüşümü ile giderilmeye çalışılmıştır [101]. Eğricik dönüşümünün bu yeni geliştirilen şekli Hızlı Ayrık Eğricik Dönüşümü (FDCT) olarak adlandırılmıştır. Bu yeni yöntem daha basit ve daha hızlıdır. Candès ve arkadaşları tarafından FDCT'nin iki uygulaması önerilmiştir. Bunlar: Eşit Olmayan Aralıklı Hızlı Fourier Dönüşümü (USFFT) ve sarma fonksiyonu olarak belirtilmiştir. Eğricik dönüşümü için ilk önce

görüntünün 2 boyutlu hızlı Fourier dönüşümü (FFT) elde edilir. Daha sonra elde edilen frekans düzleminde takozlar oluşturulur. Takozlar Fourier düzleminin, radyal ve açısıl olarak eş merkezli daireler halinde bölümlere ayrılmasıyla meydana gelmektedir [103]. Eş merkezli daireler görüntünün çok ölçekli ayrıştırılmasını sağlarken açısıl bölmeler farklı açılar veya yönelimleri belirtir. Dolayısıyla, belli bir takozu belirtmek için öncelikle ölçeğin ve açının tarif edilmesi gerekir  $(j, \theta)$  [103]. Son olarak, eğri katsayılarını bulmak için her takoz için ters FFT uygulanır [102].

*Sürekli Zaman Eğricik Dönüşümü:* Sürekli zaman eğricik dönüşümü için iki boyutlu uzayda  $(R^2)$ , bir uzaysal  $x$  değişkeni ve frekans düzleminde  $\omega$  değişkeni verilmiş olsun. Ayrıca  $r$  ve  $\theta$  'da frekans düzleminde kutupsal koordinatları ifade etsin. Buna göre; sürekli zamanda eğricik dönüşümü alınırken  $W(r)$ , 'radyal pencere' ve  $V(t)$  'açısıl pencere' olarak adlandırılan bir çift pencere kullanılmaktadır[104]. Bu pencerelerin her ikisi de pozitif ve gerçel değerlidir. Bu iki pencere  $W$  'nun  $r \in (1/2, 2)$  aralığında,  $V$  'nin de yine  $t \in [-1, 1]$  aralığında tanımlı olduğu ve pozitif reel değerler aldığı durumlarda aşağıdaki şartları sağlarlar [104].

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1 \quad r \in (3/4, 3/2) \quad (2.98)$$

$$\sum_{l=-\infty}^{\infty} V^2(t - l) = 1 \quad t \in (-1/2, 1/2) \quad (2.99)$$

Her  $j \geq j_0$  için, Fourier alanında tanımlanan frekans penceresi  $U_j$  Denklem (2.100) deki gibi tanımlanabilir [104],

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor} \theta}{2\pi}\right) \quad (2.100)$$

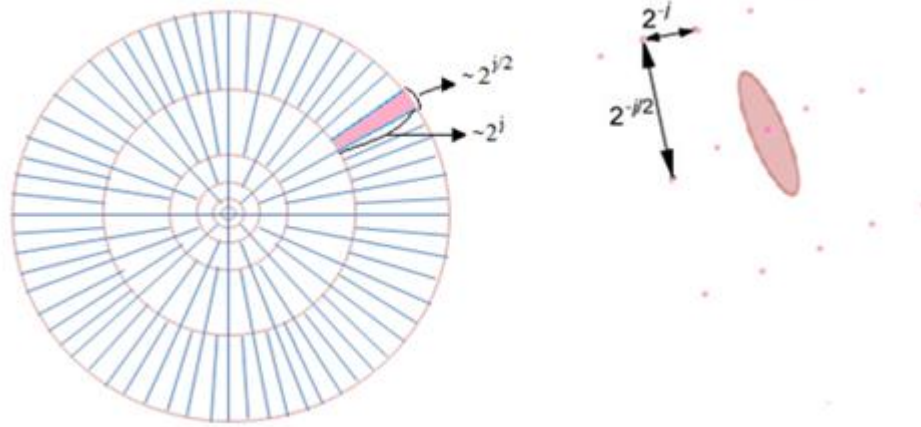
Burada  $\lfloor j/2 \rfloor$ ,  $j/2$ 'nin tam sayı kısmını ifade etmektedir ve  $U_j$ ,  $W$  ve  $V$  yardımıyla tanımlanan kutupsal bir takozdur. Bu takoz radyal ve açısıl yönlerde ölçek bağımlı pencere genişliği ile uygulanır [104]. Reel değerli eğricikler elde etmek için, Denklem (2.100)'ün simetrik versiyonuyla  $(U_j(r, \theta) + U_j(r, \theta + \pi))$  çalışılır.

Fourier dönüşümü  $\widehat{\varphi}_j(\omega) = U_j(\omega)$  aracılığıyla dalga formu  $\varphi_j(x)$  tanımlanır. Burada  $\varphi_j$  ana eğriciktir ve  $2^{-j}$  ölçeğindeki bütün eğricikler bu ana eğriciğin yönelimleri ve ötelemeleri kullanılarak elde edilmişlerdir.

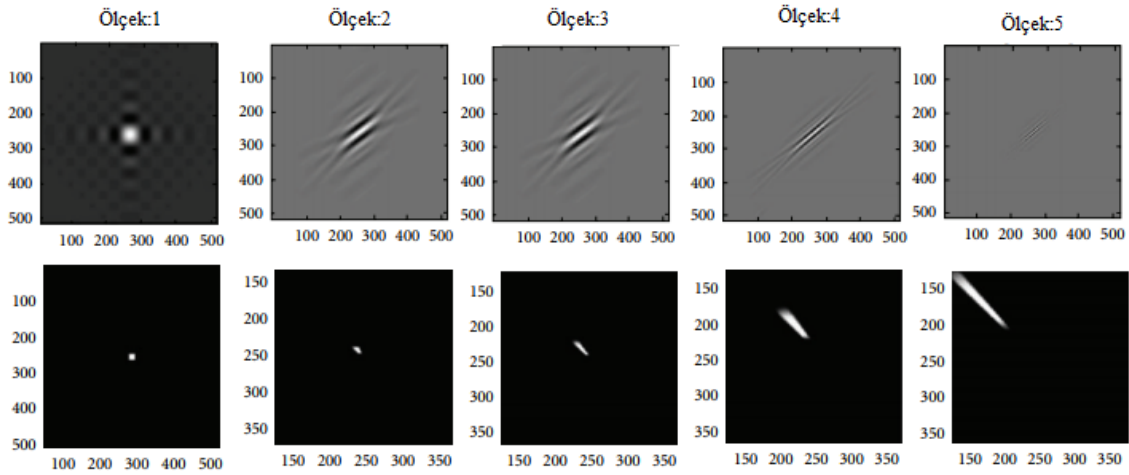
$$\varphi_{j,k,l}(x) = \varphi_j\left(R_{\theta_1}(x - x_k^{(j,l)})\right) \quad (2.101)$$

Denklem (2.101)  $\{x = (x_1, x_2)\}$  şeklindeki bir fonksiyonun  $2^{-j}$  ölçeğinde,  $\theta_1$  yöneliminde ve  $x_k(j, l)$  pozisyonunda eğricik dönüşümünü ifade eder. Burada  $R_\theta$  radyan cinsinden dönme miktarıdır.

Şekiller 2.36 ve 2.37 'de verilmiş olan bir ölçek ve yönelim için hesaplanan frekans düzlemi ve uzaysal kartezyen ızgarası ile farklı ölçeklerde takozların frekans düzleminde ve uzaysal düzlemde gösterimi sunulmuştur.



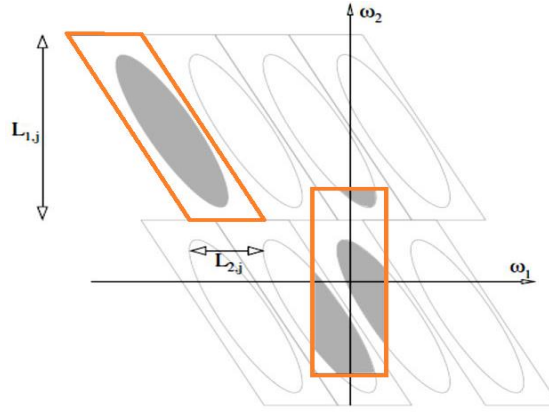
Şekil 2.36. Takoz ve eğricik gösterimi (a) Fourier düzleminde takoz gösterimi  
(b) Kartezyen ızgarada eğricik gösterimi [105].



Şekil 2.37. Farklı ölçeklerde takozların frekans düzleminde ve uzaysal düzlemde gösterimi [106].

*Sarma Yoluyla Ayrık Eğricik Dönüşümü:* Sarma yoluyla ayrık eğricik dönüşümü yöntemi uygulanırken ilk olarak iki boyutlu giriş görüntüsüne hızlı Fourier dönüşümü yapılarak Fourier örnekleri çıkarılır  $\hat{f}[n_1, n_2]$ ,  $-\frac{n}{2} \leq n_1, n_2 \leq n/2$ . Ardından pencere

fonksiyonu ile Fourier örneklerinin çarpımı her bir açı  $l$  ve ölçek  $j$  için üretilir,  $\tilde{U}_{j,l}[n_1, n_2]\hat{f}[n_1, n_2]$ . Buradan elde edilen çarpım sonucu orijin etrafında sarmalanır ve yeniden indekslenir  $\hat{f}_{j,l}[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f})[n_1, n_2]$ . Burada  $n_1$  ve  $n_2$  değerleri  $0 \leq n_1 \leq L_{1,j}$  ve  $0 \leq n_2 \leq L_{2,j}$  aralıklarında yer alırlar ve  $L_{1,j}$  ve  $L_{2,j}$  değerleri sarma işleminde paralel yüzölçüme ait ayrıtları göstermek amacıyla kullanılmıştır [9]. Ayrık eğricik katsayıları  $C^D(j, l, k)$  ise her bir  $\hat{f}_{i,j}$ 'ye ters 2B FFT uygulanarak elde edilir [9]. Paralel kenar içindeki verinin periyodik bir şekilde bir dikdörtgene sarmalanması Şekil 2.38'de sunulmuştur.



Şekil 2.38. Verinin orijin etrafında sarmalanması.

Şekilde  $\theta$  açısı  $(-\pi/4, \pi/4)$  alanındadır. Dikdörtgenin merkezi orijindedir. Periyotlandırmadan kaynaklanan kopyalar gri renkli paralel kenarlar olarak görülmektedir. Turuncu renkli paralel kenar ise eğriciğin frekans desteğini içeren fayans  $P_{j,l}$  dir. Sarılmış elips parçalara ayrılmış gibi gözükse de bu zıt kenarları tanımlanan periyodik dikdörtgende herhangi bir soruna neden olmamaktadır [9].

Bu tez çalışmasında öznitelik çıkarım amacıyla son olarak sarma yoluyla ayrık eğricik dönüşümü kullanılmıştır. Bu işlem sırasında belirlenen ölçek derecesi 4 ve 2. en kaba seviyedeki açılardan sayısı ise 16 olarak belirlenmiştir. Belirlenen kriterler ile elde edilen her bir eğricik dönüşümü seviyesi için tüm katsayıların standart sapma, ve çarpıklık değerleri hesaplanarak toplamda 162 adet öznitelik edilmiştir.

$$\text{Standart Sapma } (S_{\text{eğricik}}) = \sqrt{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (X(i, j) - \bar{X})^2} \quad (2.102)$$

$$\text{Çarpıklık } (SK_{\text{eğricik}}) = \frac{1}{NM.(S_{\text{eğricik}})^3} \left( \sum_{i=1}^N \sum_{j=1}^M (X(i, j) - \bar{X})^3 \right) \quad (2.103)$$

Burada  $N$  ve  $M$  katsayı matrisinin boyutlarını ifade ederken  $\bar{X}$ , katsayı matrisinin aritmetik ortalamasını ifade etmektedir.

Mamogramlardan görüntü işleme teknikleriyle kanser tespiti yapabilmek amacıyla özneliklerin elde edilmesi işlemi tamamlandığından tezin bir sonraki aşamasında elde edilen özneliklerin azaltılması amacıyla gerçekleştirilen işlemler anlatılmaktadır.

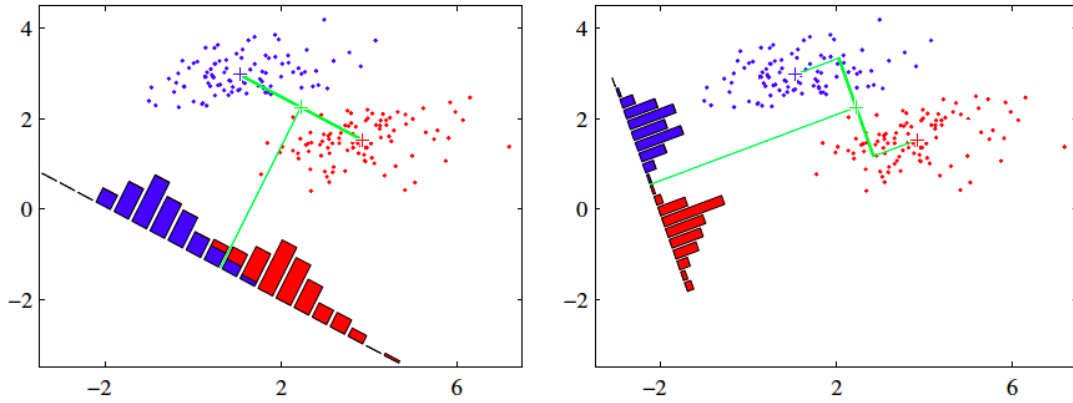
## 2.5. ÖZELLİK AZALTMA ALGORİTMALARI

Öznelik azaltma algoritmaları, sınıflandırılacak veriler arasında etkin bir ayırım yaparken, verilerle ilgili yeterli bilgiyi tutarak veri boyutunu azaltmak için kullanılırlar [107]. Öznelik azaltma için literatürde genel olarak iki farklı yöntem kullanılmaktadır. Bunlardan ilki, özgün özneliklerden bir alt grup seçimi olarak tanımlanan öznelik seçimi, diğeri ise özgün özneliklerin bilgi içeriğini özetleyen daha az sayıda yeni öznelik oluşturmak için istatistiksel dönüşümlerin uygulandığı öznelik çıkarımı tekniğidir [108]. Literatürde bu iki yaklaşımı esas alan birçok farklı öznelik azaltma algoritması bulunmaktadır. Örneğin Emary ve arkadaşları [109] yaptıkları çalışmada, öznelik azaltma amacıyla kaba küme set ve yarasa algoritmasını kullanmışlardır. Kullanıldıkları fitness fonksiyonu ile hem sınıflandırma doğruluğunu hem de seçilen özneliklerin sayısını belirleyerek sınıflandırma performansını ve küçültme boyutunu dengelemişlerdir. Yazarlar önerilen algoritmayı GA-tabanlı ve kaba-set algoritmalarına dayanan öznelik azaltma için PSO yaklaşımlarıyla karşılaştıran VCI veri setleri kullanılarak test etmişlerdir. Farklı veri setleri üzerindeki sonuçlar, yarasa algoritmasının, kaba küme set öznelik seçimi için etkili olduğunu göstermiştir. [110]'de yer alan çalışmada ise öznelik ağırlıklı entropi üzerine bulanık kümeleme algoritması ile öznelik azaltma gerçekleştirilmiştir. Bu çalışmada otomatik olarak bireysel öznelik ağırlığını hesaplayabilen ve aynı zamanda alakasız öznelik bileşenlerini azaltabilecek bulanık kümeleme algoritmalarını geliştirmek için yeni bir yöntem sunulmuştur [110]. Tran ve arkadaşları [111], yüksek boyutlu sınıflandırma için PSO tabanlı melez öznelik seçimi algoritmasını geliştirmişlerdir. Yapılan çalışmada filtreleme ve sargı yaklaşımlarını tek bir evrim sürecinde birleştirerek, daha kısa bir zamanda daha iyi sınıflandırma performansı ile daha küçük öznelik alt grupları elde etmek için yeni bir strateji önerilmiştir. Önerilen yöntem farklı zorluk derecesine sahip sekiz yüksek boyutlu problem üzerine üç yeni PSO tabanlı yöntemle karşılaştırılmıştır. Sonuçlar yeni melez PSO'nun diğeri yöntemlerden daha etkili ve verimli olduğunu

göstermektedir. Alexandre ve arkadaşları [112], botnetlerin tespiti amacıyla özellik seçimi için bir çalışma yapmışlardır. Bu çalışmada komuta kontrolü (C & C) aşamasında botnet'leri tespit etmek amacıyla öznelik seçimi yapmak için yeni bir yöntem sunmuşlardır. Makine öğrenmesi için C4.5 algoritmasını, en yüksek algılama oranını veren öznelikler kümesini seçmek için Genetik Algoritmayı kullanarak bir botnet'e ait olan veya olmayan bağlantılar arasında sınıflandırma yapılmıştır. Sonuçlar önemli oranda azalma ve daha yüksek algılama hızı elde edilmiş olduğu göstermektedir. Bu tez çalışmasında Lineer Ayırıcı Analizi(LAA) algoritması öznelik azaltmak amacıyla kullanılmıştır.

### 2.5.1. Lineer Ayırıcı Analiz(LAA)

Lineer ayırıcı analizi (LAA) , R. A. Fischer [113] tarafından geliştirilen ve sınıfları en iyi şekilde ayırmak için var olan özneliklerin lineer bir kombinasyonunu bulmak amacıyla kullanılan bir yöntemdir. Bu kombinasyon doğrusal sınıflandırma veya boyut azaltma amacıyla kullanılmaktadır. LAA algoritması, regresyon analizi ve varyans analizi ile yakından ilişkilidir [9]. LAA yönteminde kullanılan veri setinde sınıflar arası varyansın sınıf içi varyansa oranı maksimize edilerek en iyi ayrılabilirlik sağlanmaya çalışılır [12]. Şekil 2.39'da Bir Lineer ayırıcı analiz örneği görülmektedir.



Şekil 2.39. Lineer ayırıcı analiz örneği.

Şekilde sol tarafta iki sınıfa ait örnekler ve bu sınıfların ortalamalarını birleştiren doğruya göre histogramları görülmektedir. Dikkat edilirse belirtilen doğruya göre bakıldığında bu sınıfların belirli oranda çakıştıkları görülmektedir. Sağ tarafta ise aynı sınıfın LAA ile ayrılmış hali bulunmaktadır. Burada LAA sonucunda sınıf ayrımının büyük ölçüde iyileştirildiği görülmektedir [114].

LAA ile sınıf ayrımı için ilk olarak  $X$  örnekleri belirli bir doğruya yansıtılıp, skalar  $y$  elde edilmeye çalışılır [114].

$$y = \theta^T X \quad (2.104)$$

Denklemden  $\theta$  yönü ifade etmektedir. LAA iki sınıfa uygulanırken  $d$  boyutlu bir  $X$  vektörü olduğu varsayılır.  $X = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$  olmak üzere  $N_1, C_1$  sınıfına ve  $N_2, C_2$  sınıfına ait olsun.

Bu iki sınıfın  $X$  uzayında ortalama vektörü:

$$u_k = \frac{1}{N_k} \sum_{i \in C_k} x^{(i)} \quad k = 1, 2 \quad (2.105)$$

ve  $y$  uzayında ortalama vektörü şu şekilde tanımlansın:

$$\hat{u}_k = \frac{1}{N_k} \sum_{i \in C_k} y^{(i)} = \frac{1}{N_k} \sum_{i \in C_k} \theta^T x^{(i)} = \theta^T u_k \quad k = 1, 2 \quad (2.106)$$

Bu durumda sınıflar arasındaki varyans aşağıdaki gibi olacaktır.

$$\hat{u}_2 - \hat{u}_1 = \theta^T (u_2 - u_1). \quad (2.107)$$

Ayrıca her bir  $C_k$  sınıfı için sınıf içi varyans Denklem (2.108)'deki gibi tanımlanabilir.

$$\hat{s}_k^2 = \sum_{i \in C_k} (y^{(i)} - \hat{u}_k)^2 \quad k = 1, 2 \quad (2.108)$$

Bu durumda sınıf içi ve sınıflar arası varyans ile elde edilen amaç fonksiyon  $J(\theta)$  aşağıdaki gibi tanımlanır [115].

$$J(\theta) = \frac{(\hat{u}_2 - \hat{u}_1)^2}{(\hat{s}_1^2 + \hat{s}_2^2)} \quad (2.109)$$

Burada, amaç fonksiyon  $J$ 'yi maksimize ederek, en iyi ayrılabilirlik sağlanmaya çalışılır. Bunun için optimum  $\theta^*$  bulunmalıdır. Optimum  $\theta^*$ 'nın bulunması için öncelikle saçılım elde edilir. Saçılım için bazı formüller aşağıdaki gibidir:

$$X \text{ özellik uzayındaki saçılma: } S_k = \sum_{i \in C_k} (x^{(i)} - u_k)(x^{(i)} - u_k)^T \quad (2.110)$$

$$\text{Sınıf içi saçılma matrisi: } S_W = S_1 + S_2 \quad (2.111)$$

$$\text{Sınıflar arası saçılma matrisi: } S_B = (u_2 - u_1)(u_2 - u_1)^T \quad (2.112)$$

$X$  özellik uzayındaki saçılma matrisi  $y$  cinsinden ifade edilebilir. Bu durumda;

$$\begin{aligned} \hat{s}_k^2 &= \sum_{i \in C_k} (y^{(i)} - \hat{u}_k)^2 = \sum_{i \in C_k} (\theta^T x^{(i)} - \theta^T u_k)^2 \\ &= \sum_{i \in C_k} \theta^T (x^{(i)} - u_k)(x^{(i)} - u_k)^T \theta = \theta^T S_k \theta \end{aligned} \quad (2.113)$$

şeklinde elde edilir. Dolayısıyla, aşağıdaki eşitlik elde edilmiş olur.

$$\hat{s}_1^2 + \hat{s}_2^2 = \theta^T S_1 \theta + \theta^T S_2 \theta = \theta^T S_W \theta \quad (2.114)$$

Benzer şekilde,

$$(\hat{u}_2 - \hat{u}_1)^2 = (\theta^T u_2 - \theta^T u_1)^2 = \theta^T (u_2 - u_1)(u_2 - u_1)^T \theta = \theta^T S_B \theta \quad (2.115)$$

elde edilebilir. Son olarak Fisher kriteri  $S_W$  ve  $S_B$  cinsinden ifade edildiğinde amaç fonksiyon  $J(\theta)$  aşağıdaki gibi yazılabilir [114].

$$J(\theta) = \frac{\theta^T S_B \theta}{\theta^T S_W \theta} \quad (2.116)$$

Bu amaç fonksiyonu maksimize etmek için, amaç fonksiyonun türevi alınıp sıfıra eşitlendiğinde

$$\theta = \frac{(u_2 - u_1)^T \theta}{J} S_W^{-1} (u_2 - u_1) \quad (2.117)$$

olarak bulunur.

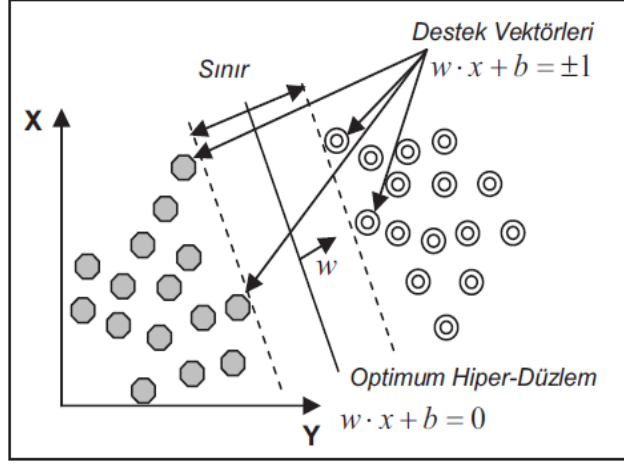
$y = \theta^{*T} X$ , Fisher'in doğrusal ayırıcılığı olarak bilinmesine rağmen aslında bir ayırt ediciden daha ziyade verilerin bir boyuta yansıtılması için özel bir yön seçimidir [114].

## 2.6. SINIFLANDIRMA

Sınıflandırmayı, verinin içerdiği benzer özelliklere göre farklı özellikte olanlardan ayırt edilmesi şeklinde tanımlayabiliriz. Günümüzde sınıflandırma veri madenciliği, tıbbi tanılama sistemleri ve istatistik gibi farklı alanlarda kullanılmaktadır. Bu tez çalışmasında iki tür sınıflandırma yapılmaktadır. Bunlardan ilki mamogramın normal/anormal olarak sınıflandırması, diğeri ise iyi huylu/kötü huylu olarak sınıflandırılmasıdır. Bu sınıflandırmalar için üç farklı sınıflandırma yöntemi kullanılmıştır. Bunlar Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA) ve K-en yakın komşuluk(K-EK) sınıflandırıcısıdır.

### 2.6.1. Destek Vektör Makineleri (DVM)

Destek Vektör Makineleri(DVM) istatistiksel öğrenme teorisine dayanan kontrollü sınıflandırma algoritmasıdır. DVM ilk olarak iki sınıflı doğrusal verilerin sınıflandırılması için tasarlanmış daha sonra çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması için genelleştirilmiştir. DVM iki sınıflı doğrusal verileri ayırmak için en uygun karar fonksiyonunu tahmin eder, iki sınıfı birbirinden en uygun şekilde ayırabilen hiper düzlemi tanımlar [115].



Şekil 2.40. Doğrusal olarak ayrılabilen veri setleri için hiper- düzlemin belirlenmesi [116].

Örneğin Şekil 2.40 da  $k$  adet örnek barındıran bir veri setimiz olsun. DVM'nin amacı bu iki sınıfı ayırabilen hiper düzlemler içinden en yakın noktalar arasındaki uzaklığı en büyük değere çıkararak hiper düzlemin bulunmasıdır. Eğitim verisini en uygun şekilde ayırabilecek hiper düzlem karar fonksiyonu kullanılarak bulunur ve optimum hiper düzlem olarak adlandırılır. Hiper düzlemden en yakın veri noktasına olan en küçük uzaklığa sınır denir. Sınır genişledikçe, daha başarılı genelleme ile ayırma işlemi yapar. Bir verinin hiper düzlem ile iki sınıfa ayrılabilmesi için uygun bir dönüşüm yeterlidir. Destek vektörleri hiper düzleme en yakın öğrenme verileri olarak adlandırılmaktadır. DVM sınıflayıcının toleransı ise ayırıcı hiper düzlem ile en yakın pozitif ve negatif veri noktaları arasındaki mesafeye denir ve aşağıdaki şekilde formüle edilir.

$$D(x) = (w, x) + b \quad (2.118)$$

Formülde;  $w$  ve  $b$  DVM'nin tahmin edeceği hiper düzlem parametrelerini,  $x$  ise yüksek boyutlu uzaya haritalanan veri kümesi vektörünü ifade etmektedir.

#### 2.6.1.1. Doğrusal Ayrılamayan Veriler için Destek Vektör Makineleri

Karşılaşılan çoğu problemde verilerin doğrusal olarak ayrılması pek de mümkün olmamaktadır. Böyle durumlarda bir yapay değişkenin ( $\xi_i$ ) tanımlanmasıyla, eğitim verilerinin bir kısmının optimum hiper düzlemin diğer tarafında kalmasından kaynaklanan problem çözülmüş olur. Ayrıca sınırı maksimize ederken yanlış sınıflandırma hatalarını minimize ederek denge sağlayan  $C$  adını verdiğimiz pozitif değerler alan bir düzenleme parametresiyle kontrol sağlanabilmektedir. Düzenleme

parametresi ve yapay deęişken kullanılarak doğrusal olarak ayırım yapılamayan veriler için optimizasyon problemi aşağıdaki gibi tanımlanır.

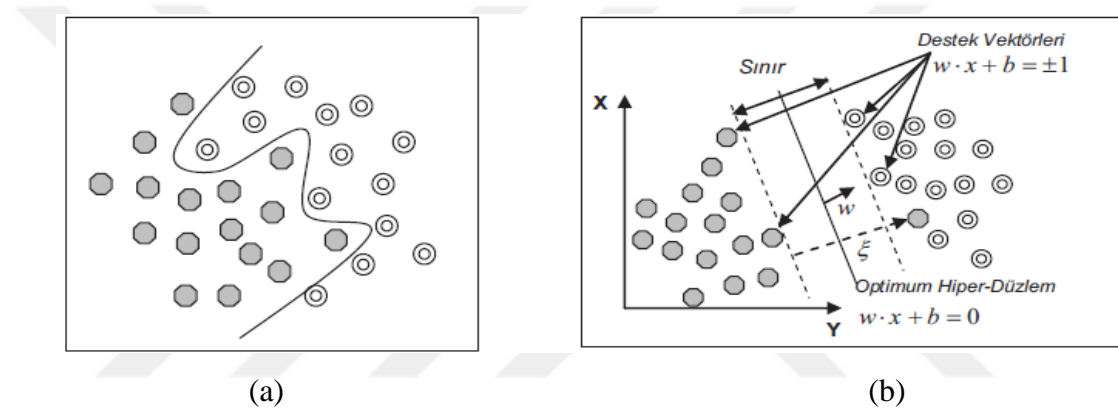
$$\min \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^t \xi_i \right] \quad (2.119)$$

Denklem (2.119)'daki  $\|w\|$  deęeri, optimum hiper düzlemin sınırının maksimize edilmesi için kullanılan deęerdir ve bu deęerin minimum olması gerekmektedir.

Denklemden yer alan sınırlamalar ise aşağıdaki gibi ifade edilirler [116].

$$y_i(w \cdot \varphi(x_i) + b) - 1 \geq 1 - \xi_i ; \xi_i \geq 0 \text{ ve } i = 1, 2, \dots, N \quad (2.120)$$

Aşağıdaki şekilde doğrusal olarak ayrılmayan veri setleri için hiper düzlemin belirlenmesi şematik olarak gösterilmiştir.



Şekil 2.41. Doğrusal olarak bölünemeyen veri setleri için hiper düzlem tespiti (a) Doğrusal olarak bölünemeyen veri seti , (b) Doğrusal olarak bölünemeyen veri setleri için hiper-düzlemin tespiti.

### 2.6.2. Yapay Sınır Ağları (YSA)

YSA'yı insan beyninin özelliklerinden yola çıkılarak geliştirilen bilgisayar sistemleri olarak tanımlayabiliriz. Yapay sınır ağlarında da öğrenme insan beynine benzer şekilde gerçekleşmektedir. İnsan beyninin çalışmasını sağlayan en temel öğelerden biri olan biyolojik sınır ağları sayesinde insanlar davranışları ve çevresini anlayabilmektedir. Bir sınır ağı milyarlarca sınır hücrenin bir araya gelmesi ile oluşmuştur [117]. Bir sınır hücresi işlediği bilgileri akson'ları yolu ile diğer hücrelere aktarırken yapay sınır hücreleri bunun için bir toplama fonksiyonu kullanır. Dışarıdan gelen bilgileri toplama fonksiyonu ile toplar ve aktivasyon fonksiyonu ile işleyerek elde ettiği çıktıyı diğer hücrelere ( proses elemanlarına) gönderir. Farklı amaçlar için uygun toplama ve aktivasyon fonksiyonları vardır. Ağırlık deęerleri yapay sınır ağlarını birbirine bağlayan

bağlantıların değerleri olarak tanımlanmaktadır. Proses elemanları girdi katmanı, ara katman ve çıkış katmanı olmak üzere birbirine paralel 3 katmandan oluşmaktadır. Ağın girdi katmanından bilgiler sunulur. Ara katmanlarda işlenerek oradan çıktı katmanına gönderilirler. Ara katmanda gerçekleştirilen işlemler aslında ağa gelen bilgilerin ağırlık değerleri ile çıktıya dönüştürülmesidir. Burada ağırlıkların doğru değerleri alması, doğru çıktı için gereklidir. Ağın eğitilmesi, doğru ağırlıkların bulunması işlemi olarak tanımlanır ve bu değerler başlangıçta rasgele atanırlar. Eğitim sırasında örnekler ağa sunuldukça öğrenme kuralına bağlı olarak ağırlıklar değiştirilir ve en doğru değerler bulunmaya çalışılır. Bütün bu işlemler ağın eğitim setindeki örneklerin tamamı için uygun ağırlıkları bularak doğru sonuca ulaşmasına kadar devam eder. Ağın eğitilmesinden sonra test setindeki örnekler ağa sunulur. Test setindeki örnekler için ağın doğru çıktılar üretmesi ağın istenen amacı gerçekleştirdiği anlamına gelmektedir. Ağın bir olayı iyi bir şekilde öğrenmesi için en doğru yapay sinir ağı modelinin seçilmesi gerekmektedir. Şimdiye kadar geliştirilen pek çok yapay sinir ağı modeli vardır. Bir yapay sinir ağının modelini kullanılan toplama fonksiyonu, ağın topolojisi, kullanılan aktivasyon fonksiyonu, öğrenme stratejisi ve öğrenme kuralı karakterize etmektedir [118]. Yapay sinir ağları sınıflandırma, genelleme, öğrenme, ilişkilendirme, özellik tespiti, optimizasyon gibi pek çok alanda başarılı bir şekilde uygulanmaktadırlar.

#### 2.6.2.1. Yapay Sinir Hücresi (İşlemci Eleman)

Yapay sinir ağı hücresi temelde beş elemandan oluşmaktadır. Bunlar:

1. *Girdiler:* Dış dünyadan yapay sinir hücresine (proses elemanına) gelen bilgilerdir. Bilgiler hem dış dünyadan hem de ağın kendisinden gelebilir.

2. *Ağırlıklar:* Hücreye gelen bilginin hücre üzerindeki etkisini ve önemini ağırlıklar gösterir. Ağırlığın küçük ya da büyük olması önemini simgelemez.

3. *Toplama Fonksiyonu:* Bu fonksiyon ile hücreye gelen net girdi hesaplanır. Toplam fonksiyonu için farklı fonksiyonlar kullanılsa da genellikle ağırlıklı toplam fonksiyonu kullanılır ve aşağıdaki şekilde ifade edilir [117].

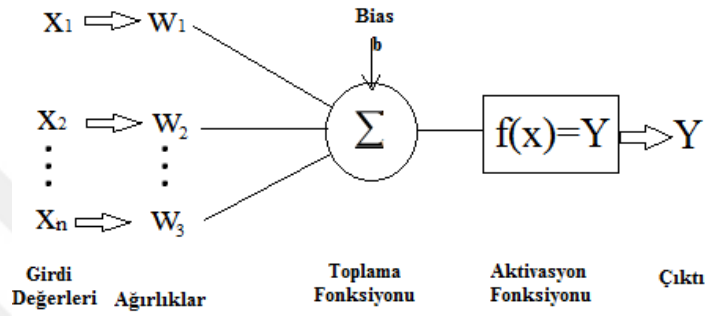
$$net = \sum_{i=1}^n x_i w_i \quad (i = 1, 2, \dots, n) \quad (2.121)$$

Burada  $n$  girdi (proses elemanı) sayısını,  $w$  ağırlık değerlerini ve  $x$  ise girdi değerlerini göstermektedir. Her proses elemanı bağımsız olarak farklı bir toplama fonksiyonuna sahip olabileceği gibi aynı toplama fonksiyonuna da sahip olabilirler.

4. *Aktivasyon Fonksiyonu*: Hücreye gelen net girdiye göre bir çıktı üreten fonksiyondur. Toplama fonksiyonu ile aktivasyon fonksiyonu çıktıyı hesaplamak için değişik fonksiyonlar kullanması ve proses elemanlarının farklı fonksiyonlar kullanabilmesi yönünden benzerdir. Çok Katmanlı Algılayıcı modelde en çok kullanılan fonksiyon sigmoid fonksiyonudur.

$$\text{sigmoid fonksiyonu: } f(\text{net}) = 1/(1 + e^{-\text{net}}) \quad (2.122)$$

Burada *net*, proses elemanına gelen net girdi değerini göstermektedir [118].



Şekil 2.42. Yapay sinir hücresinin yapısı.

Yukarıda bir yapay sinir hücresinin yapısı gösterilmiştir. Burada ilk olarak ağırlıklandırılmış girdi değerleri eşik değeri ile toplanır ve bu toplamın sonucu aktivasyon fonksiyonundan geçirilerek ağın çıktısı hesaplanır.

#### 2.6.2.2. Yapay Sinir Ağlarında Öğrenme

Yapay sinir ağlarında ağın uyarlanabilir olması iyi bir sonuç için önemlidir. Bunu sağlayabilmek için doğru bağlantılar ve uygun değerli ağırlıklar gerekir. Ağda uygun ağırlıklar ve bağlantılar elde edilirken ağ sistemin davranışlarını öğrenmeli veya kendi kendini örgütlemelidir. Yapay sinir ağlarında bilgi tüm ağda saklandığı için bir düğümün sahip olduğu ağırlık değerinin tek başına bir anlamı yoktur ağdaki tüm ağırlıkların uygun değerler alması gerekmektedir. Yapay sinir ağlarında başlangıçta ağırlık değerleri rastgele belirlenir. Zamanla ağdaki işlemci elemanlar bazı kurallar çerçevesinde en uygun ağırlık değerlerini bulurlar. İşte “ağın eğitilmesi” buna denir. Bir ağın eğitilebilir olabilmesi için ağırlık değerlerinin belirli bir kural/kurallar çerçevesinde dinamik olarak değiştirilebilmesi gerekmektedir. Yapay sinir ağlarında genel olarak öğrenme olayı iki aşamada gerçekleşir. Birinci aşama ağırlık değerlerinin rasgele belirlenerek, ağa gösterilmesi ve örnek için ağın üreteceği çıkışın bulunduğu aşamadır. Bu çıkış değerinin doğruluğuna göre ikinci aşama da ağa farklı örnekler

gösterilerek ağırlık değerleri değiştirilir. Buradaki hedef bu örnekler için doğru çıkışı elde etmeye yarayacak ağırlık değerlerini bulmaktır. Ağın doğru çıktıyı üretecek ağırlık değerlerini bulması, ağın örneklerin temsil ettiği olaylar hakkında sınıflandırma yapabilme yeteneğine sahip olduğunu gösterir. Ağın sınıflandırma yapabilmesi olayına ise “ağın öğrenmesi” denir. Temelde öğrenme yöntemleri, gözetimli, gözetimsiz ve destekleyici olmak üzere üç grupta toplanmıştır [119].

#### *Gözetimli (Öğreticili) Öğrenme Yöntemi*

Öğrenen sistemin öğrenebilmesi için bir öğreticiye ihtiyaç vardır. Öğretici öğrenilmesi istenen olay ile ilgili örneklere ait girdi değerlerini ve bu girdilere karşılık gelen çıktı değerlerini verir. Sistemin görevi girdilere göre çıktıları haritalamak ve bu sayede girdiler ile çıktılar arasındaki ilişkileri tanımlamaktır [117].

#### *Gözetimsiz (Öğreticisiz) Öğrenme Yöntemi*

Gözetimsiz öğrenme yönteminde adında da anlaşılacağı gibi ağın öğrenmesi için herhangi bir öğretici yoktur. Sisteme yalnızca girdi değerleri verilerek aralarındaki ilişkiyi sistemin kendi kendisine öğrenmesi istenir. Sistem kendini eğittikten sonra çıktıların ne anlam ifade ettiği kullanıcı tarafından belirtilmelidir. Genellikle sınıflandırma amacıyla kullanılan bu yöntemlerden en çok kullanılanları “Kohonen Ağları” ve “ART Ağları”dır [117].

#### *Destekleyici Öğrenme (Reinforcement learning)*

Bu yöntem sisteme bir öğreticinin yardımcı olması açısından gözetimli öğrenmeyle benzerlik gösterirken ağ çıkışından alınan değerlerin karşılaştırma amacıyla kullanılmaması yönünden gözetimli öğrenmeden ayrılır. Bu yöntemde üretilen çıktı doğru ya da yanlış şeklinde ağa bir sinyal ile geri döndürülür. Sistem de, öğreticiden gelen bu sinyale göre öğrenme işlemini gerçekleştirir. “LVQ Ağı” destekleyici öğrenmede kullanılan bir ağıdır [117].

#### *2.6.2.3. Yapay Sinir Ağının Yapısı*

Yapay sinir ağları birbirine bağlı pek çok yapay sinir hücrelerinin birbirine bağlanmasıyla meydana gelir. Yapay sinir ağları giriş katmanı, ara (gizli) katmanlar ve çıkış katmanı olmak üzere üç ana katmanda incelenir;

**Giriş Katmanı:** Giriş katmanı sisteme girdi verilerinin iletiği kısımdır. Bu katmanda girdiler üzerinde herhangi bir işlem yapılmaz.

Ara Katmanlar: Girdi elemanları ve belirli fonksiyonlar yardımıyla istenen çıktının üretilmesine yönelik işlemlerin gerçekleştirildiği katmandır. Kullanılan ağıın yapısına göre ara katmanda kullanılan fonksiyonlar değişiklik gösterebilir. Ara katmanın tasarlanırken tek bir katman kullanılabileceği gibi birden fazla katman da kullanılabilir.

Çıkış Katmanı: Çıkış katmanı ağıın yapısında bulunan fonksiyonlara göre sistemin ürettiği çıkışın dış dünyaya sunulduğu katmandır ve en az bir çıktıdan oluşur.

#### 2.6.2.4. Yapılarına Göre Yapay Sinir Ağları

Yapay sinir ağlarının yapısı nöronlar arasındaki bağlantılar tarafından belirlenir. Bağlantıların nasıl olacağı ise öğrenme algoritmalarına göre belirlenir. Yine öğrenme algoritmasına göre ağırlıkların değeri belirlenir. Genelde yapay sinir ağları yapılarına göre ileri beslemeli ve geri beslemeli olarak iki ana grupta incelenmektedir. İleri beslemeli ağlarda, hücreler katmanlar şeklinde düzenlenir ve bir katmandaki hücrelerin çıkışları bir sonraki katmana giriş olarak sunulur. Giriş katmanı, dış ortamlardan aldığı bilgileri hiçbir değişikliğe uğratmadan gizli katmandaki hücrelere iletir. Bilgi, gizli katmanda ve çıkış katmanında işlenerek ağ çıkışı elde edilir. İleri beslemeli yapay sinir ağlarında en sık kullanılan öğrenme algoritması geriye yayılım öğrenme algoritmasıdır. Bu algoritma ağların eğitiminde etkin olarak kullanılmaktadır. Geri beslemeli sinir ağı, gizli katman ve çıkış katmandaki çıkışların, giriş birimlerine veya önceki ara katmanlara geri beslendiği bir ağ yapısıdır. Böylece, girişler hem ileri yönde, hem de geri yönde aktarılmış olur. Geriye beslemeli ağların dinamik hafızaları vardır ve bir andaki çıkış, hem o andaki hem de önceki girişleri yansıtır. Bundan dolayı, geriye beslemeli sinir ağları önceden tahmin uygulamaları için kullanılır [119].

#### 2.6.3. K-En Yakın Komşuluk (K-EK) Sınıflandırıcı

K-en yakın komşular kuralı örüntü sınıflamada Cover ve Hart tarafından ortaya koyulmuş, en eski ve en basit yöntemlerden birisidir [120]. K-EK kuralının performansı belirlenen en yakın komşuları tanımlarken kullanılan mesafe metriğine göre değişir. Çünkü etiketlenmemiş her bir örnek eğitim setindeki en yakın K komşusunun çoğunluk etiketi ile sınıflandırılır [120]. Eğitim verisinin geniş olması durumunda, bir tek en yakın komşuluk yerine en yakın K komşuların çoğunluk değerlerinin kullanılması daha iyi sonuç vermektedir. Burada K sayısının seçimi sonucun doğruluğu açısından önem arz etmektedir. K sayısı örneklere yeterince yakın olması için mümkün olduğunca küçük, yanlış sınıflandırma olasılığını en aza indirmek için yeterince büyük seçilmelidir

[9]. Önceden bir bilgi sahibi olunamayan durumlarda genellikle K-EK sınıflandırıcıda Öklid uzaklığını kullanılır [120].

$p_1(x_1, y_1)$  ve  $p_2(x_2, y_2)$  iki boyutlu uzayda iki nokta olmak üzere bu noktalar arasındaki Öklid uzaklığı  $d(p_1, p_2)$  aşağıdaki şekilde tanımlanır.

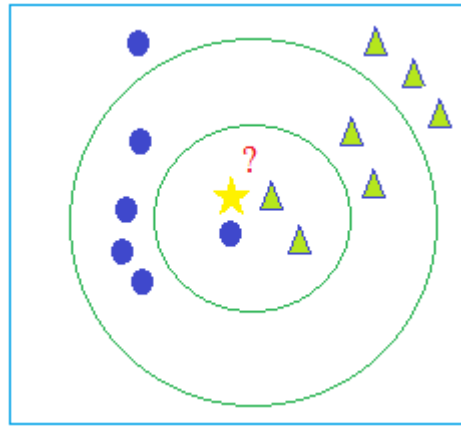
$$d(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2.123)$$

Eğer uzaklık iki boyuttan  $Z$  boyuta genişletilirse bu durumda Öklid uzaklığı aşağıdaki şekilde tanımlanır.

$$d(p, q) = \sqrt{\sum_{i=1}^Z (p_i - q_i)^2} \quad (2.124)$$

Burada  $p_i, q_i$   $i$ . boyuttaki koordinatlardır. K-EK ile sınıflandırma yapılırken ilk olarak  $K$  parametresi yani en yakın komşuların sayısı belirlenir. Daha sonra test verisi ile eğitim verisi arasındaki mesafe hesaplanır. Son olarak en yakın  $K$  komşunun sınıf çoğunluğu test örneğinin tahmini değeri olarak kullanılır.

K-EK sınıflandırmada  $K$  parametresinin seçimini bir örnekle açıklayacak olursak Şekil 2.42'de 2 farklı sınıf daire ve üçgenlerle temsil edilmiştir. Burada yıldız test verisini sınıflandırmak istediğimizde  $K=3$  seçilirse en yakın 3 komşuya bakılacaktır. İç çember en yakın 3 komşuluğu göstermek üzere bu alanda 2 üçgen ve bir daire yer aldığından ve en fazla üçgen bulunduğundan bu durumda yıldız test verisi üçgen sınıfına dahil edilecektir. Fakat  $K=9$  seçilirse dış çember içinde kalan alan en yakın 9 komşuyu kapsar. Bu alanda en fazla daire sınıfı bulunduğundan yıldız test verisi daire sınıfına dahil edilir.



Şekil 2.43. K-EK için  $K$  parametresinin seçimi.

#### 2.6.4. Sınıflandırma Performansının Değerlendirmesi

Bu bölümde, önerilen BDT sisteminin sınıflandırma sonuçlarını değerlendirmek amacıyla kullanılan performans değerlendirme yöntemleri sunulmuştur. Performans değerlendirmesi dört farklı yöntemle gerçekleştirilmiştir. Bunlar sınıflandırma başarı oranı, duyarlılık-özgüllük analizi, ROC eğrisi analizi ve k-kat çapraz doğrulamadır.

##### 2.6.4.1. Sınıflandırma Başarı Oranı

Çalışmada kullanılan test verilerinin DVM, YSA ve K-EK sınıflandırıcılar ile sınıflandırılması sonucu doğru olarak tespit ettikleri durumların (%) yüzde olarak değerleridir.

##### 2.6.4.2. Duyarlılık ve Özgüllük

Duyarlılık (Sensitivity): Duyarlılık, sağlık problemine sahip olan bireyler içinde testin sağlık problemlileri bulabilme özelliği olarak tanımlanabilir. Burada sağlık problemine sahip olan bireyler Gerçek Pozitif(GP) olarak adlandırılır.

Özgüllük (Specificity) : Özgüllük, sağlıklı olan bireyler içinde testin sağlıklı bireyleri bulabilme özelliği olarak tanımlanabilir. Burada herhangi bir sağlık problemi olmayan bireyler Gerçek Negatif olarak(GN) olarak adlandırılır.

Ayrıca bir birey hasta olmasına rağmen yapılan test sonucunda sağlıklı olarak bulunmuşsa Yanlış Negatif (YN); sağlıklı bir bireyin test sonucunda hasta olarak bulunması da Yanlış Pozitif (YP) olarak nitelendirilir.

Duyarlılık ve özgüllük analizi için, aşağıdaki ifadeleri kullanılmıştır.

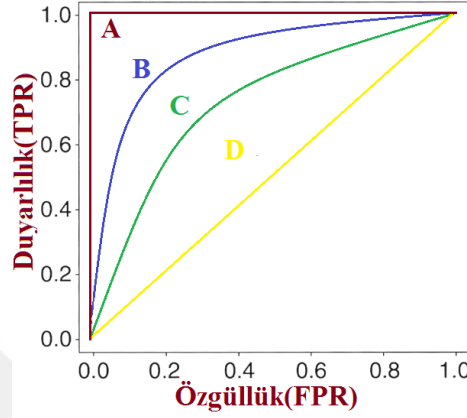
$$\text{Duyarlılık (TPR)} = GP / (GP + YN) \quad (2.125)$$

$$\text{Özgüllük (FPR)} = GN / (GN + YP) \quad (2.126)$$

##### 2.6.4.3. ROC(Receiver Operating Characteric) Analizi

ROC analizi tanı testlerinin performansını belirlemek ve testler arasında güvenilir bir karşılaştırma yapmak amacıyla sıklıkla kullanılan bir yöntemdir. Bir tanı testinin duyarlılığı ve özgüllüğü ile elde edilen grafiğe ROC eğrisi denir. Grafikteki her nokta farklı bir kesim noktası alınarak hesaplanan duyarlılık (TPR) ve özgüllük (FPR) değerlerinden oluşur [121]. ROC eğrisinin grafiksel gösterimi ölçümlerin duyarlılığı ve özgüllüğü arasındaki bağlantıları anlamayı kolaylaştırır. ROC eğrisinin oluşturulacağı koordinat sisteminde, Y ekseninde tanı testinin gerçek pozitif değeri (duyarlılık), X

ekseninde ise yanlış pozitif değeri (özgüllük) yer alır. Her kesim noktasındaki doğru pozitif ve yanlış pozitive karşılık gelen noktalar birleştirilerek ROC eğrisi çizilir. İdeal ve kötü performans gösteren testlere ilişkin ROC eğrileri Şekil 2.43 de verilmiştir. Şekilde A bölgesi mükemmel test sonucu için eğriyi göstermektedir. B eğrisi iyi bir geçerliliğe ve C eğrisi orta dereceye sahiptir. Grafikte D eğrisi kötü performansı simgelemektedir.



Şekil 2.44. Performanslarına göre ROC eğrileri.

#### 2.6.4.4. *k*-Kat Çapraz Doğrulama

Pek çok araştırmacı tarafından test sonuçlarının daha güvenilir ve belirleyici olabilmesi için verilerin sınıflandırılmasında *k* kat çapraz doğrulama yaygın olarak kullanılmaktadır. Bu yöntem ile eğitim verilerinin rastgele örnekleme mümkün olduğunca aza indirgenir. Bunun için, tüm veri seti *k* sayısı kadar yaklaşık olarak eşit büyüklükteki alt kümelerle bölünür. Her bir alt kümeden bir tanesi test verisi olarak alınırken, geriye kalan diğer katlar eğitim verisini oluşturmak için kullanılır. Sonuç itibarıyla her eğitim-test yapılandırması için *k* tane farklı test sonucu bulunmuş olur. Algoritmanın test doğruluğu, bu sonuçların ortalaması alınarak elde edilir. Örneğin, *k*=5 için *k*-kez çapraz doğrulama yöntemi bir veri kümesine uygulanmış olsun. Bu veri kümesi yaklaşık olarak 5 eşit parçaya bölünecek ve bu beş parçadan bir parçası test için geriye kalan 4 tanesi eğitim verisi için kullanılacaktır. Bu işlem her bir parça test verisi olacak şekilde 5 kez gerçekleştirilecek ve işlemler sonucunda edilen sonuçlar toplanıp beşe bölünecek ve ortalama sınıflama doğruluğu elde edilmiş olacaktır. Bu yöntem ile belirlenen test ve eğitim verilerinin gösterimi şekil 2.45' de sunulmuştur.

	1. Kat	2. Kat	3. Kat	4. Kat	5. Kat
Veri Seti	Test	Eđitim	Eđitim	Eđitim	Eđitim
	Eđitim	Test	Eđitim	Eđitim	Eđitim
	Eđitim	Eđitim	Test	Eđitim	Eđitim
	Eđitim	Eđitim	Eđitim	Test	Eđitim
	Eđitim	Eđitim	Eđitim	Eđitim	Test

Şekil 2.45. Veri setinin 5 kat çapraz doğrulama için test ve eğitim verilerine ayrılması.

Bir sonraki bölümde tez çalışması kapsamında geliştirilen BDT sistemi ile yapılan deneysel çalışmalar ve elde edilen sonuçlar sunulmuştur.



### 3. DENEYSEL ÇALIŞMALAR

Bu tez çalışmasında meme kanserinin tespiti amacıyla geliştirilen BDT sisteminin test edilmesi için sadece literatürde pek çok araştırmacı tarafından yaygın olarak kullanılan MIAS [40] veri tabanı değil aynı zamanda bu veri tabanından daha farklı özelliklere sahip mamogramlar içeren INBREAST [41] veri tabanı da kullanılmıştır. Bu sayede yapılan çalışmanın farklı veri tabanlarındaki başarısı da gösterilmek istenmiştir. MIAS ve INBREAST veri tabanı ile ilgili ayrıntılı bilgiler Materyal ve Yöntemler bölümünde sunulduğundan bu tez çalışmasında kullanılan mamogramlara ait özellikler Çizelge 3.1 ve Çizelge 3.2 'de verilmiştir.

Çizelge 3.1'de kullanılan kısaltmalar aşağıdaki bilgileri ifade etmektedir.

CIRC: İyi tanımlanmış / sınırlandırılmış kitleler

SPIC: Spiküle kitleler

MISC: Diğer, kötü tanımlanmış kitleler

ARCH: Mimari bozulma

ASYM: Asimetri

Çizelge 3.1. MIAS veri tabanından alınan mamogramlara ait özellikler.

MIAS Veri Tabanı	İyi Huylu					Kötü Huylu					Normal
	Circ	Spic	Misc	Arch	Asym	Circ	Spic	Misc	Arch	Asym	
<b>Yağlı</b>	9	0	2	4	0	1	3	5	0	3	16
<b>Bezel</b>	5	2	3	0	2	0	2	2	2	2	33
<b>Yoğun</b>	0	5	1	2	2	0	2	1	0	4	21
Ara Toplam	<b>14</b>	<b>7</b>	<b>6</b>	<b>6</b>	<b>5</b>	<b>1</b>	<b>7</b>	<b>8</b>	<b>2</b>	<b>9</b>	<b>70</b>
Toplam	<b>38</b>					<b>32</b>					<b>70</b>

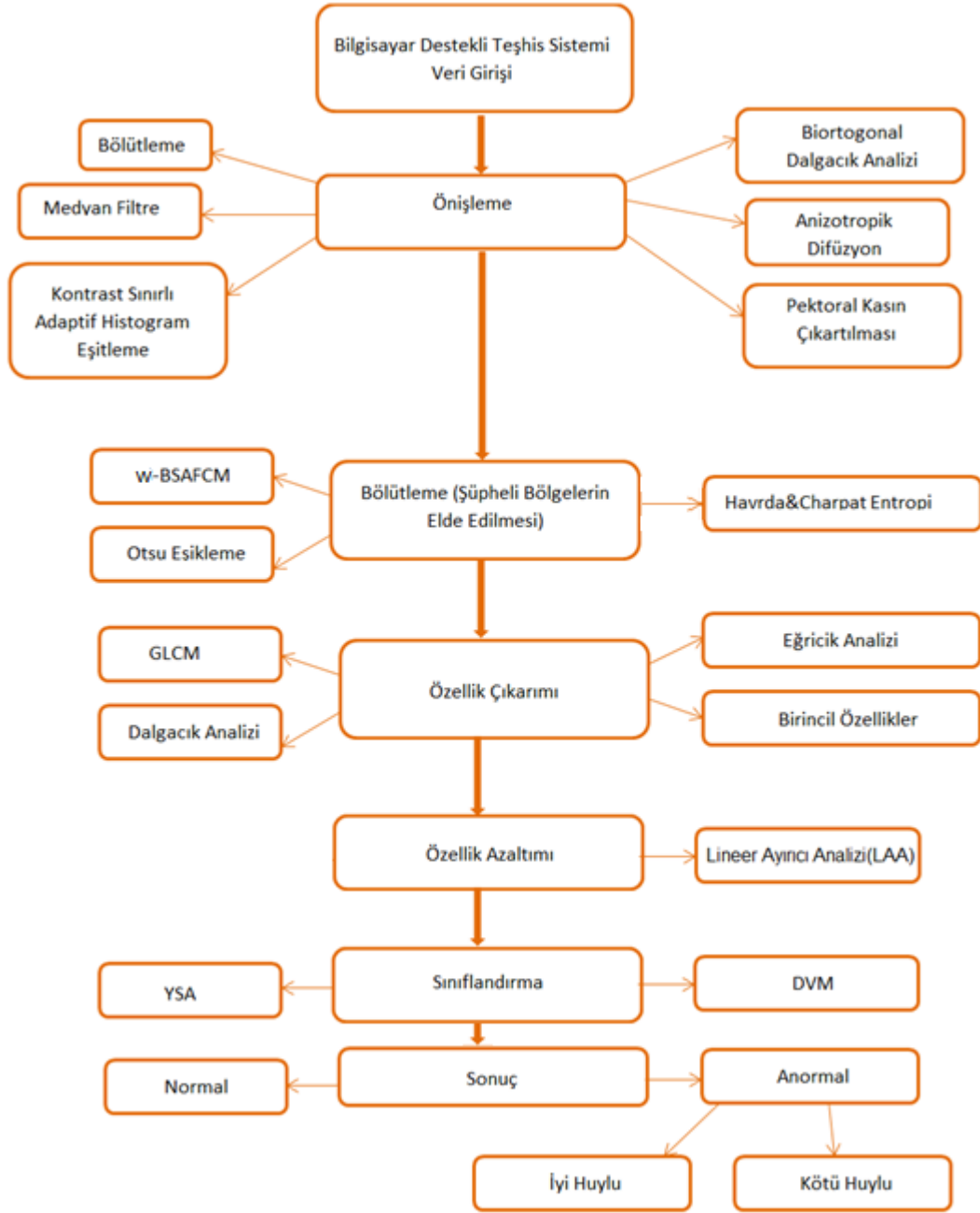
Çizelge 3.1' den de görüldüğü üzere MIAS veri tabanından 38 iyi huylu, 32 kötü huylu ve 70 normal olmak üzere toplamda 140 adet mamogram kullanılmıştır. Benzer şekilde INBREAST veri tabanından alınan mamogramlara ait özellikler Çizelge 3.2'de verilmiştir.

Çizelge 3.2. INBREAST veri tabanından alınan mamogramlara ait özellikler.

Doku Sınıflandırma	Yağlı	Bezel	Yoğun	Toplam
İyi Huylu	7	3	10	20
Kötü Huylu	9	22	2	33
Normal	21	21	11	53

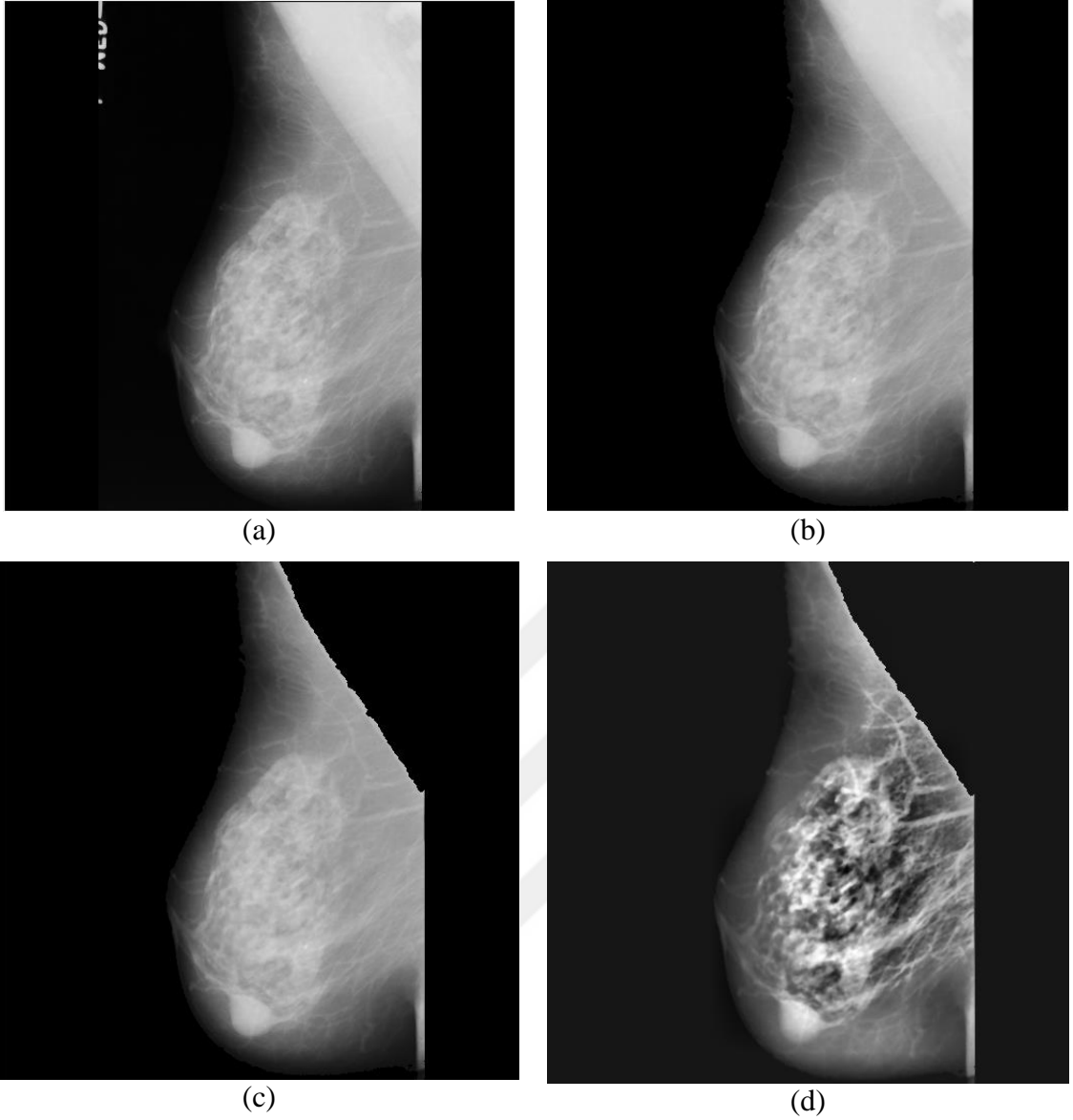
Çizelge 3.2'den de görüldüğü gibi INBREAST veri tabanından 20 iyi huylu, 33 kötü huylu ve 53 normal olmak üzere toplamda 106 adet mamogram kullanılmıştır. Ayrıca her iki veri tabanından alınan mamogramlara ait dokuların yapısı da (yağlı-bezel-yoğun) Çizelge 3.1 ve 3.2'den görülebilmektedir. Bu tez çalışmasında toplam 246 adet mamogram sınıflandırılmıştır.

Bu bölümde seçilen örnek mamogramlar kullanılarak geliştirilen BDT sistemi test edilmiştir. Bunun için ilk olarak geliştirilen BDT sistemi özet olarak bir şema yoluyla Şekil 3,1'de verilmiştir. Şekilde de görüldüğü üzere BDT sistemi sırasıyla; önışleme, bölütleme, öznitelik çıkarımı, öznitelik azaltımı ve sınıflandırma olmak üzere 5 bölümden oluşmaktadır.



Şekil 3.1. Geliştirilen BDT sistemi ve kullanılan yöntemler.

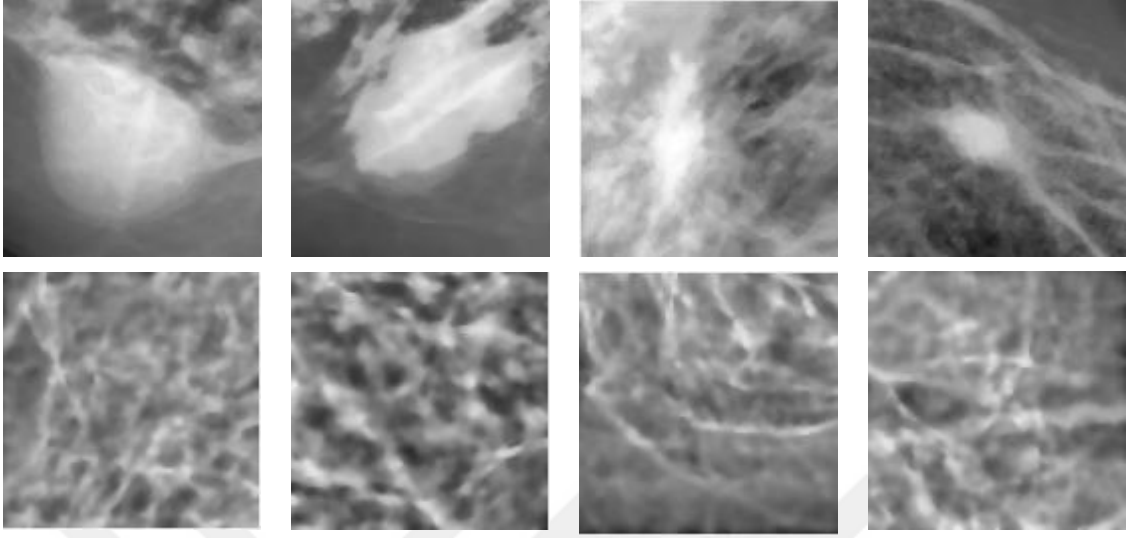
Çalışmanın ön işleme aşamasında veri bankasından alınan mamogramlar gürültü giderme, görüntünün iyileştirilmesi ve istenmeyen bölgelerin çıkartılması amacıyla sırasıyla bölütleme, medyan filtre, pektoral kasın çıkartılması, kontrast sınırlı adaptif histogram eşitleme(CLAHE), biortogonal dalgacık analizi ve anizotropik difüzyon yöntemi işlemlerine tabii tutulmuştur. Bu aşamada gerçekleştirilen işlemler materyal ve yöntemler bölümünde anlatıldığından burada ayrıntılara yer verilmemiştir. Yapılan ön işlemlere dair örnek görüntüler Şekil 3.2’de verilmiştir.



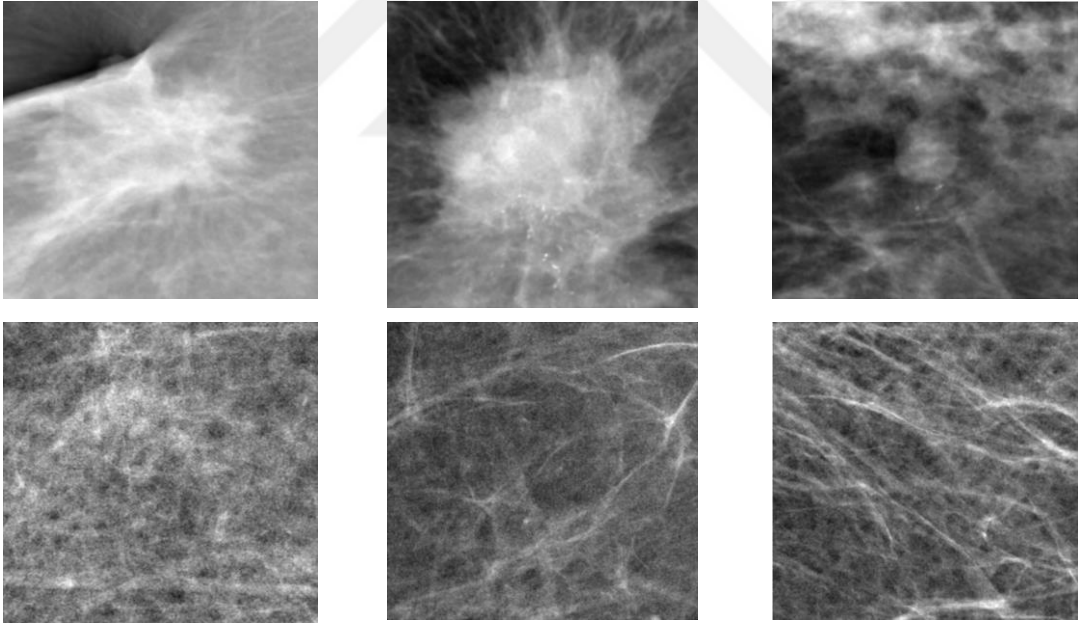
Şekil 3.2. Önışlem aşamasına ait mamogram görüntüleri (a) Orijinal mamogram (b) meme bölgesi bölütlenmiş ve medyan filtre uygulanmış mamogram (c) Pektoral kas çıkartılmış mamogram (d) Biortogonal analiz, anizotropik difüzyon ve kontrast sınırlı adaptif histogram eşitleme işlemleri uygulanmış mamogram.

Şekil 3.2’de de görüldüğü üzere mamogram üzerindeki artefektler kaldırılmış, gürültü giderilmiş, mamogram görüntüsü iyileştirilmiş ve pektoral kas bölgesi başarıyla çıkartılmıştır.Şüpheli bölgelere ait görüntülerin tam otomatik bir sistem içinde elde edilmesi amacıyla pek çok farklı deneysel çalışma yapılmıştır. Geliştirilen sistem için en iyi sonuç Otsu Eşikleme, Havrda & Charvat Entropi ve bu tez kapsamında geliştirilen w-BSAFCM algoritmasının birlikte kullanılmasıyla elde edilmiştir. Bu yöntemlerin birlikte kullanımıyla belirlenen alanlar MIAS veri tabanı için 128x128 boyutlarında ve

INBREAST veri tabanı için 500x500 boyutlarında kırılmıştır. Kırılan bazı örnek görüntüler Şekil 3.3 ve Şekil 3.4’de sunulmuştur.



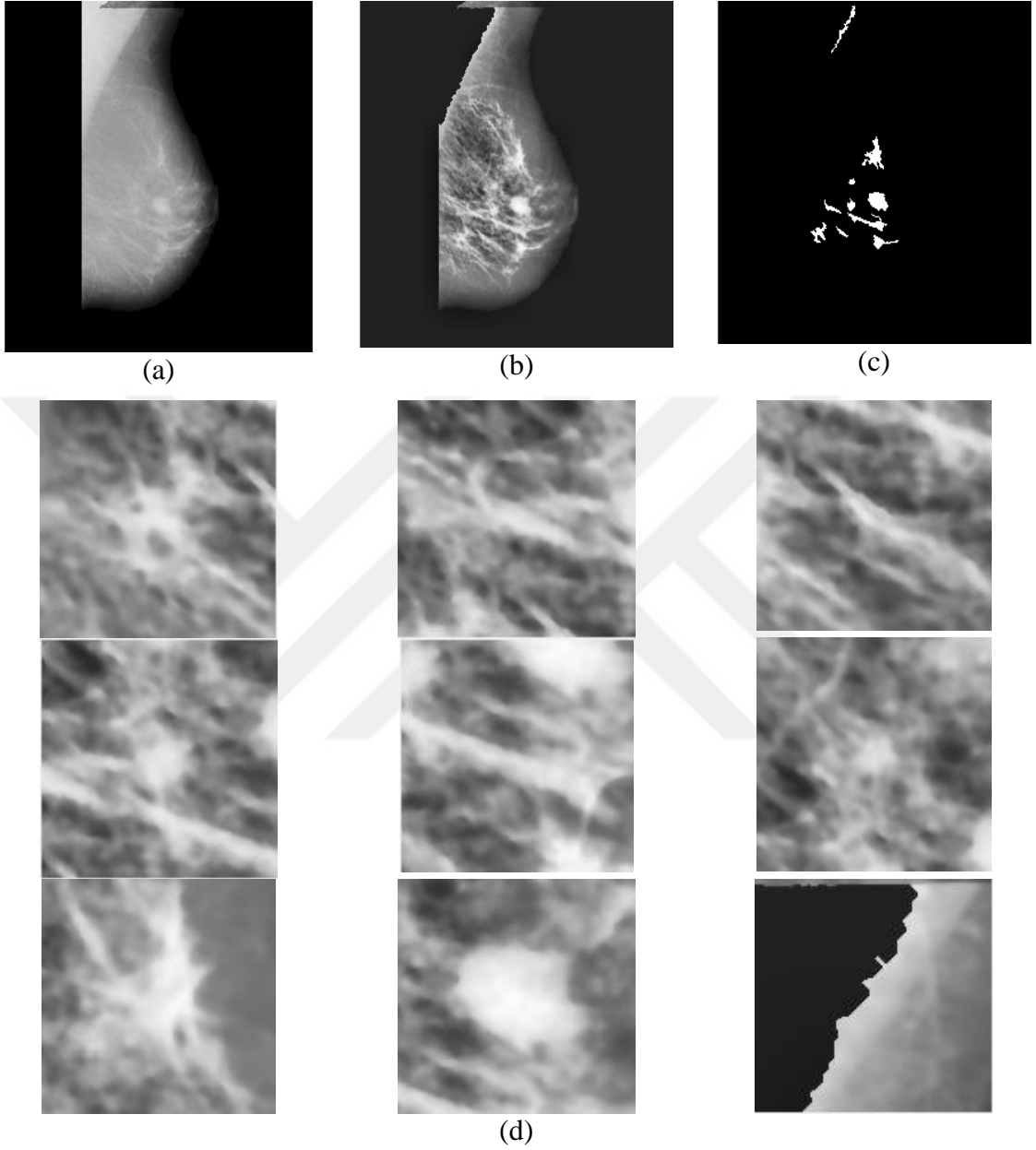
Şekil 3.3. MIAS veri tabanından alınan 128x128 kırılmış mamogram görüntüleri anormal(üstte) ve normal(altta).



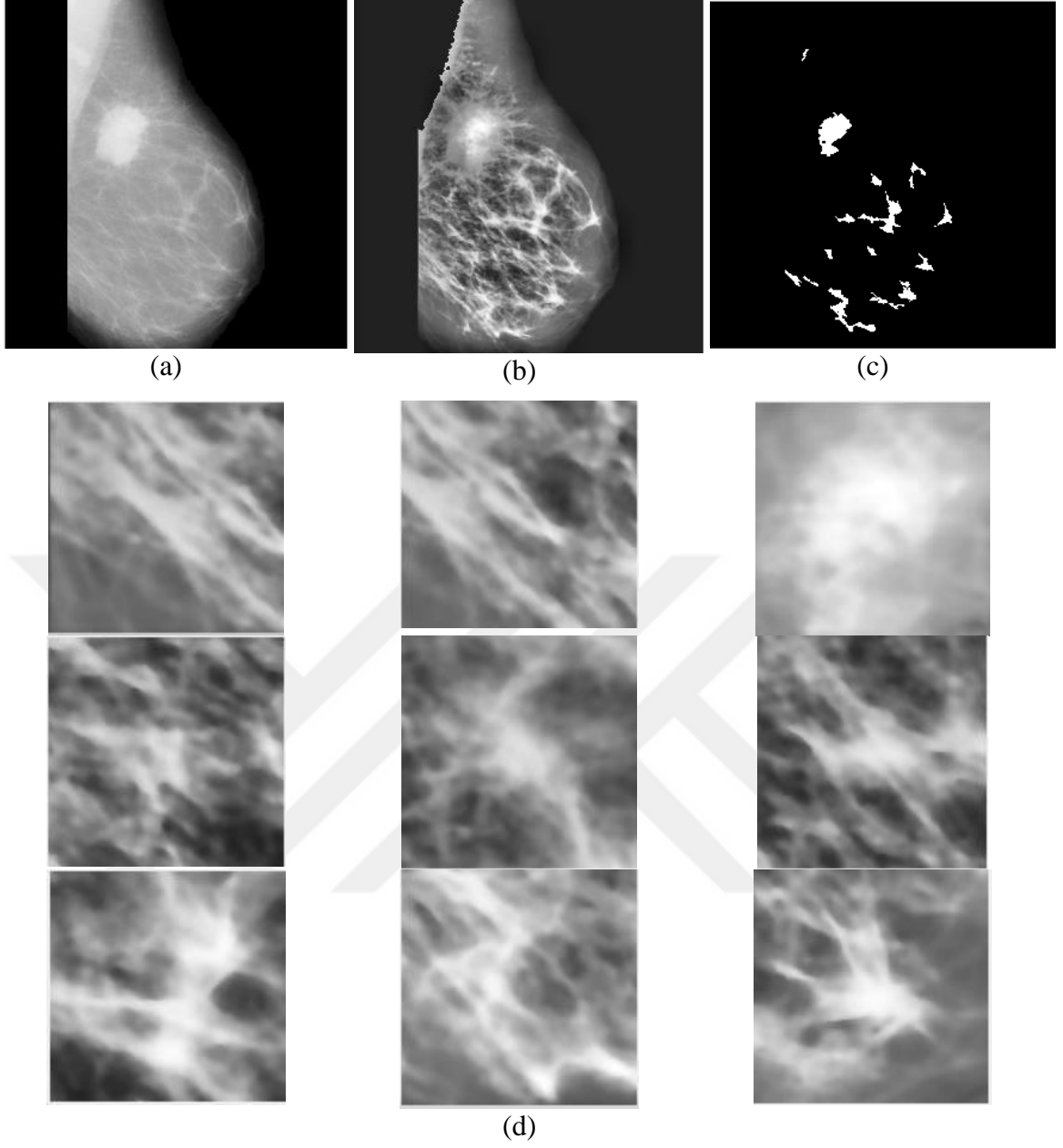
Şekil 3.4. INBREAST veri tabanından alınan 500x500 kırılmış mamogram görüntüleri anormal(üstte) ve normal(altta).

Elde edilen şüpheli bölgelerin kırılması işlemi için boyut belirlenirken kullanılan veri tabanlarındaki mamogramların özellikleri ve var olan anormalliklerin büyüklükleri göz önünde bulundurulmuştur. Ayrıca normal mamogramlar için ROI’ler meme bölgesi dâhilinde rastgele olmak üzere MIAS veri tabanı için 128x128, INBREAST veri tabanı

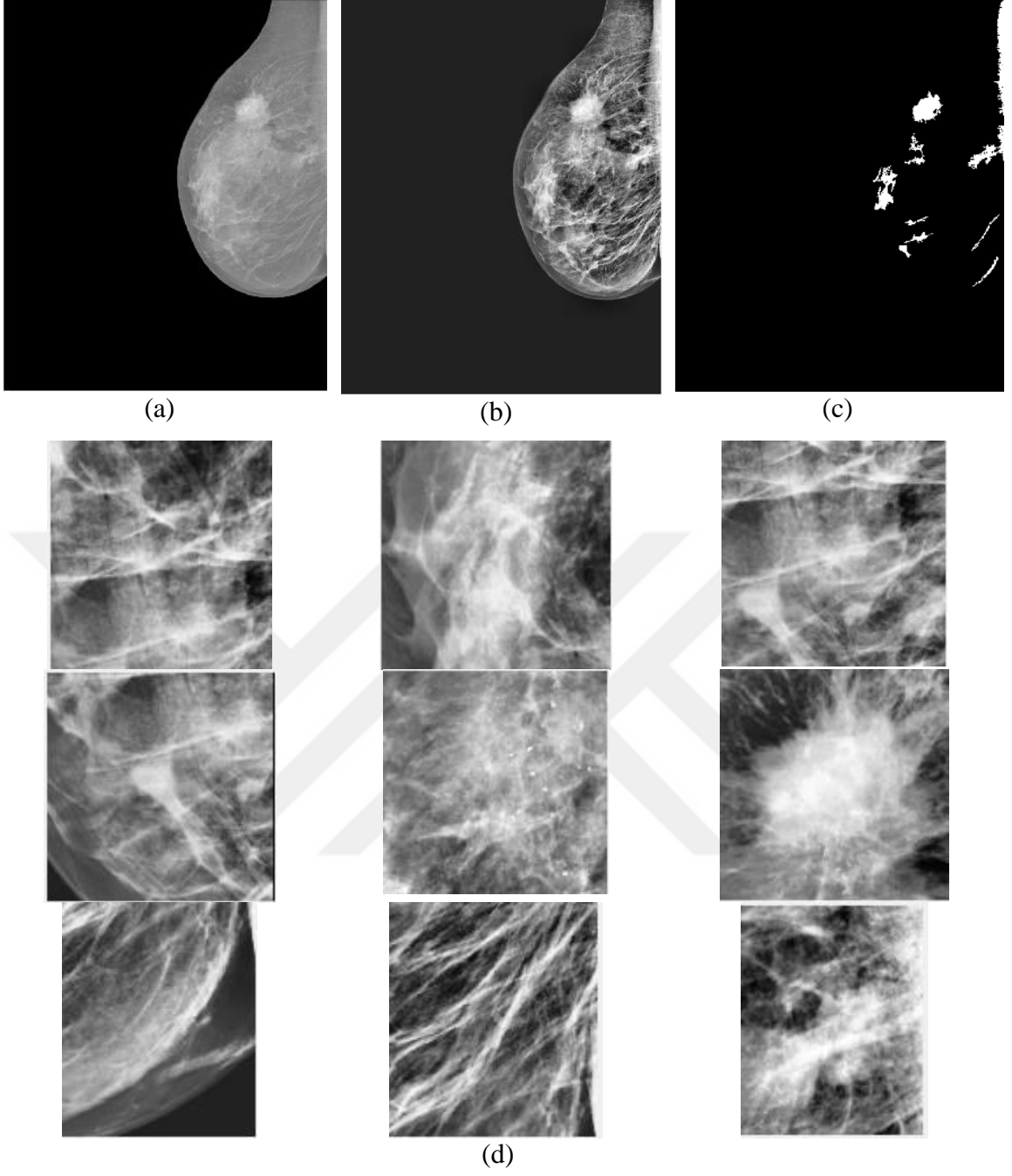
için 500x500 boyutunda kırılarak elde edilmiştir. Önerilen sistem hem normal hem de anormal mamogramlar için bu kırılmış görüntüler kullanılarak eğitilmiştir. Şüpheli bölgeleri elde edilmiş mamogram örnekleri şekil 3.5-3-8’de sunulmuştur.



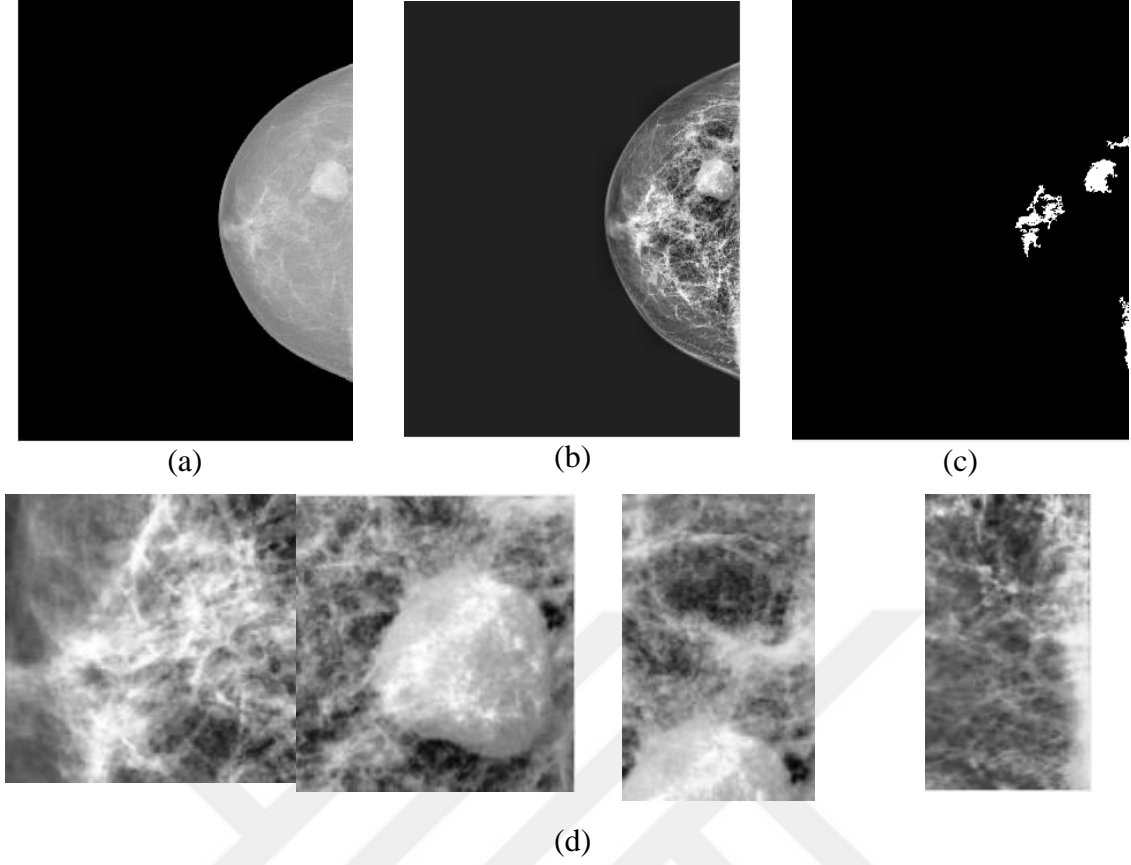
Şekil 3.5. MIAS veri tabanından Mdb010 isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önışlemeden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 128x128 elde edilmiş yeni görüntüler.



Şekil 3.6. MIAS veri tabanından Mdb184 isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önışlemeden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 128x128 elde edilmiş yeni görüntüler.



Şekil 3.7. Inbreast veri tabanından 22613822\_45c7f44839fd9e68\_MG\_R\_ML\_ANON isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önişlemeden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 500x500 elde edilmiş yeni görüntüler.



Şekil 3.8. Inbreast veri tabanından Şekil 3.8 Inbreast veri tabanından 50996406\_6aba0b402889a16f\_MG\_R\_CC\_ANON isimli mamogram için şüpheli bölgelerin bulunması (a) Orijinal mamogram (b) Önışlemeden geçirilmiş mamogram (c) Şüpheli bölgeleri tespit edilmiş mamogram (d) Şüpheli her bir bölge için 500x500 elde edilmiş yeni görüntüler.

Şüpheli bölgeler elde edildikten sonra öznitelik çıkarımı işlemi gerçekleştirilmektedir. Öznitelik çıkarımı için birincil öznitelikler, gri seviye eş-oluşum matrisi, dalgacık dönüşümü ve eğricik dönüşümü kullanılmıştır. Bu özniteliklerden birincil öznitelikler ilgili alanlarının konumu, büyüklüğü gibi bilgilerin elde edilmesi amacıyla kullanılmıştır. Diğer öznitelikler olan GLCM, dalgacık ve eğricik yöntemleri için sırasıyla 17, 21 ve 162 adet öznitelik elde edilmiş ve mamogramların normal-anormal, iyi huylu-kötü huylu sınıflandırılmasında kullanılmışlardır. Çıkarılan özniteliklerin boyutunu azaltmak amacıyla Lineer Ayırıcı Analiz kullanılmıştır. Son olarak seçili özniteliklere sınıflandırma işlemi uygulanarak, normal-anormal ve iyi huylu-kötü huylu ayrıştırması gerçekleştirilmiştir. Sınıflandırıcı olarak Yapay Sinir Ağları, Destek Vektör Makineleri ve K-En Yakın Komşuluk sınıflandırıcı olmak üzere üç farklı sınıflandırıcı kullanılmıştır. YSA için İleri beslemeli geri yayılım ağı tasarlanmış ve öğrenme

algoritması olarak Levenberg-Marquardt kullanılmıştır. Kullanılan gizli katman sayısı 10 olarak belirlenmiştir. Destek Vektör Makinesi için kullanılan çekirdek fonksiyonu Radyal Tabanlı Fonksiyon (RBF) ve RBF Sigma değeri de 0.5 ile 10 arasında 0.5 aralıklarla değişen değerler arasından en yüksek DVM sonucunu üreten RBF sigma değeri sonuç değerlendirilmesinde kullanılmıştır. K-EK sınıflandırıcı için kullanılan k değeri ise bu tez çalışmasında 3 olarak alınmıştır. Elde edilen her grup öznitelik ayrı ayrı sınıflandırıcılara sunularak performansları ölçülmüştür. Öznitelikler sınıflandırılırken özniteliklerin tamamı veya özniteliklerin LAA ile boyut azaltılmış formu sınıflandırıcılara sunulmuştur. Ayrıca kullanılan öznitelikler 5 kat çapraz doğrulama kullanılarak eğitim ve test verileri elde edilmiş ve sınıflandırma işlemi gerçekleştirilmiştir. Sonuçlar bu 5 katın ortalaması alınarak elde edilmiştir. Her 3 öznitelik grubu için LAA ile boyut indirgenmiş ve LAA ile boyut indirgenmemiş verilerin üç sınıflandırıcı ile elde edilen performans verileri Çizelge 3.3-3.10'da sunulmuştur.

Çizelge 3.3. MIAS veri tabanı 140 (normal/anormal) mamogram için sınıflama başarısı (%)

<b>GLCM</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	100	100	100	100	100	<b>100</b>
<b>Dalgacık Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	92,85	92,85	100	96,43	<b>96,43</b>
<b>DVM</b>	96,43	96,43	96,43	96,43	96,43	<b>96,43</b>
<b>K-EK</b>	96,43	96,43	96,43	96,43	96,43	<b>96,43</b>
<b>Eğricik Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	100	100	100	100	100	<b>100</b>

Çizelge 3.3 incelendiğinde MIAS veri tabanından alınan 140 mamogram için normal-anormal sınıflandırmasında her üç sınıflandırıcı için % 100 başarının hem GLCM 'den hem de eğricik dönüşümünden elde edilen özniteliklerden elde edildiği görülmektedir. Çizelgede en kötü sonuçların % 96,43 ile dalgacık dönüşümünden çıkartılan özniteliklerden elde edildiği görülmektedir.

Çizelge 3.4. Lineer ayırıcı analiz kullanılarak MIAS veri tabanı 140 (normal/anormal) mamogram için sınıflama başarısı (%).

<b>GLCM + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	96,43	100	100	96,43	100	<b>98,57</b>
<b>DVM</b>	96,43	96,43	96,43	96,43	96,43	96,43
<b>K-EK</b>	96,43	96,43	96,43	96,43	96,43	96,43
<b>Dalgacık Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	98,57	100	100	99,71
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	96,43	96,43	96,43	96,43	96,43	96,43
<b>Eğricik Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	100	100	100	100	100	<b>100</b>

Çizelge 3.4 incelendiğinde MIAS veri tabanından elde edilen 140 mamogram için çıkarılan öznitelikler LAA kullanılarak sınıflandırıldığında Normal-Anormal ayırımında YSA sınıflandırıcı ile GLCM için % 98.57, dalgacık dönüşümü için SVM sınıflandırıcı ile % 100 ve eğricik dönüşümü için ise her üç sınıflandırıcı için % 100 başarı elde edildiği görülmektedir.

MIAS veri tabanından alınan 70 anormal mamogram için iyi huylu-kötü huylu sınıflandırması sonuçları Çizelge 3.5’de verilmiştir.

Çizelge 3.5. MIAS veri tabanı 70 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).

<b>GLCM</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	57,14	78,57	57,14	50	42,86	<b>57,14</b>
<b>DVM</b>	50	50	50	50	50	50
<b>K-EK</b>	50	50	50	50	50	50
<b>Dalgacık Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	50	57,14	64,29	64,29	64,29	<b>60,02</b>
<b>DVM</b>	57,14	57,14	57,14	57,14	57,14	57,14
<b>K-EK</b>	57,14	57,14	57,14	57,14	57,14	57,14
<b>Eğricik Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	57,14	57,14	71,42	64,28	50	<b>61,43</b>
<b>DVM</b>	42,86	55,55	50	50	42,86	45,71
<b>K-EK</b>	57,14	57,14	57,14	42,86	42,86	51,43

Çizelge 3.5 incelendiğinde iyi huylu-kötü huylu sınıflandırmada bütün öznelilikler için normal-anormal sınıflandırmaya göre çok daha düşük başarı değerleri elde edildiği görülmektedir. İyi huylu-kötü huylu sınıflandırma için en iyi sonuçlar YSA sınıflandırıcı ile eğricik analizi için % 61.43, dalgacık dönüşümü için % 60.02 ve GLCM analizi için % 57.14 şeklinde elde edilmiştir. Aynı mamogramlar için LAA kullanıldığında elde edilen sonuçlar Çizelge 3.6’da verilmiştir.

Çizelge 3.6. Lineer ayırıcı analiz kullanılarak MIAS veri tabanı 70 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).

<b>GLCM + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	71,42	64,28	71,42	64,29	64,29	<b>67,14</b>
<b>DVM</b>	50,00	50	50	50	50	50
<b>K-EK</b>	50	50	50	50	50	50
<b>Dalgacık Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	78,57	64,29	57,14	78,57	50	<b>65,71</b>
<b>DVM</b>	42,85	50	57,14	50	42,86	48,57
<b>K-EK</b>	57,14	50	57,14	57,14	57,14	55,71
<b>Eğricik Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. Kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	100	100	100	100	100	<b>100</b>

Çizelge 3.6 incelendiğinde MIAS veri tabanından alınan 70 mamogram için iyi huylu-kötü huylu sınıflandırmasında en başarılı sonuçların yine YSA sınıflayıcısıyla elde edildiği görülmektedir. Çizelgede en kötü sonuçların % 65.71 ile dalgacık analizinden çıkartılan özneliliklerden elde edildiği, en iyi sonuçların ise her üç sınıflandırıcı için % 100 ile Eğricik Dönüşümünden elde edildiği görülmektedir.

Yapılan tez çalışmasının farklı veri tabanlarındaki başarısını test etmek amacıyla INBREAST veri tabanından alınan mamogramlar da kullanılmıştır. Bu veri tabanı için elde edilen sonuçlar Çizelge 3.7-3.10 arasında sunulmuştur. Çizelge 3.7 ve 3.8 106 (Normal-Anormal) mamogram için sınıflama başarısını gösterirken Çizelge 3.9 ve 3.10 53 iyi huylu-kötü huylu sınıflama başarısını göstermektedir.

Çizelge 3.7. INBREAST veri tabanı 106 (normal/anormal) mamogram için sınıflama başarısı (%).

<b>GLCM</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	80,95	80,95	76,19	90,47	86,36	<b>82,98</b>
<b>DVM</b>	85,71	80,95	76,19	90,47	72,73	81,21
<b>K-EK</b>	80,95	76,19	61,9	90,47	72,73	76,448
<b>Dalgacık Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	76,19	80,95	76,19	80,95	72,64	<b>77,38</b>
<b>DVM</b>	71,43	80,95	57,14	80,95	68,18	71,73
<b>K-EK</b>	61,9	80,95	47,61	71,43	63,64	65,106
<b>Eğricik Dönüşümü</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	90,47	85,71	80,95	90,47	81,82	<b>84,74</b>
<b>DVM</b>	76,19	66,67	76,19	85,71	81,82	77,316
<b>K-EK</b>	90,47	80,95	71,43	71,43	72,73	74,135

Çizelge 3.7 incelendiğinde INBREAST veri tabanından alınan 106 mamogram için normal-anormal sınıflandırmasında en düşük sonuçların dalgacık dönüşümü için % 77.38 ve en başarılı sonuçların eğricik dönüşümü için % 84.74 olarak elde edildiği görülmektedir.

Çizelge 3.8. Lineer ayırıcı analiz kullanılarak INBREAST veri tabanı 106  
(normal/anormal) mamogram için sınıflama başarısı (%).

<b>GLCM + LAA</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	95,24	90,48	90,48	95,24	90,91	<b>92,47</b>
<b>DVM</b>	66,67	59,09	59,09	59,09	59,09	60,606
<b>K-EK Sınıflama</b>	57,14	68,18	68,18	68,18	68,18	65,972
<b>Dalgacık Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	95,45	90,47	85,71	80,95	95,45	<b>90,58</b>
<b>DVM</b>	63,64	66,67	63,64	66,98	63,64	64,914
<b>K-EK Sınıflama</b>	68,18	68,18	68,18	68,18	68,18	68,18
<b>Eğricik Dönüşümü + LAA</b>						
Sınıflandırıcı	1.Kat	2. kat	3. Kat	4.Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>DVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK Sınıflama</b>	100	100	100	100	100	<b>100</b>

Çizelge 3.8 incelendiğinde INBREAST veri tabanında 106 normal-anormal sınıflandırmada eğricik analizi ve LAA yöntemleri birlikte kullanıldığında her üç sınıflandırıcı için % 100 başarı oranı elde edilmiştir. YSA sınıflandırıcı için GLCM + LAA yönteminde en iyi başarı oranı % 92.47 olarak bulunurken dalgacık analizi + LAA yöntemleri için en iyi başarı oranı % 90.58 olarak kaydedilmiştir.

Çizelge 3.9. INBREAST veri tabanı 53 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).

<b>GLCM</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	70	70	72,73	63,63	72,72	69,77
<b>SVM</b>	72,73	72,73	72,73	72,73	72,73	<b>72,73</b>
<b>K-EK</b>	45,45	45,45	54,55	54,55	45,45	50
<b>Dalgacık Dönüşümü</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	80	70	72,73	54,55	81,82	<b>71,82</b>
<b>SVM</b>	72,73	54,55	72,73	54,55	72,73	65,45
<b>K-EK</b>	54,55	54,55	54,55	54,55	54,55	54,55
<b>Eğricik Dönüşümü</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	70	80	81,82	63,63	72,73	<b>73,64</b>
<b>SVM</b>	70	70	81,82	63,63	54,55	68
<b>K-EK</b>	60	70	72,73	63,63	54,55	64,18

Benzer şekilde 53 iyi huylu-kötü huylu sınıflandırma için elde edilen sonuçlar incelendiğinde (Çizelge 3.9) SVM sınıflandırıcı için, GLCM’de % 72.73, YSA sınıflandırıcı için dalgacık analizinde % 71.82 ve eğricik analizinde % 73.64 başarı oranı elde edilmiştir.

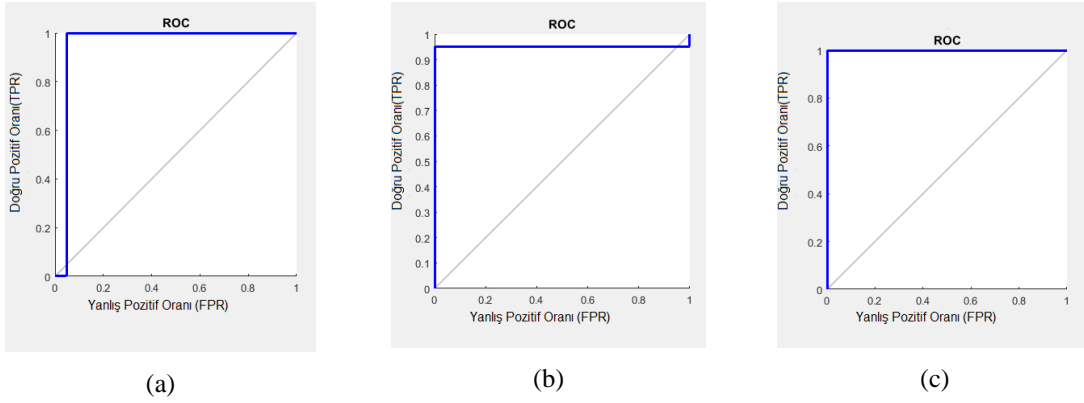
Çizelge 3.10. Lineer ayırıcı analiz kullanılarak INBREAST veri tabanı 53 anormal mamogram için iyi huylu/kötü huylu sınıflama başarısı (%).

<b>GLCM + LAA</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	80	80	100	81,82	90,91	<b>83,09</b>
<b>SVM</b>	72,73	72,73	72,73	72,73	72,73	72,73
<b>K-EK</b>	45,45	45,45	72,73	72,73	45,45	72,73
<b>Dalgacık Dönüşümü + LAA</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	80	90	72,73	81,82	90,91	<b>83,09</b>
<b>SVM</b>	72,73	72,73	72,73	72,73	72,73	72,73
<b>K-EK</b>	72,73	72,73	72,73	72,73	72,73	72,73
<b>Eğricik Dönüşümü + LAA</b>						
Sınıflandırıcı	1. Kat	2. Kat	3. Kat	4. Kat	5.Kat	Ortalama
<b>YSA</b>	100	100	100	100	100	<b>100</b>
<b>SVM</b>	100	100	100	100	100	<b>100</b>
<b>K-EK</b>	100	100	100	100	100	<b>100</b>

Çizelge 3.10 incelendiğinde YSA sınıflandırıcı için GLCM ve LAA ile Dalgacık analizi ve LAA'nın birlikte kullanılmasıyla elde edilen sınıflama başarı oranı % 83.09 olarak elde edilmiştir. Eğricik analizi ile LAA birlikte kullanıldığında ise her üç sınıflandırıcı için % 100 sınıflama başarısı elde edilmiştir.

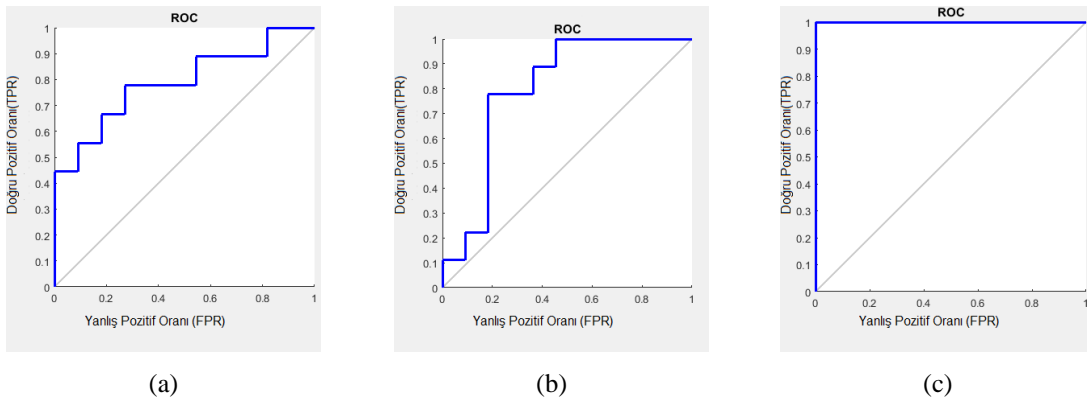
Bu çalışmada sınıflandırma başarısının yanında ayrıca ROC analizi de kullanılarak geliştirilen sistemin performansı değerlendirilmiştir. ROC analizi için sınıflandırma başarısız yüksek olan YSA ve üç ayrı öznel grup için boyut indirgenmiş veriler kullanılmıştır. Elde edilen sonuçlar Şekiller 3.9-3.12'de sunulmuştur.

Şekil 3.9’da MIAS veri seti 140 mamogram için normal-anormal sınıflandırmasında test verisi ROC analizi sonuçları sunulmuştur.



Şekil 3.9. MIAS 140 mamogram için normal-anormal sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA.

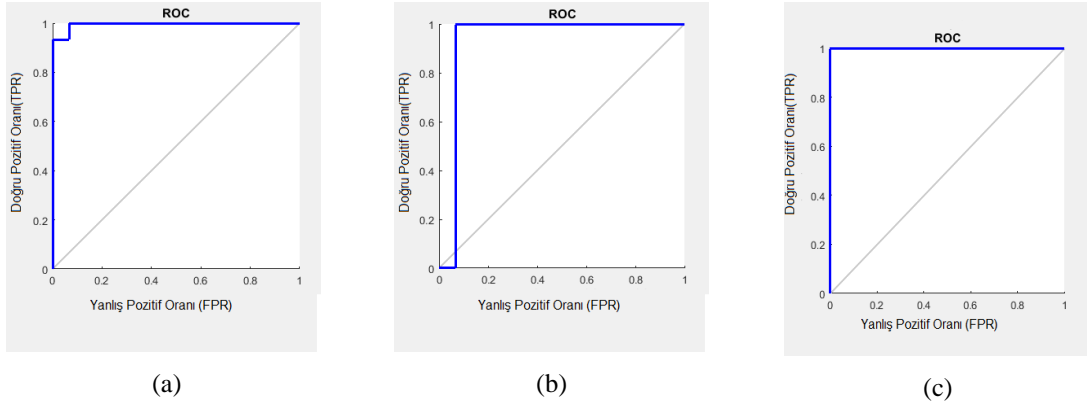
Şekil 3.10 incelendiğinde MIAS veri tabanı için her üç öznelik grubunda normal-anormal sınıflandırmada başarılı olduğu görülmektedir. En yüksek başarı oranının ise eğricik analizi ve LAA ile elde edildiği açıkça görülmektedir. MIAS veri seti 70 mamogram için iyi huylu-kötü huylu sınıflandırmasında test verisi ROC analizi sonuçları Şekil 3.7’de sunulmuştur.



Şekil 3.10. MIAS 70 mamogram için iyi huylu-kötü huylu sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA.

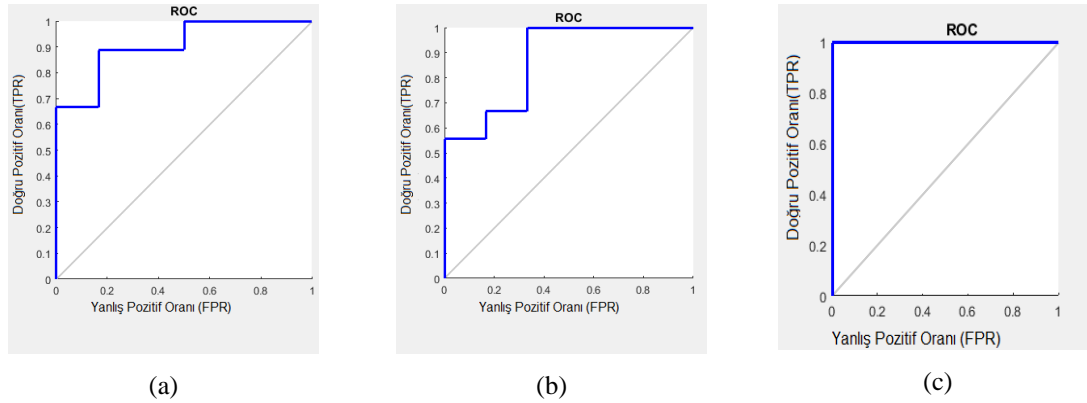
Şekil 3.10 incelendiğinde ise MIAS veri tabanı için iyi huylu-kötü huylu sınıflamasında GLCM ve dalgacık dönüşümünün kısmen başarılı, eğricik analizinin ise % 100 başarılı

olduğu görülmektedir. INBREAST veri seti 106 mamogram için normal-anormal sınıflandırmasında test verisi ROC analizi sonuçları Şekil 3.11’de sunulmuştur.



Şekil 3.11. INBREAST 106 mamogram için normal-anormal sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA.

Benzer şekilde INBREAST veri tabanı için normal-anormal ayırımında her üç öznitelik grubunda LAA ile birlikte kullanıldığında oldukça başarılı olduğu görülmekle birlikte eğricik dönüşümünün yine % 100 başarıya ulaştığı görülmektedir. INBREAST veri seti 53 mamogram için iyi huylu-kötü huylu sınıflandırmasında test verisi ROC analizi sonuçları Şekil 3.12’de sunulmuştur.



Şekil 3.12. INBREAST 53 mamogram için iyi huylu-kötü huylu sınıflandırması ROC analizi sonuçları (a) GLCM + LAA (b) Dalgacık dönüşümü + LAA (c) Eğricik dönüşümü + LAA.

INBREAST veri tabanı 53 iyi huylu-kötü huylu ayırımında GLCM + LAA yöntemi ile dalgacık analizi + LAA yöntemleri kısmen başarılıyken en yüksek başarı yine eğricik dönüşümü ve LAA yöntemlerinin birlikte kullanılmasıyla elde edilmiştir.

ROC analiziyle doğrudan ilgili olan duyarlılık ve özgüllük analizleri de Çizelge 3.11 de sunulmuştur. Çizelge 3.11’de de görüldüğü üzere MIAS veri tabanında iyi huylu-kötü huylu sınıflandırmasında GLCM analizi ve dalgacık dönüşümünün LAA ile birlikte kullanılması sonucunda özgüllük düşük çıkarken, INBREAST veri tabanında iyi huylu-kötü huylu sınıflandırmasında GLCM analizi ve dalgacık dönüşümünün LAA ile birlikte kullanılması sonucunda duyarlılık nispeten düşük çıkmıştır. En yüksek duyarlılık ve özgüllük değerlerinin eğricik analizi ve LAA yöntemlerinin birlikte kullanılmasıyla elde edildiği açıkça görülmektedir.

Çizelge 3.11. Kullanılan tüm mamogramlar için duyarlılık ve özgüllük analizi.

Duyarlılık- Özgüllük Analizi	Duyarlılık			Özgüllük		
	GLCM + LAA	Dalgacık Analizi +LAA	Eğricik Analizi +LAA	GLCM + LAA	Dalgacık Analizi + LAA	Eğricik Analizi + LAA
MIAS 70 (İyi Huylu-Kötü Huylu)	0,91	0,82	1	0,56	0,44	1
MIAS 140 (Normal-Anormal)	0,95	0,95	1	0,95	0,95	1
INBREAST 53(İyi Huylu-Kötü Huylu)	0,66	0,66	1	0,88	0,77	1
INBREAST 106 (Normal-Anormal)	0,86	0,86	1	1	1	1

Yapılan deneysel çalışmaları daha iyi değerlendirebilmek amacıyla bu bölümde benzer çalışmalar incelenmiş ve bu çalışmalarından kısaca bahsedilmiştir. Eltoouky ve arkadaşları [122] meme kanseri tanısı için yaptıkları çalışmada veri seti olarak MIAS veri tabanını kullanmışlardır. Uzmanlar tarafından belirtilen koordinatları kullanarak anormal bölgeleri 128x128 büyüklüğünde kırpılmışlar ve işlemlerini bu kırpılmış görüntüler üzerinden gerçekleştirmişlerdir. Öznitelikleri elde ederken bu kırpılmış görüntülere eğricik dönüşümü uygulamışlar ve çıkarılan en büyük katsayıların özel bir kümesini kullanmışlardır. Yaptıkları deneysel sonuçlar neticesinde eğricik dönüşümü ile dijital mamogramların analizi ve sınıflandırılmasında % 98.59'lık bir sınıflandırma doğruluğu elde etmişlerdir. Yine Eltoouky ve arkadaşlarının [123] yaptığı bir diğer çalışmada dijital mamografide meme kanseri tanısı için dalgacık ve eğricik dönüşümü karşılaştırmalı bir çalışma olarak sunulmuştur. Yaptıkları çalışmada çok çözünürlüklü analiz kullanılarak, mamografi görüntüleri, farklı frekans bantlarına duyarlı olan farklı çözünürlük seviyelerine ayırmışlardır. Her ayırma seviyesinden en büyük

katsayıların bir kümesini çıkararak öznitelik olarak kullanılmıştır. Veri seti olarak yine MIAS veri tabanından elde ettikleri görüntüleri kullanmışlardır. Bu görüntüler uzmanlar tarafından belirtilen koordinatlar merkez alınarak 128x128 boyutlarında kırılarak elde edilmiştir. Elde ettikleri deneysel sonuçlara göre eğricik dönüşüm katsayılarına dayalı normal, anormal sınıflandırmada başarı oranı % 94.07 olurken, dalgacık katsayıları ile elde edilen en yüksek oran ise % 90.05'dir. İyi huylu-kötü huylu sınıflar için, eğricik dönüşüm katsayıları ile % 94.28 başarı oranı elde edilirken dalgacık fonksiyonları (db8, Bior3.7 ve sym8) dönüşüm katsayıları ile elde edilen en yüksek oran % 87.83 tür. Bir başka çalışma Dhahbi ve arkadaşlarının [124] yaptığı çalışmadır. Bu çalışma meme kanserini tespit etmek amacıyla kullanılan çok ölçekli analizlerden elde edilen katsayıların boyut problemine bir çözüm bulmak amacıyla yapılmıştır. Çalışmalarında ayırık eğricik dönüşümü ile elde ettikleri katsayılar için ortalama, standart sapma, çarpıklık ve basıklık olmak üzere dört adet birinci seviye moment hesaplamışlardır. Veri seti olarak MIAS ve DDSM veri bankalarını kullanmışlardır. Her bir veri setinden görüntüler uzmanlar tarafından belirtilen koordinatlar yardımıyla 128x128 boyutlarında kırılarak yapılan işlemler bu görüntüler üzerinden gerçekleştirilmiştir. MIAS veri tabanından alınan 252 ROI üzerinde yapılan deneyler, meme kanseri teşhisi için eğricik seviyesi momentlerinin verimliliğini ve etkililiğini göstermişlerdir. Anormallik tespiti için 10 öznitelik ile % 91.27'lik bir doğruluk ve kötü huylu tespiti için 8 öznitelik ile % 81.35'lik bir doğruluk elde etmişlerdir. Ayrıca önerilen yöntem ile DDSM veri tabanından 11553 ROI üzerinde karşılaştırma yapmışlar ve önerdikleri yöntemin istatistiksel olarak anlamlı üstünlüğünü göstermişlerdir. Gedik ve Atasoy [125], dalga atomu dönüşümünü kullanarak dijital mamografide meme kanseri tanısı için bir yaklaşım sunmuşlardır. Veri seti olarak MIAS ve DDSM veri tabanı kullanılmıştır. Kullanılan veri tabanında yer alan mamogramlar uzmanlar tarafından belirtilen koordinatlar kullanılarak anormal bölgeleri 128x128 büyüklüğünde kırılmış ve işlemler bu kırılmış görüntüler üzerinden gerçekleştirmişlerdir. Öznitelik elde etmek amacıyla çok çözünürlüklü gösterim yöntemlerinin yakın bir üyesi olan dalga atomu yöntemini kullanmışlardır. Çalışmalarında, kırılmış görüntüleri dalga atomları temelinde ayrıştırılmış ve daha sonra dalga atom dönüşümünden en büyük katsayıların özel bir kümesini bir öznitelik vektörü olarak kullanmışlardır. Öznitelik azaltım amacıyla Temel Bileşen Analizi, sınıflandırıcı olarak da Destek Vektör Makinelerini kullanmışlardır. Normal-anormal sınıflandırma ve iyi huylu-kötü huylu sınıflandırma da % 100 doğruluk oranı elde etmişlerdir.

Bu tez çalışmasında da dijital mamogramlar üzerinden görüntü işleme teknikleriyle meme kanserinin teşhisi amacıyla bir BDT sistemi geliştirilmiştir. Bu sistemde veri seti olarak MIAS ve INBREAST veri tabanları kullanılmıştır. Kullanılan veri tabanlarında yer alan mamogramlar MIAS veritabanı için 128x128 büyüklüğünde, INBREAST veri tabanı için 500x500 boyutlarında kırılarak ROI bölgeleri elde edilmiş ve işlemler bu ROI'ler üzerinden gerçekleştirilmiştir. Öznitelikler elde edilirken GLCM analizi için 17, dalgacık dönüşümü için 21 ve eğricik dönüşümü için 162 öznitelik elde edilmiştir. Eğricik analizi için 4. seviyeye kadar katsayılar elde edilmiş ve toplamda 120153 adet katsayı üzerinden istatistiksel olarak hesaplanan standart sapma ve çarpıklık değerleri öznitelik olarak kullanılmıştır. Öznitelik azaltım amacıyla Lineer Ayırıcı Analizi, sınıflandırıcı olarak da Yapay Sinir Ağları, DVM ve K-EK sınıflandırıcılar kullanılmıştır. Her iki veri seti için hem normal-anormal sınıflandırmada hem de iyi huylu-kötü huylu sınıflandırma da eğricik analizi ve LAA'nın birlikte kullanılması sonucu her üç sınıflandırıcı için % 100 doğruluk oranı elde edilmiştir.

#### 4. BULGULAR VE SONUÇ

Bilgisayar Destekli Teşhis Sistemleri anormal yapıları ön plana çıkararak radyoloğun dikkatini şüpheli bölgelere çeken, radyologlara " ikinci bir okuyucu " olarak hizmet etmesi amaçlanan sistemlerdir. Nihai karar radyoloğa aittir. Geliştirilen BDT sisteminin de temel amacı budur. Mamografi meme kanseri ile mücadelede etkin bir araç olmasına rağmen mamografilerin doğru okunması her zaman mümkün olamamaktadır. Meme kanserlerinin yaklaşık % 85-90'ı mamografi incelemeleri sonucu elde edilirken, görüntülenen mamografilerdeki tümörlerin yaklaşık % 10-15'i çeşitli sebeplerden radyologlar tarafından gözden kaçırılmaktadır [11]. Ayrıca radyolojik anormalliklerin yaygınlığı, dikkati kaybetme ve yüksek miktarda mamografinin okunduğu taramalarda radyoloğun yorgunluğu, algısını bozabilmekte ve tümörlerin gözden kaçmasına neden olabilmektedir. Bu nedenle BDT sistemlerin geliştirilmesinin ve amaca uygun kullanılmasının gözden kaçan tümör oranlarını düşürmesi hedeflenmektedir. Bu kapsamda geliştirilen BDT sisteminde ilk olarak mamogramlar çeşitli ön işlemlere tabi tutulmuştur. Bu işlemler gürültü giderilmesi, görüntünün iyileştirilmesi ve bu tez kapsamında geliştirilen bir yöntemle pektoral kasın çıkartılması olarak sıralanabilir. Pektoral kas kitle vb. anormal yapılarla benzer özellikler gösterdiğinden bilgisayar destekli teşhis çalışmalarında şüpheli bölgelerle karıştırılmakta ve doğru teşhisi zorlaştırmaktadır. Bu nedenle ön işlem aşamasında pektoral kasın mamogramlardan çıkartılması sağlanarak doğru teşhis olasılığı yükseltmek istenmiştir. İkinci olarak radyoloğun dikkatini anormal yapıların üzerine çekmek amacıyla şüpheli bölgelerin elde edilmesi işlemleri gerçekleştirilmiştir. Şüpheli bölgeler yine bu tez kapsamında geliştirilen w-BSAFCM, Otsu eşikleme, Havrda & Charvat yöntemleri birlikte kullanılarak tespit edilmiş ve MIAS veri tabanı için 128x128 ve INBREAST veri tabanı için 500x500 boyutlarında kırılarak elde edilmiştir. Elde edilen şüpheli bölgelerden özneliklerin elde edilmesi için ise GLCM, dalgacık dönüşümü ve eğricik dönüşümü yöntemleri kullanılmıştır. GLCM, dalgacık dönüşümü ve eğricik dönüşümü için sırasıyla 17, 21 ve 162 adet özneliğin LAA ile boyut indirgenmiş formu sınıflandırıcılara sunulmuş ve normal-anormal, iyi huylu-kötü huylu ayrımı yapılmıştır. Sınıflandırıcı olarak YSA, DVM ve K-EK sınıflandırıcılar kullanılarak performansları

değerlendirilmiştir. Yapılan tez çalışmasında her iki veri tabanında hem normal-anormal hem de iyi huylu-kötü huylu sınıflamasında en iyi sonuçların %100 başarı oranı ile eğricik dönüşümü ve LAA'nın birlikte kullanılması sonucunda elde edildiği görülmüştür. Literatürde yer alan diğer çalışmalar ve bu tez çalışmasında elde edilen sonuçlar göz önünde bulundurulduğunda eğricik dönüşümünün meme kanserinin teşhisinde GLCM ve dalgacık analizine göre çok daha iyi sonuçlar verdiği görülmektedir. Bu tez çalışmasında önerilen yöntemlerin farklı veri tabanlarındaki sonuçlarını incelemek amacıyla iki farklı veri tabanı kullanılmıştır. Bunlardan ilki literatürde sıklıkla kullanılan ve erişime açık olarak bulunan MIAS veri tabanıdır. İkincisi ise güncel bir veri tabanı olan ve erişime açık olmayan INBREAST veri tabanıdır. INBREAST veri tabanından verilerin elde edilmesi amacıyla gerekli yazışmalar yapılarak veri tabanına erişim sağlanmıştır. Bu tez çalışması kapsamında geliştirilen yöntemler her iki veri tabanından alınan mamogramlar kullanılarak test edilmiştir.

Önerilen yeni BDT sisteminin literatüre katkıları şu şekilde özetlenebilir.

1. SSEM adı verilen pektoral kasın çıkartılmasına yönelik yeni bir yöntem geliştirilmiştir. Bu yöntem sayesinde BDT Sistemlerinde sıklıkla kitle vb. anormal yapılarla karıştırılmakta olan ve doğru teşhisi zorlaştıran pektoral kasın çıkartılması mümkün olmaktadır.
2. Bu tez kapsamında w-BSAFCM adı verilen yeni bir görüntü kümeleme algoritması geliştirilmiştir. Mamogram görüntüleri üzerinde Otsu N Algoritması, Havrda & Charvat Entropi ve w-BSAFCM kümeleme algoritması birlikte kullanılarak şüpheli bölge tespiti gerçekleştirilmiştir. Bu üç yöntemin birlikte kullanılmasıyla elde edilen ROI sayısı azaltılmış böylece işlem yükü de hafifletilmiştir.
3. Özniteliklerin elde edilmesi aşamasında üç farklı yöntem olan GLCM, Dalgacık ve Eğricik analizi kullanılmış ve bu yöntemlerin hem öznitelik azaltılmadan önce hem de öznitelik azaltıldıktan sonra mamogramların sınıflandırılması üzerindeki etkileri incelenmiş ve sonuçlar karşılaştırmalı olarak sunulmuştur. Bu üç yöntemin karşılaştırmalı olarak sunulduğu bir çalışmaya literatürde rastlanmamıştır.
4. Literatürde sıklıkla eğricik dönüşümü için elde edilen özniteliklerde katsayılar kullanılmış ve bu boyut problemine neden olmuştur. Boyut problemini azaltmak

amacıyla arařtırmacılar genellikle her seviye için en büyük katsayıların ilk yüzünü almıřlardır. oęunlukla Eęricik analizi ile öznitelik elde edilirken 4. Seviye seçilmekte bu da öznitelik matris boyutunun NX400 (N:kullanılan mamogram adeti) olması anlamına gelmektedir. Bu alıřmada ise eęricik dönüşümlerinde katsayıların kullanılması yerine her bir eęricik dönüşümü seviyesi için tüm katsayılar için bazı istatistiksel hesaplamalar kullanılmaktadır. Bunlar standart sapma ve arpıklık olarak belirlenmiştir. Toplamda 4. Seviye eęricik analizi için NX162 (N:kullanılan mamogram adeti) adet öznitelik elde edilerek boyut daha da indirgenmiş ve tüm katsayıların içerdiği veriler hesaplamalara dahil edilmiştir.

5. Yapılan tüm bu alıřmalar doğrutusunda mamogramların normal-anormal sınıflandırmasında en iyi sonuçlara eęricik dönüşümü ve lineer ayırıcı analizin birlikte kullanılması sonucunda ulařılmıştır. MIAS veri tabanından alınan 140 mamogram ve INBREAST veri tabanından alınan 106 mamogram için normal-anormal sınıflandırma başarısı % 100 olarak görülmüřtür. Benzer şekilde mamogramların iyi huylu–kötü huylu sınıflandırmasında MIAS veri tabanından alınan 70 anormal ve INBREAST veri tabanından alınan 53 anormal mamogram için eęricik dönüşümü ve lineer ayırıcı analizin birlikte kullanılması sonucunda % 100 başarı elde edilmiştir.
6. Bu tez kapsamında geliştirilen alıřmalar 2 tane bilimsel makale ve 3 tane bildiri olarak literatürde yer almıştır.

İleri alıřmalarda mamogramların BIRADS(Breast Imaging Reporting and Data Systems) kategorisine uygun şekilde sınıflandırılmasına yönelik alıřmalar yapılması planlanmaktadır.

## 5. KAYNAKLAR

- [1] V. Özmen, Z. Cantürk, V. Çelik, N. Güler, M. Kapkaç, A. Koyuncu, M. Müslümanoğlu, Z. Utkan, *Meme Hastalıkları Kitabı*, 2. Baskı, İstanbul, Türkiye: Güneş Tıp Kitabevleri, 2012, böl. 1, ss. 3-13.
- [2] S. Ateş, "Kadınların kendi kendine meme muayenesi uygulamalarının değerlendirilmesi," Yüksek lisans tezi, Hemşirelik Bölümü, İstanbul Bilim Üniversitesi, İstanbul, Türkiye, 2014.
- [3] L. Cıvcık, "Görüntü zenginleştirme ve hücreselel YSA kullanarak meme kanseri teşhisi," Doktora tezi, Elektrik-Elektronik Mühendisliği Bölümü, Selçuk Üniversitesi, Konya, Türkiye, 2013.
- [4] Halk Sağlığı Genel Müdürlüğü Kanser Dairesi Başkanlığı.(2017). [Online]. Erişim:www.kanser.gov.tr.
- [5] American College of Radiology (ACR), "Illustrated breast imaging reporting and data system (BI-RADS)," America, Rep. 3, 1998.
- [6] B. M. Güllüoğlu, "Meme hastalıklarına yaklaşım: "Meme kanseri için risk değerlendirmesi ve tarama stratejileri"," *Türkiye Aile Hekimliği Dergisi*, c. 12, s. 1, ss. 1-9, 2008.
- [7] H. Alpteker, D. Gümüş, S. Doğan, S. Bilir, M. Önal, "Kız öğrencilerin meme kanseri ve kendi kendine meme muayenesi bilgi ve uygulamalarının incelenmesi," *Meme Sağlığı Dergisi*, c. 7, s. 3, ss. 176-181, 2011.
- [8] V. Özmen, "Ulusal meme kanseri veritabanı verilerinin analizi," Türkiye Meme Hastalıkları Dernekleri Federasyonu, İstanbul, Türkiye, Rap., 2012.
- [9] N. Gedik, "Bilgisayar destekli teşhis sistemi ile mamografi görüntülerinin sınıflandırılması," Doktora tezi, Elektrik-Elektronik Mühendisliği Bölümü, Karadeniz Teknik Üniversitesi, Trabzon, 2013.
- [10] A. Açıkgöz, R. Çehreli, H. Ellidokuz, "Kadınların kanser konusunda bilgi ve tutumları ile erken tanı yöntemlerine yönelik davranışları," *Dokuz Eylül Üniversitesi Tıp Fakültesi Dergisi*, c. 25 s. 3 ss. 145-154, 2011.
- [11] J. Dheeba, N. Albert Singh, S. Tamil Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *Journal of Biomedical Informatics*, vol. 49, pp. 45-52, 2014.
- [12] B. Kurt, "Mamografi görüntülerinin yorumlanmasıyla bilgisayarlı teşhis sisteminin tasarımı," Doktora tezi, Bilgisayar Mühendisliği Bölümü, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye, 2014.
- [13] S. Ciatto, D. Ambrogetti, G. Collini, A. Cruciani, E. Ercolini, G. Risso, M. Rosselli Del Turco, "Computer-aided detection (CAD) of cancers detected on double reading by one reader only," *The Breast*, vol. 15, no. 4, pp. 528-532, 2006.

- [14] D. S. Gowri and T. Amudha, "A review on mammogram image enhancement techniques for breast cancer detection," *2014 International Conference on Intelligent Computing Applications*, Coimbatore, 2014, pp. 47-51.
- [15] V. Vijikala and D. A. S. Dhas, "Identification of most preferential denoising method for mammogram images," *2016 Conference on Emerging Devices and Smart Systems (ICEDSS)*, Namakkal, 2016, pp. 173-179.
- [16] M. Kumar, V. M. Thakkar, H. S. Bhadauria and I. Kumar, "Mammogram's denoising in spatial and frequency domain," *2016 2nd International Conference on Next Generation Computing Technologies*, Dehradun, 2016, pp. 654-659.
- [17] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio and F. Caselli, "Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1422-1430, 2008.
- [18] J. Kook Kim, J. Mi Park, K. Sik Song and H. Wook Park, "Adaptive mammographic image enhancement using first derivative and local statistics," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 495-502, 1997.
- [19] S. Amutha, D. R. Ramesh Babu, M. Ravi Shankar and N. Harish Kumar, "Mammographic image enhancement using modified mathematical morphology and Bi-orthogonal wavelet," *2011 IEEE International Symposium on IT in Medicine and Education*, Cuangzhou, 2011, pp. 548-553.
- [20] H. Moradmand, S. Setayeshi, A. R. Karimian, M. Sirous & M. E. Akbari, "Comparing the performance of image enhancement methods to detect microcalcification clusters in digital mammography" *Iranian Journal of Cancer Prevention*, vol. 5, no. 2, pp. 61-68, 2012.
- [21] K. Akila, L.S. Jayashree, A. Vasuki, "Mammographic image enhancement using indirect contrast enhancement techniques – A comparative study" *Procedia Computer Science*, vol. 47, pp. 255-261, 2015.
- [22] A. Strauss, A. Sebbar, S. Désarnaud, P. Mouillard and M. Le Gal, "Automatic detection and segmentation of microcalcifications on digitized mammograms," *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Paris, France, 1992, pp. 1938-1939.
- [23] Huai-Dong Li, M. Kallergi, L. P. Clarke, Wei Qian and R. A. Clark, "Markov random field model for mammogram segmentation," *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ*, 1993, pp. 54-55.
- [24] A. Sahakyan, H. Sarukhanyan, "Segmentation of the breast region in digital mammograms and detection of masses," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 2, pp. 102-105, 2012.
- [25] N.A.M. Isa, S.S. Ting. *Automatic segmentation and detection of mass in digital mammograms* [Online]. Available: [http:// http://www.wseas.us/e-library/conferences/2012/SaintMalo/SITE/SITE-23.pdf](http://www.wseas.us/e-library/conferences/2012/SaintMalo/SITE/SITE-23.pdf).
- [26] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, and H.N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition*, vol. 39, no. 4 , pp. 646-668, 2006.

- [27] T. O. Gulsrud, K. Engan and T. Hanstveit, "Watershed segmentation of detected masses in digital mammograms," *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, 2005, pp. 3304-3307.
- [28] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features of image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, 1973.
- [29] S. Lahmiri and M. Boukadoum, "Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction," *2011 IEEE 9th International New Circuits and systems conference*, Bordeaux, 2011, pp. 53-56.
- [30] H. Al-Shamlan and A. El-Zaart, "Feature extraction values for breast cancer mammography images," *2010 International Conference on Bioinformatics and Biomedical Technology*, Chengdu, 2010, pp. 335-340.
- [31] D. Saraswathi, D. Dharani and E. Srinivasan, "An efficient feature extraction technique for breast cancer diagnosis using curvelet transform and swarm intelligence," *2016 International Conference on Wireless Communications, Signal Processing and Networking*, Chennai, 2016, pp. 441-445.
- [32] X. Liu and J. Tang, "Mass classification in mammograms using selected geometry and texture features and a new SVM-based feature selection method," *IEEE Systems Journal*, vol. 8, no. 3, pp. 910-920, 2014.
- [33] R. J. Nandi, A. K. Nandi, R. Rangayyan and D. Scutt, "Genetic programming and feature selection for classification of breast masses in mammograms," *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, 2006, pp. 3021-3024.
- [34] R. Rabidas, A. Midya and J. Chakraborty, "Neighborhood structural similarity mapping for the classification of masses in mammograms," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 826-834, 2018.
- [35] M. E. Elmanna and Y. M. Kadah, "Implementation of practical computer aided diagnosis system for classification of masses in digital mammograms," *2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering*, Khartoum, 2015, pp. 336-341.
- [36] A. Šerifović-Trbalić, A. Trbalić, D. Demirović, N. Prljača and P. C. Cattin, "Classification of benign and malignant masses in breast mammograms," *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics*, Opatija, 2014, pp. 228-233.
- [37] Chun-Chu Jen, Shyr-Shen Yu, "Automatic detection of abnormal mammograms in mammographic images," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3048-3055, 2015.
- [38] R. Nithya and B. Santhi, "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer," *International Journal of Computer Applications* vol. 28 no. 6, pp. 21-25, 2011.
- [39] M.M. Eltoukhy, I. Faye, I. and B.B. Samir, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multi resolution representation," *Computers in Biology and Medicine*, vol. 42, pp. 123-128, 2012.

- [40] Pilot European Image Processing Archive. (2015). *The mini-MIAS database of mammograms* [Online]. Available: <http://peipa.essex.ac.uk/info/MIAS.html>.
- [41] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, J. S. Cardoso, "INBREAST", *Academic Radiology*, vol. 19, no. 2, pp. 236-248, 2012.
- [42] P. Athira, K. K. Fasna, A. Krishnan, "An overview of mammogram noise and denoising techniques," *International Journal of Engineering Research and General Science*, vol. 4, no. 2, 2016.
- [43] B. Ergen, M. Baykara. "Dalgacık ve dalgacık paket ayrıştırması yöntemleriyle imgelerden gürültü temizlemesi analizi," *e-Journal of New World Sciences Academy*, c. 6, s. 2, ss. 518-526, 2011.
- [44] E. A. Açoğlu, "Sayısal süzgeç tasarımı ve uygulamaları," Yüksek lisans tezi, Elektrik ve Elektronik Mühendisliği Bölümü, Başkent Üniversitesi, Ankara, 2008.
- [45] C. Tepe and H. Sezgin, "EMG artifact removing in ECG signal by wavelet transform," *National Conference on Electrical, Electronics and Computer Engineering*, Bursa, 2010, pp. 507-511.
- [46] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, New Jersey, USA: Prentice-Hall Inc., 1995.
- [47] S. Jose and D. A. Chandy, "Content based mammogram retrieval using biorthogonal wavelet filters in DDSM database," *2014 International Conference on Green Computing Communication and Electrical Engineering*, Coimbatore, 2014, pp. 1-6.
- [48] M. Misiti, Y. Misiti, G. Oppenheim, J. M. Poggi. (2016). *Wavelet Toolbox for Use with MATLAB* [Online]. Available: <http://www.mathworks.com>.
- [49] M. Bhat and P. M. S. Tarun, "Adaptive clip limit for contrast limited adaptive histogram equalization (CLAHE) of medical images using least mean square algorithm," *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, 2014, pp. 1259-1263.
- [50] M. Sundaram, K. Ramar, N. Arumugam et al., "Histogram modified local contrast enhancement for mammogram images," *Applied Soft Computing*, vol. 11, no. 8, pp. 5809–5816, 2011.
- [51] P. Rahmati, G. Hamarneh, D. Nussbaum et al., "A new preprocessing filter for digital mammograms," *Image and Signal Processing, ser. Lecture Notes in Computer Science*, vol. 6134, pp. 585–592, 2010.
- [52] M. Shelda and M. Ravishankar, "Optimized histogram based contrast limited enhancement for mammogram images," *ACEEE International Journal on Information Technology*, vol. 3, pp. 66–71, 2013.
- [53] S. Wu, S. Yu, Y. Yang et al., "Feature and contrast enhancement of mammographic image based on multiscale analysis and morphology," *Computational and Mathematical Methods in Medicine*, pp. 1–8, 2013.
- [54] I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram," *Computer Methods and Programs in Bio Medicine*, vol. 107, pp. 175–188, 2012.

- [55] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *The Journal of VLSI Signal Processing*, vol. 38 no.1, pp. 35-44, 2004.
- [56] J. Weickert, *Anisotropic Difusion in Image Processing*, Stuttgart., Denmark: Teubner, 1998.
- [57] M. Lee, Y. Chen, S. Kim and J. Kim, "Segmentation of the pectoral muscle boundary in breast MR images," *2011 IEEE 11th International Conference on Bioinformatics and Bioengineering*, Taichung, 2011, pp. 323-326.
- [58] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, R. A. Borges and A. F. Frere, "Automatic identification of the pectoral muscle in mammograms," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 232 – 245, 2004.
- [59] A. Sultana, M. Ciuc, R. Strungaru, "Detection of pectoral muscle in mammograms using a mean-shift segmentation approach," *2010 8th International Conference on Communications*, 2010, pp. 165–168.
- [60] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, "Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection," *Intelligent Information Systems Conference, The Seventh Australian and New Zealand*, 2001, pp. 67–72.
- [61] M. Yam, M. Brady, R. Highnam, C. Behrenbruch, R. English, Y. Kita, "Three-dimensional reconstruction of microcalcification clusters from two mammographic views," *IEEE Transactions on Medical Imaging*, vol 20. pp. 479–489, 2001.
- [62] X. Weidong, L. Lihua, L.A. Wei, "Novel pectoral muscle segmentation algorithm based on polyline fitting and elastic thread approaching," *The 1st International Conference on Bioinformatics and Biomedical Engineering*, 2007, pp. 837–840.
- [63] P. Miller, S. Astley, "Classification of breast tissue by texture analysis," *Image and Vision Computing*, vol. 10, no. 5, pp. 277-282, 1992.
- [64] A. Khademi, S. Farhang, A. Venetsanopoulos, K. Sridhar, "Region, lesion and border-based multiresolution analysis of mammogram lesions," *Lecture Notes in Computer Science*, vol. 5627, pp. 802–813, 2009.
- [65] V. Manik and Z. Andrew, "Classifying images of materials: achieving viewpoint and illumination independence," *Lecture Notes in Computer Science*, pp. 255-271, 2002.
- [66] K. Ganesan, U. Rajendra Acharya, K. Chua Chua, L. Choo Min, K. Thomas Abraham, "Pectoral muscle segmentation: A review," *Computer Methods and Programs in Biomedicine*, vol. 110, no. 1, pp. 48-57, 2013.
- [67] S. Deepa, , & V. S. Bharathi, "Efficient ROI segmentation of digital mammogram images using Otsu's N thresholding method," *Indian Journal of Automation and Artificial Intelligence*, vol. 1 no.2, pp. 51-56, 2013.
- [68] S. Sharma and P. Khanna, "ROI segmentation using local binary image," *2013 IEEE International Conference on Control System, Computing and Engineering*, Mindeb, 2013, pp. 136-141.

- [69] D. Cascio et al., "Mammogram segmentation by contour searching and mass lesions classification with neural network," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2827-2833, 2006.
- [70] B. W. Hong and B. S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 129-139, 2010.
- [71] J. Nagi, S. Abdul Kareem, F. Nagi and S. Khaleel Ahmed, "Automated breast profile segmentation for ROI detection using digital mammograms," *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences*, Kuala Lumpur, 2010, pp. 87-92.
- [72] F. Fauci et al., "Mammogram segmentation by contour searching and massive lesion classification with neural network," *IEEE Symposium Conference Record Nuclear Science 2004*, Rome, 2004, pp. 2695-2699.
- [73] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193-199, 2002.
- [74] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 737-752, 1999.
- [75] N. R. Pal, K. Pal, J. M. Keller and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517-530, 2005.
- [76] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," in *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1328-1337, 2010.
- [77] P. Parvathi and R. Rajeswari, "A hybrid FCM-ALO based technique for image segmentation," *2016 IEEE International Conference on Advances in Computer Applications*, Coimbatore, 2016, pp. 342-345.
- [78] Y. Gao, S. Wang and S. Liu, "Automatic clustering based on GA-FCM for pattern recognition," *2009 Second International Symposium on Computational Intelligence and Design*, Changsha, 2009, pp. 146-149.
- [79] K. L. Hemalatha, S. Ranjitha and H. N. Suresh, "Image segmentation based on modified centroid weight particle swarm optimization and spatial fuzzy C-means clustering algorithm," *2015 International Conference on Applied and Theoretical Computing and Communication Technology*, Davangere, 2015, pp. 723-727.
- [80] S. Shaik Parveen and C. Kavitha, "Segmentation of CT lung nodules using FCM with firefly search algorithm," *2015 International Conference on Innovations in Information, Embedded and Communication Systems*, Coimbatore, 2015, pp. 1-6.
- [81] J. Zang and B. Li, "Image segmentation using fast fuzzy c-means based on particle swarm optimization," *2010 Third International Conference on Intelligent Networks and Intelligent Systems*, Shenyang, 2010, pp. 370-373.

- [82] P. Civicioglu, "Backtracking search optimization algorithm for numerical optimization problems." *Applied Mathematics and Computation*, vol. 219 no.15, pp. 8121-8144, 2013.
- [83] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981, 95-107.
- [84] J. C. Bezdek, E. Robert, W. Full. "FCM: The fuzzy c-means clustering algorithm," *Computers and Geosciences*, vol 10, pp. 191-203, 1984.
- [85] X. Yong-Feng, Z. Shu-Ling, "Fuzzy particle swarm clustering of infrared images," *Second International Conference on Information and Computing Science*, 2009. pp. 122-124.
- [86] J. Hao, Y. Shen, H. Xu and J. Zou, "A region entropy based objective evaluation method for image segmentation," *2009 IEEE Instrumentation and Measurement Technology Conference*, Singapore, 2009, pp. 373-377.
- [87] B.S. Khehra, A.P.S. Pharwaha, "Digital mammogram segmentation using non-Shannon measures of entropy," *Proceedings of the World Congress on Engineering*, London, 2011, vol. 2, pp. 6-8.
- [88] R. Tiwari and B. Sharma, "A comparative study of Otsu and entropy based segmentation approaches for lesion extraction," *2016 International Conference on Inventive Computation Technologies*, Coimbatore, 2016, pp. 1-4.
- [89] S. Mathur and M. Gupta, "An analysis on color preservation using non-shannon entropy measures for gray and color images," *2014 Fourth International Conference on Advances in Computing and Communications*, Cochin, 2014, pp. 109-112.
- [90] A.P.S. Pharwaha, B. Singh, "Shannon and non-shannon measures of entropy for statistical texture feature extraction in digitized mammograms," *Proceedings of the World Congress on Engineering and Computer Science*, vol. 2, 2009, pp. 20-22.
- [91] B. Kurt, V. V. Nabyev, K. Turhan, "A novel automatic suspicious mass regions identification using Havrda & Charvat entropy and Otsu's N thresholding," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 3, pp. 349-360, 2014.
- [92] N. Otsu, "A threshold selection method from gray-level histograms," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.
- [93] R. Chaieb, A. Bacha, K. Kalti and F. Ben Lamine, "Image features extraction for masses classification in mammograms," *2014 6th International Conference of Soft Computing and Pattern Recognition*, Tunis, 2014, pp. 203-208.
- [94] A. Demirhan, İ. Güler, "Öz örgütlemeli harita ağları ve gri düzey eş oluşum matrisleri ile görüntü bölütleme," *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, c. 25 s. 2 ss. 285-291, 2013.
- [95] F. Albgretsen, *Statistical Texture Measures Computed From Grey Level Co-Occurrence Matrices*, 2008.
- [96] W. Gómez, W. Pereira & A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast

- ultrasound," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 1889–1899, 2012.
- [97] F. Doğru, "Gravite yönteminde dalgacık (wavelet) dönüşümü uygulamaları," Yüksek lisans tezi, Jeofizik Mühendisliği Bölümü, Dokuz Eylül Üniversitesi, İzmir, 2015.
- [98] C. Tepe, H. Sezgin, "EKG sinyallerinde gürültü gidermede ayrık dalgacık dönüşümünde farklı ana dalgacıkların ve ayrıştırma seviyelerinin karşılaştırılması," *12. Elektrik Elektronik Bilgisayar ve Biyomedikal Mühendisliği Ulusal Kongresi*, Eskişehir, 2007, ss. 383-387.
- [99] J. F. Ramirez-Villegas, D. F. Ramirez-Moreno, "Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions," *Neurocomputing*, vol. 77, no. 1, pp. 82-100, 2012.
- [100] E. Candès and D. Donoho, *Curvelets—A surprisingly effective nonadaptive representation for objects with edges*, in *Curves and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schumaker, Eds. Nashville: Vanderbilt Univ. Press, 2000.
- [101] M. A. Engin, "Eğricik dönüşümü kullanılarak imgelerde gürültü giderme ve içerik tabanlı görüntü erişimi," Doktora tezi, Haberleşme Bölümü, Atatürk Üniversitesi, Erzurum, 2016.
- [102] D. Saraswathi, D. Dharani and E. Srinivasan, "An efficient feature extraction technique for breast cancer diagnosis using curvelet transform and swarm intelligence," *2016 International Conference on Wireless Communications, Signal Processing and Networking*, Chennai, 2016, pp. 441-445.
- [103] N. Gedik, A. Atasoy, "A computer-aided diagnosis system for breast cancer detection by using a curvelet transform," *Turkish Journal of Electrical Engineering & Computer Sciences*, c. 21, s. 4 , pp. 1002-1014, 2013.
- [104] M. M. Eltoukhy, I. Faye and B. B. Samir, "Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram," *2010 International Conference on Intelligent and Advanced Systems*, Kuala Lumpur, Malaysia, 2010, pp. 1-5.
- [105] A. Tzaniş, "The Curvelet Transform in the analysis of 2-D GPR data: Signal enhancement and extraction of orientation-and-scale-dependent information," *Journal of Applied Geophysics*, vol. 115, pp. 145-170, 2015.
- [106] S. Alzubi, N. Islam and M. Abbod, "Multiresolution analysis using wavelet, ridgelet and curvelet transforms for medical imaging segmentation," *International Journal of Biomedical Imaging*, vol. 2011, pp. 1-18, 2011.
- [107] M. A. Akbar et al., "An empirical study for pca- and lda-based feature reduction for gas identification," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5734-5746, 2016.
- [108] F. E. Fassnacht et al., "Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central european test sites," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2547-2561, 2014.

- [109] E. Emary, W. Yamany and A. E. Hassanien, "New approach for feature selection based on rough set and bat algorithm," *2014 9th International Conference on Computer Engineering & Systems*, Cairo, 2014, pp. 346-353.
- [110] M. S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 817-835, 2018.
- [111] B. Tran, M. Zhang and B. Xue, "A PSO based hybrid feature selection algorithm for high-dimensional classification," *2016 IEEE Congress on Evolutionary Computation*, Vancouver, 2016, pp. 3801-3808.
- [112] F. V. Alexandre, N. C. Cortés and E. A. Anaya, "Feature selection to detect botnets using machine learning algorithms," *2017 International Conference on Electronics, Communications and Computers*, Cholula, 2017, pp. 1-7.
- [113] R. Fisher, "The Use of multiple measurements in taxonomic problems," *Annals of Eugenics*, pp. 179–188, 1936.
- [114] C. Li, B. Wang. (2017). *Fisher linear discriminant analysis*. [Online]. Access: [http://www.ccs.neu.edu/home/vip/teach/MLcourse/5\\_features\\_dimensions/lecture\\_notes/LDA/LDA.pdf](http://www.ccs.neu.edu/home/vip/teach/MLcourse/5_features_dimensions/lecture_notes/LDA/LDA.pdf)
- [115] V.N. Vapnik, *The Nature of Statistical Learning Theory*, New York , USA : Springer-Verlag, 1995.
- [116] T. Kavzoğlu , İ. Çölkesen, "Karar destek makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi," *Harita Dergisi*, c. 144, ss. 73–82, 2010.
- [117] F. Adıyaman, "Talep tahmininde yapay sinir ağlarının kullanılması," Yüksek lisans tezi, Mühendislik Yönetimi Bölümü, İstanbul Teknik Üniversitesi, İstanbul, 2007.
- [118] E. Öztemel, *Yapay Sinir Ağları*, 3. Baskı, İstanbul, Türkiye: Papatya Yayıncılık, 2003.
- [119] L. Oral , "Uyarlamalı sinir-bulanık sınıflayıcı ile parmak izi tanıma ve dalgacık tabanlı momentlerle özellik çıkarma," Yüksek lisans tezi, Biyometri Bölümü, Süleyman Demirel Üniversitesi, Isparta, 2010.
- [120] K.Q. Weinberger and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, pp. 207–244, 2009.
- [121] M. S. İyisoy, "Tanı test ölçütlerinde ROC eğrisi ve sınıflama analizlerinin karşılaştırılmasında kullanımı," Yüksek lisans tezi, İstatistik Bölümü, Selçuk Üniversitesi, Konya, 2014.
- [122] M. M. Eltoukhy, I. Faye, B. B. Samir, "Breast cancer diagnosis in digital mammogram using multiscale curvelet transform," *Computerized Medical Imaging and Graphics*, vol. 34, no. 4, pp. 269-276, 2010.
- [123] M. M. Eltoukhy, I. Faye, B. B. Samir, " A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 384-391, 2010.

- [124] S. Dhahbi, W. Barhoumi, E. Zagrouba, "Breast cancer diagnosis in digitized mammograms using curvelet moments," *Computers in Biology and Medicine*, vol. 64 pp. 79-90, 2015.
- [125] N. Gedik, A. Atasoy, "Performance evaluation of the wave atom algorithm to classify mammographic images," *Turkish Journal of Electrical Engineering & Computer Sciences.*, vol. 22, pp. 957-969, 2014.



# ÖZGEÇMİŞ

## KİŞİSEL BİLGİLER

Adı Soyadı : Güliz TOZ  
Doğum Tarihi ve Yeri : 25/12/1978-Balıkesir  
Yabancı Dili : İngilizce  
E-posta : glz.toz@gmail.com

## ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Y. Lisans	Elektrik Elektronik Müh.	Düzce Üniversitesi	2014
Lisans	Bilgisayar Öğretmenliği	Kocaeli Üniversitesi	2006
Ön Lisans	Bilgisayar Programcılığı	Gaziosmanpaşa Üniversitesi ZMYO	1999
Lise		Ankara Fatih Lisesi	1995