

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**MEME KANSERİ BİYOPSİ ÖRNEKLERİNDE
ÖSTROJEN HORMONU ALICI VARLIĞININ
ZEKİ YÖNTEMLERLE TESPİTİ**

**Hazırlayan
Fatih SARIKOÇ**

**Danışman
Prof. Dr. Adem KALINLI**

Doktora Tezi

**Ağustos 2012
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**MEME KANSERİ BİYOPSİ ÖRNEKLERİNDE
ÖSTROJEN HORMONU ALICI VARLIĞININ
ZEKİ YÖNTEMLERLE TESPİTİ
(Doktora Tezi)**

**Hazırlayan
Fatih SARIKOÇ**

**Danışman
Prof. Dr. Adem KALINLI**

**Bu çalışma; Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından
FBD-08-35 kodlu proje ile desteklenmiştir.**

**Ağustos 2012
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim. Ayrıca, ERÜ Klinik Araştırma Etik Kurulunun 2012/89 karar numaralı ve 10.02.2012 tarihli kararı uyarınca tez konusu çalışmanın gerçekleştirilmesinde etik ve bilimsel sakınca bulunmadığını beyan ederim.

Fatih SARIKOÇ

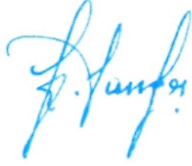


YÖNERGEYE UYGUNLUK

“Meme Kanseri Biyopsi Örneklerinde Östrojen Hormonu Alıcı Varlığının Zeki Yöntemlerle Tespiti” adlı Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

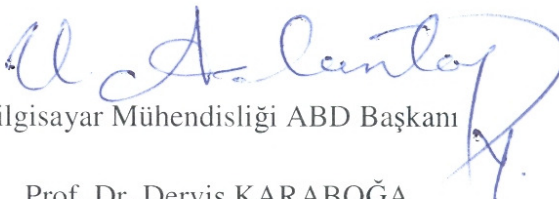
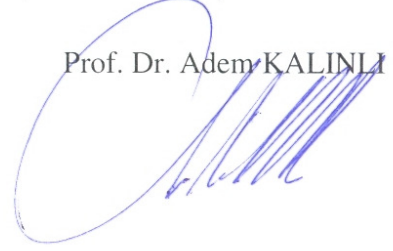
Tezi Hazırlayan

Yük. Müh. Fatih SARIKOÇ



Danışman

Prof. Dr. Adem KALINLI



Bilgisayar Mühendisliği ABD Başkanı

Prof. Dr. Derviş KARABOĞA

Prof. Dr. Adem KALINLI danışmanlığında **Fatih SARIKOÇ** tarafından hazırlanan “Meme Kanseri Biyopsi Örneklerinde Östrojen Hormonu Alıcı Varlığının Zeki Yöntemlerle Tespiti ” adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Mühendisliği** Anabilim Dalında **Doktora** tezi olarak kabul edilmiştir.

07.08.2012
(Tez Savunma Sınav Tarihi)

JÜRİ:

Başkan: Prof. Dr. Vasif NABİYEV

Üye : Prof. Dr. Kenan DANIŞMAN

Üye : Prof. Dr. Derviş KARABOĞA

Üye : Doç. Dr. Veysel ASLANTAŞ

Üye : Prof. Dr. Adem KALINLI

ONAY:

Bu tezin kabulü, Enstitü Yönetim Kurulunun 11/09/2012 tarih ve 2012/42-06 sayılı kararı ile onaylanmıştır.

11/09/2012


Prof. Dr. Necmettin MARAŞLI
Enstitü Müdürü

TEŞEKKÜR

Hayatımın tüm evrelerinde bana sınırsız özveri ile yürekten destek olan anneme, babama, kardeşlerime ve kendisinden esirgediğim zamana katlanan eşime teşekkür ediyorum. Ailemin gösterdiği anlayış ve verdiği destek olmaksızın bu tez çalışmasını bitirmek mümkün olmayacaktı.

Doktora eğitimi, geçimini akademisyenlik mesleğinden sağlamayan ve tez çalışmasını bir gönül işi olarak arta kalan zamanlarında yapan şahsım için zorlu bir süreçti. Her fırsatta bana duyduğu güveni hissettiren hocam Prof. Dr. Kenan Danışman'a, sezgisel yöntemler ve yapay öğrenmenin temelleri için hocam Prof. Dr. Derviş Karaboğa'ya ve görüntü işleme temelleri için hocam Doç. Dr. Veysel Aslantaş'a teşekkürlerimi sunarım. Bu tez çalışmasının yazımında nazik eleştirileriyle birçok eksiğin giderilmesine katkı sağlayan hocam Prof. Dr. Vasif Nabiye'ye de ayrıca teşekkür ederim.

Tıp Fakültesi Patoloji Bölümü arşivini bu tez çalışması için bizlere açan Prof. Dr. Figen Öztürk'e, tıbbi rehberlik ve görüntü işaretlemeleri için Doç Dr. Hülya Akgün'e, östrojen alıcı puanlaması ve işaretlemeleri için Patolog Uz. Dr. Mehtap Kala'ya, mikroskop okumaları için Dr. Fatma Aykaş'a, istatistiksel tıbbi veri analizinde sağladığı destek için Yrd. Doç. Dr. Ahmet Öztürk'e, mühendislik fakültesinde kullanıma sunduğu mikroskop için Yrd. Doç. Dr. Alper Baştürk'e, tez savunma sınavımda katkı sağlayan Yrd. Doç. Dr. Mete Çelik'e, tez çalışmasının baskısı ve teslim edilmesinde verdiği emekler için kardeşim Arş. Gör. Selçuk Sarıkoç'a teşekkür borçluyum.

Son olarak, 2002 yılından bugüne kadar hayatın bizlere getirdiği üzüntüleri ve sevinçleri beraber yaşadığımız hocam Prof. Dr. Adem Kalınlı'ya teşekkürlerimi sunarım.

Fatih SARIKOÇ

Ankara, 2012

MEME KANSERİ BİYOPSİ ÖRNEKLERİNDE ÖSTROJEN HORMONU ALICI VARLIĞININ ZEKİ YÖNTEMLERLE TESPİTİ

Fatih SARIKOÇ
Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Doktora Tezi, Ağustos 2012
Danışman: Prof. Dr. Adem KALINLI

KISA ÖZET

Östrojen hormonu alıcı varlığının değerlendirilmesi, meme kanserinin tanısında başvuru ve doktor tarafından öznel olarak gerçekleştirilen bir işlemdir. Östrojen alıcı durum değerlendirmesinde yeniden üretilebilir ve nesnel sonuçlar almak için, bu tez çalışmasında çeşitli yapay öğrenme yöntemleri ve farklı öznitelik kümelerinin sınıflama ve tıbbî tanı başarımları incelenmiştir.

Yarıçap fonksiyonlu ağlar, k-en yakın komşuluk metodu, destek vektör makineleri, sade bayes sınıflayıcı, fonksiyonel karar ağaçları ve k-orta nokta algoritması test veri kümesi üzerinde deneysel olarak incelenmiştir. Çeşitli öznitelik kümeleri üzerinde her bir metodun sınıflama başarımları araştırılmıştır. Bu metodlarla test resimleri üzerinde elde edilen tanı kararları uzman patoloğların kararları ile karşılaştırılmıştır.

Elde edilen deneysel sonuçlara göre, östrojen alıcı tanısında fonksiyonel karar ağaçları ve sade bayes sınıflayıcısının birleşiminden oluşan tümleşik model en iyi tanı puanlamasını ve östrojen alıcı negatif/pozitif ayırımını gerçekleştirmiştir. Bu iki yöntem diğer metotlara göre daha iyi genelleme yapabilme özelliğini ortaya koymuştur. Elde edilen sonuçlara göre, dalgacık desen öznitelikleri Laws desen özniteliklerinden ve eş oluşum matrisi özniteliklerinden daha üstün sınıflama kabiliyeti göstermektedir.

Anahtar Kelimeler: görüntü işleme, sınıflama, hücre bölütleme, östrojen hormonu alıcı durum değerlendirmesi, meme kanseri tanısı

USING INTELLIGENT METHODS TO DETECT ESTROGEN HORMONE RECEPTOR EXISTANCE IN BREAST CANCER BIOPSY SAMPLES

Fatih SARIKOÇ

Erciyes University, Graduate School of Natural and Applied Sciences

Ph. D. Thesis, August 2012

Supervisor: Prof. Dr. Adem KALINLI

ABSTRACT

In medical practice for prognosis of breast cancer, estrogen receptor status evaluations have been done subjectively. To get objective and reproducible analysis of estrogen receptor, this thesis examined classification and prognostic scoring performances of several computer methods on different feature sets.

Radial basis function network, k-nearest neighborhood search, support vector machines, naive bayes classifier, functional trees and k-means clustering algorithm were experimented on the test datasets. Several features were employed and classification accuracies of each method on these features were examined. Assessment results of the methods on test images were also experimentally compared with that of two pathologist experts.

According to results our experimental work, combination of functional trees and naive bayes classifier gave the best prognostic scores and dichotomization results for assessment of estrogen receptor status. These two methods showed better generalization behavior than the other methods. Wavelet color features provided better classification accuracy than Laws texture energy and co-occurrence matrix features.

Keywords: image processing, classification, nucleus segmentation, estrogen hormone receptor (ER) status evaluation, breast cancer prognosis

İÇİNDEKİLER

MEME KANSERİ BİYOPSİ ÖRNEKLERİNDE ÖSTROJEN HORMONU ALICI VARLIĞININ ZEKİ YÖNTEMLERLE TESPİTİ

Sayfa

BİLİMSEL ETİĞE UYGUNLUK	ii
YÖNERGEYE UYGUNLUK	iii
ONAY	iv
TEŞEKKÜR	v
KISA ÖZET	vi
ABSTRACT	vii
İÇİNDEKİLER.....	viii
TABLolar LİSTESİ	xii
ŞEKİLLER LİSTESİ.....	xiii
SİMGELER	xv

GİRİŞ

G.1. Tezin Konusu.....	1
G.2. Tezin Amacı ve Önemi.....	2
G.3. Yöntemler ve Tezin Bölümleri	6
G.4. Terimler Sözlüğü	8

1. BÖLÜM

GENEL BİLGİLER VE ÖSTROJEN ALICI TANISI

1.1. Meme Kanserinin Sınıflandırılması	11
1.2. Doku Yapısında Meme Kanserinin Yaygın Türleri	14

1.2.1. Yayılmacı Olmayan	14
1.2.2. Yayılmacı.....	16
1.3. Tümör Derecesi	17
1.4. Meme Kanserinin Evreleri	19
1.5. Kanserli Hücrelerin Gen Yapısı	20
1.6. Hormon Alıcı Durumu	20
1.6.1. Östrojen Hormonu Alıcı Durumu Puanlaması	21
1.6.2. Östrojen Alıcı Tanısında Literatür Taraması	22
1.6.3. Tez Çalışmasının Farklılığı.....	23

2. BÖLÜM

GÖRÜNTÜ İŞLEME YÖNTEMLERİ

2.1. Bölütleme Yöntemleri	25
2.1.1 Büyüyen Bölge Bölütlemesi	25
2.1.2. Bölge Parçalama ve Birleştirme Bölütlemesi	26
2.1.3. Sutaşımı Bölütleme Algoritması.....	26
2.1.4. <i>K</i> -Orta Nokta Bölütleme Yöntemi.....	28
2.1.5. Otsu Eşikleme Yöntemi	28
2.2. Bölütleme Yöntemlerinin Görsel Sınaması.....	31

3. BÖLÜM

ÖZNİTELİK ELDE ETME YÖNTEMLERİ

3.1. Birinci Dereceden İstatiksel Öznitelikler	40
3.2. İkinci Dereceden İstatiksel Öznitelikler	42
3.2.1. Birliktelik Matrisi Öznitelikleri	42
3.2.2. Laws Desen Enerji Öznitelikleri.....	45

3.2.3 Dalgacık Dönüşümü Enerji Öznitelikleri	46
3.3. Öznitelik Seçimi	48

4. BÖLÜM

YAPAY ÖĞRENME VE SINIFLAMA YÖNTEMLERİ

4.1. Öklid Uzaklık Ölçütü	55
4.2. K-Orta Nokta Algoritması	55
4.3. Yarıçap Taban Fonksiyonlu Ağlar	57
4.4. K-En Yakın Komşuluk Algoritması	61
4.4.1. Olasılık Yoğunluk Fonksiyonu Tahmini	61
4.4.2. Parzen Pencere Fonksiyonları	61
4.4.3. Olasılık Tabanlı Sinir Ağı Modeli	64
4.4.4. K-En Yakın Komşuluk Algoritması	66
4.5. Destek Vektör Makineleri	67
4.6. Sade Bayes Sınıflayıcı	72
4.7. Karar Ağaçları Sınıflayıcısı	73
4.8. Sınıflayıcı Modeli Başarım Ölçütleri	76

5. BÖLÜM

FONKSİYONEL KARAR AĞAÇLARI İLE OTOMATİK TANI BULGULARI

5.1. Uygulama Özeti	79
5.2. Malzemeler ve Yöntemler	80
5.2.1. Bölümleme ve Biçimbilim İşlemleri	83
5.2.2. Öznitelik Çıkarımı ve Elemeşi	86
5.2.3. Sınıflayıcı	86
5.2.4. Bilgisayar Tabanlı Puanlama Yöntemi	90

5.3. Sonular ve Tartışma.....	90
5.3.1. Öznitelik Çıkarım ve Eleme Sonuları	90
5.3.2. Sınıflama Sonuları.....	92
5.3.3. Tanı Kararları ve Puanlama Sonuları.....	94
5.3.5. Tartışma ve Sonular	96

6. BÖLÜM

ÇEŞİTLİ ÖZNİTELİK GURUPLARI VE SINIFLAYICILARIN OTOMATİK TANI BAŞARIMINDA BULGULAR

6.1. Özet	103
6.2. Malzeme ve Yöntemler	104
6.3. Sonular.....	105
6.3.1. Öznitelik Sonuları	105
6.3.2. Sınıflayıcı Sonuları	107
6.3.3. Puanlama Sonuları	110
6.3.4. Tartışma ve Sonu	112

7. BÖLÜM

TARTIŞMA - SONUÇ VE ÖNERİLER

7.1. Tartışma Sonular ve Öneriler	117
--	-----

KAYNAKLAR	122
------------------------	------------

ÖZGEÇMİŞ.....	130
----------------------	------------

TABLÖLER LİSTESİ

Tablo G.1. Türkçe terimler ve İngilizce karşılıkları	8
Tablo 1.1. Allred puanı, Toplam Puan (TP) = YP + AP (değerler 0, 2–8)	22
Tablo 3.1. Laws desen maskeleri	45
Tablo 4.1. Çekirdeğin ait olabileceği sınıf türleri	76
Tablo 5.1. Azaltılmış renkli öznitelikler ve kısaltma adları	86
Tablo 5.2. Doğrusal fonksiyonlarla FT modeli	87
Tablo 5.3. Doğrusal SVM modeli ve katsayıları (destek vektörleri değil)	88
Tablo 5.4. Farklı öznitelik grupları üzerinde FT başarımı	91
Tablo 5.5. FT hata matrisi	93
Tablo 5.6. Kappa katsayılarının yorumlanması	94
Tablo 5.7. Test resimlerinde puanlamalar	95
Tablo 5.8. Puanlar arasında örtüşme, ilinti ve farklılık	96
Tablo 6.1. WEKA kütüphanesinden çağrılan sınıflayıcılar ve ilk değer atamaları	104
Tablo 6.2. Özniteliklere göre öğrenme yöntemlerinin doğru sınıflama yüzdeleri	105
Tablo 6.3. Öznitelik gruplarının öğrenme yöntemlerinde doğru sınıflama yüzdeleri	106
Tablo 6.4. Öğrenme yöntemlerinin duyarlılık ve özgüllük değerleri	107
Tablo 6.5. Öğrenme yöntemleri ve gözlemciler arasında	111

ŞEKİLLER LİSTESİ

Şekil G.1. Doktor tarafından verilen östrojen alıcı tanısı.....	2
Şekil G.2. Bilgisayar tabanlı tanı sistemi işlem basamakları	3
Şekil G.3. Östrojen alıcı boyamasını gösterir örnek bir sayısal görüntü	4
Şekil G.4. Bilgisayar tabanlı kanser tanısı/teşhisi aşamaları [13].....	7
Şekil 1.1. Genel olarak meme kanser teşhis işlemleri [15].....	12
Şekil 1.2. Meme kanserinde tümör [16].....	12
Şekil 1.3. Kanser sınıflandırılması.....	13
Şekil 1.4. HE boyamasında normal bir meme dokusunun görünümü [18].....	15
Şekil 1.5. HE boyamasında DCIS meme kanseri doku görüntüsü [18].....	15
Şekil 1.6. HE boyamasında IDC meme kanseriyle bozulmuş doku [18].....	16
Şekil 2.1. Hücre çekirdeği bölütleme işlem basamakları.....	31
Şekil 2.2. Asıl resim, 2048x1536x24 piksel, 40x ölçekte.....	32
Şekil 2.3. Griton dönüşümü	32
Şekil 2.4. Eşikleme işlemi çıktısı olan ikilik resim.....	33
Şekil 2.5. Bağlı bileşen hesaplama çıktısı resim.....	34
Şekil 2.6. Otsu eşikleme yöntemiyle bölütleme çıktısı	35
Şekil 2.7. k -orta nokta öbekleme yöntemiyle bölütleme çıktısı	36
Şekil 2.8. Sutaşımı algoritması ile elde edilen bölütleme çıktısı	36
Şekil 3.1. Her bir renk kanalında üretilen özniteliklerin türü ve sayısı	38
Şekil 3.2. Öznitelik türleri ve hesaplanan öznitelikler	39
Şekil 3.3. Öznitelik eleme işlemi sonucu	49
Şekil 3.4. Öznitelik eleme sonucunda seçilen 11 adet öznitelik ve kısaltmaları.....	49
Şekil 4.1. Sınıflayıcı giriş ve çıkışı	54
Şekil 4.2. Basit bir RBFN ağı.....	58
Şekil 4.3. Olasılık tabanlı yapay sinir ağı modeli	64
Şekil 4.4. Farklı doğrusal ayırıcılar.....	68
Şekil 4.5. En uygun doğrusal ayırıcı	68
Şekil 4.6. Noktasal çarpımla sınıf tespiti	69
Şekil 4.7. Hata matrisi	77
Şekil 5.1. Sitoplazmik boyama	81
Şekil 5.2. Tanı sistemini oluşturan ana ve alt yazılım modülleri.....	82

Şekil 5.3. Tanı sisteminde ana modüllere ait giriş ve çıkış verileri	82
Şekil 5.4. Bölümlendirme aşaması alt işlemleri; (a) asıl resimden bir kesit, (b) kesit üzerinde Otsu eşikleme işlemi sonucu, (c) biçimbilim işlemlerinin sonucu, (d) tespit edilen hücre çeperinin asıl resim kesiti üzerine giydirilmesi işlemi sonucu ..	85
Şekil 5.5. C4.5 karar ağacı	89
Şekil 5.6. FT sınıflayıcısı alıcı işletim karakteristik (ROC) eğrileri	93
Şekil 6.1. Sınıflayıcıların ROC eğrileri; (a) tümleşik sınıflayıcı (b) fonksiyonel karar ağacı, (c) en yakın-k komşuluk arama yöntemi, (d) destek vektör makineleri	108
Şekil 6.2. Sınıflayıcıların ROC eğrileri; (e) yarıçap fonksiyonlu ağlar, (f) sade bayes sınıflayıcı, (g) k-orta nokta öbekleme.....	109
Şekil 6.3. Birleşik sınıflayıcı modellerinin doğru sınıflama yüzdeleri	110

SİMGELER

<u>Sembol</u>	<u>Anlamı</u>
N_p	Pozitif boyanmış hücre çekirdeği sayısı
N_n	Negatif boyanmış hücre çekirdeği sayısı
YP	Allred puanlama yordamına göre yüzdelik puanı
AP	Allred puanlama yordamına göre aydınlık puanı
$\varphi(x, y)$	Dalgacık dönüşümünde ölçekleme fonksiyonu
$\psi(x, y)$	Dalgacık dönüşümünde dalgacık fonksiyonu
$\overline{r_{cf}}$	Öznitelik-sınıf arasında ortalama ilinti değeri
$\overline{r_{ff}}$	Öznitelik-öznitelik arasında ortalama ilinti değeri
D_E	Öklid uzaklığı
o	Yapay sinir ağı çıkışı
E	Yapay sinir ağı çıkışı hata değeri
d	Yapay sinir ağı çıkışında arzu edilen değer
w	Yapay sinir ağı ağırlık katsayısı
f	Yapay sinir ağı iletim fonksiyonu
η	Yapay sinir ağı gradiyent yönünde en dik iniş katsayısı
δ	Yapay sinir ağı lokal gradiyent
α	Yapay sinir ağı momentum katsayısı

$p(C_i \vec{x})$	Vektör $\vec{x}$ bilindiğinde C_i sınıfına ait olma olasılığı
∇	Birinci sınıflayıcının oluşturduğu verisi kümesinde ikinci sınıflayıcı eğitilmesini gösterir işlemci
TP	Doğru Pozitif
FP	Yanlış Pozitif
TN	Doğru Negatif
FN	Yanlış Negatif
ρ	İlinti katsayısı
κ	Kappa örtüşme katsayısı
p	Wilcoxon farklılık sınamasında varsayım hipotezi doğruluk olasılığı

GİRİŞ

G.1. Tezin Konusu

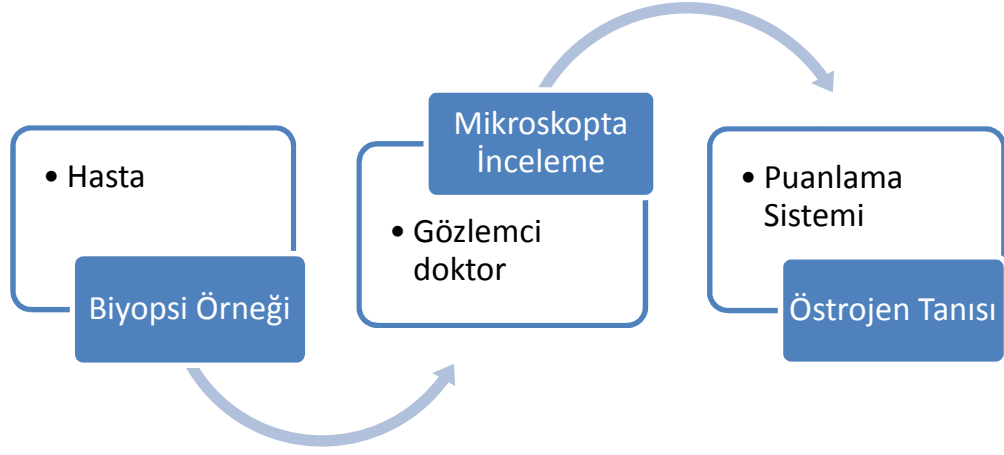
Özellikle gelişmiş ülkelerde meme kanseri bayanlarda en sık rastlanan kanser türüdür. Meme kanserinin çeşitli türleri bulunmaktadır [1]. Bunlardan en yaygın olanı meme dokusunun süt kanallarında oluşan yayılmacı süt kanalı kanseridir. Bu tez çalışması, yayılmacı süt kanalı meme kanseri biyopsi örneklerinde, östrojen alıcısı üzerine bilgisayar tabanlı tanı konulmasına odaklanmıştır.

Östrojen alıcı (ER) tanısı hastalığın tedavi süresini, yöntemini, maliyetini ve kullanılacak kimyasal ilacın belirlenmesini etkilemektedir [2, 3]. Hasta için östrojen alıcı tanısı konurken uygulanan çeşitli puanlama yöntemleri bulunmaktadır [4-6]. Bunlardan birisi de Allred puanlama yöntemidir [7]. Bu yöntem östrojen boyamasında düşük seviyeli pozitif boyanmaları da dikkate alarak daha hassas tanı sonuçları üretmekte ve bu nedenle özellikle son yıllarda giderek yaygınlık kazanmaktadır [8, 9].

Tıbbi görüntü işlemede birçok farklı donanımla çeşitli görüntü türleri elde edilmektedir [10]. Bunlardan bir tanesi de tümörlü bölgeden alınan doku örneklerinin ışın mikroskopunda incelenmesi ile elde edilen görüntülerdir. Meme kanserinde ışın mikroskopunda histolojik ve sitolojik görüntüler elde edilebilir. Sitolojik görüntüler sadece hücreyi ve hücre gruplarını gösterir. Sitolojik görüntü elde edilmesi sürecinde hücre dokusu bozulur. Çekirdeğin büyüklüğü, biçimi ve renklendirici sıvının hücre içerisindeki yoğunluğu sitolojik uygulamaların en kullanışlı ölçütleridir. Histolojik görüntüler dokuya odaklanır, dokuyu ve doku yapısı içerisinde hücre çekirdeklerini gösterir [11].

Genel olarak östrojen alıcı tanısı aşağıdaki şekilde verildiği biçimde elde edilir (Şekil G.1). Meme kanseri şüphesi görülen veya meme kanseri teşhisi konmuş olan hastadan

bir biyopsi örneği alınır. Bu biyopsi örneği östrojen alıcı tanısı amacıyla östrojen alıcı boyamasına tabi tutulur. Patolog hekim, biyopsi örneğinin östrojen alıcı boyanmasını ışın mikroskobu altında inceler. Patolog, kahverengi tonda boyanmış hücre çekirdeklerinin östrojen alıcı pozitif olarak, mavi ve mor tonlarda boyanmış hücre çekirdeklerini östrojen alıcı negatif olarak tespit eder. Mikroskop altında tümörü temsil ettiği düşünülen bir biyopsi kesiminde toplam östrojen alıcı pozitif ve negatif hücre çekirdekleri sayılır. Patolog ayrıca pozitif boyanma durumunda boyanmanın aydınlık şiddetini de dikkate alır. Bu veriler doğrultusunda uygulanmakta olan bir puanlama sistemine göre tanı puanı verilerek östrojen tanısı koyma işlemi tamamlanmış olur.



Şekil G.1. Doktor tarafından verilen östrojen alıcı tanısı

Bu tez çalışmasında, immünohistokimyasal resimlerde östrojen hormonu alıcısı varlığı bilgisayar tabanlı olarak incelenmiştir. Bu inceleme sonucunda hastadan alınan örnek üzerinde östrojen alıcısı varlığı hakkında tanı konulur. Bu tanı tedavinin sürecini, maliyetini ve uygulanacak kimyasal ilaç türünü belirlemede temel oluşturmaktadır.

G.2. Tezin Amacı ve Önemi

Bu tez çalışması, yayılcı süt kanalı meme kanserinde östrojen alıcı durum tanısı amacıyla kullanılabilecek çeşitli mühendislik araçlarının (yöntemlerinin) tıbbi tanı elde edilmesinde kullanımını ve bunların başarımlarını incelemesini amaçlar.

Bu amaç doğrultusunda, Şekil G.2’de gösterildiği gibi giriş verisi olarak östrojen alıcı boyaması ile boyanmış sayısal bir resmi alan ve çeşitli işlem basamaklarını uyguladıktan sonra çıkış olarak Allred puanlama sistemine göre östrojen tanısı koyan

bilgisayar tabanlı bir sistem gerçekleştirilmiştir. Bu sistemde giriş verisi, östrojen boyaması ile boyanmış biyopsi örneğinin sayısal görüntüsüdür. Bu görüntüyü elde etmek için tümörü temsil eden bölgede patolog hekim tarafından kayıt edilir.

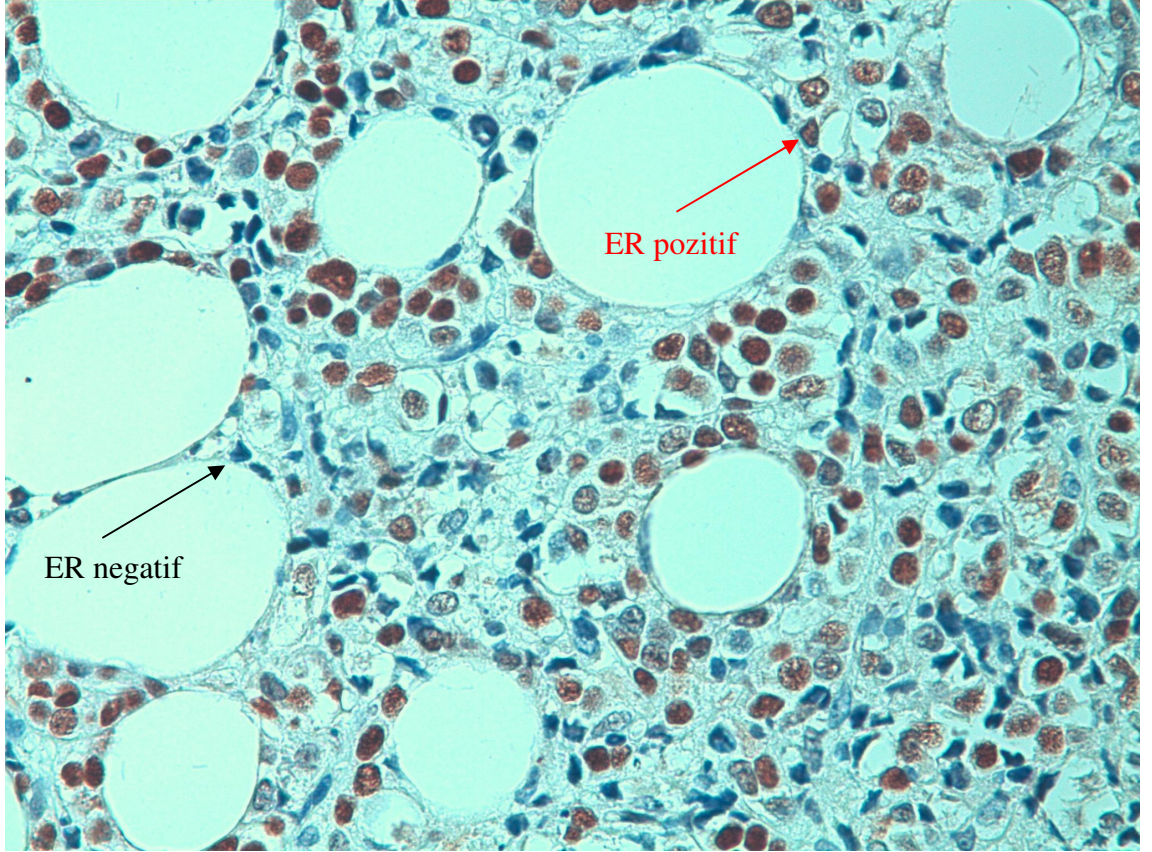


Şekil G.2. Bilgisayar tabanlı tanı sistemi işlem basamakları

Östrojen alıcı boyamasını gösterir örnek bir sayısal görüntü aşağıda verilmektedir (Şekil G.3). Bu görüntü üzerinde tez çalışması kapsamında oluşturulan bilgisayar tabanlı tanı sistemi sırasıyla bölütleme, öznitelik çıkartma, öznitelik eleme, hücre çekirdeği sınıflama ve Allred puanı hesaplama işlemlerini uygular ve sonuçta östrojen alıcı varlığına dair tanı elde edilir. Bilgisayar tabanlı tanı sistemi, östrojen alıcı boyamasında östrojen alıcısı pozitif ve negatif çekirdekleri ayırt ederken kahve, mavi veya mor tonlarda boyanma özelliğinin yanı sıra hücre çekirdeklerinin desen özniteliklerini de dikkate alır.

Görünüşlerin karar vermeye yeterli olduğu alanlarda, patolojik incelemenin tanıya ve uygun tedavi yönteminin belirlenmesine katkısı da çok büyüktür. Günümüzde, tümörlerin tanısı başta olmak üzere, pek çok hastalığın kesin tanısı için patolojik inceleme gerekli ve zorunludur. Özellikle kanser tedavisinde kanserli dokunun değişiminin belirlenmesi, hastalığın seyri ve uygulanacak tedavi şeklinin belirlenmesinde büyük önem taşımaktadır. Elde edilen resimlerde çok sayıdaki organik

yapı, şekil, renk ve desen bakımından kaynaşmış bir görüntü sergilemektedir. Mikrometre ölçeğinde çok küçük boyutlu resimlerde dahi çok sayıda hücrenin sayılması ve nicelik bakımından ayrı ayrı değerlendirilmesi zaman alıcı, yorucu ve hataya açık bir işlemdir [11].



Şekil G.3. Östrojen alıcı boyamasını gösterir örnek bir sayısal görüntü

Bir patolog aynı görüntüyü farklı zamanlarda farklı yorumlayabilmekte veya aynı görüntü için farklı hekimler farklı inceleme raporları sunabilmektedir. Kısacası değerlendirmeler soyuttur ve kişiye, zamana göre değişim gösterebilmektedir [11].

Almanya’da yapılan bir çalışmaya göre [12] meme kanserinde östrojen alıcı tanısının yüzde %24’ü yanlış belirlenmiş negatiftir. Bir başka deyişle gelen her 100 hastadan 24’ünün aslında pozitif olduğu halde yanlışlıkla negatif östrojen alıcı tanısı aldığını gösterir. Kanser tanısında yaşanan ve deneysel klinik çalışmasıyla belgelenen bu kusur, hastalar açısından hayati derecede önemlidir. Çünkü yanlış tanı yanlış tedaviyi veya yapılması gerektiği halde zamanında yapılamayan tedaviyi beraberinde

getirecektir. Sonuç olarak hastanın ya yaşam konforu ya da yaşam süresi kaçınılmaz olarak etkilenecektir.

Almanya gibi gelişmiş bir ülkenin çok sayıda laboratuvarını içine alan bu çalışmada ortaya çıkan dikkat çekici bu sonuç daha kısıtlı teşhis ve tanı olanaklarına sahip gelişmekte olan ülkelerde de beklenebilir ve belki de daha büyük bir oranda yaşanmaktadır. Gelişmiş teknolojiye ve laboratuvar olanaklarına rağmen ortaya çıkan bu kusur temelde tıbbî kararın öznel olarak verilmesinden kaynaklanmaktadır. Öznellik, tanı kararının incelemeyi yapan hekime veya aynı hekimin inceleme zamanına bağlı olarak ortaya çıkar.

Öte yandan nesnel olarak hücrelerin sınırlarının belirlenmesi, boyanmış hücrelerin tespit edilmesi ve sayılması, boyanma miktarı bakımından bu hücrelerin birbirleriyle ilişkilerinin belirlenmesi, karar verme mekanizmasında doktora katkıda bulunacak süreçleri önemli ölçüde kolaylaştıracaktır. Bu yetenekleri kısmen bile olsa karşılayacak görüntü işleme ve sınıflama tekniklerinin uygulanması hem hastalara daha doğru tanı konulmasını, hem de tedavinin daha başarılı olmasını sağlayacaktır [11].

Tümüyle bir patolog hekimin yerini alabilecek otomatik bir tanı yöntemi ve yazılımının oluşturulması yakın gelecek için gerçekçi bir hedef olmayacaktır. Fakat mevcut görüntü işleme tekniklerinin seçilen problem üzerinde deneysel olarak incelenmesi, başarıyı artıran sınıflama yöntemlerin kullanımının araştırılması, probleme özel ve kısmen hekimden destek alan yarı otomatik bir yöntemin geliştirilmesi makul bir hedef olarak benimsenebilir. Bu genel tespitler doğrultusunda tez çalışmasının amacı üç madde altında özetlenebilir:

- Yeni özniteliklerin kullanımının araştırılması ve başarımlarının deneysel olarak incelenmesi
- Yeni sınıflama algoritmalarının kullanımının araştırılması sınıflama başarımının farklı sınıflayıcılar için deneysel olarak incelenmesi
- Allred puanlama sistemine dayalı bir bilgisayar destekli karar sisteminin östrojen alıcı durumunu otomatik olarak değerlendirilmesi ve tanı verebilmesi amacıyla tasarlanması, geliştirilmesi ve başarımının deneysel olarak incelenmesi

Bu tez çalışması, Erciyes Üniversitesinin FBD-08-355 numaralı araştırma fonu ile desteklenmiştir. Yürütülen çalışmalar için Erciyes Üniversitesi Etik Kurulunun izni alınmıştır.

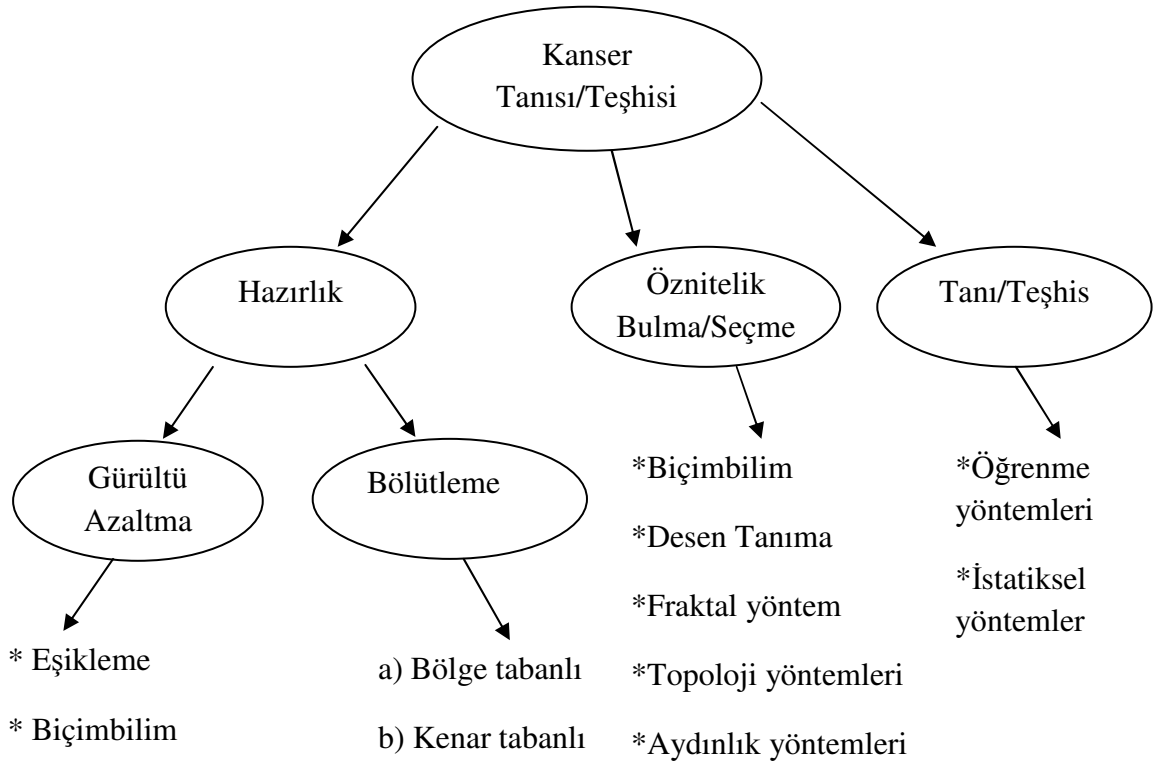
Bu tez çalışması kapsamında gerçekleştirilen uygulamalar sonucunda mühendislik yöntemlerinin meme kanseri ER durum tanısında Allred puanlama protokolüne göre nesnel sonuçlar verdiği görülmüştür. Deneysel inceleme sonuçlarına göre; dalgacık desen enerji öznitelikleri, Laws desen enerji özniteliklerinden ve birliktelik matrisi özniteliklerinden daha başarılı sınıflama kabiliyeti sağlamaktadır. Renkli özniteliklerin ayırt etme kabiliyeti gri ton özniteliklerden daha yüksektir. Tüm öznitelikler üzerinde yapılan denemelerde, sade bayes sınıflayıcı ve fonksiyonel karar ağaçlarından oluşan tümleşik bir sınıflayıcı en başarılı sınıflama ve tanı kararlarını vermiştir. Yapay öğrenme yöntemlerinin kullanılması ile elde edilen yazılım araçlarının doktorlara üçüncü bir fikir olarak destek sağlaması mümkündür.

G.3. Yöntemler ve Tezin Bölümleri

Geçmişte yapılan çalışmalara bakıldığında [13], meme kanserinde bilgisayar tabanlı bir teşhis veya tanı sisteminde Şekil G.4'te gösterilen belirli işlem basamaklarının tamamlanmış olması gerekir.

Benzer süreçler ve işlem basamakları bu tez çalışmasında izlenmiştir. Tezin bu bölümünde tezin konusu, amacı ve önemi açıklanmıştır. İkinci bölümünde, meme kanserinin sınıflandırılmasında kullanılan veriler ve inceleme amacıyla yapılan işlemler tanıtılmıştır. Bu işlemler içerisinde östrojen alıcı tanısının yeri ve tedaviye etkisi belirtilmiştir. Daha sonra, östrojen alıcı tanısı hakkında literatürde yayınlanan çalışmalar incelenmiş ve bu tez çalışmanın yeri anlatılmıştır.

Üçüncü bölümde, mikroskopla elde edilen sayısal resimler üzerinde uygulanan görüntü işleme yöntemleri, bölütleme ve biçimbilim işlemleri kısaca tanıtılmıştır. Bölütleme yöntemi olarak Otsu eşikleme uygulanmış ve biçimbilim işlemcileri içerisinde standart açma ve kapama işlemcileri kullanılmıştır.



Şekil G.4. Bilgisayar tabanlı kanser tanısı/teşhisi aşamaları [13]

Dördüncü bölümde özniteliklerin üretilmesi üzerinde durulmuştur. Pozitif ve negatif boyalı hücre çekirdeklerinin birbirlerinden ayrılmasında sınıflayıcının giriş verisi olarak kullanılacak olan öznitelikler ve bunların hesaplanması ele alınmıştır. Birinci ve ikinci dereceden istatiksel özniteliklerin yanı sıra dalgacık dönüşümü gibi frekans uzayında üretilen desen öznitelikleri de kullanılmıştır. Birinci dereceden öznitelikler arasında ortalama aydınlık, ortalama karşıtlık, yumuşaklık, üçüncü moment, düzgünlük ve düzensizlik ölçütlerinden faydalanılmıştır. İkinci derece öznitelikler olarak eş-oluşum matrisi öznitelikleri, Laws ve dalgacık örüntü enerji öznitelikleri kullanılmıştır. Bu özniteliklerin sınıflayıcı başarımı açısından bilgi değeri, renkli ve renksiz resimler üzerinde ayrı ayrı incelenmiştir. Ayrıca bu bölümde, kullanılan boyut azaltma yöntemi açıklanmıştır. Elde edilen çok sayıda öznitelik sınıflayıcının giriş veri uzayını çok büyötmektedir. Bu nedenle hem sınıflayıcı başarımı azalmakta hem de artan işlem karmaşası nedeniyle işlem zamanı artmaktadır. Bu bölümde başarımı düşürmeden giriş veri boyutunu azaltacak ilinti tabanlı bir boyut azaltma yöntemi tanıtılmıştır.

Beşinci bölümde, tez kapsamında oluşturulan tanı sisteminin sınıflama modeli, başvuru sınıflama ölçütleri ve yöntemleri tanıtılmıştır. Bu yöntemler; yarıçap fonksiyonlu sinir ağları, k -en yakın komşuluk algoritması, k -orta nokta öbekleme yöntemi, destek vektör makineleri, sade bayes sınıflayıcı, fonksiyonel karar ağaçlarıdır. Ayrıca bu bölümde sınıflama başarımını değerlendirmek için başvuru ölçütler ve geçerleme düzeni kısaca tanıtılmıştır.

Altıncı bölümde, bilgisayar tabanlı östrojen alıcısı durum tespiti yapan ve tanı kararı otomatik olarak veren bir tanı sistemi sunulmuştur.. Bu sistem içerisinde, fonksiyonel karar ağaçlarına dayanan karar modelinin sınıflama ve tanı başarımı incelenmiştir.

Yedinci bölümde; yarıçap fonksiyonlu sinir ağları, k -en yakın komşuluk algoritması, k -orta nokta öbekleme yöntemi, sade bayes sınıflayıcı, fonksiyonel karar ağaçları, destek vektör makineleri ve tümleşik bir sınıflayıcı modelin östrojen alıcı durum tanısı verme uygulamasında başarımları deneysel ve karşılaştırmalı olarak incelenmiştir. Çeşitli sınıflayıcıların yanı sıra aynı zamanda bu sınıflayıcıların çeşitli öznelilik kümeleri üzerinde sınıflama ve tanı başarımları bu bölümde ayrıca deneysel olarak incelenmiştir. Bu tez çalışması kapsamında önerilen dalgacık dönüşümü öznelilik türünün literatürde geçmişte önerilen öznelilik türlerine kıyasla başarımı bu bölümde açıklanmıştır.

Son olarak sekizinci bölümde bu tez çalışmasının çıkış noktası olan savlar ve savların geçerliliği değerlendirilmiştir.

G.4. Terimler Sözlüğü

Tezinde içerisinde adı geçen fakat henüz Türkçede yaygın kullanılmadığı düşünülen bazı Türkçe terimlerin anlamlarına açıklık getirmek amacıyla İngilizce karşılıkları parantez içinde aşağıda verilmiştir (Tablo G.1).

Tablo G.1. Türkçe terimler ve İngilizce karşılıkları

ağırlık fonksiyonunu (kernel function)
alıcı işletim karakteristik eğrisi (receiver operating characteristics curve, ROC curve)
aydınlık (intensity)
bağlı bileşen (connected component)
benzemezlik (dissimilarity)
biçimbilim (morphology)

Tablo G.1. (devamı)

biçimsel açma ve kapama (opening and closing)
biçimsel aşındırma-erozyon (erosion)
biçimsel genişletme (dilation)
bilgisayarlı kesityazar taraması (computed tomography scan, CT, CAT)
birleşik saçılım (composite variance)
birliktelik veya eş-oluşum matrisi (co-occurrence matrix)
boyut azaltma (dimension reduction)
bölütleme (segmentation)
büyüyen bölge (region growing)
çapraz geçerleme (crossvalidation)
çok biçimli çekirdek yapısı (nuclear pleomorphism)
çok değişkenli ağaçlar (multivariate trees)
çok katmanlı algaç ağı (multilayer perceptron network, MLP)
dalgacık dönüşümü (wavelet transform)
destek vektör makineleri (support vector machines, SVM)
duyarlılık (sensitivity)
düzensizlik, belirsizlik (entropy)
düzgünlük, enerji (uniformity)
evrilim (convolution)
fonksiyonel ağaçlar (functional trees, FT)
gri ton değeri (gray level value)
hata matrisi (confusion matrix)
hormon alıcı durumu (hormone receptor status)
iç-ilinti (auto-correlation)
ikilik resim (binary image)
ileri doğru seçme araması (forward selection search)
ilinti (correlation)
immünohistokimyasal (immunohistochemical, IHC)
ince iğne uçlu biyopsi (fine needle aspiration biopsy)
ince zar (membrane)
kalın iğne uçlu biyopsi (core needle biopsy)
karar ağacı budama (decision tree pruning)
karşıtlık (contrast)
kemik taraması (bone scan)
k-en yakın komşuluk sınıflayıcısı (k nearest neighborhood classifier)
k-orta nokta öbekleme (k-means clustering, KM)
k-parçalı çapraz geçerleme (k-fold cross validation)
magnetik eşitresim görüntüsü (magnetic resonance imaging, MRI)
mamografi, meme röntgeni (mammography)
model ağaçları (model trees)
nöron iletim fonksiyonu (neuron activation function)
olasılık tabanlı yapay sinir ağları (probabilistic neural networks)
olasılık yoğunluk fonksiyonu (probability density function)
öbekleme (clustering)
östrojen alıcı (estrogen receptor, ER)
özgüllük (specificity)

Tablo G.1. (devamı)

öznitelik elde etme (feature extracting)
öznitelik seçme, eleme (feature elimination)
pozitron salımlı kesityazar taraması (positron emission tomography scan, PET)
saçılma, saçılım, varyans (variance)
sade bayes sınıflayıcısı (naive bayes classifier)
sıfır hipotezi (null hypothesis)
sıklık dağılımı (histogram)
sınıflama (classification)
sınıflayıcı (classifier)
sınıflayıcı birleşimi (ensemble classifier)
sutaşımı bölütleme (watershed segmentation)
süt bezi kanseri (lobular carcinoma)
süt kanalı kanseri (ductal carcinoma)
süt kanalının ve bezinin biçimsel yapısı (tubule and glandular structural formation)
spearman'ın sıralı ilinti katsayısı (spearman's rank order correlation coefficient)
tekdüzelik (homogeneity)
tohum (seed)
tüm olasılıkları arayan (exhaustive)
tümör derecelendirme (tumor grading)
tümör farklılaşması iyi (tumor well differentiated)
tümör farklılaşması kötü (tumor poorly differentiated)
tümör farklılaşması orta (tumor moderately differentiated)
tümör puanlama (tumor scoring)
ultrason, ses ötesi (ultrasound)
üstdüzlem (hyperplane)
üstküp (hypercube)
x-ışınları (x-rays)
yarıçap taban fonksiyonlu ağlar (radial basis functions networks, RBFN)
yayılmacı (invasive)
yayılmacı olmayan (noninvasive)
yayılmacı süt bezi kanseri (invasive lobular carcinoma, ILC)
yayılmacı süt kanal kanseri (invasive ductal carcinoma, IDC)
yerinde, yerleşik (in situ)

1. BÖLÜM

GENEL BİLGİLER VE ÖSTROJEN ALICI TANISI

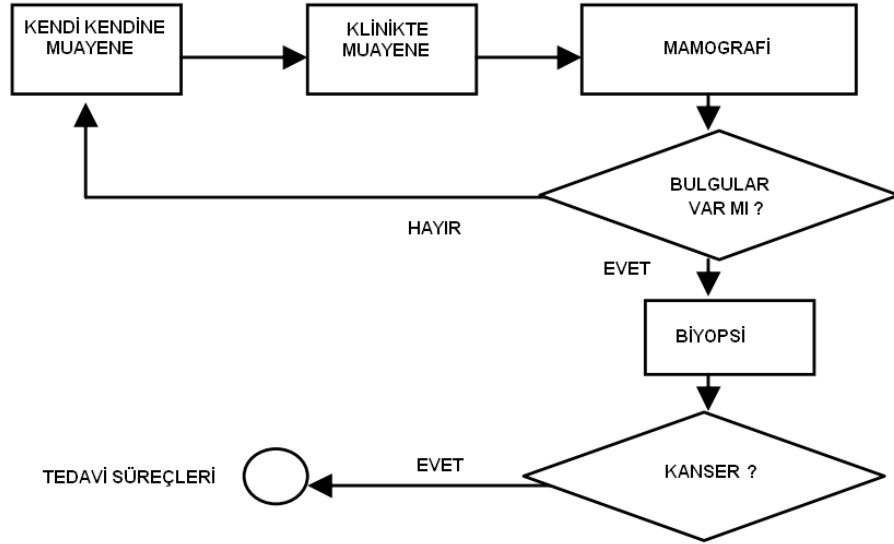
1.1. Meme Kanserinin Sınıflandırılması

Bayanlarda meme kanserinin teşhisinde ilk aşamada hastanın kendi kendini muayene etmesi ile başlar. Hasta, meme dokusu içerisinde sert bir kitle, ciltte içine çökme veya şişme, meme ucunda akıntı vb. belirtiler tespit ettiğinde doktor incelemesinde klinik muayenesini yaptırır. Gelişmiş ülkelerde hasta kendi kendisine bir belirti tespit etmese bile 40 yaşını geçen bayanlarda her yıl bir kez düzenli olarak klinikte doktorun elle muayenesi ve mamografi testinin yapılması önerilmektedir [14]. Doktor elle muayene ve mamografi sonucunu değerlendirir, kanseri çağrıştıran bulgular tespit etmişse hastadan biyopsi alınmasını ister. Şekil 1.1’de gösterildiği üzere, biyopsi incelemesinden sonra kesin olarak meme kanseri teşhisi konulabilir.

Doktor mamografi dışında çeşitli görüntüleme tekniklerini kullanarak meme dokusu üzerinde inceleme yapabilir. Bu teknikler, ultrason, x-ışınları, bilgisayarlı kesityazar taraması (CT, CAT), magnetik eşitleşim görüntülemesi (MRI), pozitron salımlı kesityazar taraması (PET), kemik taraması olarak listelenebilir. Bu görüntüleme yöntemlerinden biri veya bir kaç kullanılarak meme üzerinde Şekil 1.2’de temsili olarak gösterildiği gibi şüpheli bir kanser bölgesi belirlenir [16].

Doktor bu bölge üzerinden bir doku örneği alır. Bu işleme biyopsi adı verilir. En yaygın kullanılan biyopsi türleri ince iğne uçlu ve kalın iğne uçlu biyopsidir. İnce iğne uçlu biyopside içi oyuk çok küçük bir iğne meme dokusu içerisine batırılır. Küçük bir parça hücre örneği alınarak mikroskop altında incelenir ve işlem yara izi bırakmaz. Kalın iğne uçlu biyopside daha büyük bir iğne kullanılarak birkaç farklı doku örneği şüpheli

bölgeden alınır. Küçük bir yara izi birkaç haftalığına kalır. Bu işlem sonrasında hormon ve protein alıcı testleri için yeterince doku örneği alınmış olur.



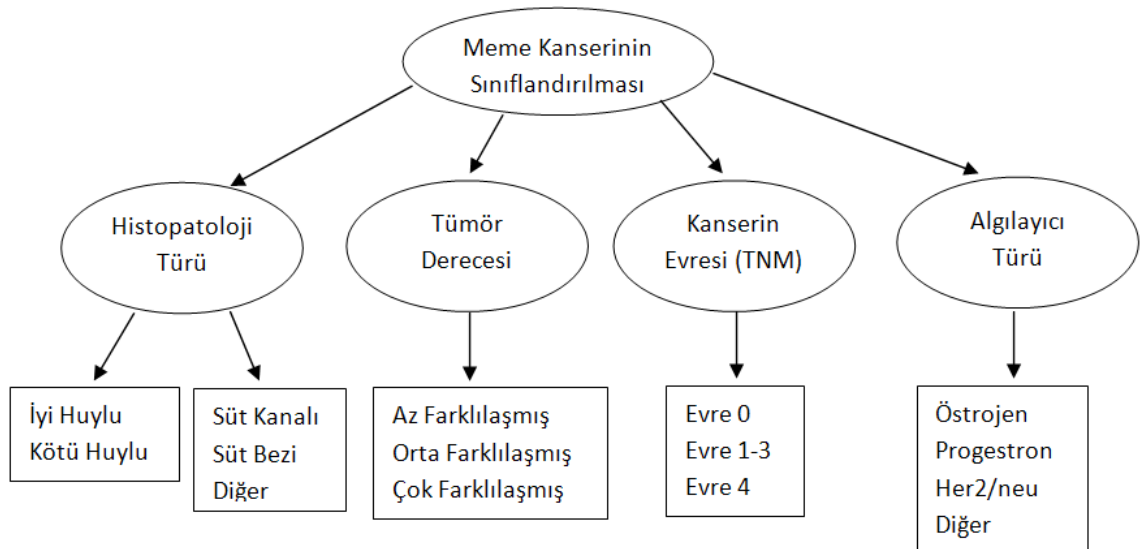
Şekil 1.1. Genel olarak meme kanser teşhis işlemleri [15].



Şekil 1.2. Meme kanserinde tümör [16]

Doktor bu bölge üzerinden bir doku örneği alır. Bu işleme biyopsi adı verilir. En yaygın kullanılan biyopsi türleri ince iğne uçlu ve kalın iğne uçlu biyopsidir. İnce iğne uçlu biyopside içi oyuk çok küçük bir iğne meme dokusu içerisine batırılır. Küçük bir parça hücre örneği alınarak mikroskop altında incelenir ve işlem yara izi bırakmaz. Kalın iğne uçlu biyopside daha büyük bir iğne kullanılarak birkaç farklı doku örneği şüpheli bölgeden alınır. Küçük bir yara izi birkaç haftalığına kalır. Bu işlem sonrasında hormon ve protein alıcı testleri için yeterince doku örneği alınmış olur.

Şüpheli bölgeden alınan doku örneği eğer kanser gösterirse, elde edilen sonuçlar Şekil 1.3'te gösterilen konu başlıklarında kanserin türü ve tedavi seçenekleri hakkında bilgi verir.



Şekil 1.3. Kanser sınıflandırılması

Genel olarak biyopsi sonuçlarından aşağıdaki bilgileri elde etmek mümkündür [17].

- Kanser yayılmacı olup olmadığı,
- Kanser derecesi; kanserli hücrelerin ne oranda sağlıklı hücrelere benzediği, başka bir deyişle tümör derecesi,

- Kanserin hormonsal tedavilere duyarlı olup olmadığı, kanserli hücrelerin çok sayıda Her-2 büyüme proteinleri içerip içermediği.

Bu sonuçların yanı sıra, kanserli doku hücrelerinin genetik özelliklerini ve moleküler yapısını inceleyen ileri düzey laboratuvar testleri de yapılmaktadır.

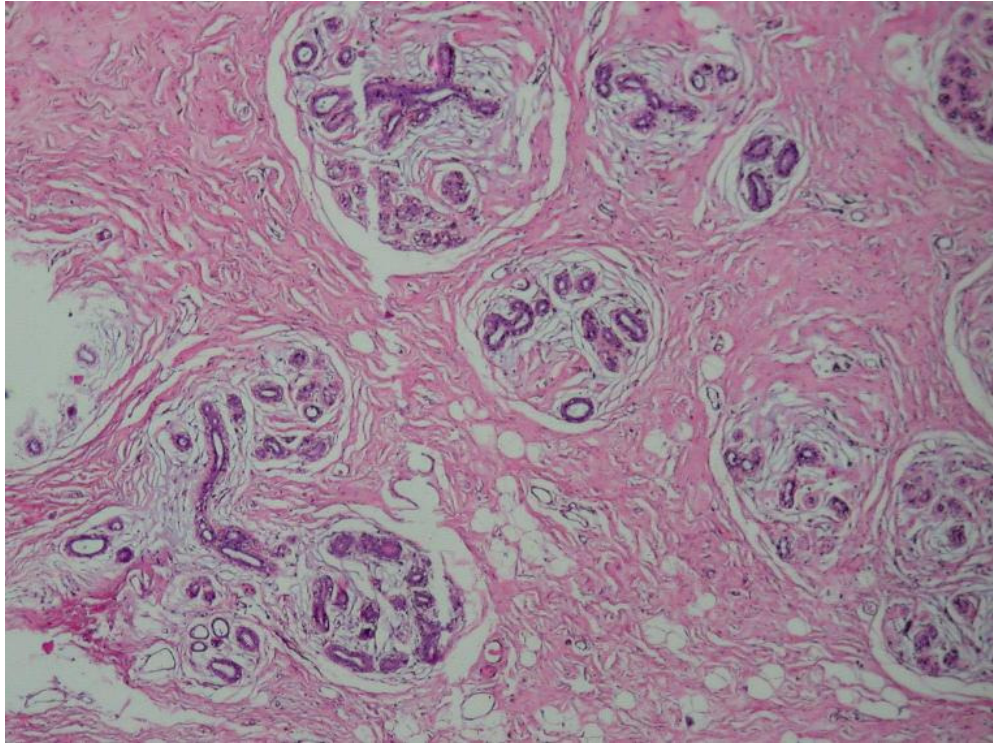
Bu bölümde meme kanseri hakkında genel bilgiler vermek amaçlanmıştır. Genel bilgiler, Şekil 1.3'te verilen şemaya göre sırayla verilmektedir. Bu bölümde ayrıca, tanı amaçlı yapılan incelemeler arasında östrojen alıcı durum tanısının yeri gösterilmektedir. Bölüm sonunda, östrojen alıcı tanısında kullanılan Allred puanlama sistemi tanıtılmakta, otomatik östrojen tanısına yönelik literatür taraması ve bu tez çalışmasının önceki çalışmalara göre farklılığı açıklanmaktadır.

1.2. Doku Yapısında Meme Kanserinin Yaygın Türleri

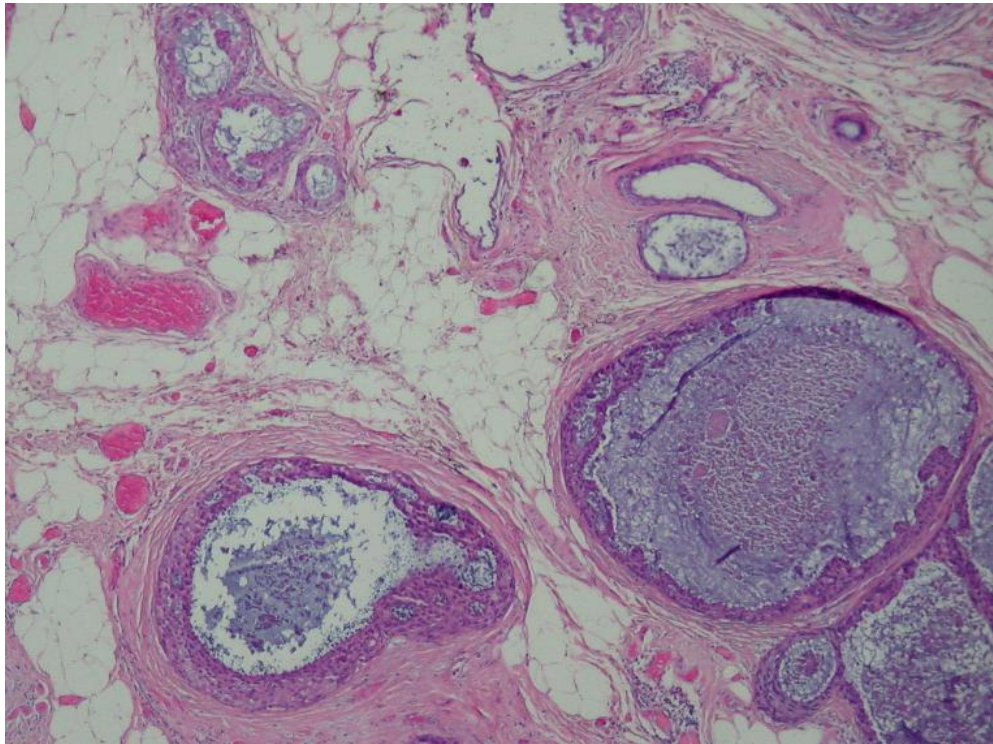
Meme kanserinin en yaygın türleri; ya süt taşıyan kanallarda oluşan süt kanalı kanseri ya da süt bezlerinde oluşan süt bezi kanseri olarak bilinir. Kanserin meme dokusu üzerinde nerede olduğu mikroskop incelemesi ile belirlenir. Şekil 1.4'te, 40x büyütmeli ışın mikroskobu altında Hemotoksilin Eosin (HE) boyanmasında normal bir meme dokusuna örnek görüntü verilmiştir [18]. Meme kanserinin adlandırılması yapılırken kanserin nerede olduğunun yanı sıra kanserli hücrenin davranış biçimine de bakılır.

1.2.1. Yayılmacı Olmayan

Latince'de yerinde manasına gelen (in situ) kelimesiyle de tanımlanan bu davranış biçiminde, kanserli hücreler oluştukları yerde kalıcıdır, süt bezleri yahut kanalları dışındaki meme dokusuna yayılmazlar. Yayılmacı olmayan meme kanseri türleri içerisinde en çok görüleni yayılmacı olmayan süt kanalı kanseridir (DCIS). Bu kanser türünde, kanser hücreleri süt kanallarının iç duvar yüzeyinde hapis kalırlar ve dışarı yayılmazlar (Şekil 1.5). Uygun bir tedavi yöntemiyle bu kanser türünde başarılı sonuçlar alınabilmektedir.



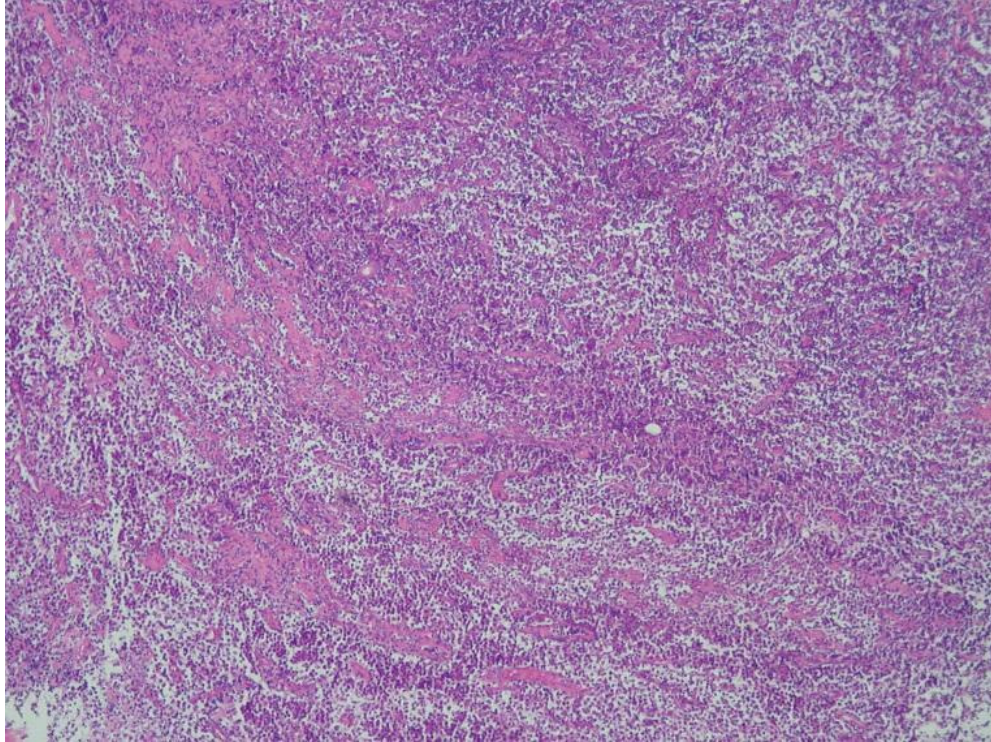
Şekil 1.4. HE boyamasında normal bir meme dokusunun görünümü [18]



Şekil 1.5. HE boyamasında DCIS meme kanseri doku görüntüsü [18].

1.2.2. Yayılmacı

Bu tür kanserde, kanserli hücreler süt kanalını veya bezini diğer meme dokusundan ayıran ince zarın dışına geçerek meme dokusu içerisinde yayılırlar (Şekil 1.6). Kanser hücreleri vücudun diğer kesimlerine, örneğin lenf düğümlerine, geçebilirler. Yaygın iki türü bulunmaktadır [19].



Şekil 1.6. HE boyamasında IDC meme kanseriyle bozulmuş doku [18].

Yayılmacı Süt Kanalı Kanseri (IDC): Meme kanseri vakalarının %70'i bu kanser türündendir. Kanser hücreleri süt kanalının iç yüzey duvarlarında oluşur, süt kanalı duvarı içerisinden yakındaki meme dokusuna geçer. Bu aşamadan sonra kanser hücreleri belirli bölgede kalabilir veya kan dolaşım sistemi ve lenf sistemi vasıtasıyla vücudun diğer bölgelerine yayılabilirler.

Yayılmacı Süt Bezi Kanseri (ILC): Her ne kadar IDC den daha az rastlansa da, bu tür kanser süt bezlerinde başlar ve benzer biçimde çevreleyen meme dokusu içerisine yayılır. ILC vücudun daha uzak bölgelerine de yayılabilir. Bu kanser türünde ayırık sert bir parça hissi yerine meme üzerinde bir bölgenin veya tamamının katılaşması hissi

vardır. Bu yaygın türlerin yanı sıra nadir rastlanan farklı bazı meme kanseri türleri de vardır.

1.3. Tümör Derecesi

Eğer kanser yayılmacı ise patolojik kanseri derecelendirir. Bu derecelendirme örnek doku parçasındaki hücrelerin ne oranda sağlıklı doku hücrelerine benzediğine bağlıdır. Kanser türü ve düzeyi tedavi seçeneklerinin belirlenmesine yardımcı olur. Kanser hücreleri 1-3 arası bir ölçekle derecelendirir [17].

Derecelendirme aşağıdaki hususların değerlendirilmesine ve puanlanmasına bağlıdır [20].

1-) Süt kanallarının ve bezlerinin biçimsel yapısı

- Benzeme oranı $< \%10$ ise 3 puan
- Benzeme oranı $\%10-75$ ise 2 puan
- Benzeme oranı $> \%75$ ise 1 puan

Süt kanallarının ve bezlerinin yapısı sağlıklı meme dokusuna ne kadar çok benzer ise inceleme sonrasında verilen puan ters orantılı olarak azalacaktır. Biçimsel bozulma arttıkça daha çok puan verilecektir.

2-) Çok biçimli çekirdek yapısı

- Küçük ve düzgün dağılımlı hücreler 1 puan
- Hücre boyutunda ve biçimsel farklılaşmasında hafif veya orta ise artış 2 puan
- Hücre boyutunda ve biçimsel farklılaşmasında göze çarpan, belirgin artış 3 puan

3-) Mitoz (bölünen) hücre sayısı: x40 mikroskop büyütmesi altında incelenir. Puanlama 1-3 arasında değişir.

- 11 adet bölünme için 1 puan verilir.
- 11-21 adet bölünme için 2 puan verilir.
- 22 ve üzeri bölünme sayıldığında en fazla 3 puan verilir.

Yukarıda verilen üç unsurun değerlendirilmesi sonucunda toplam puan hesaplanır. Tümörün derecesi toplam puana göre belirlenir.

- Derece-1 (Toplam Puan 3 ve 5 arasında): Hücreler oldukça normal görünümlüdür, farklılaşma azdır veya iyidir denilebilir.
- Derece-2 (Toplam Puan 6 ve 7 arasında): Hücreler normal dışı görünümündedir. Orta düzey farklılaşma vardır.
- Derece-3 (Toplam Puan 8 ve 9 arasında): Hücreler olması gereken yapı ve fonksiyonlarını yitirmişlerdir. Kötü ve sağlıklı bir farklılaşma oluşmuştur.

Patolog, derecelendirme yaparken hücrenin ve hücre çekirdeğinin şekline, yapısına bakar, ne kadar hücrenin bölünme sürecinde olduğunu sayar. Dikkat edilirse önce puanlama yapılır ve bu işlem sonrasında elde edilen sonuçlara göre tümör derecelendirme yapılmaktadır. Kanserin derecesi büyüdükçe, daha hızlı yayılan ve büyüyen bir hastalık söz konusu olur.

Tümör derecelendirmesinde çeşitli yöntemler kullanılmaktadır. Bloom-Richardson [21] ve Elston-Ellis modifikasyon yöntemleri [22] bu konuda örnek olarak verilebilir.

1.4. Meme Kanserinin Evreleri

Meme kanserine dair patoloji raporunda kanserin evresi hakkında bilgi bulunmaktadır. Evre, kanserin ne kadar yayıldığıнын ölçüsüdür. Kanserin, meme içerisinde sağlıklı doku kesimlerine ve diğer organlara yayılıp yayılmadığına bakılır. Evreyi belirlerken tümörün boyutu, kanserin yayılmacı olup olmadığı, lenf düğümlerinde olup olmadığı ve vücudun diğer organlarına sıçrayıp sıçramadığı araştırılır. Kanserin evrelendirilmesinde bazı ülkelerde bu bölümde verilenden farklı olarak TNM evrelendirme sistemi kullanılmaktadır [23].

Evre 0: Aynı zamanda “in-situ” olarak da adlandırılır. Bu kanserler yerlerinde kalmış ve çevre dokulara sıçramamış kanserlerdir

Evre 1: Tümör 20 mm. ve daha küçüktür. Bu durumda kanser lenf bezlerine sıçramamıştır. Tedavide meme koruyucu yöntemle lenf bezlerinin alınmasından sonra radyasyon tedavisi uygulanır. Desteklemek için kemoterapi ve/veya hormonoterapi eklenir. Bir diğer uygulama da mastektomi yöntemidir. Bu yöntemde kanserli meme alınarak koltuk altı lenf bezleri çıkarılır.

Evre 2-A: Tümörün 20 - 50 mm. arasında olup, lenf bezlerine sıçramamış halidir.

Evre 2-B: Tümörün 2-A evresindeki gibi bir boyut aralığında olup (50 mm. den büyük olabilir), koltuk altı lenf bezlerine sıçramamış halidir.

Evre 2: Evre 1 ile aynı tedavi yöntemi uygulanmakla birlikte, eğer tümör aşırı büyümüş ya da lenf bezlerine sıçramışsa kemoterapi, hormonoterapi ve radyasyon tedavisi tamamlayıcı olarak önerilir.

Evre 3-A: Tümör koltuk altı lenf bezlerine ve göğüs dışı dokulara sıçramış durumdadır ve bu halde mastektomi yöntemiyle tedavi yoluna gidilir. Cerrahi müdahaleden sonra kemoterapi ve hormon tedavisi uygulanır.

Evre 3-B: Bu aşamada tümörün boyutu dikkate alınmaz; tümör göğüs duvarına bağlıdır ve lenf bezlerine sıçramıştır. "Neoadjuvant" adı verilen tümörün boyunun küçültülmesi

amaçlı kemoterapi uygulanmasının ardından tümörün boyunun küçülmesinden sonra “lampektomi” veya “mastektomi” yapılır.

Evre 4: Bu aşamada kanser göğüs dışındaki vücut bölümlerine yayılmıştır. Bu evre tedavisinde hastanın yaşam süresini artırmak ve yaşam kalitesini yüksek düzeyde tutmak hedeflenir. Kemoterapi ve hormonoterapi yapılır. Hasta şikâyetlerine bağlı olarak mastektomi de uygulanabilir [24, 20, 25].

1.5. Kanserli Hücrelerin Gen Yapısı

Kanserli hücrelerin çok sayıda Her-2 genine sahip olduğunun bilinmesi tedavi kararını etkileyen bir bilgidir. Bu gen fazla olması halinde Her-2 proteinin üretimi de fazla olur. Bu protein hücre büyümesini tetikler ve hızlandırır. Meme kanseri türlerinin yaklaşık %20’si Her-2 pozitifdir. Her-2 pozitif kanser hücreleri daha çok büyürler ve daha saldırganca yayılırlar. Genellikle ileri düzey iki tür laboratuvar testi kanser hücrelerinde Her-2 protein varlığını tespit eder [26].

İmmünohistokimya Testi: Her-2 proteinine bağlanabilen kimyasallar doku örneğine uygulanır, çok sayıda Her-2 proteinine sahip olan hücreler renk değiştirir.

FISH (Florescent In Situ Hybridizaiton) Testi: DNA’ya ait “Florescent” parçaları Her-2 genine ait fazladan kopyaları tespit eder. Benzer bir yöntem “chromogenic in situ hybridization (CISH)” olarak bilinir.

Her-2 pozitif kanser hücreleri “Trastuzamab (Herceptin)” veya “Lapatinib (Tykerb)” adı verilen ilaçlarla tedavi edilir [26].

1.6. Hormon Alıcı Durumu

Sağlıklı meme dokusu hücreleri ve bazı kanser hücreleri dişilik hormonları olan östrojen ve progesteron hormonlarına tutunabilen alıcılara sahiptir. Meme dokusu üzerinde tümör hücreleri östrojen ve progesteron hormonlarına ait alıcıların var olup olmadığı belirlemek üzere test edilirler. Alıcı, hücre dışında bulunan bir proteindir. Bu protein

kan dolaşım sisteminde bulunan ilaçlara hormonlara ve özel kimyasallara tutunabilme özelliğine sahiptir.

Bu dişilik hormonları hücre büyümesini arttıran yahut açıp kapatan sinyallerdir. Kanser hücreleri hormon negatif (HR-) veya hormon pozitif (HR+) olabilir. HR+ olduğu belirlenen hücreler şu biçimde sınıflandırılabilir; östrojen alıcı pozitif (ER+) veya progesteron alıcı pozitif (PR+). HR+ hücreler tamoksifen adı verilen bir ilaçla hormon engellemesi yapmak suretiyle durdurulabilir ve kanserin büyümesi yavaşlatılabilir. HR+ kanser hücreleri HR- kanser hücrelerinden daha yavaş büyürler [27].

1.6.1. Östrojen Hormonu Alıcı Durumu Puanlaması

Meme kanserinin tedavisinde en önemli tanı etkenlerinden birisi tümörde östrojen alıcısının var olması durumudur [2, 3]. İmünohistokimyasal boyama yöntemi, biyopsi kesitinde östrojen alıcı varlığını değerlendirmek için kullanılacak ucuz ve kolay bir yöntemidir [8]. Bu tez çalışmasında östrojen alıcı tanısı amacıyla uygulanan çeşitli puanlama yöntemleri [4-6] arasında Allred puanlama yöntemi [7] tercih edilmiştir. Bu puanlama sistemi kullanım kolaylığının yanı sıra düşük negatif boyamaları tespit etme ve böylece yanlış negatif tanıları azaltma özelliğine sahiptir [5, 9].

Bu protokole göre toplam puanlamayı belirleyen iki tür puanlama bulunmaktadır. Bunlardan birincisi yüzdeler puanı (YP),

$$yüzdelikoranı = \frac{N_p}{N_p + N_n} \quad (1.1)$$

oranına göre hesaplanır (Denklem 1.1.). Burada N_p ve N_n pozitif ve negatif boyalı çekirdek sayısıdır. YP puanı, Allred puanlama tablosu (Tablo 1.1) ile hesaplanan yüzdeler oranı karşılaştırılarak belirlenir.

İkinci puan türü aydınlık puanıdır (AP). Bu puan pozitif boyalı hücre çekirdeğinin boyanma şiddetine göre 0 ile 3 arasında değişen değerlerde olabilir. Tümüyle negatif boyamanın olduğu bir kesit için aydınlık puanı 0 olur. Pozitif boyanmanın var olduğu

hallerde zayıf, orta ve güçlü boyama için sırasıyla 1,2 ve 3 puanı verilir. Toplam puan (TP), Allred puan tablosunda gösterildiği gibi (Tablo 1.1) YP ve AP puanlarının toplanması ile elde edilir.

Tablo 1.1. Allred puanı, Toplam Puan (TP) = YP + AP (değerler 0, 2–8)

Boyanma yüzdesi puanı (YP)	Pozitif boyalı çekirdek oranı	Aydınlık puanı (AP)	Pozitif boyalı çekirdeklerin ortalama aydınlık şiddeti
0	Yok	0	Hiç
1	< 1/100	1	Zayıf
2	1/100 ile 1/10	2	Orta
3	1/10 ile 1/3	3	Güçlü
4	1/3 ile 2/3		
5	>2/3		

1.6.2. Östrojen Alıcı Tanısında Literatür Taraması

Bilgisayar destekli olarak kötü huylu süt kanalı meme kanserinde geçmişte yapılan çalışmalara bakıldığında iki ana gruba ayırmak mümkündür. Bunlardan bir kısmı sitolojik incelemeler üzerine özelleşir. Sitoloji, tek bir hücreyi ve hücre grubunu inceleyen bilim dalıdır [1]. Sitolojik inceleme yöntemlerinde incelenecek çekirdeği seçmek için doku tahrip edilir. İkinci grup çalışmalar histolojik incelemelere dayanır. Histolojik yöntem kanseri doku düzeyinde inceler, doku mimarisine ve doku içerisinde hücre özelliklerine bakar [1]. Bu tez çalışmasında uygulanan bilgisayar tabanlı yaklaşım histolojik görüntülere üzerinde gerçekleştirilmiştir.

Meme kanseri hakkında yayınlanmış olan bilgisayar destekli tanı çalışmaları hücre düzeyinde ve doku düzeyinde olmak üzere ikiye ayrılabilir. Sitolojik görüntüler genellikle tek bir hücre üzerinde elips yaklaştırması vb. parametrik yöntemlerle hücre bölütlemesine odaklanır. Bu tür görüntüler üzerinde bilgisayar destekli çalışmalar etkileşimli-yarı otomatik veya etkileşimsiz-otomatik yöntemlerdir [28-31].

Doku üzerinde bilgisayar destekli karar sistemlerinin geliştirilmesine yönelik çalışmalar daha yaygındır [13]. Histokimyasal görüntülerde doku mimarisinin görülebilir olması, bu yöntemi daha çok tercih edilen ve altın standart olarak kabul edilen rutin yöntemler haline getirmektedir [11]. Bu kapsamda çeşitli kanser türleri üzerinde çok sayıda çalışma yayınlanmıştır [13]. Tez çalışmasının konusu, yayılmacı meme kanserinde östrojen alıcı tanısına odaklı olduğundan burada sadece östrojen alıcı tanısına yönelik çalışmalar belirtilecektir.

Histolojik incelemenin bilgisayar destekli olarak yapıldığı çalışmalar kullanılan puanlama yöntemine göre ve bilgisayar tabanlı analiz yaklaşımına göre sınıflandırılabilir. Puanlama yöntemi ele alındığında, doku resminde hücre sayan ve Allred puanlama yöntemine göre östrojen alıcı tanısını bilgisayar tabanlı olarak gerçekleştirmiş iki çalışma bulunmaktadır [32, 33]. Öte yandan, doku resminde hücre tespiti yapan ve Allred puanlamasında farklı puanlama yordamı kullanan bilgisayar tabanlı tanı sistemleri de geçmişte yayınlanmıştır [34-37].

Bazı bilgisayar tabanlı yaklaşımlarda doku resmi incelenmesine karşın hücre tespiti ve sayımı yapılmamış bunu yerine histolojik resim üzerinde piksel veya resim alanı tabanlı bir inceleme gerçekleştirilmiştir [38-47, 48-50].

Geçmişte yayınlanan bu çalışmalar içerisinde bulunan bazı ticari sistemler; SAMBA 200 [39], CAS 200 [35], WinROOF [43] ve Photoshop of Adobe Systems [41] diğerlerinden ayrılmaktadır. Çünkü bu çalışmalar görüntü üzerinde yapılan bilgisayar tabanlı analizi biyokimyasal veya sitometrik inceleme sonuçları ile karşılaştırmaktadır.

1.6.3. Tez Çalışmasının Farklılığı

Yukarıda belirtilen piksel sayısı veya alan ölçümüne veya resim üzerinde genel aydınlık seviyesi eşiklemeye dayanan yöntemler gerçek tıbbi uygulamadan farklıdır. Gerçek tıbbi yaklaşımda tanı konulması amacıyla hücre çekirdeğinin tasnifi ve sayımı yapılmaktadır [7]. Dolayısıyla, bu tez çalışmasında hücre çekirdeği tespiti ve sayımı gerçekleştirilmiş ve tanıyı belirleyen karar süreci buna göre tasarlanmıştır.

Hücre çekirdeğinin otomatik olarak sınıflandırılması amacıyla çok sayıda öznitelik üretilmiştir. Boyut azaltma yöntemi kullanılarak karar probleminin veri uzayı küçültülmüş ve sınıflama başarımı arttırılmıştır. Geçmişte bu problem üzerinde yapılan diğer çalışmalardan farklı olarak çeşitli örüntü ve istatistiksel resim özniteliklerinin yanı sıra dalgacık dönüşümü adı verilen frekans uzayı öznitelikleri de araştırılmış ve sınıflama amacıyla kullanılmıştır. Bu tez çalışmasında elde edilen deneysel çalışma sonuçlarına göre, dalgacık dönüşümü öznitelikleri literatürde önerilen diğer özniteliklerden daha başarılı sonuçlar vermektedir.

Tanı amacıyla Allred puanlama sisteminin kullanılmış olması bu çalışmayı önceki çalışmalardan ayıran bir başka özelliştir. Son olarak bu tez çalışması kapsamında literatürde önerilen çeşitli makine öğrenmesi yöntemlerinin sınıflama ve genelleme başarımları ilk kez aynı veri ve test kümesi üzerinde incelenmiştir.

2. BÖLÜM

GÖRÜNTÜ İŞLEME YÖNTEMLERİ

2.1. Bölütleme Yöntemleri

2.1.1 Büyüyen Bölge Bölütlemesi

Adından da anlaşılacağı üzere büyüyen bölge yöntemi, pikselleri veya alt resim bölgelerini önceden tanımlı bir ölçüye göre birleştirerek genişleten bir yordamdır. Bu yöntem tohum olarak nitelendirilen noktalardan başlar. Her tohum tanımlanmış ölçü uyarınca kendisiyle benzerlik gösteren komşu piksellerle genişler ve bölgelere dönüşür. Tanımlanan ölçü bir gri ton değeri veya bir renk aralığı olabilir.

Tek bir tohumla veya çok sayıda tohumla yordamın başlayacak olması probleme bağlıdır. Örneğin uydu görüntülerinin analizi büyük ölçüde renklere bağlıdır. Eğer problem hakkında başlangıçta bir bilgi yoksa büyüme süreci içerisinde, yordam her pikselde büyüme ölçüsünü tanımlayan özellikleri hesaplar. Bu işlem sonrasında çeşitli kümeler oluşursa, seçilen büyüme ölçütüne göre değeri bu kümelerden birinin orta noktasına denk gelen pikseller birer tohum olarak kullanılabilir. Benzerlik ölçütünün seçimi sadece probleme bağlı kalmayıp ayrıca elde bulunan resim verisine de bağlıdır. Eğer resim tek renkli ise analiz işlemi gri ton değerlerine, boyut, moment, desen vb. gibi uzay-konum özelliklerine dayalı olabilir.

Bölütleme işleminde ölçütü sağlamak tek başına yeterli değildir. Komşuluk ve bağlantılı olma çok önemlidir. Örneğin 3 farklı tonda gri ton seviyeli piksellerin rastgele dağıtılmasıyla elde edilmiş bir resim düşünelim. Böyle bir resimde aynı gri ton değerinde pikseller birbirleriyle bağlantılı olmadığından bu resmin bölümlenmiş olduğunu söylenemez.

Büyüyen bölge yordamı için son olarak gerekli olan unsur bir durdurma koşuludur. Bir bölgedeki hiçbir piksel büyüme sağlama koşulu karşılamıyorsa, başka bir değişle büyüme artık mümkün olmuyorsa, o bölge için hesaplama durdurulur. Durdurma koşulu olarak bölgesel bilgiler gri ton değeri, desen veya renk bilgileri kullanılabilir. Yordamı güçlendirmek için büyüyen bölgenin biçimi, boyutu yahut aday noktalar ile büyüyen bölgenin ortalama aydınlığının kıyaslanması kullanılabilir [51].

2.1.2. Bölge Parçalama ve Birleştirme Bölütlemesi

Bölge tabanlı yaklaşımlardan bir tanesi şu ilkeye dayanır: Önce resim birbirinden bağlantısız keyfi bölgelere parçalanır ve parçalar belirli ölçüleri sağlamak koşuluyla yeniden birleştirilir. Parçala ve birleştir algoritması belirlenen kısıtları sağlayana dek tekrarlamalı olarak koşturulur.

B bir resmi ve V de seçtiğimiz bir vasfı tanımlasın. Önce B birbiri ardına 4 küçük parçaya ayrılır. İlk adımda B_1, B_2, B_3, B_4 olmak üzere 4 küçük bölge oluşur. Eğer bölgelerden birisi daha önceden tanımlı vasfı karşılamazsa ($V(B_i) = YANLIS$) o zaman bu bölge kendisi içerisinde yeniden 4'lü bir parçaya ayrılır. Bu yordam aşağıdaki biçimde adımlarla özetlenebilir [51].

1. $V(B_i) = YANLIS$ ise i bölgesini kendi içerisinde parçala
2. $V(B_j \cup B_k) = YANLIS$ ise ortak vasfa sahip olan komşu B_j ve B_k bölgelerini birleştir.
3. Parçalanma ve birleşme artık meydana gelmiyorsa dur.

2.1.3. Sutaşımı Bölütleme Algoritması

$M_1, M_2, \dots, M_R$ noktaları $g(x, y)$ resminin bölgesel minimum noktaları olsun. $C(M_i)$, M_i bölgesel minimum değerine yakınsayan bir noktalar kümesi olsun. Böyle bir noktayı gerçek dünyada tek bir su çukurunun veya havzasının eğimli yüzeyinde bulunan bir noktaya benzetmek mümkündür. $T[n]$, resim üzerinde $g(s, t) < n$ koşulunu sağlayan

(s, t) koordinatlar kümesi olsun. Bu küme su çukuru benzetmesinde suyun en dip noktasına göre belirli bir n kotunun altında kalan $(T[n] = \{(s, t) \mid g(s, t) < n\})$ tüm noktaları ifade eder. Bu algoritmada resim bir coğrafi yüzeye benzetilir. Resim üzerinde bölgesel minimumlar çukurları veya göl yataklarını, aydınlık değeri yüksek olan noktalar dağları veya tepe gibi yükseltileri temsil eder. Resim üzerindeki bu izafi coğrafya yüzeyine su verildiğinde çukurlar su ile dolmaya başlayacaktır. Bu durumda çukurun bazı noktaları yükselen su seviyesinin altında kalacaktır. İzafi yüzey üzerinde yükseltisi en düşük olan noktalar resim üzerinde aydınlık değeri en düşük olan $M_1, M_2, \dots, M_R$ noktalarıdır. Bunlar arasında en küçük aydınlık değeri $\min\{g(x, y)\}$ ve maksimum olanı ise $\max\{g(x, y)\}$ dir.

Algoritmaya göre izafi yüzey üzerinde su seviyesi tam sayı artışları ile $\min\{g(x, y)\} + 1$ değerinden başlatılır ve $\max\{g(x, y)\} + 1$ değerine kadar yükseltilir. Algoritma içinde bu artışın sayacı olan n izafi yüzey üzerinde yükselmekte olan su seviyesini temsil eder. İzafi yüzey üzerinde su seviyesinin altına kalan noktalar, resim üzerinde piksel aydınlık değeri n değerinden küçük olan bir başka değişle $g(x, y) < n$ koşulunu sağlayan noktalardır. Su seviyesinin yükseltilmesi sürecinde her yükseliş değerinde su seviyesinin altında kalan bu noktalar $C_n(M_i)$ siyah renklendirilir ve üstünde kalan noktalar beyaz renklendirilir. M_i bölgesel minimumuna yakınsayan ve siyah renklendirilen bu noktalar kümesi $C_n(M_i)$ siyah ve beyaz renklerden oluşan ikilik bir pikseller kümesi olarak görülebilir. Bu küme Denklem 2.1’de verilen bağıntıyla elde edilmektedir. İzafi yüzey üzerinde birden fazla bölgesel en küçük söz konusu ise algoritmanın n . aşamasında yükselen su seviyesi altında kalan tüm noktaların kümesi Denklem 2.2’deki gibi tanımlanır.

$$C_n(M_i) = C(M_i) \cap T[n] \quad (2.1)$$

$$C[n] = \bigcup_{i=1}^R C_n(M_i) \quad (2.2)$$

İzafi coğrafya yüzeyinde elde edilen $C[n]$ noktalar kümesi gerçek resimde elde edilen bölütleme çıktısıdır. İzafi yüzeyde sanal havzalar arasında suyun üstünde kalan kesimler ise resim üzerinde bölütleme sonucunda ortaya çıkan bağlı bileşen sınırlarını göstermektedir [51].

2.1.4. *K*-Orta Nokta Bölütleme Yöntemi

K-orta nokta algoritmasına dayanan bir bölütleme yöntemidir. Bu yöntem genel amaçlı bir sınıflama yöntemidir. Bu tez çalışması kapsamında *K*-orta nokta algoritması sadece hücre çekirdeği bölütleme yöntemi olarak değil ayrıca öznitelik vektörüne bağlı olarak hücre çekirdeği sınıflama yöntemi olarak da kullanılmıştır. *K*-orta nokta algoritmasının açıklaması diğer sınıflayıcı yöntemlerle birlikte Bölüm-5'te yapılmıştır.

Bu algoritma, bir bölütleme yöntemi olarak kullanıldığında griton 8 bitlik bir resmin aydınlık değerlerine ait sıklık dağılımı üzerinde uygulanır. Bu tez çalışmasında bölütleme aşamasında hücre çekirdeklerini (östrojen alıcı pozitif veya negatif) diğer doku kesimlerinden ayırmak amaçlanmıştır. Bu işlemde dikkate alınan ipucu, hücre çekirdeklerinin diğer doku kesimlerine göre daha karanlık renklerde ve düşük aydınlık değerlerinde olmasıdır. Bu nedenle, aydınlık sıklık dağılımında bulunan griton aydınlık değerlerinin iki öbek etrafında, aydınlık ve karanlık değerler olarak, öbeklenmesi hedeflenmiştir. *K*-orta nokta öbekleme yöntemi, aydınlık ve karanlık olarak 2 adet öbek varlığını giriş parametresi olarak alır ve resimde bulunan tüm pikselleri buna göre otomatik olarak öbekler. Aydınlık öbeği içerisinde yer alan piksellere 1 aydınlık değeri, karanlık öbeği içerisinde yer alan piksellere 0 aydınlık değeri atanır. Sonuç olarak ortaya ikilik bir resim çıkar. Bu resimde, hücre çekirdeğini oluşturduğu tahmin edilen piksellere siyah, diğer doku kesimlerini oluşturduğu tahmin edilen pikseller beyaz renkte olacaktır.

2.1.5. Otsu Eşikleme Yöntemi

Bu yöntemde aydınlık sıklık dağılımının iki farklı aydınlık düzeyine kümeleneceği varsayılır. Kümelerden birisi resimde bölünmek istenen nesneyi diğeri ise resmin geri kalan kısmını, arka fonu temsil eder. Böylece, eşikleme yöntemi gri ton bir resmi giriş olarak alır ve çıkışta ikilik (siyah-beyaz) bir çıkış resmi üretir. Dönüştürme işlemi

sırasında aydınlık sıklık dağılımının eğrisine bakılarak en uygun eşik düzeyi araştırılır. Uygunluk ölçüsü, aydınlık dağılımında sınıf varyansları arasındaki farkın en yüksek olmasına dayanır. Otsu çalışmasında sınıf için varyansın en küçük olması ile sınıflar arası varyansın en yüksek olmasının eşdeğer sonuç verdiğini göstermiştir [52]. Bu yaklaşıma göre sınıf içi varyans, $\sigma_w^2(t)$, şu şekilde tanımlanır [53].

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \quad (2.3)$$

Burada (Denklem 2.3), $w_1(t)$ ve $w_2(t)$ eşik düzeyi t ile belirlenen iki sınıfa ait olasılık değerlerini gösterir. $\sigma_1^2(t)$ ve $\sigma_2^2(t)$ her iki sınıfa ait aydınlık dağılımından hesaplanmış varyans değerleridir. Bir sınıfa ait olasılık değeri $w_1(t)$, t eşik düzeyine göre o sınıf içerisinde sayılan piksellerin aydınlık değer olasılıklarının toplamıdır (Denklem 2.4).

$$w_1(t) = \sum_{i=0}^t p(i) \quad (2.4)$$

Diğer sınıf için de aynı biçimde olasılık hesaplaması yapılır. Burada, i aydınlık değerine sahip pikselin olasılık değeri $p(i)$, i düzeyinde aydınlığa sahip olan piksellerin sayısının $n(i)$ resimdeki tüm piksellerin sayısına, N , oranlanması ile bulunur (Denklem 2.5).

$$p(i) = \frac{n(i)}{N} \quad (2.5)$$

Otsu, sınıf içi varyansı en küçük yapmanın sınıflar arası varyansı en büyük yapmaya eşdeğer olduğunu şu bağıntı (Denklem 2.6) ile tanımlamıştır.

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2 \quad (2.6)$$

Bu bağıntıda, σ^2 aydınlık dağılımının tümünün (remin) varyansını gösterir. $\mu_1(t)$, t eşik düzeyine göre birinci sınıfın ortalama sıklık değeridir ve bu ortalama

$$\mu_1(t) = \sum_{i=0}^t p(i)x(i) \quad (2.7)$$

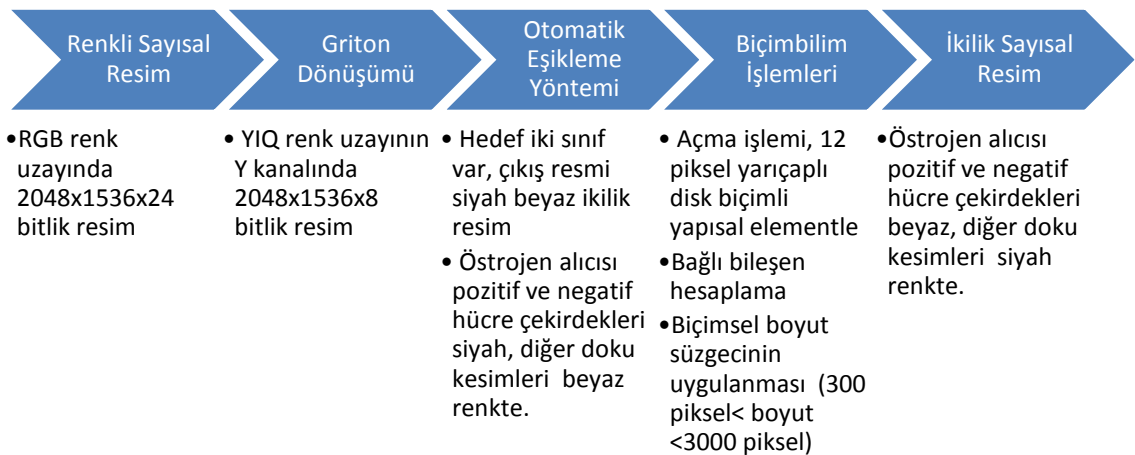
Bağıntısı ile hesaplanır. Burada $x(i)$, i aydınlık şiddetinin sıklığını ifade etmektedir. Yukarıda verilen bağıntılara uygun biçimde $w_2(t)$ ve $\mu_2(t)$ değişkenleri de hesaplanır. En uygun eşik düzeyini araştırmaya, Denklem 2.6'de verilen amaç fonksiyonu en büyük yapan eşik düzeyi bulununcaya kadar devam edilir. Bu araştırma yönteminin algoritması aşağıda verilen temel işlem basamakları ile özetlenmiştir.

1. Resmin aydınlık sıklık dağılımını oluştur ve her bir aydınlık düzeyinin bu dağılıma göre bulunma olasılığını hesapla.
2. Sınıf olasılıklarının $(w_1(0), w_2(0))$ ve ortalamalarının $(\mu_1(0), \mu_2(0))$ ilk değer atamalarını yap.
3. $t = 1$ 'den başlayarak dağılımda bulunan en büyük aydınlık seviyesine kadar bütün aydınlık değerlerine aşağıdaki hesaplamaları yap.
 - a. Sınıf olasılıklarını $(w_1(t), w_2(t))$ ve ortalamalarını $(\mu_1(t), \mu_2(t))$ hesapla.
 - b. Sınıflar arası varyans değerini $\sigma_b^2(t)$ hesapla.
4. En büyük $\sigma_b^2(t)$ değerini veren t eşik düzeyini seç ve bölütlemeyi bu eşik düzeyine göre yap.

Algoritmadan da anlaşılacağı üzere, Otsu eşikleme yöntemi araştırma uzayında tüm olasılıkları arayan bir yöntemdir. Ancak sadece birinci ve ikinci moment düzeyinde hesaplamalar içerdiğinden karmaşası azdır. Parametrik bir yöntem olmaması nedeniyle kolaylıkla genellenebilen bir yöntemdir [52, 53].

2.2. Bölütleme Yöntemlerinin Görsel Sınaması

Bölütleme amacıyla; kullanımı kolay olan, kullanıcıdan giriş parametresi istemeyen ve hızlı sonuç üretebilen danışmansız bir bölütleme yönteminin uygulanması hedeflenmiştir. Bu tez çalışmasında kullanılan danışmansız bölütleme yaklaşımları aşağıda verilen işlem basamaklarından oluşmaktadır (Şekil 2.1).

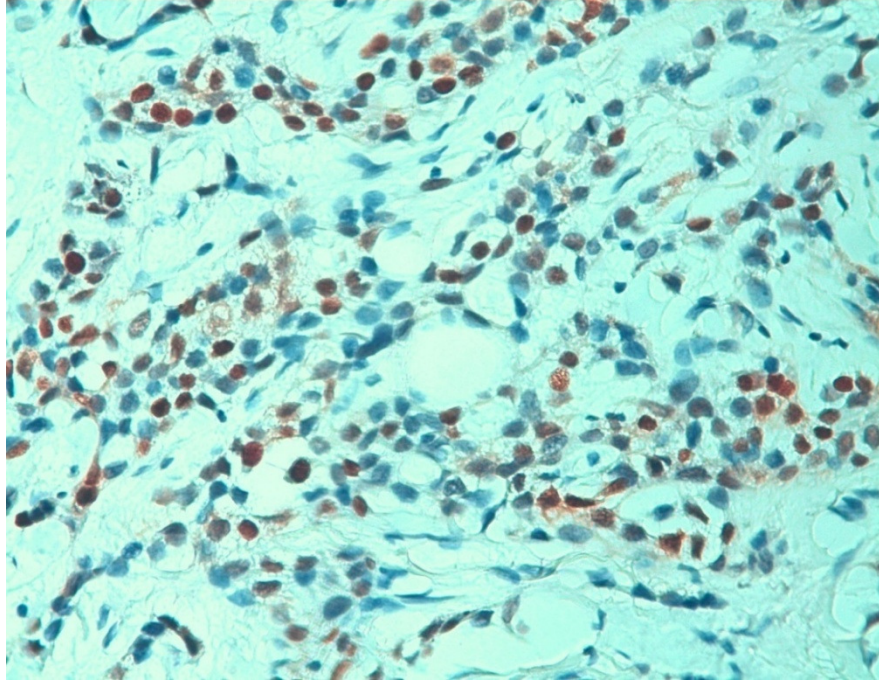


Şekil 2.1. Hücre çekirdeği bölütleme işlem basamakları

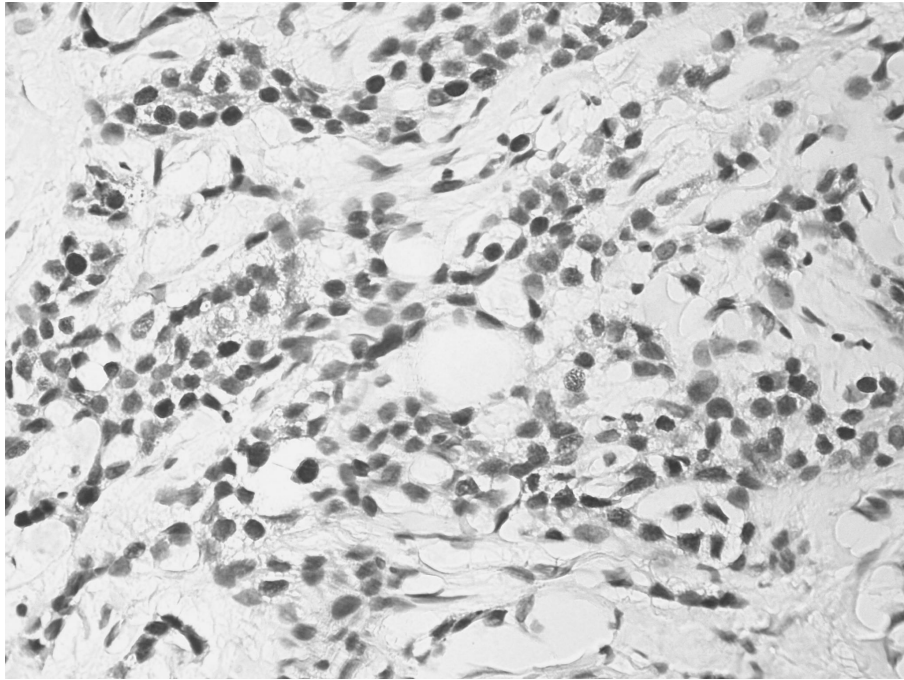
Bölütleme yöntemleri genel olarak piksel tabanlı, kenar seçme tabanlı ve bölge tabanlı olarak sınıflandırılabilir [54]. Kullanılan tıbbî resimler 2048x1536x24 piksel boyutundadır. Büyük boyutlu giriş verisi resimler dikkate alındığında, tercih edilen bölütleme yönteminin az işlem yükü olması ve ayrıca önceden bir eğitim verisi gerektirmemesi önemlidir. Çünkü işlem karmaşası ve başlangıç veri ihtiyacı arttıkça sistemin pratikte kullanım değeri azalmaktadır.

Bölütleme aşamasında öncelikle RGB renk uzayında kaydedilmiş 2048x1536 boyutunda ve her pikselin 24 bitlik bilgi ile temsil edildiği JPEG sıkıştırılmış resim, 8 bitlik bir griton resme dönüştürülmüştür. Bu tez çalışmasında kullanılan östrojen alıcı

boyamasını gösterir renkli örnek bir RGB resmi Şekil 2.2’de ve bu resmin elde edilen griton dönüşümü Şekil 2.3’te gösterilmektedir.



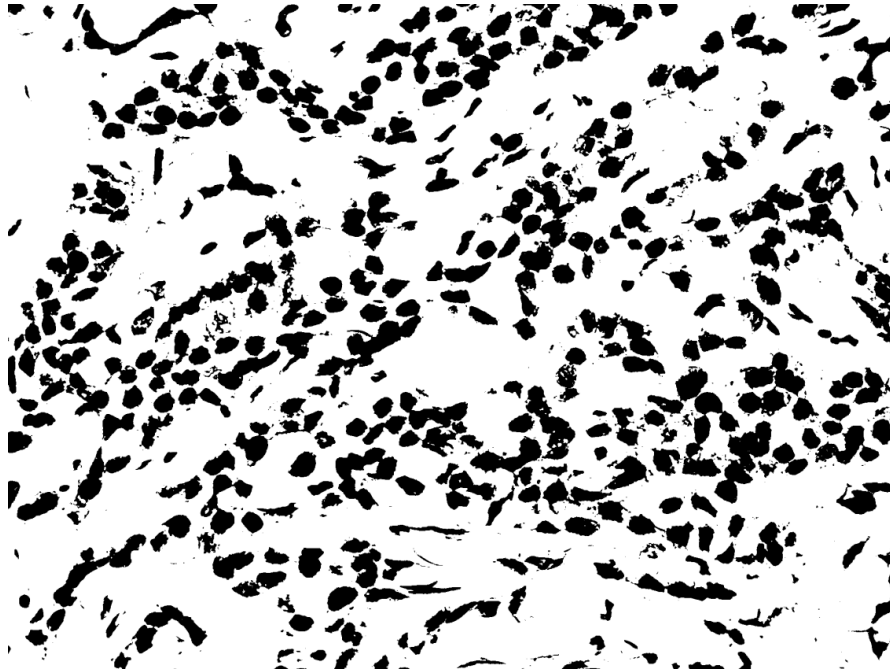
Şekil 2.2. Asıl resim, 2048x1536x24 piksel, 40x ölçekte.



Şekil 2.3. Griton dönüşümü

Griton dönüşümü esnasında, RGB renk kanalında her bir renk bileşeninin ağırlandırılmış ortalaması kullanılır ($Griton = 0.2989R + 0.5870G + 0.1140B$). Bu ortalama, YIQ renk uzayında Y bileşenidir. Y bileşeni renkli görüntülerin siyah beyaz televizyonlarda gösterilmesinde kullanılmaktadır [55].

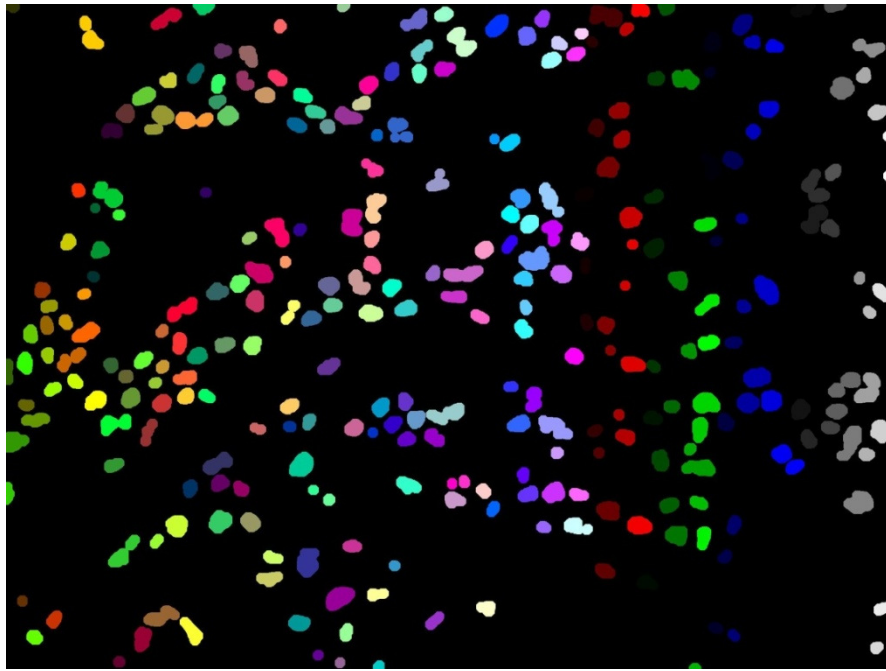
Bölütleme işleminde, griton dönüşümünden sonra otomatik eşikleme yöntemi devreye girer. Eşikleme işlemi, griton resim içerisinde nispeten koyu renkli olan hücre çekirdekleri ile diğer doku kesimlerini ayırt etmeyi amaçlar. Eşikleme işlemi sonunda, belirli bir aydınlık eşik seviyesine göre griton resim ikilik siyah beyaz resme dönüşür. Karar eşik seviyesi olan aydınlık değeri kullanılan bölütleme yöntemine göre farklı olabilir. Eşikleme aşamasında östrojen alıcı varlığı pozitif ve negatif hücre çekirdekleri arasında bir ayırım yapılmaz. Sadece hücre çekirdekleri bölütlenerek diğer doku kesimlerinden ayrılır. Bu yönüyle bölütleme işlemi, aydınlık seviyesine bağlı olarak karar veren iki sınıflı bir sınıflama olarak düşünülebilir. Örnek griton resim üzerinde Otsu eşikleme sonrasında elde edilen çıktı resim Şekil 2.4'te gösterilmiştir.



Şekil 2.4. Eşikleme işlemi çıktısı olan ikilik resim

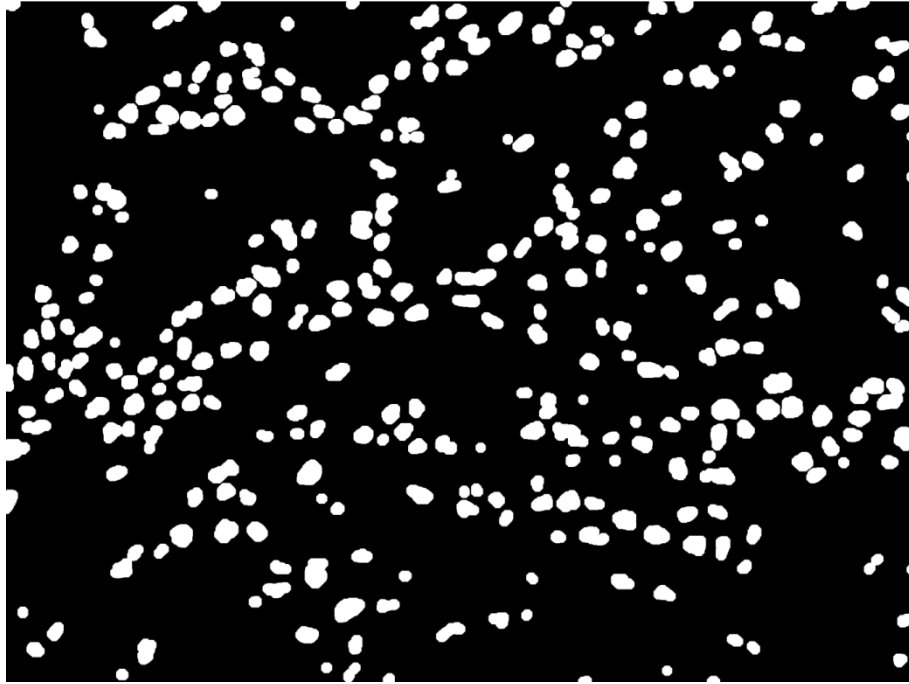
Eşikleme işleminin çıktısı olan resimde hücre çekirdekleri siyah, diğer doku kesimleri beyaz renklidir. Otomatik eşikleme yöntemleri aydınlık düzeyi düşük (koyu renkli) hücre çekirdeklerine siyah, aydınlık düzeyi nispeten yüksek diğer doku kesimlerine beyaz renk ataması yapar. Biçim bilim işlemcileri uygulanmadan önce, eşikleme çıktısı ikilik resim aydınlık düzeyi bakımından terslenir. Böylece biçimsel işlemcilere hedef teşkil eden hücre çekirdekleri beyaza, diğer doku kesimleri siyah renge dönüşür.

Bölütlemenin son işlem basamağında biçimbilim işlemcileri uygulanır ve resim üzerinde bağlı bileşenler hesaplanır. Resim üzerinde kaynaşan, çapaklı nesnelerin varlığını ortadan kaldırmak için 12 piksel yarıçaplı disk biçimli bir yapısal element kullanılarak biçimsel açma işlemi resme uygulanır. Daha sonra, resim üzerinde bağlı bileşenler hesaplanarak her bileşen bir hücre çekirdeği olarak kaydedilir. Küçük boyutlu stroma hücre çekirdeklerini elemek ve hücre çekirdeklerinin kaynaşmasıyla oluşan sıra dışı büyük boyutlu bileşenleri ortadan kaldırmak için biçimsel boyut eleme uygulanır. Buna göre, boyutu 300 pikselden küçük ve 3000 pikselden büyük olan bileşenler elenir. Bu değerler deneyime bağlı olarak tespit edilmiştir. Örnek resim üzerinde, bağlı bileşen hesaplama sonrasında ayrıca bir dosyaya kaydedilen bağlı bileşenleri gösterir resim ve nihai bölütleme çıktısı aşağıda Şekil 2.5'te verilmiştir.



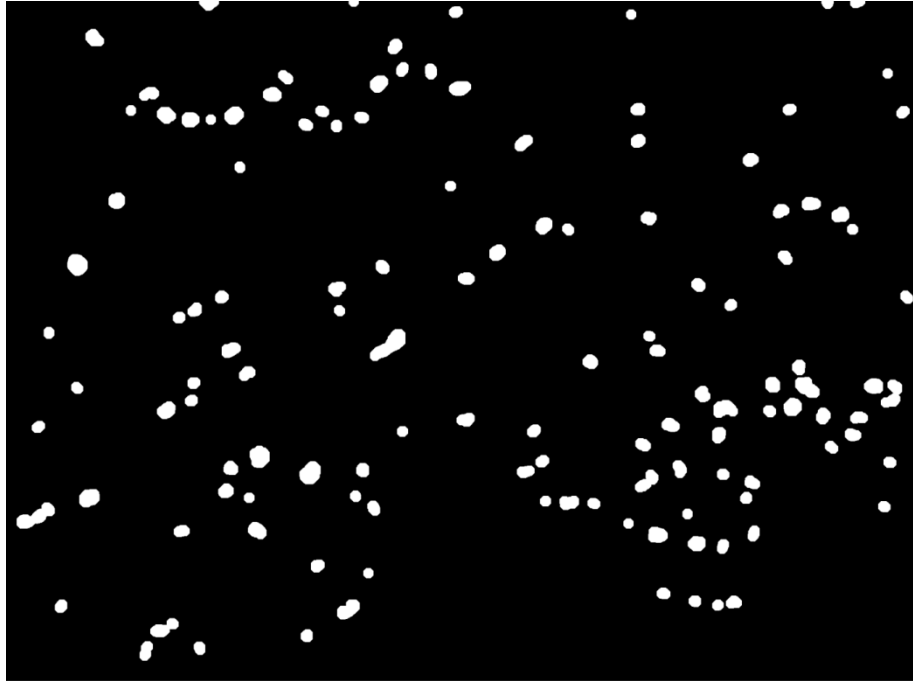
Şekil 2.5: Bağlı bileşen hesaplama çıktısı resim

Bölütleme aşamasında, kullanılacak eşikleme yönteminin seçimi deneyime dayalı olarak yapılmıştır. Seçilen yöntemlerin karşılaştırılması görsel sınama ile gerçekleştirilmiştir. Örnek bir resim üzerinde (Şekil 2.2), Otsu eşikleme yöntemiyle elde edilen bölümle çıktısı (Şekil 2.6), k -orta nokta öbekleme algoritması ile elde edilen bölütleme çıktısı (Şekil 2.7), su taşıma algoritması ile elde edilen bölütleme çıktısı (Şekil 2.8) gösterilmektedir. Her bölütleme yönteminden sonra, 12 piksel yarıçaplı disk görünümü biçimbilim açma işlemcisi ikilik resme uygulanmıştır.

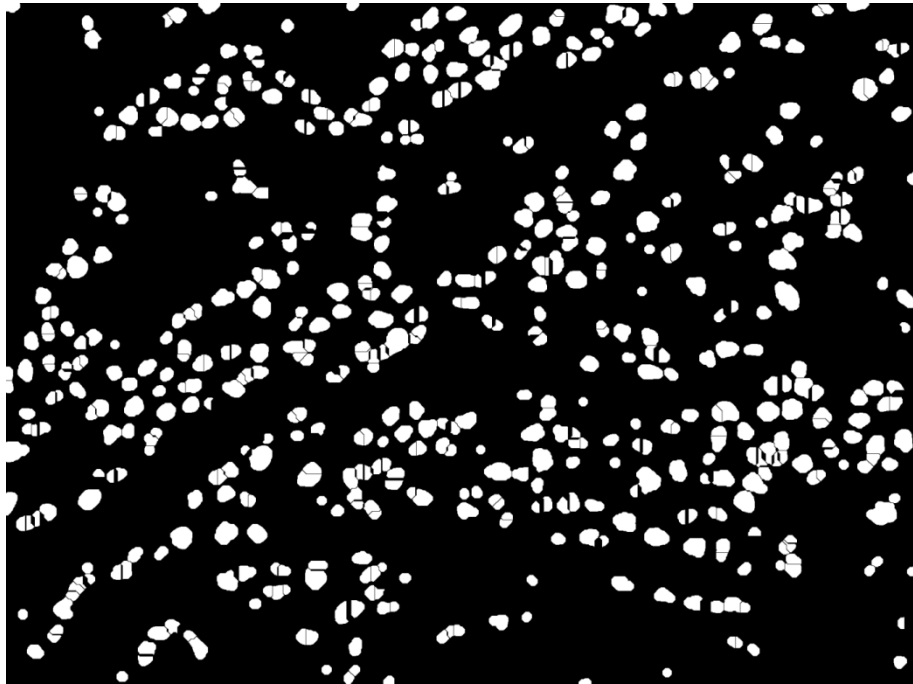


Şekil 2.6. Otsu eşikleme yöntemiyle bölütleme çıktısı

Elde edilen sonuçlar görsel olarak karşılaştırıldığında; k -orta nokta öbekleme algoritması bölütleme sonrasında az sayıda hücrenin bölütlendiği gözlenmiştir. Diğer taraftan, sutaşımı bölütleme algoritmasında aşırı bölütlemenin gerçekleştiği, çok sayıda hücrenin birkaç parçaya bölündüğü görülmüştür. Biçimsel süzgeç, dokuda bulunan stroma hücrelerini incelemeye esas hücre çekirdeklerinden ayırt etmek için zorunlu olarak uygulanmaktadır. Bu durum, sonradan gelecek olan boyuta dayalı biçimsel eleme (300 pikselden küçük ve 3000 pikselden büyük bileşenlerin silinmesi) işleminin birçok parçalanmış hücreyi tümüyle ortadan kaldırmasına sebep olmaktadır.



Şekil 2.7. k -orta nokta öbekleme yöntemiyle bölütleme çıktısı



Şekil 2.8. Sutaşımı algoritması ile elde edilen bölütleme çıktısı

Otsu eşikleme yöntemi hızlı ve makul sonuçlar sergileyebilmektedir. Boyutu 2048x1536x24 piksel olan örnek resim (Şekil 2.2) üzerinde yapılan 20 adet deneme sonucunda; RGB resmin griton dönüşümü yaklaşık olarak ortalama 0.26 (standart

sapma 0.03) saniye, griton resim üzerinde Otsu eşikleme yapılması yaklaşık olarak ortalama 0.05 (standart sapma 0.02) saniye, biçimsel açma-bağlı bileşen hesaplama-biçimsel eleme işlem silsilesi yaklaşık olarak ortalama 3.44 (standart sapma 0.20) saniye zaman almaktadır. Diğer taraftan, K -orta nokta öbekleme algoritması ile yapılan bölütleme ortalama 18.15 (standart sapma 0.18) saniye ve sutaşımı algoritması yapılan bölütleme işlemi aynı koşullarda 4.73 (standart sapma 0.10) saniye sürmektedir. Bu denemeler, 32 bit komut setli Intel Pentium M 1.73 GHz işlemcili ve 1.25 GB DDR2 RAM özelliklerinde 2006 üretim tarihli bir PC üzerinde gerçekleştirilmiştir. Bu süreler bilgisayarın anlık işlem yüküne göre azalış veya artış gösterebilmektedir.

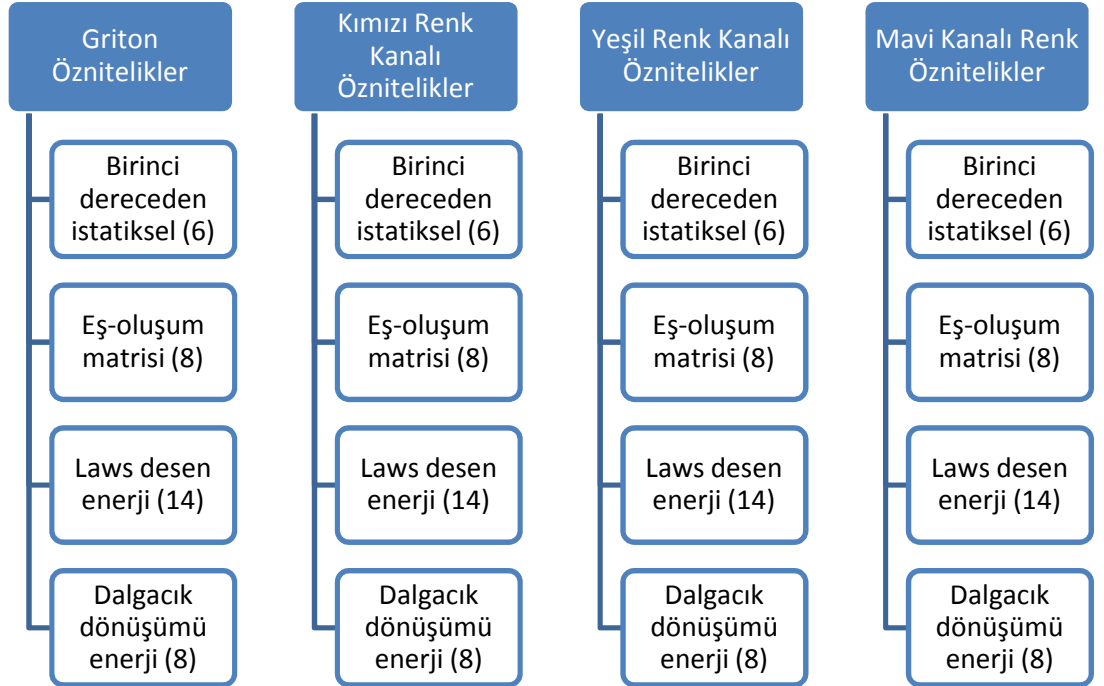
Otsu eşikleme algoritmasına dayalı hücre çekirdeği bölütleme yöntemi, östrojen alıcı tanısında daha önce kullanılmıştır [35]. Meme kanseri biyopsi örneklerinde Her-2/neu (c-erbB2) protein alıcısı tanısına yönelik bir çalışmada hücre çekirdeği bölütleme amacıyla Otsu eşikleme algoritması kullanılmış ve elde edilen sonuç sutaşımı bölütleme algoritması ile iyileştirilmiştir [56]. İmmünohistokimyasal resimlerde Otsu eşikleme algoritmasına dayalı hücre çekirdeği bölütleme yöntemini öneren bir başka çalışmada, Otsu otomatik eşikleme algoritması sonrasında bölütlenmiş hücre çekirdeklerinin şekli ve boyutunda hataların olabildiği ancak bu hataların tespit edilen hücre sayısını etkilemediği bildirilmiştir [57]. Literatürde belirtilen bu görüşler ve gerçekleştirilen deneyimler doğrultusunda, bölütleme aşamasında Otsu eşikleme yönteminin uygulanması tercih edilmiştir.

3. BÖLÜM

ÖZNİTELİK ELDE ETME YÖNTEMLERİ

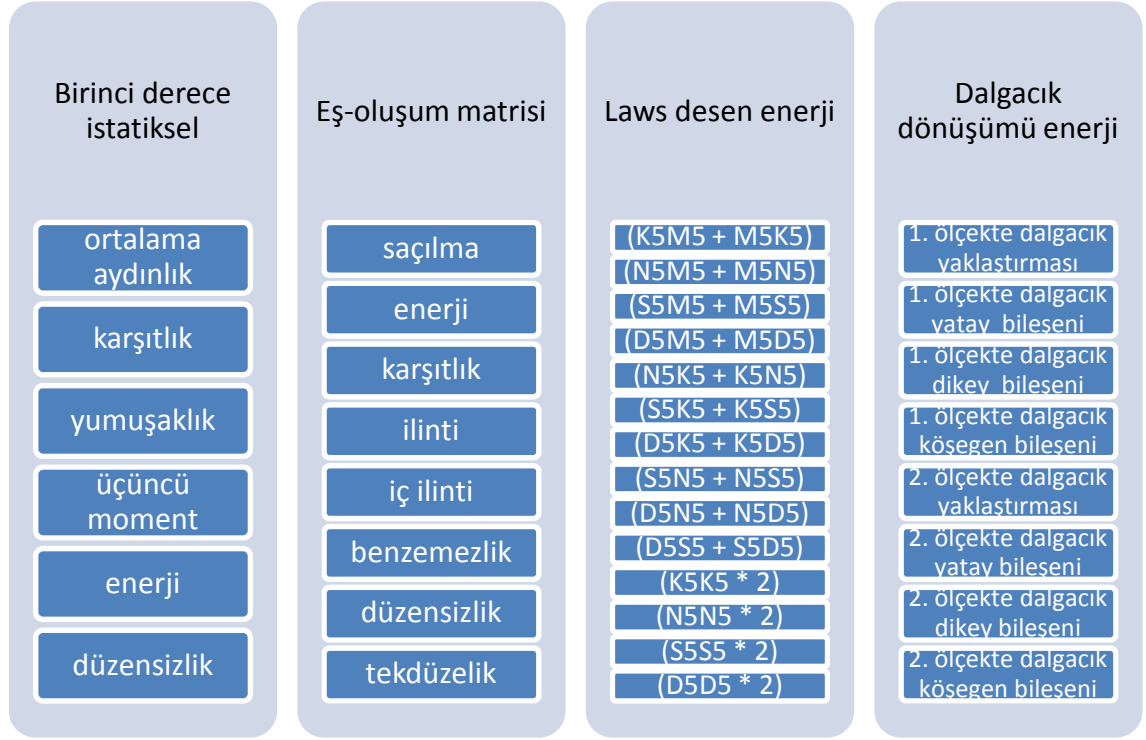
Bölütleme aşamasında tespit edilen hücre çekirdeğinden bu bölümde açıklanan ölçütlere göre öznitelikler üretilir. Bu tez çalışması kapsamında her bir hücre çekirdeğinden toplamda 144 adet öznitelik çıkarılmıştır.

Üretilen öznitelikler birinci dereceden istatiksel öznitelikler, eş oluşum matrisi öznitelikleri, Laws desen enerji öznitelikleri ve dalgacık dönüşümü öznitelikleri olarak gruplandırılmıştır. Öznitelik türüne göre hesaplanan öznitelik sayısı Şekil 3.1’te gösterilmiştir.



Şekil 3.1. Her bir renk kanalında üretilen özniteliklerin türü ve sayısı

Her bir öznitelik türünde hangi özniteliklerin hesaplandığı Şekil 3.2’de özet olarak verilmiştir. Bu bölüm içerisinde türüne göre özniteliklerin hangi bağıntılarla hesaplandığı açıklanmaktadır. Özetle, her bir hücre çekirdeğinden toplamda 144 adet öznitelik çıkarılmıştır.



Şekil 3.2. Öznitelik türleri ve hesaplanan öznitelikler

Ayrıca, ilinti tabanlı öznitelik eleme yöntemi sayesinde 144 adet öznitelik içerisinde ayırt etme kabiliyeti en yüksek olan 11 adet öznitelik seçilmiştir. Kullanılan ilinti tabanlı öznitelik eleme yöntemi bu bölüm içerisinde tanıtılmaktadır. Seçilen bu 11 adet öznitelik, sınıflayıcı modelin giriş değişkenlerini oluşturmaktadır. Bu sayede, sınıflayıcı girişinde azalan giriş değişkeni miktarına karşın sınıflama başarımında artış hedeflenmiştir.

Sınıflama işleminde, öznitelikler sınıflayıcı modelin giriş değişkeni olacağından, sınıflama başarımını düşürmeden sınıflama modelin işlem yükünü azaltmak ve bu yükü uygulanabilir düzeye düşürmek gereklidir. Ayrıca, mevcut öznitelik topluluğunda hangi özniteliklerin ayırt etme kabiliyetinin üstün olduğu tez çalışmasını önemli bir amacdır. Bu nedenle, temel bileşen analizi vb. mevcut öznitelik değerlerini değiştiren yaygın boyut

azaltma yöntemleri yerine sınıflayıcıdan bağımsız olan ve mevcut öznelilik topluluğunda en etkili ayırt edici öznelilikleri seçmeyi amaçlayan bir öznelilik eleme yöntemi bu tez çalışmasında kullanılmıştır. Kullanılan öznelilik eleme yöntemi süzgeç türünde bir öznelilik eleme yöntemidir ve ilinti tabanlı bir yaklaşımla seçilecek özneliliklere karar vermektedir.

3.1. Birinci Dereceden İstatiksel Öznelilikler

Birinci derece istatiksel öznelilikler; renk kanallarına, renk kanallarında sıklık dağılımından (histogram) elde edilen istatiksel verilere dayanır. Bu çalışmada kullanılan ve sıklık dağılımından elde edilen istatiksel öznelilikler şunlardır; ortalama aydınlık, karşıtlık, yumuşaklık, üçüncü moment, enerji ve düzensizliktir [51].

Gri ton değerlerini ifade eden rastgele değişken z olsun ve her z aydınlık seviyesi $p(z_i)$ sıklık dağılımı ile ifade edilsin ($i = 0, 1, 2, 3, \dots, L - 1$) ve L , farklı aydınlık seviyelerini gösteren bir sayı olmak üzere z 'nin aydınlık ortalaması m

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (3.1)$$

bağıntısı (Denklem 3.1) ile hesaplanır. Ortalama aydınlık değeri etrafında n 'inci moment

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \quad (3.2)$$

bağıntısı ile belirlenir (Denklem 3.2). Burada sıfırcı momentin 1 ($\mu_0 = 1$) ve birinci momentin 0 ($\mu_1 = 0$) olduğu görülmektedir. Ancak ikinci moment aydınlık dağılımının varyansını gösterir. Varyansın karekökü olan standart sapma, aydınlık dağılımında ortalama karşıtlığın bir ölçüsüdür (Denklem 3.3).

$$\sigma = \sqrt{\mu_2(z)} \quad (3.3)$$

Yumuşaklık resim bölgesinde göreceli olarak aydınlık geçişlerinin yumuşaklığını ifade eder (Denklem 3.4). Sabit aydınlık değeri olan bölgede R değeri 0 olur, aydınlık değerlerinde sapmaların çok olduğu bölgede bu değer 1 sınırına yaklaşır. Uygulamada bu ölçüt içinde kullanılan saçılma (varyans), $\mu_2(z)$, $(L - 1)^2$ değerine bölünerek normalleştirilir.

$$R = 1 - \frac{1}{(1 + \sigma^2)} \quad (3.4)$$

Üçüncü moment, $\mu_3(z)$, aydınlık sıklık dağılımında çarpıklığın bir ölçüsüdür. Simetrik aydınlık dağılımında bu değer 0, sağa kayan dağılımında bu değer pozitif ve sola kayan dağılımında bu değer negatif olmaktadır (Denklem 3.5).

$$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \quad (3.5)$$

Düzgünlük ölçütü tüm gri ton değerleri eşit olduğunda maksimum değerini alır (Denklem 3.6). Üçüncü moment $\mu_3(z)$, $(L - 1)^2$ değerine bölünerek normalleştirilir. Düzensizlik ölçütü gri ton dağılımında rastgeleliğin matematiksel bir ifadesidir (Denklem 3.7).

$$U = \sum_{i=0}^{L-1} p^2(z) \quad (3.6)$$

$$e = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (3.7)$$

Bu öznitelik ölçütlerinden faydalanarak, RGB ve gri ton kanalların her birinden 6 adet olmak üzere, toplamda 24 adet birinci dereceden öznitelik elde edilmiştir.

3.2. İkinci Dereceden İstatiksel Öznitelikler

3.2.1. Birliktelik Matrisi Öznitelikleri

Birinci derece öznitelikler sadece bir aydınlık değerinin resimde var oluşunu ve oluş sıklığını gösterir. Çeşitli aydınlık değerinde piksellerin resimde birlikte konumlanması veya bir araya geliş sıklığı hakkında fikir edinmek için birliktelik (eş-oluşum) matrisleri resim üzerinde hesaplanır.

Sıklık dağılımı ve birliktelik matrisi arasındaki farklılık şu şekilde açıklanabilir. Birliktelik matrisi, bir resimde belirli bir yönde ve uzaklıkta bir araya gelen piksel değerlerinin sıklık dağılımını gösterir bir matristir. Tek boyutlu sıklık dağılımı, diğer adıyla histogram, bir resimde belirli bir aydınlık şiddetinin ne kadar sıklıkla var olduğunu gösterirken, birliktelik matrisi belirli bir piksel ikilisinin ne sıklıkla var olduğunu gösterir. Sıklık dağılımı tek boyutlu bir veridir, birliktelik dağılımı iki boyutlu bir veridir. Sıklık dağılımı, resmin aydınlık değerlerinin tek başına varlığı hakkında bilgi verirken, birliktelik dağılımı resimde aydınlık değerlerinin birlikte yerleşimi hakkında bilgi verir. Birliktelik dağılımının verdiği bilgiler iki boyutlu yerleşim veya dizilimi gösteren bilgilerdir.

Birliktelik matrisini tanımlayan ve bu matris üzerinden desen özelliklerinin hesaplanmasını öneren Haralick ve arkadaşları, bu matris için kendi makalelerinde gri-ton yerleşimi bağımlılık matrisi adlandırmasını kullanmışlardır [58].

Bir gri ton resmin N_x ve N_y pikselden oluşan dikdörtgen yapıda bir kesitten oluştuğunu varsayalım. Bu resimde, piksellere ait gri ton aydınlık değerleri N_g ayrı aydınlık

düzeyinde örneklenmiş (bu düzeylerden en yakın olana atanmış) olsun. Matematiksel olarak birliktelik matrisi $P(i, j, d, 0^0)$; yatay ekseninde (0^0 yönünde) ve birbirlerine d adet piksel adımıyla tanımlı uzaklıkta komşuluğu olan i ve j aydınlık değerlerinde iki pikselin birlikte olma sayısı aşağıdaki gibi tanımlanır (Denklem 3.8) [58].

$$\begin{aligned} P(i, j, d, 0^0) &= adet\{((k, l), (m, n)) \in (N_g \times N_g) | (k - m) \\ &= 0, |l - n| = d, I(k, l) = i, I(m, n) = j\} \end{aligned} \quad (3.8)$$

Bu yaklaşımla oluşturulan birliktelik matrisi üzerinde $p(i, j) = P(i, j, d, 0^0)/R$ olmak üzere (R normalleştirme sabitidir.) aşağıda verilen desen öznitelikleri elde edilmiştir.

$$saçılma = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i, j) \quad (3.9)$$

$$enerji = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)^2 \quad (3.10)$$

$$karşıtlık = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \right\}, |i - j| = n \quad (3.11)$$

$$ilinti = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.12)$$

İlinti hesaplamasında kullanılan μ_x , μ_y , σ_x ve σ_y sırasıyla $p_x(i)$ ve $p_y(j)$ olasılık dağılımlarından elde edilen ortalama ve standart sapma değerleridir. Bu dağılımların hesaplanması aşağıda gösterilmiştir.

$$p_x(i) = \sum_{j=1}^{N_g} P(i, j) \quad (3.13)$$

$$p_y(j) = \sum_{i=1}^{N_g} P(i, j) \quad (3.14)$$

Haralick özelliklerinin bir kısmı; tekdüzelik, karşılıklık, engebelik, düzgünlük gibi tanımlara karşılık ölçümler yapmaktadır. Diğer bazı özellikler doğrudan bir desen durumu tanımlamazken farklı desenlerin karşılaştırılmasında sayısal bir ayırım sağlamaktadır [58]. Bu özniteliklerin yanı sıra kullanılan diğer birliktelik matrisi öznitelikleri; iç-ilinti, benzemezlik, düzensizlik, tekdüzelik aşağıda verilmiştir [59].

$$i\check{c} \text{ ilinti} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij)p(i, j) \quad (3.15)$$

$$benzemezlik = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j|p(i, j) \quad (3.16)$$

$$d\ddot{u}zensizlik = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j)) \quad (3.17)$$

$$tekd\ddot{u}zelik = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i, j) \quad (3.18)$$

Sıklık dağılımının çözünürlüğü, sıklık miktarı aranan farklı aydınlık seviyelerinin sayısına bağlıdır. Benzer biçimde birliktelik dağılımının çözünürlüğü, birlikteliği araştırılan farklı aydınlık seviyelerinin sayısına bağlı olmaktadır. Bu çalışmada gri ton resim 8 farklı gri ton seviyesine göre örneklenmiş ($N_g=8$) ve her bir resim için 8×8 boyutunda bir birliktelik matrisi hesaplanmıştır. Bu hesaplama sürecinde, hedef piksel ve bu pikselin yatay doğrultuda (0^0) ve 1 adım ($d=1$) yanındaki piksellerin aydınlık değerlerine bakılarak birliktelik matrisinde farklı aydınlık ikililerine ait sıklık değerleri sayılmıştır.

Özetle, birliktelik matrisi üzerinde 8 ayrı öznitelik; saçılım, enerji (açısal ikinci moment), karışıklık, ilinti, [58]; iç-ilinti, benzemezlik, düzensizlik, tekdüzelik [59] hesaplanmıştır [60, 61]. Sonuçta, RGB ve gri ton renk kanallarına ait birliktelik matrislerinin her birinden 8 adet olmak üzere toplamda 32 adet ikinci derece öznitelik birliktelik matrisleri üzerinden hesaplanarak kullanılmıştır.

3.2.2. Laws Desen Enerji Öznitelikleri

Bir resimde bazı karakteristik özelliklerin (miktar, kenar, nokta, salınım ve damla) ortaya çıkarılması için küçük evrişim maskeleri kullanılabilir. Tipik olarak aşağıda verilen tek boyutlu (1-B) 5 elemanlı maskeler bu amaca yönelik olarak seçilir [54, 62]. Tablo 3.1’de her bir fiziksel kavramın temsil ettiği fiziksel hal tek boyutlu 5 elemanlı matris üzerinde matematiksel olarak tanımlanmıştır.

Tablo 3.1. Laws desen maskeleri

$$\begin{aligned} M5 &= [1 \ 4 \ 6 \ 4 \ 1], M=\text{miktar} \\ K5 &= [-1 \ -2 \ 0 \ 2 \ 1], K=\text{kenar} \\ N5 &= [-1 \ 0 \ 2 \ 0 \ -1], N=\text{nokta} \\ S5 &= [-1 \ 2 \ 0 \ -2 \ 1], S=\text{salınım} \\ D5 &= [1 \ -4 \ 6 \ -4 \ 1], D=\text{damla} \end{aligned}$$

Her bir tek boyutlu matrisi bir başka 1-B matris ile evrişim işlemine tabi tutarak iki boyutlu (2-B) maskeler elde etmek mümkündür. Oluşan 2-B maske, kendisini meydana

getiren her iki 1-B matrisin fiziksel özelliklerini temsil eder. Böylece, 5 adet 1-B maskeyi ikili gruplar halinde evrişim yaparak toplamda çok sayıda 2-B maske elde edilir. Örneğin, K5S5 hem kenar hem de salınım özelliklerini içinde barındıran 5x5 boyutlu 2-B bir maskedir ve bu maske, 1-B dikey K5 ile 1-B yatay S5 matrislerinin evrişiminin sonucudur.

Yön bilgisinin önemli olmadığı uygulamalarda, aynı 1-B maskelerin farklı dikey ve yatay konumu ile oluşturulan benzer 2-B maskeler toplanarak yön bilgisinden bağımsız toplamda 15 adet maske elde edilebilir. Bu çalışmada, M5M5 maskesi dışında kalan 14 adet 2-B maske kullanılmıştır. M5M5 maskesi ile oluşturulan resmin aydınlık ortalaması sıfırdan farklıdır, bu maskenin diğer maskelerle oluşturulan resimleri normalleştirme amacıyla kullanılması ve onun dışında hesaba katılmaması tavsiye edilmektedir [62]. RGB ve gri ton kanallarından her birinden 14 adet olmak üzere toplamda 56 adet Laws desen resmi üretilmiştir. Daha sonra bu resimlerin her birinden Denklem 3.19’da tanımlanan enerji özniteliği hesaplanmıştır.

$$Enerji = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f^2(x, y) \quad (3.19)$$

Burada; $f(x, y)$, $M \times N$ boyutlu olan f resminin x ve y ile pikselleri indekslenen aydınlık değeridir. Enerji özniteliğinin elde edilmesinde, sadece hücre çekirdeğinin çeperi içerisinde kalan piksel değerleri hesaba katılmıştır.

3.2.3 Dalgacık Dönüşümü Enerji Öznitelikleri

Eğer bir resmin örüntüsünde, periyodiklik, kendine benzerlik, süreksizlik gibi özellikler mevcutsa o zaman resmin frekans içeriğinde ayırıcı öznitelikler elde edilebilir. Bu amaç doğrultusunda çalışmada 2-B dalgacık dönüşümü kullanılmıştır.

Bu dönüşüm, dalgacık adı verilen ötelenmiş ve ölçeklenmiş dikgen fonksiyonların doğrusal toplamı türünden resmi ifade eder. Fourier dönüşümlerinden farklı olarak bu dönüşüm, zaman ve konum ekseninde ölçeklenebilir. Bu sayede, bir resmin frekans

içeriğini farklı ölçeklerde konumsal yerleşimi bilgisini kaybetmeksizin görmek mümkündür.

Dalgacık dönüşümünü hesaplamak amacıyla bir ölçekleme $\varphi(x, y)$ ve bir de dalgacık $\psi(x, y)$ fonksiyonu kullanılır. Tek boyutlu sonuç üreten $\varphi(x) \cdot \varphi(x)$ çarpımını dışarıda bırakmak kaydıyla, ölçek ve dalgacık fonksiyonlarını çarparak aşağıda verildiği gibi çeşitli 2-B dalgacık fonksiyonları üretilebilir.

$$\varphi(x, y) = \varphi(x)\varphi(y) \quad (3.20)$$

$$\psi^H(x, y) = \psi(x)\varphi(y) \quad (3.21)$$

$$\psi^V(x, y) = \varphi(x)\psi(y) \quad (3.22)$$

$$\psi^D(x, y) = \psi(x)\psi(y) \quad (3.23)$$

Ortaya çıkan farklı dalgacık fonksiyonları, çeşitli yönlerde aydınlık değişimini ifade eder. Ölçekleme fonksiyonu, $\varphi(x, y)$, düşük frekans bileşenleri $\psi^H(x, y)$, $\psi^V(x, y)$ ve $\psi^D(x, y)$ sırasıyla yatay, dikey ve köşegen yönünde yüksek frekans bileşenlerinin varlığını ölçer. Yukarıda verilen çarpımlar sonucu elde edilen 2-B dalgacık fonksiyonları aşağıdaki biçimde gösterilir.

$$\varphi_{j,m,n}(x, y) = 2^{j/2} \varphi(2^j x - m, 2^j y - n) \quad (3.24)$$

$$\psi_{j,m,n}(x, y) = 2^{j/2} \psi^i(2^j x - m, 2^j y - n), \quad i = \{H, V, D\} \quad (3.25)$$

Denklem 3.25’de, i yatay (H) , düşey (V) ve köşegen (D) yönünü ifade eden ve sayısal olmayan bir indekstir. Bu fonksiyonlar verildiğinde, $M \times N$ boyutlu $f(x, y)$ fonksiyonunun 2-B ayrık dalgacık dönüşümü

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (3.26)$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x, y) \quad i = \{H, V, D\} \quad (3.27)$$

olarak tanımlanır. j_0 keyfi bir başlangıç ölçeği olup $W_{\varphi}(j_0, m, n)$ katsayıları $f(x, y)$ fonksiyonunun j_0 ölçeğinde dalgacık dönüşümü yaklaştırmasıdır. $W_{\psi}^i(j, m, n)$ katsayıları $j \geq j_0$ ölçeğinde sırasıyla yatay, düşey ve köşegen detaylarını tanımlar [51].

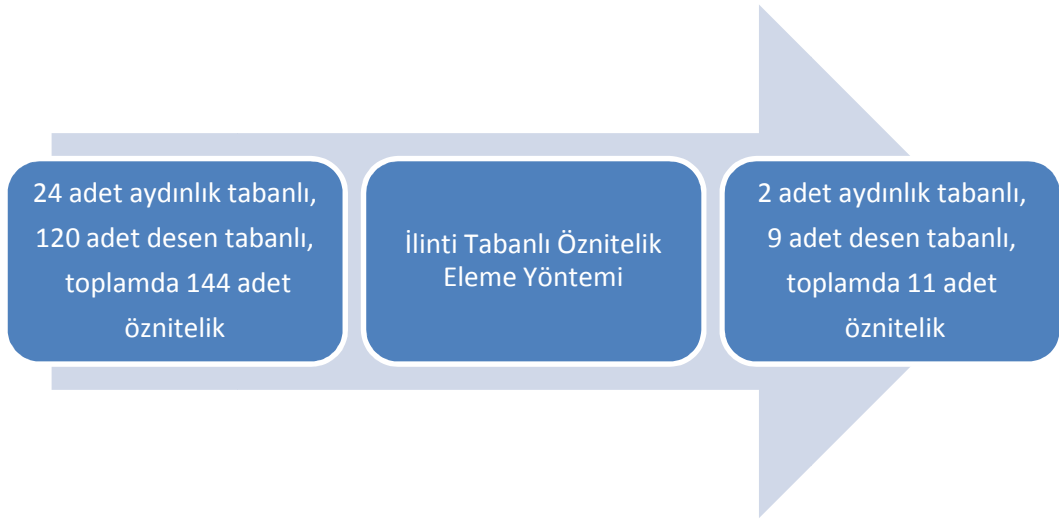
Bu çalışmada, 2 farklı ölçekte asıl resmin toplam 8 farklı dalgacık dönüşümü oluşturulmuş, her bir dönüşümün Denklem-3.19’de verildiği gibi enerjisi hesaplanarak öznitelik olarak kaydedilmiştir. Böylece, RGB ve gri ton kanalların her birinden 8 adet olmak üzere toplamda 32 adet dalgacık dönüşümü enerji özniteliği elde edilmiştir.

3.3. Öznitelik Seçimi

Her bir hücreden 144 adet olmak üzere eğitim veri kümesinde bulunan 384 hücreden toplamda 55296 adet öznitelik hesaplanarak çalışmanın sonraki aşamalarında kullanılmak üzere kayıt altına alınmıştır.

Öznitelik elde etme işlemlerinin sonucunda sınıflayıcıların girişine verilmek üzere 144 adet öznitelik hesaplanmıştır. Ancak giriş veri kümesinin bu kadar yüksek sayıda oluşu sınıflama işleminin hem zamanını hem de işlem karmaşasını arttırmaktadır. Ayrıca sınıflayıcı eğitim veri kümesinde veri miktarı (384) ile değişken sayısı (144) arasında da sınıflayıcı başarımı bakımından bir ilişki bulunmaktadır. Genellikle veri miktarının

değişken sayının 10-30 katı olması gerektiği belirtilmektedir [63, 64]. Değişken sayısının çok olması halinde sınıflayıcı aşırı öğrenme hatasına düşecektir, az olması halinde ise sınıflama başarımı olumsuz etkilenebilir. Bu bilgiler dikkate alındığında sınıflayıcı giriş veri miktarının azaltılması gerektiği açıktır. Bu nedenle çalışmada ilinti tabanlı bir öznitelik eleme yöntemi kullanılmıştır. Bu sayede, Şekil 3.3'te görüldüğü gibi sınıflayıcı için kullanılacak veri giriş uzayı 144 değişkenden sadece 11 değişkene düşürülmüştür. Elde edilen 11 adet seçilmiş öznitelik ve seçilen özniteliklerin kısaltmaları Şekil 3.4'te listelenmiştir.



Şekil 3.3. Öznitelik eleme işlemi sonucu

Seçilmiş Öznitelik Alt Kümesi
• Ortalama aydınlık kırmızı renk kanalında (R-AvInt) • Ortalama aydınlık mavi renk kanalında (B-AvInt) • Birliktelik matrisi tekdüzeliği kırmızı renk kanalında (R-Hom) • Birliktelik matrisi iç-ilinti mavi renk kanalında (B-Auto) • Dalgacık dönüşümü birinci ölçekte resim yaklaştırmasının enerjisi kırmızı kanalda (R-EnAp1) • Dalgacık dönüşümü ikinci ölçekte resim yaklaştırmasının enerjisi kırmızı kanalda (R-EnAp2) • Dalgacık dönüşümü birinci ölçekte resim yaklaştırmasının enerjisi mavi kanalda (B-EnAp1) • Dalgacık dönüşümü ikinci ölçekte resim yaklaştırmasının enerjisi mavi kanalda (B-EnAp2) • Dalgacık dönüşümü ikinci ölçekte köşegensel detay resim enerjisi mavi kanalda (B-EnDg2) • R5L5 Laws desen maskesi sarmaşım enerjisi kırmızı renk kanalında (R-R5L5) • E5L5 Laws desen maskesi sarmaşım enerjisi mavi renk kanalında (B-E5L5)

Şekil 3.4. Öznitelik eleme sonucunda seçilen 11 adet öznitelik ve kısaltmaları

Bu tez çalışmasında kullanılan ilinti tabanlı eleme yönteminin dayandığı hipotez şudur; “iyi bir öznitelik alt kümesi sınıfla yüksek fakat kendi aralarında düşük ilinti gösteren öznitelikleri içerir” [65]. Denklem 3.28’de tanımlanan bu yöntemde,

$$Fayda_s = \frac{\overline{kr_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (3.28)$$

$Fayda_s$, k adet öznitelik içerisinde oluşturulan bir S öznitelik alt kümesinin değerinin bir ölçüsüdür. $\overline{r_{cf}}$ öznitelik ile sınıf arasında ortalama ilinti değeridir ve $\overline{r_{ff}}$ öznitelik-öznitelikler arasında ortalama ilinti değeridir. Eğer, iki öznitelik birbirleri ile yüksek ilinti gösterirse bu özniteliklerin ikisini birden kullanmaya gerek yoktur, bir tanesi atılabilir. Bu eşitlikte, pay bir grup özniteliğin sınıfı tahmin kabiliyetini gösterirken payda öznitelikler arasında ne kadar fazlalığın olduğunun ölçüsüdür. Gerçekte yukarıda verilen formülün Pearson’ın ilinti formülünde değişkenlerin standartlaştırılmış hali olduğu ifade edilmektedir [65].

Çok sayıda özniteliğin birleşik saçılımı, (Denklem 3.29) bağıntısı ile tanımlanmaktadır [66].

$$\sigma_b = k\bar{\sigma} + k(k-1)\overline{r_{ff}\sigma_f\sigma_f} \quad (3.29)$$

Bu bağıntıda; σ_b birleşik saçılmayı, $\bar{\sigma}$ ortalama saçılmayı, k öznitelik sayısını, $\overline{r_{ff}\sigma_f\sigma_f}$ ortalama eş saçılmayı ifade eder. Bir eş saçılma matrisinde k adet saçılma değerinin (matris köşegeninde bulunan eleman sayısı) ve $k(k-1)$ adet eş saçılma değerinin (köşegen dışı elemanların sayısı) hesaplanmış olduğu hatırlanırsa Denklem 3.29 bağıntısı daha kolay anlaşılabilir. Burada, σ_f tek başına bir öznitelik saçılmasını, $\overline{r_{ff}}$ değişkeni ise ikili olarak öznitelikler arasında ortalama ilinti değerini göstermektedir. Eğer, öznitelik değişken değerleri standart Z puanlamasına göre standartlaştırılmış ise ortalama saçılma $\bar{\sigma}$ ve her bir özniteliğin tek başına saçılması σ_f sayısal 1 değerini

alacağından Denklem 3.29 bağıntısı aşağıda verildiği biçimde sadeleştirilebilir (Denklem 3.30).

$$\sigma_b = k + k(k - 1)\overline{r_{ff}} \quad (3.30)$$

Bu bilgi ile Denklem 3.28 bağıntısı değerlendirildiğinde, bağıntının payında bulunan ortalama öznitelik sınıf ilintisinin, paydada bulunan birleşik toplam saçılmaya oranlandığı fark edilebilir. Bağıntı bu haliyle Pearson'ın ilinti formülüne benzemektedir. Ancak farklı olarak bağıntının payında bulunan k katsayısı, aynı sınıflama kabiliyetini vermesine karşın öznitelik sayısı fazla olan öznitelik alt kümesinin tercih edilmesini sağlamaktadır.

Denklem 3.28'de verilen ölçüte göre bir öznitelik kümesinin değerinin ölçülmesi için iki önemli ilinti değişkeninin, $\overline{r_{ff}}$ ve $\overline{r_{cf}}$, incelenen öznitelik kümesi üzerinde hesaplanması gerekmektedir. Sayısal veride ilinti kolaylıkla hesaplanabilmektedir. Ancak, hem sembolik (nominal) hem de sayısal veri türleri arasında ilintinin tekdüze bir yaklaşımla belirlenebilmesi amacıyla simetrik düzensizlik adı verilen bir kazanç bağıntısı kullanılmaktadır [65]. Bu kazanç katsayısı, bir öznitelikteki düzensizlik durumunun bir başka öznitelik bilindiğinde ne kadar değiştiğinin göstergesidir. Bu kazanç katsayısını hesaplamada temel fikir şudur: Gözlenen bir özniteliğin değerleri ikinci bir özniteliğe göre tasnif edildiğinde (kümelendiğinde), birinci özniteliğin düzensizlik değerinde bir azalma olursa iki öznitelik arasında bir ilinti olduğu kabul edilir.

Bir Y öznitelik değişkeni, $y \in Y$ olmak üzere, y değerlerinden oluştuğunda, Y özniteliğinin içinde barındırdığı düzensizlik,

$$H(Y) = - \sum_{y \in Y} p(y) \log_2(p(y)) \quad (3.31)$$

bağıntısı ile hesaplanır (Denklem 3.31). Gözlenen Y özniteliğinin değerleri ikinci bir X özniteliğine göre tasnif edildiğinde Y özniteliğindeki düzensizlik,

$$H(Y|X) = - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)) \quad (3.32)$$

Denklem 3.32’de verilen bağıntıya göre tespit edilir. Eğer Y özniteliği, X özniteliğine göre kümelendiğinde belirsizlik değerinde bir azalma oluyorsa, o zaman Y ve X öznitelikleri arasında bir ilinti vardır. Bu ilintinin düzeyi,

$$kazanç = H(Y) - H(Y|X) \quad (3.33)$$

Denklem 3.33 bağıntısı ile hesaplanır. Bu bağıntı ile hesaplanan ilinti değeri simetriktir. Dolayısıyla,

$$kazanç = H(X) - H(X|Y) \quad (3.34)$$

bağıntısı da (Denklem 3.34) geçerlidir. Simetri, iki öznitelik arasında ilintinin belirlenmesinde dikkate alınmak istenen önemli bir etkidir. Bu nedenle simetrik düzensizlik katsayısı hesaplanırken iki öznitelik arasında görülen *kazanç* sabit 2 katsayısı ile çarpılır. Diğer taraftan, düzensizlik bağıntısına göre hesaplanan bilgi kazancı, değer sayısı fazla olan öznitelik değişkeni lehine ağırlık etkisi oluşturmaktadır. Bu istenmeyen ağırlık etkisini dengelemek için *kazanç* değeri her iki özniteliğin düzensizlik toplamına bölünür. Sonuçta simetrik belirsizlik katsayısı,

$$simetrik\ belirsizlik\ katsayısı = 2 \left(\frac{kazanç}{H(Y) + H(X)} \right) \quad (3.35)$$

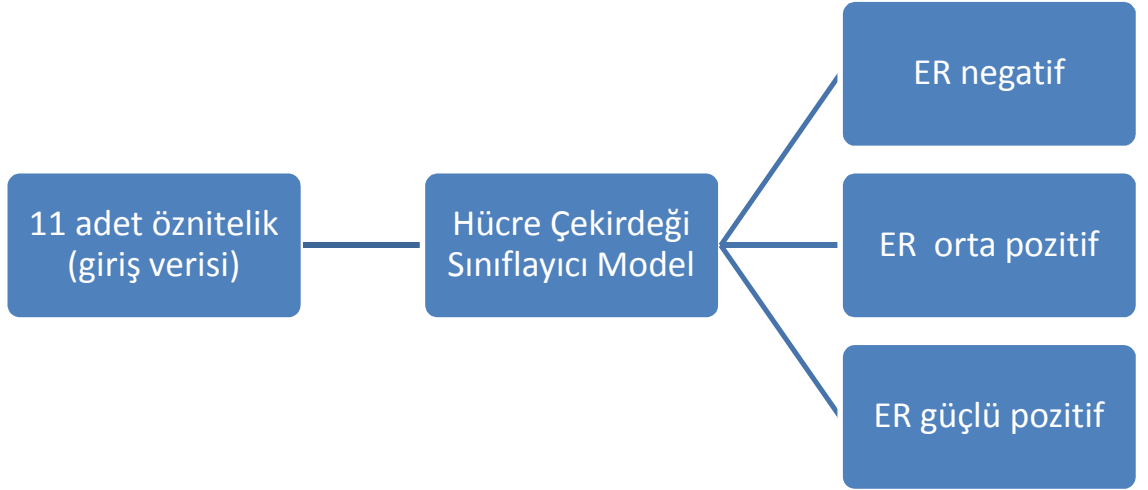
bağıntısı ile belirlenir (Denklem 3.35). Özetle, bu katsayısı, sembolik iki öznitelik arasında hesaplanmak istenen $\overline{r_{ff}}$ ilintisidir. Aynı biçimde, öznitelik ve sınıf değişkeni arasında ilinti değeri $\overline{r_{cf}}$ hesaplanabilir. Böylece tüm öznitelikler arasında ikili olarak ve her bir öznitelik ile sınıf değişkeni arasında ikili olarak ilinti değerleri hesaplanır. Bu ilinti değişkenleri, Denklem 3.28'de yerine konarak tek bir özniteliğin mevcut bir öznitelik altkümesine faydası, kazanç etkisi, ölçülmektedir.

Bu tez çalışmasında, uygun öznitelik alt kümesinin araştırılması, ileri doğru seçme araması yapılarak gerçekleştirilmiştir. Buna göre önce boş bir alt küme oluşturulur. Daha sonra sırayla her bir özniteliğin tek başına mevcut kümeye faydası araştırılır. En yüksek faydayı sağlayan öznitelik boş kümeye eklenerek tek elemanlı bir öznitelik alt kümesi elde edilir. Sonra, geriye kalan diğer özniteliklerin tek tek mevcut alt kümeye faydası araştırılır. Yine en yüksek fayda sağlayan öznitelik mevcut öznitelik alt kümesine eklenerek seçilen öznitelik kümesi genişletilir. Mevcut öznitelik alt kümesinin genişletilmesine geriye kalan öznitelikler elde edilen en yüksek fayda değerini arttıramaz hale gelinceye kadar devam edilir. Böylece en uygun öznitelik alt kümesi oluşturulur ve öznitelik eleme işlemi tamamlanmış olur.

4. BÖLÜM

YAPAY ÖĞRENME VE SINIFLAMA YÖNTEMLERİ

Bölütleme aşamasında immünohistokimyasal resim üzerinde tespit edilen hücre çekirdeklerinin östrojen alıcı (ER) pozitif veya negatif oluşuna sınıflayıcı model karar vermektedir. Bu tez çalışmasında kullanılan sınıflayıcı model, öznelilik eleme sonrasında 11 değişkene indirgenen öznelilik veri kümesini giriş değişkenleri olarak alır ve çıkışta ER negatif, ER orta pozitif ve ER güçlü pozitif olmak üzere 3 sınıftan birisine atama yapar (Şekil 4.1).



Şekil 4.1. Sınıflayıcı giriş ve çıkışı

Bilgisayar tabanlı tanı sistemi, tüm hücre çekirdekleri için sınıflayıcı modelin kararını alır. Buna göre o resmin temsil ettiği biyopsi örneği hakkında Allred puanı hesaplanarak östrojen alıcı durum tanısı elde edilir.

Tez çalışmasının bu bölümünde sınıflayıcı model olarak başarıımı incelenen yöntemler tanıtılmıştır. Bu yöntemler şunlardır; k -orta nokta sınıflayıcı, yarıçap taban fonksiyonlu ağlar, k -en yakın komşuluk algoritması, destek vektör makineleri, sade bayes sınıflayıcı, ve fonksiyonel karar ağaçları. Bu sınıflayıcı modellerin tanı sisteminde kullanılması ile elde edilen bulgular sonraki bölümlerde verilecektir.

4.1. Öklid Uzaklık Ölçütü

Birçok yapay öğrenme yönteminde karar fonksiyonlarında kararı etkileyen ölçülebilir bir etken olarak Öklid uzaklık ölçütü kullanılmaktadır. Bu nedenle, yapay öğrenme yöntemlerinden önce kısaca bu uzaklık ölçütü sunulmuştur. Öklid mesafesi D_E , genellikle N boyutlu bir vektör uzayında mesafe ölçümü yapmak için kullanılır ve Denklem 4.1’de verilen bağıntı ile tanımlanır. Üç boyutlu RGB renk uzayında mesafe hesaplaması Denklem 4.2’de gösterilmiştir.

$$D_E(\vec{v}_1, \vec{v}_2) = \|\vec{v}_1 - \vec{v}_2\| \quad (4.1)$$

$$D_E(\vec{v}_1, \vec{v}_2) = \sqrt{(\vec{v}_{1,1} - \vec{v}_{2,1})^2 + (\vec{v}_{1,2} - \vec{v}_{2,2})^2 + (\vec{v}_{1,3} - \vec{v}_{2,3})^2} \quad (4.2)$$

Bu eşitlikte, $\vec{v}_1 = [v_{1,1} \ v_{1,2} \ v_{1,3}]^T$ olmak üzere her bir bileşeni ile R, G ve B renk kanallarında aydınlık şiddetini tanımlayan bir vektördür. RGB renk uzayında Öklid mesafesi aydınlık şiddeti değişimlerine karşı duyarlıdır. Bu yüzden bu yöntemin bir tanıma ve sınıflandırma aracı olarak kullanılabilmesi için standart bir aydınlatma ortamı içerisinde verilerin elde edilmiş olması gerekir [67, 68].

4.2. K-Orta Nokta Algoritması

K -orta nokta algoritması danışmansız bir öğrenme algoritmasıdır [69]. Bu yöntem, bir kümeyi k adet farklı sınıfa ayırmayı hedefler. Bunun için veri kümesi k adet merkez etrafında öbeklenir. Başlangıçta, algoritma bu merkez noktalarını rastgele fakat mümkün olduğunca birbirinden uzak noktalarda belirler. Daha sonra kümenin her bir

noktası mevcut merkezlere olan uzaklığı dikkate alınarak kendisine en yakın öbek merkezine atanır. Buna göre yeni öbekler oluşturulur. Oluşan yeni öbeklerin merkezleri yeniden hesaplanır ve bir sonraki işlem basamağı için yeni merkezler belirlenir. Bir döngü içerisinde yeni öbeklerin ve bu öbeklere ait yeni merkez noktalarının hesaplanmasına devam edilir. Merkez noktaları değişmez oluncaya kadar bu döngü sürdürülür.

Algoritmayı, bir en küçük değer bulma problemi olarak tanımlamak mümkündür [68]. $X = \{x_1, x_2, \dots, x_n\}$ olmak üzere sınıflandırılacak noktalar kümesini temsil etsin. $V = \{v_1, v_2, \dots, v_c\}$ kümesi her bir öbeğin -sınıfın- merkez noktalarını göstere. Bu durumda K -orta nokta algoritmasının amaç fonksiyonu $J(V)$ aşağıda verilen karesel hata fonksiyonu biçiminde tanımlanabilir (Denklemler 4.3).

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} \|x_{ij} - v_i\|^2 \quad (4.3)$$

Bu amaç fonksiyon içerisinde, $\|x_{ij} - v_i\|^2$ değeri, x_{ij} ve v_i noktaları arasındaki Öklid mesafesinin karesidir. c_i her bir i sınıfındaki noktaların sayısıdır. i 'inci merkez noktası v_i şu formülle hesaplanır.

$$v_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_{ij}, \quad i = 1 \dots c \quad (4.4)$$

Öbek merkezi v_i 'nin hesaplanış biçimi Denklem 4.4'de görüldüğü üzere öbek ağırlık merkezinin hesaplanmasına eşittir. Algoritma yordamı kısaca aşağıdaki adımlarla özetlenebilir.

Adım-1) Araştırma uzayında rastgele k adet merkez noktası belirle.

Adım-2) Her nokta ile bu merkez noktaları arasındaki mesafeyi hesapla.

Adım-3) Her noktayı en kısa mesafe ölçüsüne bir merkez noktasına ata.

Adım-4) Merkez noktaları yeni öbeklere göre yeniden hesapla.

Adım-5) Merkez noktalar değişmez oluncaya kadar adım 2, 3 ve 4'ü tekrarla.

4.3. Yarıçap Taban Fonksiyonlu Ağlar

Bir merkez noktasından uzaklığına bağlı olarak çıkış üreten fonksiyonlara yarıçap tabanlı fonksiyonlar denir. Giriş değişkeni bir vektör olabilir, uzaklık ölçümü genellikle Öklid mesafesi ile tanımlanır.

Giriş vektörü $\vec{x}$ 'in bir $\vec{c}$ merkezinden uzaklığına bağlı çıktı veren yarıçap tabanlı bir uzaklık fonksiyonu ϕ , aşağıda verildiği gibi gösterilebilir (Denklem 4.5). Burada $\|\vec{x} - \vec{c}\|$, Öklid mesafesi olarak belirtilen uzaklığı ifade eder.

$$\phi(\vec{x}, \vec{c}) = \phi(\|\vec{x} - \vec{c}\|) \quad (4.5)$$

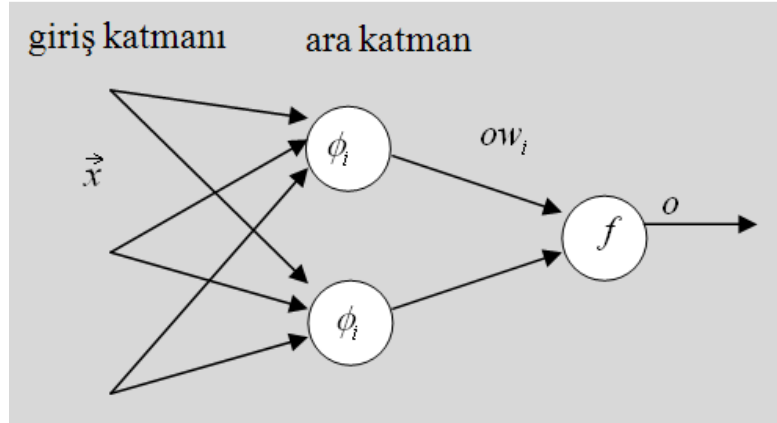
İletim fonksiyonu, yarıçap tabanlı fonksiyonlarından oluşan bir sinir ağına, Yarıçap Tabanlı Fonksiyon Ağı (RBFN) adı verilir. Bir RBFN yapısında; giriş, ara ve çıkış katmanları olarak üç katman bulunur. Tipik sinir ağlarından farklı olarak ara katman sayısı sadece birdir. Ara katmanda bulunan her bir nörona bir tür yarıçap tabanlı iletim fonksiyonu atanmıştır. Genellikle, Gauss fonksiyonu yarıçap tabanlı iletim fonksiyonu olarak kullanılır.

Bu ağ yapısında, giriş katmanı sinirlerinin ağırlık katsayı değeri yoktur. Ara katman nöronları giriş değişkeninin veri uzayında temsil ettiği noktanın konuma göre çıkış üretirler. Giriş verisinin temsil ettiği nokta, nöronun yarıçap fonksiyon merkezine ne kadar yakın olursa o kadar büyük bir çıkış değeri üretilmektedir. Sonuçta her bir nöronun yarıçap fonksiyon merkezi değiştirilerek giriş verisinin Gauss fonksiyonlarından oluşan bir yaklaştırması yapılmaktadır. Temel bir RBFN ağında ara katmanda bulunan nöron sayısı toplam veri sayısına eşittir. Geliştirilmiş bir RBFN ağında nöron sayısı veri miktarından çok daha az olabilir [68].

RBFN ağları eğitim esnasında; (a) yarıçap fonksiyon merkezi ve genişliğini (b) ara katman ve çıkış katmanları arasında bulunan sinirlerin ağırlık katsayılarını öğrenir. Nöron

sayısı, yarıçap fonksiyon merkezi ve varyansı K -orta nokta adıyla bilinen danışmansız öğrenme yöntemi ile belirlenebilir. Daha sonra, yarıçap fonksiyon parametreleri sabit tutularak, sinir ağırlık katsayıları doğrusal veya logaritmik tabanda doğrusal (logistic) yaklaştırma yöntemleri ile hesaplanır. Yarıçap taban fonksiyonu parametreleri ve sinir ağırlık katsayı hesaplamaları birbirinden bağımsızdır [70].

Ayrıca, Geri Yayılım Algoritması'nda olduğu gibi bahse konu tüm parametre değerleri, hata fonksiyonunun kısmi türevleri üzerinden en dik iniş yöntemiyle belirlenebilir [64]. İletim fonksiyonu olarak Gauss fonksiyonu kullanan Şekil 4.2'deki gibi bir ağ yapısında çıkış yarıçap fonksiyonu aşağıda verilmiştir (Denklem 4.6).



Şekil 4.2. Basit bir RBFN ağı

$$\phi_i(\vec{x}) = e^{-\frac{1}{2} \frac{\|\vec{x} - \vec{\mu}_i\|^2}{\sigma_i^2}} = \exp\left(-\frac{1}{2} \frac{\|\vec{x} - \vec{\mu}_i\|^2}{\sigma_i^2}\right) \quad (4.6)$$

Burada $\vec{\mu}_i$, i 'inci yarıçap fonksiyonu merkez noktasını ve σ_i^2 varyansını gösterir. Giriş değişkeni $\vec{x}$ vektörüdür. Elde edilen $\phi_i(\vec{x})$ çıktısı, çıkış katmanını ağırlık katsayıları ile çarpılır ve sonuçta ağ çıktısı Denklem 4.7'te verildiği gibi elde edilir.

$$o = \sum_{i=1}^K ow_i \cdot \phi_i(\vec{x}) \quad (4.7)$$

Burada ow_i , çıkış katmanı sinirlerinin ağırlık katsayılarını ve o ise ağ çıkışını ifade eder. Ağırlık katsayılarının güncellenmesi için gradiyent vektörüne dayalı bir yaklaşım benimsenebilir. Bu durumda hata yüzeyinin ağırlık katsayılarına göre gradiyent vektörü elde edilir (Denklem 4.8).

$$\frac{\partial E}{\partial ow_i} = \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial o} \cdot \frac{\partial o}{\partial ow_i} = -e \cdot \phi_i(\vec{x}) = (d - o) \cdot \phi_i(\vec{x}) \quad (4.8)$$

Tıpkı MLP ağlarında olduğu gibi; E hata yüzeyini, e hatayı, o ağ çıkışını ve d arzu edilen ağ çıkış değerini ifade etmektedir. Hata yüzeyinin bayır vektörü yönündeki değerini elde ettikten sonra ağırlık katsayısındaki değişim

$$\Delta ow_i = -\eta \cdot \frac{\partial E}{\partial ow_i} = \eta \cdot (d - o) \cdot \phi_i(\vec{x}) \quad (4.9)$$

olarak hesaplanır (Denklem 4.9). Sonuçta ağırlık katsayıları Denklem 4.10 bağıntısına göre hesaplanır.

$$\begin{aligned} ow_i(n+1) &= ow_i(n) + \Delta ow_i \\ &= ow_i(n) + \eta \cdot (d - o) \cdot \phi_i(\vec{x}) \end{aligned} \quad (4.10)$$

Benzer biçimde danışmanlı öğrenme yaklaşımı ile yarı çap fonksiyonu merkez noktası aşağıda tanımlanan Denklem 4.11 uyarınca güncellenebilir.

$$\frac{\partial E}{\partial \vec{\mu}_i} = \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial o} \cdot \frac{\partial o}{\partial \phi_i(\vec{x})} \cdot \frac{\partial \phi_i(\vec{x})}{\partial \vec{\mu}_i} \quad (4.11)$$

Denklem 4.11 bağıntısında yarıçap fonksiyonu $\phi_i(\vec{x})$ 'in yarıçap merkez noktası μ_i 'e göre kısmi türevinin

$$\frac{\partial \phi_i(\vec{x})}{\partial \vec{\mu}_i} = \frac{\|\vec{x} - \vec{\mu}_i\| \cdot \phi_i(\vec{x})}{\sigma_i^2} \quad (4.12)$$

olduğu (Denklem 4.12) dikkate alınır,

$$\frac{\partial E}{\partial \mu_i} = -e \cdot ow_i \cdot \frac{\|\vec{x} - \vec{\mu}_i\|}{\sigma_i^2} \cdot \phi_i(\vec{x}) \quad (4.13)$$

gradiyent değeri Denklem 4.13 bağıntısı ile elde edilir. Bu durumda, merkez değeri güncellemesi Denklem 4.14 bağıntısına göre ve yeni merkez değeri ise Denklem 4.15 bağıntısına göre yapılır.

$$\Delta \mu_i = -\eta \cdot \frac{\partial E}{\partial \mu_i} = \eta \cdot e \cdot ow_i \cdot \frac{\|\vec{x} - \vec{\mu}_i\|}{\sigma_i^2} \cdot \phi_i(\vec{x}) \quad (4.14)$$

$$\begin{aligned} \mu_i(n+1) &= \mu_i(n) + \Delta \mu_i \\ &= \mu_i(n) + \eta \cdot (d - o) \cdot ow_i \cdot \frac{\|\vec{x} - \vec{\mu}_i\|}{\sigma_i^2} \cdot \phi_i(\vec{x}) \end{aligned} \quad (4.15)$$

RBFN ağlarının zayıf tarafı, tüm özniteliklere eşit ağırlık değeri verilmesidir. Uzaklık ölçütü her öznitelik için eşit biçimde işlemektedir. Bu nedenle RBFN ağlarının ilgisiz veya etkisiz özniteliklerin üstesinden gelmede MLP ağlarına göre yetersiz kaldığı görülmektedir. Destek Vektör Makineleri (SVM) de aynı zayıflığı göstermektedir. Aslında, Gauss ağırlık fonksiyonunu kullanan bir SVM, RBFN ağının özel bir türüdür ve bu türde öğrenim kümesindeki her bir veri için bir yarıçap fonksiyonu merkezi konumlandırılmış ve ağ çıkışı bu fonksiyonların doğrusal toplamı olarak elde edilmiştir. Bu türde sadece bazı YTF fonksiyonları sıfır-olmayan sonuçlar üretir, bu fonksiyonlar destek vektörleri üzerine konumludur [70].

RBFN ağları, K -orta nokta öbekleme yöntemine ve olasılık tabanlı sinir ağlarına benzerlik göstermektedir. Olasılık tabanlı ağlar her bir giriş verisi için ara katmanda ayrı bir nöron kullanırken RBFN ağları giriş verisine oranla daha az sayıda nöron kullanmaktadır.

4.4. K-En Yakın Komşuluk Algoritması

4.4.1. Olasılık Yoğunluk Fonksiyonu Tahmini

Danışmanlı öğrenme yöntemlerinde olasılık tabanlı yaklaşımlardan faydalanılmaktadır. Eğer giriş ve çıkış arasında olasılık tabanlı bir bağıntı varsa, bu bağıntıyı sağlayan olasılık yoğunluk fonksiyonun biçimi önceden bilinmesi halinde parametrik yöntemlerle olasılık yoğunluk fonksiyonu (oyf) tahmin edilebilir. Biçimi önceden bilinmeyen bir oyf söz konusu ise parametrik olmayan yöntemler sayesinde oyf tahmin edilebilir. Parametrik olmayan yöntemlerde iki ayrı temel yaklaşım bulunmaktadır. Birincisinde, sıklık diyagramları diğesinde ise ağırlık fonksiyonları kullanılır. Her iki parametrik olmayan yaklaşımda da, veri uzayının oyf biçimi, eğitim veri kümesini oluşturan sınırlı sayıda veriden, -örneklem verisinden- faydalanarak tahmin edilir.

Sıklık diyagramlarında; örneklemin alındığı veri uzayı eşit aralıklara bölünür ve her aralığa düşen örnek miktarı kayıt edilerek bir dağılım elde edilir. Elde edilen bu dağılım basittir ancak süreksizdir ve ayrık değerlidir. Bu nedenle söz konusu dağılım üzerinde türeve dayalı işlemler gerçekleştirilemez.

Öte yandan ağırlık fonksiyonuna dayalı yaklaşımda, örnekleme oluşturan her bir verinin temsil ettiği nokta bir ağırlık fonksiyonun merkezi olmak kaydıyla çok sayıda ağırlık fonksiyonu toplanarak genel bir fonksiyon elde edilir. Bu genel fonksiyon örneklem verisi üzerinden tahmin edilen oyf olur.

4.4.2. Parzen Pencere Fonksiyonları

Sürekli bir oyf, $p(x)$, aşağıda verilen şu özellikleri sağlamaktadır [64, 71].

- Bir x noktasının a ve b noktaları arasında bulunma olasılığı $P(a < x < b) = \int_a^b p(x)dx$ olur.
- Bütün x değerleri için oyf pozitif olmalıdır.
- Olasılık yoğunluk fonksiyonun toplam değeri $\int_{-\infty}^{+\infty} p(x)dx = 1$ olmalıdır.

En çok bilinen oyf fonksiyonu, aşağıda verilen Gauss fonksiyonudur (Denklem 4.16).

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \quad (4.16)$$

Burada, oyf'nin merkez noktasını c , standart sapmasını ise σ belirlemektedir. Mevcut tanımlamalar çok boyutlu bir vektör uzayı için uyarlandığında,

- Bir $\vec{x}$ noktasının çok boyutlu bir R kapalı bölgesinde bulunma olasılığı

$$P = \int_R p(\vec{x}) d\vec{x} \text{ olur.}$$

- Bütün $\vec{x}$ değerleri için oyf pozitif olmalıdır.
- Olasılık yoğunluk fonksiyonun toplam değeri $\int p(\vec{x}) d\vec{x} = 1$ olmalıdır.

Mevcut örneklemin n adet veriden oluşan $\vec{x}$ vektör kümesi olması halinde, oyf'yi tahmin etmek mümkündür. Böylece, örnek bir $\vec{x}$ vektörünün $p(\vec{x})$ olasılık değeri tahmin edilebilir. Bir $\vec{x}$ vektörünün kapalı bir R bölgesinde bulunma olasılığı P ,

$$P = \int_R p(\vec{x}) d\vec{x} \quad (4.17)$$

formülü ile belirlenir (Denklem 4.17). Eğer R bölgesinin $p(\vec{x})$ değişiminin ihmal edileceği kadar çok küçük bir bölgeyi kapsadığı varsayılırsa,

$$P = \int_R p(\vec{x}) d\vec{x} \approx p(\vec{x}) \int_R d\vec{x} = p(\vec{x})V \quad (4.18)$$

eşitliği elde edilir (Denklem 4.18). Bu eşitlikte V , çok boyutlu uzayda R bölgesinin sınırladığı kesimdir. Söz konusu kesim tek boyutlu uzayda bir uzaklık $-h-$, iki boyutlu uzayda bir alan $-h^2-$, boyutu d olan daha yüksek boyutlu uzayda bir hacim $-h^d-$ tanımlar.

R bölgesinin kenar uzunluğu h olan bir küp hacmini kapsadığı varsayalım. Bir örneklemin bu kapalı küp hacmi içerisinde bulunup bulunmadığını ağırlıklandıran, bulunma durumunda 1 ve bulunmama durumunda 0 sonucunu veren bir ağırlık fonksiyonu şu şekilde tanımlanabilir (Denklem 4.19).

$$\phi\left(\frac{\vec{x}_i - \vec{x}}{h}\right) = \begin{cases} 1 & \frac{|\vec{x}_i - \vec{x}|}{h} \leq \frac{1}{2} \\ 0 & \text{diğer durumda} \end{cases} \quad i = 1, \dots, n \quad (4.19)$$

Bu eşitlikte ağırlık fonksiyonu ϕ , üstküpün (hypercube) merkez noktası $\vec{x}$ ve n adet örneklemden her biri $\vec{x}_i$ ile gösterilmiştir. Kapalı R bölgesine düşen toplam örnek sayısı tüm örneklem için ağırlık fonksiyonu çıktılarının toplamı alınarak belirlenebilir (Denklem 4.20).

$$k = \sum_{i=1}^n \phi\left(\frac{|\vec{x}_i - \vec{x}|}{h}\right) \quad (4.20)$$

Diğer taraftan, birbirinden bağımsız n adet veriden oluşan örneklemde k adet veri noktasının kapalı bir R bölgesinde bulunduğu varsayılırsa yeni bir örnek vektör $\vec{x}$ 'in bahse konu kapalı R bölgesinde bulunma olasılığı örneklem verisi k ve n üzerinden tahmin edilebilir (Denklem 4.21).

$$P = k/n \quad (4.21)$$

Denklem 4.18, 4.20 ve 4.21 birleştirilerek tahmin edilmek istenen oyf iki boyutlu bir veri uzayında

$$p(\vec{x}) = \frac{k/n}{V} = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} \phi\left(\frac{\vec{x}_i - \vec{x}}{h}\right) \quad (4.22)$$

olarak hesaplanır (Denklem 4.22). Eğer ağırlık fonksiyonu olarak sürekli olan ve türevlenebilen bir fonksiyonu kullanılırsa bu fonksiyona Parzen penceresi adı verilir. Tipik bir Parzen penceresi olan Gauss fonksiyonu ile tek boyutlu uzayda oyf tahmini,

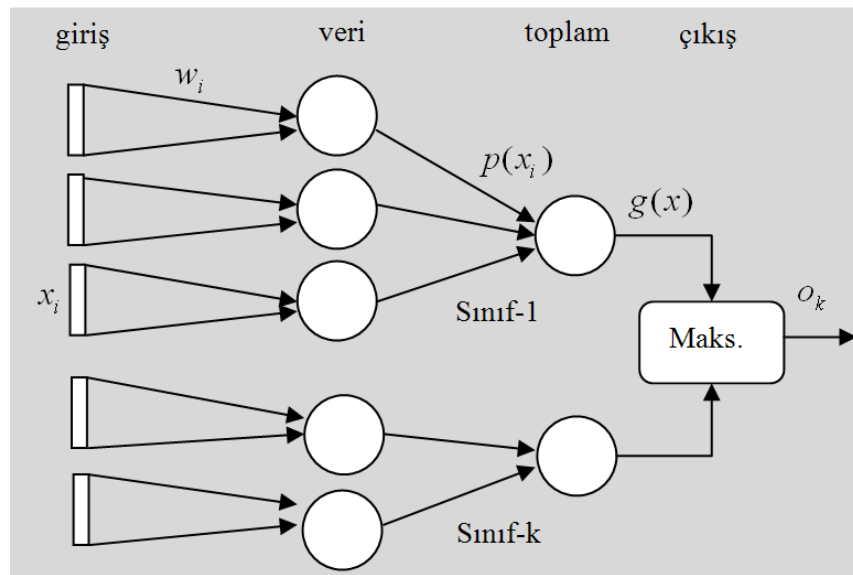
$$p(\vec{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\vec{x}_i - \vec{x})^2}{2\sigma^2}\right) \quad (4.23)$$

elde edilebilir (Denklem 4.23). Burada elde edilen sonuç Gauss eğrilerinin toplamından oluşan bir yoğunluk fonksiyonudur. Her bir Gauss eğrisinin merkezi $\vec{x}_i$ veri noktasıdır ($i = 1, \dots, n$). Örneklem kümesinde n adet veri bulunduğundan yoğunluk fonksiyonu n adet Gauss eğrisinin birleşimiyle oluşturulur. Bu yaklaşım d boyuttan oluşan çok boyutlu veri uzayı için aşağıdaki gibi genelleştirilebilir (Denklem 4.24).

$$p(\vec{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(2\pi)^{1/d} \sigma^d} \exp\left(-\frac{(\vec{x}_i - \vec{x})^T (\vec{x}_i - \vec{x})}{2\sigma^2}\right) \quad (4.24)$$

4.4.3. Olasılık Tabanlı Sinir Ağı Modeli

Olasılık tabanlı sinir ağı, eğitim veri kümesinde tanımlanan veri noktalarını merkez alan Parzen pencereleri ile olasılık tahmini yapan bir modeldir. Yaygın olarak Denklem 4.24'te verilen Parzen penceresi bilinmeyen bir oyf'yi Gauss eğrileriyle tahmin etmekte kullanılmaktadır. Bu model dört temel katmandan oluşur; giriş, veri, toplam ve çıkış katmanı (Şekil 4.3).



Şekil 4.3. Olasılık tabanlı yapay sinir ağı modeli

Giriş katmanında, her biri d boyutlu n elemandan oluşan giriş veri kümesinden biri giriş $\vec{x}_n$ alınır, normalleştirilerek ($\vec{x}_n = \vec{x}_n / \|\vec{x}_n\|$) veri katmanına aktarılır. Giriş ve veri katmanı arasında bulunan sinir bağlantıları normalleştirilmiş giriş verisine eşitlenir. Bu bağlantılar sadece bilgiyi bir sonraki katmana iletme görevini üstlenir. MLP ya da RBFN ağ yapısında olandan farklı olarak sinir ağırlık katsayılarının bir daha değiştirilmesine gerek duyulmamaktadır.

Veri katmanında, eğitim kümesinde bulunan her bir veri için bir nöron oluşturulur. Her nöron çıkışında bir olasılık yoğunluk fonksiyonu oluşturulur. Bu yoğunluk fonksiyonun merkez noktası, nörona tahsisli olarak iletilen giriş verisiyle temsil edilen noktadır. Veri katmanında bulunan nöronlar temsil ettikleri verinin sınıfına göre alt gruplara ayrılır. Dolayısıyla giriş veri kümesinde c adet sınıfın tanımlı olduğu bir durumda veri katmanında da c adet gruplanmış nöron kümesi bulunur.

Toplam katmanında aynı sınıfa ait nöronların oluşturduğu yoğunluk fonksiyonları toplanarak sınıfı tümleşik olarak temsil eden genel bir oyf elde edilir. Bu toplam, sınıfta bulunan veri adedine bölünerek sınıf için toplamın ortalaması hesaplanır. Bu işlem, veri katmanında bulunan her sınıf için ayrı ayrı tekrarlanır.

Çıkış katmanında, toplam katmanında elde edilen ortalama sonuçları karşılaştırılarak bilinmeyen veriyi en yüksek olasılıkla içeren oyf fonksiyonu dolayısıyla sınıf belirlenir. Belirlenen sınıf olasılık tabanlı ağı sonuç çıktısı olur.

Bu ağ modelinde Parzen pencereleri elde edilen oyf fonksiyonlarını ortalaması karşılaştırıldığından karar çıktısı bir kıyaslama işlemidir. Bu nedenle Parzen pencerelerini kullanılan tüm fonksiyonlarda ortak olan, değeri eşit ve sabit olan katsayılar ihmal edilerek karar fonksiyonu basitleştirilebilir. Eşitlik 4.24'te tanımlanan Parzen penceresinde normalleştirme işleminin üstlü ifadede oluşturduğu ve aşağıda verilen sadeleştirme

$$\begin{aligned}
 (\vec{x}_i - \vec{x})^T (\vec{x}_i - \vec{x}) &= \|\vec{x}\|^2 + \|\vec{x}_i\|^2 - 2\vec{x}_i^T \vec{x} \\
 &= 2 - 2\vec{x}_i^T \vec{x}
 \end{aligned} \tag{4.25}$$

dikkate alınır ve karar oyf tahminlerinde sabit katsayılar göz ardı edilebilir (Denklem 4.25). Böylece, aynı sınıfa ait nöronların tümünün birlikte oluşturduğu sadeleştirilmiş karar fonksiyonu ortaya çıkar [64, 71].

$$g(\vec{x}) = \frac{1}{m} \sum_{i=1}^m \exp\left(\frac{\vec{x}_i^T \vec{x} - 1}{\sigma^2}\right) \quad (4.26)$$

Burada, (Denklem 4.26), m aynı sınıfta bulunan veri (nöron) miktarını ifade etmektedir. Bu sabit her sınıf için farklı değerler alabilir.

4.4.4. K -En Yakın Komşuluk Algoritması

Eleman sayısı n olan örneklem verisinden oluşan bir uzayda, merkezi $\vec{x}$ vektörü ile tanımlanan V hacimli bir kesim içerisinde k adet örneklem bulunduğu ve bunlardan k_i adet örneklemin toplamda n_i elemanı olan C_i sınıfına ait olduğu biliniyor olsun.

Bu durumda, örneklem verisi içinde bir veri noktasının C_i sınıfına ait olma olasılığı

$$p(C_i) = n_i/n \quad (4.27)$$

bağıntısı ile tahmin edilir (Denklem 4.27). Bir veri noktasının kapalı V bölgesinde bulunma olasılığı Parzen pencere yöntemi ile oyf tahmini konusunda açıklandığı üzere

$$p(\vec{x}) = k/n.V \quad (4.28)$$

bağıntısı (Denklem 4.28) ile yaklaşık olarak belirlenir [71, 72]. Ayrıca, C_i sınıfına ait bir verinin x merkezli V bölgesinde bulunma olasılığı (Denklem 4.29)

$$p(\vec{x}, C_i) = k_i/n.V \quad (4.29)$$

olarak ifade edilebilir [71]. Bu temel veriler Bayes teorimi ile yorumlanarak bir veri noktasının ait olduğu C_i sınıfı bilindiğinde bu noktanın kapalı V bölgesinde bulunma

olasılığı $p(\vec{x}|C_i)$ ve veri noktasının V bölgesinde bulunduğu bilindiğinde bu veri noktasının bir C_i sınıfına ait olma olasılığı $p(C_i|\vec{x})$ hesaplanabilir.

$$p(C_i|\vec{x}) = p(\vec{x}, C_i)/p(\vec{x}) = (k_i/n.V)/(k/n.V) = k_i/k \quad (4.30)$$

Koşullu olasılık bağıntıları ile elde edilen bu çıkarım (Denklem 4.30), herhangi bir x noktasının C_i sınıfına ait olma olasılığını kapalı bölgede o sınıfa eleman sayısı k_i ile doğru orantılı olduğunu göstermektedir. En yakın komşu sınıflayıcısı en yüksek olasılığı sağlayan i indeksli sınıfı karar olarak seçecektir.

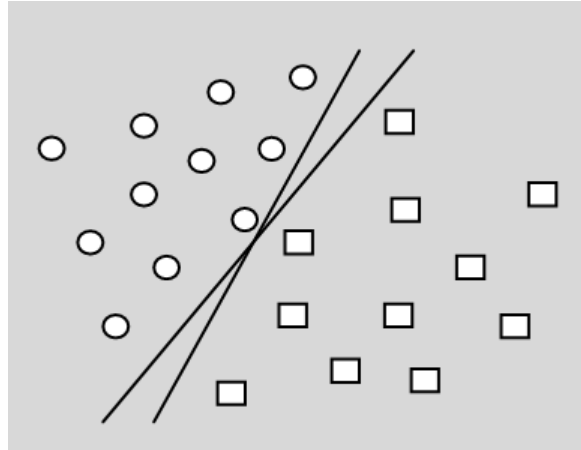
V bölgesinin hacmi değiştirilerek k sayısını değiştirmek mümkündür. Parzen penceresi yaklaşımında genel olarak bu yol benimsenir ve $V = 1/\sqrt{n}$ olarak seçilir. Öte yandan, istenen k miktarında elemanı içine alacak biçimde bir V hacmi genişletilebilir. En yakın k -komşu sınıflayıcısında benimsenen yaklaşım budur. Bu yaklaşımda genellikle $k = \sqrt{n}$ eşitliği uygulamada kullanılır [71].

4.5. Destek Vektör Makineleri

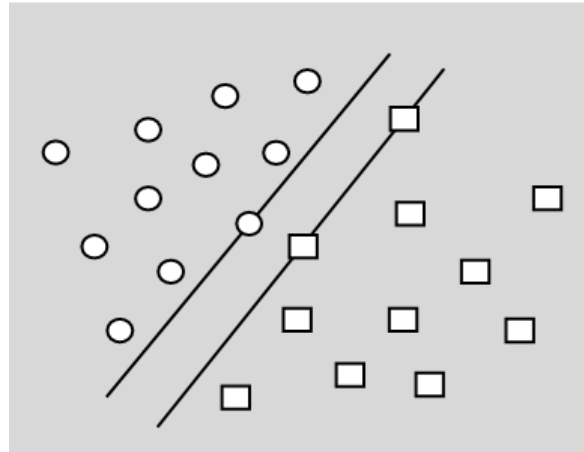
Destek vektörleri doğrusal olarak birbirinden ayrılabilen iki sınıf veri kümesini sınıflandırmak amacıyla kullanılır. Doğrusal ayırıcı; iki boyutlu uzayda bir doğrudur, üç boyutlu uzayda bir düzlem halini alır. Daha yüksek boyutlu uzaylarda doğrusal ayırıcı görüntü olarak hayal edilemez ancak kuramsal olarak varlığı kabul edilir. Kuramsal olarak varlığı kabul edilen bu doğrusal ayırıcılara üstdüzlem denilmektedir [71].

Her bir doğrusal ayırıcı, çok boyutlu uzayda, uzay boyutuna eşit boyutlu bir vektörle temsil edilir. Bu nedenle doğrusal ayırıcıyı hesaplama gayreti aslında uygun vektörü hesaplama arayışıdır. Fakat temelde, doğrusal olarak ayrılabilen iki farklı tür veriyi birbirinden ayırabilecek sonsuz sayıda farklı ayırıcı bulmak mümkündür. Bu durum Şekil 4.4'te gösterilmiştir.

O halde en uygun olan ayırıcı nasıl bulunabilir? Bunun için kullanılabilecek bir yaklaşım; Şekil 4.5'te gösterildiği gibi her iki veri kümesini aralarında en fazla boşlukla ayıran doğrusal ayırıcı ile ayırmak olabilir.



Şekil 4.4. Farklı doğrusal ayırıcılar

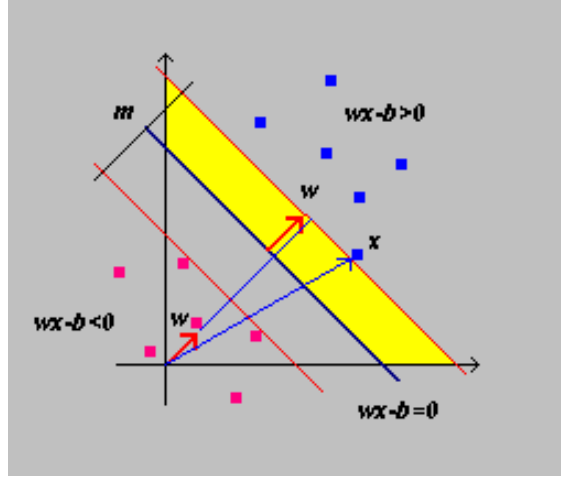


Şekil 4.5. En uygun doğrusal ayırıcı

Bu durumda Şekil 4.6'da gösterildiği üzere doğrusal ayırıcıyı ifade eden vektör $\vec{w}$ 'nin yanı sıra, arada kalan boşluğu (merkez ve ayırıcı doğru arasında dik uzaklığı) ifade eden m 'in birer değişken olarak bir amaç fonksiyonu içinde tanımlanması ve bu fonksiyon için en uygun $\vec{w}$ ve m değerlerinin hesaplanması gerekmektedir.

Amaç fonksiyonda hesaba katılacak kısıt, iki farklı veri kümesinin sınıflanmış olmasıdır. Bir doğruyu, düzlemi yahut bir üstdüzlemi temsil eden vektör bir veri kümesinin tüm noktalarını diğerinden ayırıyorsa vektörün o veri kümesini desteklediği kabul edilir. O halde sınıflama fonksiyonunda kullanılacak olan vektör, destek vektörü olma vasfına sahip olmalıdır. Sınıflama fonksiyonu her iki veri kümesini geometrik ve cebirsel işlemler sonrasında birbirinden ayırt edebilmelidir. Bunun için vektörler arası

noktasal çarpımı kullanılabilir. Bir noktanın ayırt edicinin ne tarafında kaldığı başka bir ifadeyle hangi veri sınıfına ait olduğu bu yolla hesaplanabilir.



Şekil 4.6. Noktasal çarpımla sınıf tespiti

Veri kümesinde bir noktayı ifade eden $\vec{x}$ özellik vektörü ayırt edici düzlem vektörü $\vec{w}$ ile çarpılır. Bilindiği üzere noktasal çarpım, başvuru alınan vektörün uzunluğu ile diğer vektörün başvuru vektörü üzerine izdüşüm uzunluğunun çarpımıdır ve sonucu yön­süz bir büyüklüktür. Eğer düzlemi temsil eden vektör düzlemin normali ise, başka bir de­ğişle düzleme dik birim vektör ise, o zaman bu çarpım özellik vektörünün düzlem normali yönünde merkeze olan uzaklığını belirtir. Elde edilen bu uzaklık değeri düzlemin merkeze uzaklığı ile kıyaslanır. Kıyaslama yöntemi çıkarma işlemidir. Noktasal çarpım sonucundan düzlemin merkeze uzaklığından çıkarılır ve işlem sonucunun işaret değerine bakılır. Sonuç pozitif işaretli ise nokta pozitif sınıfa değilse negatif sınıfa aittir.

Kısacası sınıflama fonksiyonun temelde bir işaret fonksiyonu olacağı söylenebilir. Aynı zamanda bu fonksiyon $\vec{w}$ düzlem normali vektörü ile bir $\vec{x}$ özellik vektörünün noktasal çarpımını ve ayırıcı düzlemin merkeze uzaklığı olan b mesafe değişkenini içermektedir. Bu verilere göre sınıflama fonksiyonu Denklem 4.31’te gösterildiği gibi tanımlanabilir [73].

$$h(\vec{x}) = \text{sign}(\vec{w}^T \cdot \vec{x} - b) \quad (4.31)$$

Bir sınıf için $\vec{w}^T \cdot \vec{x}_1 - b = d$ olduğunu kabul edelim. Diğer sınıf için $\vec{w}^T \cdot \vec{x}_2 - b = -d$ olduğunu kabul edelim. İki farklı veri sınıfına ait $\vec{x}_1$ ve $\vec{x}_2$ arasındaki uzaklığın $\vec{w}$ yönündeki değeri,

$$\frac{\vec{w}^T}{\|\vec{w}\|} (\vec{x}_1 - \vec{x}_2) = \frac{2d}{\|\vec{w}\|} \quad (4.32)$$

eşitliğinde gösterildiği gibi $2d/\|\vec{w}\|$ olacaktır. Daha önce tanımlanan amaç fonksiyon ölçütüne göre bu aralık değeri maksimum olmalıdır. Aralığın maksimum değeri alması içinse $\|\vec{w}\|^2$ vektör büyüklüğünün en küçük olması gerekmektedir.

Artık bu tespitlere göre kısıtlarla birlikte amaç fonksiyonu karesel bir en uygunu bulma problemi olarak aşağıda verildiği gibi tanımlamak mümkündür.

Amaç fonksiyon,

$$\frac{1}{2} \|\vec{w}\|^2$$

Kısıtlar,

$$\vec{w}^T \cdot \vec{x} - b > 0, \text{ (sınıf pozitif ise)}$$

$$\vec{w}^T \cdot \vec{x} - b < 0, \text{ (sınıf negatif ise)}$$

Kısıt fonksiyonunu cebirsel bir hile ile daha basit ifade etmek mümkündür. Bir vektörün hangi sınıfa ait olduğunu gösterir y değeri +1 ya da -1 cebirsel değerini alabilsin. Bu kabulden sonra pozitif sınıf verisi için $y = +1$ değerini alır, negatif sınıf verisi için $y = -1$ değerini alır. Kısıt fonksiyonu, $y \cdot (\vec{w}^T \cdot \vec{x} - b) > 0$ halini alacak ve kısıt sayısı azaltılmış olacaktır. Son haliyle amaç fonksiyonu tek bir kısıtlama eşitsizliği ile ifade etmek aşağıda gösterildiği gibi mümkün olacaktır.

Amaç fonksiyon,

$$\frac{1}{2} \|\vec{w}\|^2$$

Kısıtlar,

$$y . (\vec{w}^T . \vec{x} - b) > 0$$

Literatürde noktasal çarpımla elde edilen izdüşümün düzlemin hangi yönünde kaldığını tespit etme amacıyla sınıflayıcı (karar) fonksiyonu olarak $(\vec{w}^T . \vec{x} - b) > 0$ ve $(\vec{w}^T . \vec{x} - b) < 0$ kullanmak yerine $(\vec{w}^T . \vec{x} - b) > +1$ ve $(\vec{w}^T . \vec{x} - b) < -1$ kullanılmaktadır. Her iki farklı kullanımda bulunacak b değerleri farklı olsa da geometrik olarak hesaplanacak ayırıcı düzlemin yeri değişmeyecektir. Bu durumda

Amaç fonksiyon,

$$\frac{1}{2} \|\vec{w}\|^2$$

Kısıtlar,

$$y . (\vec{w}^T . \vec{x} - b) > 1$$

elde edilir.

Özetleyecek olursak, destek vektörleri iki farklı veri sınıfını çok boyutlu uzayda doğrusal olarak ayırmak amacıyla kullanılır. Destek vektörleri yapay sinir ağlarına benzemektedir, hatta sigmoid iletim fonksiyonu kullanan bir destek vektör yapısı iki katmanlı perceptron yapay sinir ağının eş değeridir.

Destek vektörlerinin işlev olarak yerini alabilecek diğer yöntemler yapay sinir ağları ve karar ağaçlarıdır. Destek vektörleri bu iki yapıdan farklı olarak tek bir doğrusal ayırıcının sınıflama işlemini yapabileceğini garanti edemez.

4.6. Sade Bayes Sınıflayıcı

Sade (Naive) Bayes sınıflayıcısı, Bayes teoremine dayalı olasılık tabanlı bir sınıflayıcıdır. Bayes teoremi, koşulu olasılıklar arasında bağıntı kurar. Bir C sınıfının bilinmesi koşulunda x özniteliğinin var olma olasılığı $P(x|C)$ ile x özniteliğinin bilinmesi koşulunda C sınıfının var olma olasılığı $P(C|x)$ arasında Bayes Teoremine göre şu bağıntı bulunmaktadır (Denklem 4.33).

$$P(C|x) = \frac{P(C).P(x|C)}{P(x)} \quad (4.33)$$

Çok sayıda öznitelik söz konusu olması halinde bu bağıntı aşağıda verildiği biçimde adım adım açılabilir (Denklem 4.34) [74].

$$\begin{aligned} P(C, x_1, \dots, x_n) &= P(C).P(x_1, \dots, x_n, |C) \\ &= P(C).P(x_1|C).P(x_2, \dots, x_n, |C, x_1) \\ &= P(C).P(x_1|C).P(x_2|C, x_1)P(x_3, \dots, x_n, |C, x_1, x_2) \\ &= P(C).P(x_1|C).P(x_2|C, x_1).P(x_3|C, x_1, x_2)P(x_4, \dots, x_n, |C, x_1, x_2, x_3, \dots) \end{aligned} \quad (4.34)$$

Burada verilen açılım, $\vec{A}_k = (C, x_1, \dots, x_n)$ olarak $n + 1$ elemanlı bir $\vec{A}_k$ vektörünü tanımlamak üzere zincir kuralı ile şu şekilde özetlenebilir (Denklem 4.35).

$$P\left(\bigcap_{k=1}^{n+1} \vec{A}_k\right) = \prod_{k=1}^{n+1} P(\vec{A}_k | \bigcap_{j=1}^{k-1} \vec{A}_j) \quad (4.35)$$

Özniteliklerin birbirlerinden istatistiksel olarak bağımsız olduğu varsayılırsa

$$\begin{aligned} P(x_2|C) &= P(x_2|C, x_1) \\ P(x_3|C) &= P(x_3|C, x_1, x_2) \\ P(x_n|C) &= P(x_n|C, x_1, \dots, x_{n-1}) \end{aligned} \quad (4.36)$$

eşitlikleri sağlanır (Denklem 4.36). Bu durumda çok değişkenli uzayda Bayes bağıntısı aşağıda verilen biçimde sadeleşir (Denklem 4.37).

$$P(C|x_1, \dots, x_n) = \frac{P(C) \cdot \prod_{k=1}^n P(x_k|C)}{P(x_1, \dots, x_n)} \quad (4.37)$$

Sonuçta bilinen bir öznitelik vektörü üzerinden bu öznitelik vektörünün herhangi bir C sınıfına ait olma olasılığı hesaplanabilir. Sade Bayes sınıflayıcı, en yüksek olasılık hangi sınıf için hesaplandıysa karar sınıfı olarak o sınıfı seçer.

Kısaca, bu sınıflayıcı özniteliklerin istatistiksel bağımsızlığına ve Bayes Teoremine dayanır. Her ne kadar gerçek hayatta sınıflama işlemlerinde tüm giriş değişkenlerinin istatistiksel bağımsızlığı söz konusu değilse de, bu sınıflayıcının çok iyi tahminler ürettiği bilinmektedir [74, 64].

4.7. Karar Ağaçları Sınıflayıcısı

Karar ağacı, ağaç yapısında veri uzayında dallanmalar yaparak arama gerçekleştiren grafik tabanlı bir modeldir. Karar ağacı, araştırma uzayını alt bölümlere ayırır ve bu bölümlerin her birisi bir düğüm ya da dal yapısı ile temsil eder. Karar ağacı parametrik olmayan bir modeldir, bu nedenle önceden tanımlı bir yapısı olmaz veya veriyi temsil eden başlangıç değişkenlerine ihtiyaç duymaz. Ancak, giriş ve çıkış verisine ihtiyaç duyar. Bu yönüyle danışmanlı bir öğrenme modelidir.

Öğrenme süreci sonunda, oluşan ağaç yapısı sıralı bir sorgulama ve karar üretme yapısı sunar. Bu yapı yapay sinir ağlarına göre daha basit ve açıktır [72]. Ayrıca karar ağacı modeli, insanlar tarafından daha kolay anlaşılan ve koşul cümleleri ile temsil edilebilen bir yapıya sahiptir [63]. Bu yönüyle özellikle tıbbi uygulamalarda kullanım yeri bulunmaktadır.

Bu artıların yanında karar ağacı modelinin bazı eksileri de bulunmaktadır. Bunlardan ilk akla geleni dallanmalar esnasında araştırma uzayının eğrisel değil doğrusal olarak bölünmesidir. Özellikle ikili dallanma yapan bir karar ağacı modeli, öğrenme işlemini tamamladığında veri uzayını üstküplerden oluşan kesimlere bölmüş olacaktır. Bu

durumda, tek bir eğri ile yapılan bir sınıflama işlemi ancak çok sayıda düğüm ve dal içeren bir ağaç yapısı ile modellenebilir. Bu nedenle, ağaç yapısının faydaları korurken veri uzayının eğrisel olarak bölümlenmesini de sağlayacak bazı yeni ağaç modelleri geliştirilmiştir.

Bunlardan birisi fonksiyonel ağaçlar (FT) modelidir. Sınıflama problemleri için çok değişkenli ağaçlar modeli önerilmiştir. Çok değişkenli model, değişkenlerin bileşimlerini düğümlerde kullanır. Kararlar düğümlerde birçok değişkene bakılarak verilir. Çok değişkenli bir diğer karar ağacı yapısı, model ağaçları olarak bilinir. Model ağaçları değişkenlerin birleşimini ağaç yapısının dal/yaprak kısmında bir yaklaştırma eğrisi gibi kullanır. Bu yapıda kararlar yapraklarda verilir, her yaprakta verilen karar bir yaklaştırma eğrisinin sabitlerini oluşturur.

Fonksiyonel ağaçların temel fikri, değişkenlerin birleşimini hem karar ağaçlarında hem de yapraklarda kullanmaktır. Böylelikle birden çok değişken için karar testleri iç düğümlerde yapılır, sınıf tahminleri ise yaprak düğümlerde yaklaştırma eğrileri kullanarak gerçekleştirilir. Böylelikle, FT tek değişkenli ağaç yapısını doğrusal fonksiyonlarla birleştirir. Bunu yaparken mevcut veri kümesinden yeni veriler üretir [75]. Yeni veri üretme operatörü daha önce çok değişkenli ağaçlar içerisinde kullanılmıştır [76]. FT bu yaklaşımın bir uzantısıdır, ilaveten yaprak düğümlerde fonksiyonların varlığına müsaade edilmiştir [77]. FT sadece sınıflama değil aynı zamanda eğri yaklaştırma amacıyla kullanılır.

Bir veri kümesi $D = (\vec{x}_n, y_n)$ üzerinde, $\vec{x}_n$ giriş vektörü ($n=1..N$) ve y_n çeşitli sınıflardan $y_n \in \{Cl_1, Cl_2, ..., Cl_c\}$ birisine ait olan bir çıkış değeri tanımlasın. Sınıflayıcı fonksiyon Cf giriş veri kümesi üzerinde eğitilir ve bir sınıflayıcı model $Cf(D)$ oluşur. Bu model girişleri alır ve $Cf(\vec{x}, D)$ çıkışını üretir. Burada $\vec{x}$ giriş vektörünü D eğitim verisini ve çıkış değeri $\vec{p}$ vektörünü tanımlar. Çıkış vektörü, $[p_1, p_2, ..., p_c]$, koşullu olasılık dağılımını tanımlayan değerlerdir. Bu değerler, $\vec{p} = P(y = Cl_i | \vec{x})$ giriş değerine göre çıkışın bir sınıfa ait olma olasılığını tanımlar. En yüksek olasılıkla tanımlanan sınıf karar sınıfı olarak seçilir.

FT yönteminin bir başka özelliği yeni değişken üretme operatörüdür. Bu operatör, Φ , tüm veri kümesi D üzerine uygulanır ve daha geniş yeni bir veri kümesi D' oluşturulur. Bu yeni veri kümesi önceki ile eşit veri miktarına sahiptir ama yeni üretilenlerle birlikte daha çok sınıf içerir

FA yapısı içerisinde birden fazla sınıflayıcı kullanılabilir. Bir eğitim kümesi L , test kümesi T ve iki farklı sınıflayıcı Cf_1 ve Cf_2 verildiğinde;

$$L_{new}^1 = \Phi(L, A(Cf_1(L), L)) \quad (4.38)$$

$$T_{new}^1 = \Phi(T, A(Cf_1(L), T)) \quad (4.39)$$

Yeni değişken üretme operatörü ile yeni bir eğitim kümesi L_{new}^1 (Denklem 4.38) ve yeni bir test kümesi T_{new}^1 (Denklem 4.39) oluşturulur. Burada, $Cf_1(L)$ modeli tahmin ve yeni değişkenler üretmek için kullanılır. Bir başka sınıflayıcı Cf_2 üretilen yeni eğitim veri kümesi üzerinde eğitilir ve test veri kümesi üzerine uygulanır. Birden çok sınıflayıcı bu yordam uyarınca birbiri ardı sıra aşağıda verildiği gibi kullanılabilir (Denklem 4.40).

$$Cf_2 \nabla Cf_1 = A(Cf_2(L_{new}^1), T_{new}^1) \quad (4.40)$$

Burada ∇ işlemcisi iki sınıflayıcının birbiri ardı sıra uygulandığını ve birisinin oluşturduğu veri kümeleri üzerinde diğerinin eğitildiğini koşutulduğunu ifade eder. Çok sayıda sınıflayıcını bir arada kullanılması durumunda çok sayıda sınıflayıcının ardışık uygulanması aşağıda verilmiştir (Denklem 4.41).

$$Cf_n \nabla Cf_{n-1} \nabla Cf_{n-2} \dots \nabla Cf_1 = A(Cf_n(L_{new}^{n-1}), T_{new}^{n-1}) \quad (4.41)$$

Nihai model $(n-1)$ uygulamadan sonra Cf_n sınıflayıcısı ile belirlenir. Bu biçimde çeşitli sınıflayıcılar kullanarak araştırma uzayının olabildiğince geniş taranması ve farklı sınıflayıcıların üstünlüklerinden faydalanılması öngörülmüştür.

4.8. Sınıflayıcı Modeli Başarım Ölçütleri

Sınıflayıcı için giriş verisi olacak özniteliklerin hesaplanmasından sonra, sınıflayıcı modelin hücre çekirdeğinin türünü, boyanma şiddetini ve aydınlık değerini tanımlaması beklenmektedir. Bu değerlendirme sonucunda dikkate alınacak 4 farklı durum bulunmaktadır (Tablo 4.1). Burada, ikinci türde hücre çekirdeği mevcut veri kümesinde bulunmamaktadır, bu nedenle sınıflayıcı üç ayrı çıkıştan birisini verecek biçimde tasarlanmıştır.

Tablo 4.1. Çekirdeğin ait olabileceği sınıf türleri

Negatif çekirdek (Sınıf-1)
Pozitif çekirdek zayıf boyanmış (veri kümesinde yok)
Pozitif çekirdek orta şiddette boyanmış (Sınıf-2)
Pozitif çekirdek kuvvetli boyanmış (Sınıf-3)

Sınıflayıcı başarımını ölçmek için 5 farklı araç kullanılmıştır. Bunlar, doğru sınıflandırma yüzdesi, hata matrisi, duyarlılık, özgüllük, ve alıcı işletim karakteristik eğrisi (ROC) olarak adlandırılmaktadır.

Adından da anlaşılacağı üzere doğru sınıflama yüzdesi kaç adet örneğin sınıflayıcı tarafından doğru olarak tasnif edildiğini belirtir. Hata matrisi, sınıflayıcı modelin eğitim veri kümesi üzerindeki sınıflama tahminlerini bir matris biçiminde ortaya koyar. Bu matrisin köşegen değerleri doğru sınıflanan veri miktarını, köşegen dışı değerleri ise yanlış sınıflanan veri miktarını göstermektedir. Matriste; P pozitif ve N negatif veri sayısını ifade eder. DP , DN , YP , YN , sırasıyla doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif sınıflayıcı tahminlerini göstermektedir. Örnek olarak 2x2 boyutlarında bir hata matrisi aşağıda gösterilmiştir (Şekil 4.7).

		gerçek değer		tahmin toplam
		<i>p</i>	<i>n</i>	
tahmin edilen değer	<i>p'</i>	doğru pozitif (DP)	yanlış pozitif (YP)	P'
	<i>n'</i>	yanlış negatif (YN)	doğru negatif (DN)	N'
toplam		P	N	

Şekil 4.7. Hata matrisi

Duyarlılık ve özgüllük ise incelenen sınıf için sınıflayıcının tahmin soncuna göre aşağıda verilen bağıntılarla (Denklem 4.42 ve 4.43) hesaplanır. Duyarlılık, hedef (pozitif) veri içerisinde hedefi doğru tahmin etme kabiliyetini gösterir. Özgüllük, hedef dışı (negatif) veri içerisinde, hedef dışı olanı doğru ayırt etme kabiliyetini yansıtır [63, 64, 70-72].

$$duyarlilik = \frac{DP}{DP + YN}(\%) \quad (4.42)$$

$$ozgulluk = \frac{DN}{DN + YP}(\%) \quad (4.43)$$

Alıcı işletim karakteristik eğrisi, sınıflama modeli karar fonksiyonunun sınıf türüne göre hatadan ne kadar bağışık olduğunun bir ölçüsüdür. Karar fonksiyonunda kararı belirleyen eşik düzeyi belirli adımlarla değiştirilerek, değişen eşik düzeyine bağlı olarak sınıfı doğru tahmin etme başarımının ne kadar etkilendiği grafik olarak gözlenir. Bu grafikte, başarımın ölçüsü olarak kabul edilen iki değişken bulunmaktadır. Bu değişkenlerden birincisi doğru pozitif oranı, $DPO = DP/(DP + YN)$, ikincisi ise yanlış pozitif, $YPO = YP/(YP + DN)$, değerleridir. Grafikte, değişen karar eşik değerine bağlı olarak yanlış pozitif oranı yatay eksene, doğru pozitif oranı ise dikey eksene

yansıtılır. Böylece elde edilen iki boyutlu grafik eğrisinde, eğri sağ alt köşeye yaklaştıkça sınıflama modelinin değişen eşik düzeyine göre verdiği kararlarda yanlış pozitiflerin etkin olduğu ve sınıflama başarımının düşüş gösterdiği anlaşılır. Öte yandan, oluşan eğri sol üst köşeye yaklaştıkça sınıflama başarımının farklı karar eşik düzeylerine karşı bağışık olduğu, doğru pozitif oranının başarımında etkin olduğu anlaşılır [63, 64, 70-72].

Bu başarımlar ölçütleri sınıflayıcıyı değerlendirmek için yeterli olmaz. Ayrıca sınıflayıcının eğitim kümesinde olmayan farklı veri girişlerinde ne kadar başarılı sonuçlar verdiğini bilmek ve nasıl bir genelleme kabiliyetine sahip olduğunu görmek de gerekir. Bu amaçla geçerleme yordamları kullanılmaktadır. Geçerleme işlemi için genellikle eğitim kümesinden başka ayrıca bir veri kümesi oluşturulur. Ancak, ayrı bir geçerleme veri kümesi yerine eğitim kümesi verisi bölümlere ayrılarak k -parçalı çapraz geçerleme tekniği de kullanılabilir. Genellikle eğitim verisini elde etmenin zor olduğu tıbbi uygulamalarda çapraz geçerleme yöntemi kullanılır. Bu nedenle, bu çalışmada 10-parçalı çapraz geçerleme yöntemi kullanılmıştır. Bu yöntemin icrasında, genel olarak eğitim veri kümesi 10 parçaya bölünür ve bu parçalardan birisi test verisi, kalanı da eğitim verisi olarak ayrılarak sınıflayıcı başarımları ölçülür. Bu işlem basamağı k defa tekrarlanarak her bir parçanın bir kez test verisi olarak işlem görmesi sağlanır. Böylece bir çapraz geçerleme döngüsü tamamlanmış olur. Aynı eğitim veri kümesi içerisinde çok farklı şekilde k adet parça oluşturmak mümkündür ve bu oluşuma göre sınıflayıcı başarımları farklılaşabilir. Bu farklılaşma etkenini dikkate alarak bir çapraz geçerleme döngüsü 10 defa tekrarlanmış ve başarımlar ölçütü olarak her bir sınıflayıcı modelinin (toplamda 100 adet) başarımlar değerlerinin ortalaması alınmıştır. Böylece, geçerleme işlemi esnasında oluşan parçaların sınıflayıcı başarımları üzerindeki olası sapma etkisi telafi edilmiştir. Ayrıca, eğitim veri kümesi parçalanırken sınıf değerlerinin her bir parça içerisinde eşit miktarda dağılması gözetilmiştir.

5. BÖLÜM

FONKSİYONEL KARAR AĞAÇLARI İLE OTOMATİK TANI BULGULARI

5.1. Uygulama Özeti

Östrojen alıcı (ER) durum değerlendirmesi meme kanserinin tanısında yaygın olarak kullanılmaktadır. Ancak bu işlem gözlemci doktorun öznel değerlendirmesine tabiidir. Bu bölümde, fonksiyonel karar ağaçları (FT) yapay öğrenme yöntemi ile bilgisayar tabanlı tanı sisteminin invaziv duktal karsinom teşhisli meme kanseri örneklerinde kullanımı denenmiş ve başarımların incelenmesi yapılmıştır.

Bu amaçla 27 adet biyopsi örneği incelenmiş ve ışın mikroskobu aracılığıyla sayısal resim dosyası olarak kayıt edilmiştir. Bu resimler üzerinde 5150 adet hücre çekirdeği otomatik olarak çeşitli görüntü işleme teknikleri kullanılarak bölümlenmiştir. Bölümleme sonucunda bağlı bileşen olarak hafızaya alınan her bir hücre çekirdeği üzerinde birinci ve ikinci dereceden istatistiksel öznitelikler, eş-oluşum matrisi öznitelikleri, Laws desen enerji öznitelikleri ve dalgacık dönüşümü enerji öznitelikleri hesaplanmıştır. Bu yolla çıkarılan 144 adet özniteliğin arasından ilinti tabanlı bir eleme yöntemi kullanılarak 11 tanesi seçilmiştir. Azaltılmış sayıda öznitelik değerlerini giriş olarak alacak biçimde tasarlanan FT türü sınıflayıcı başarımları, 10-kümelili çapraz geçerleme yöntemiyle ölçülmüştür. Eğitim kümesinden elde edilen sınıflayıcı modeli, eğitim kümesinde tamamen bağımsız test resimleri üzerinde çalıştırılmıştır. Bu yolla bilgisayar tabanlı tanı sonuçları ve test resimleri üzerinde uzman doktorların verdiği kararlar karşılaştırılmıştır. Sonuç olarak, FT yapay öğrenme modeline dayalı bilgisayar tabanlı otomatik tanı sisteminin puanlamasının uzmanların puanlaması ile örtüştüğü ve önerilen sistemin

uzman doktor kararlarından istatistiksel anlamda önemli bir fark göstermediği gözlenmiştir.

5.2. Malzemeler ve Yöntemler

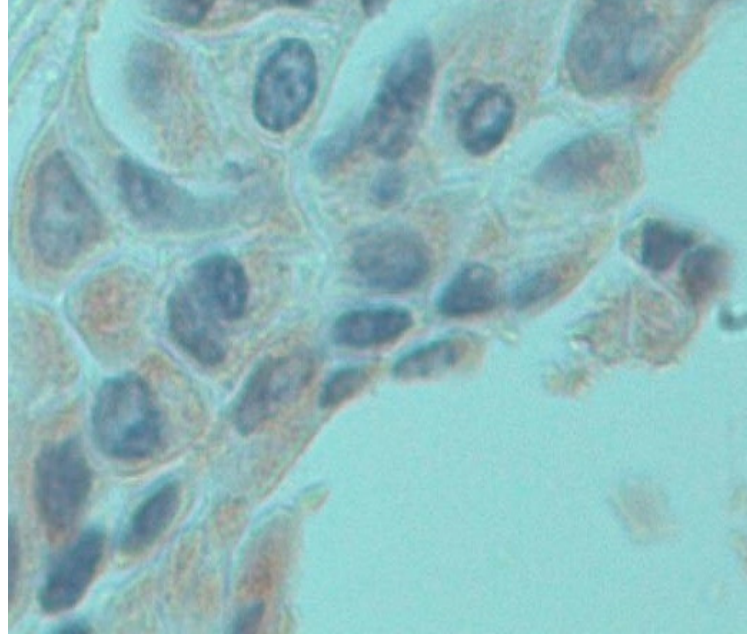
Erciyes Üniversitesi Tıp Fakültesinin arşivinde invaziv duktal karsinom örnekleri 2007-2008 yıllarını kapsayacak şekilde araştırılmıştır ve toplam 40 adet örnek toplanmıştır. Hemoksilin-Eosin (HE) ile boyalı, formalinle sabitlenmiş ve parafinle tümleşik (formalin-fixed paraffin-embedded) 4–5 µm kalınlığında tümör kesitleri histolojik inceleme için kullanılmıştır.

Bu örneklerde, tümörün türü Dünya Sağlık Örgütü'nün tümör sınıflama yordamına göre belirlenmiştir [21] ve histolojik derecelendirme Elston-Ellis yordamına [22] göre yapılmıştır. Tümörü temsil eden kesim ve normal meme dokusu (dâhili kontrol yapılarak) östrojen alıcı immünohistokimyasal (IHC) boyaması ile boyanmıştır. Tüm immünohistokimyasal işlemlerde, parafin bloklarda bulunan 4–5 µm kalınlığında doku kesitleri kullanılmış ve immünohistokimyasal incelemeler üretici firmanın talimatlarına göre gerçekleştirilmiştir. Diaminobenzidinetetrahydrochloride (Dako liquid DAB plus, K3468, Denmark) kromojen -renklendirici- olarak kullanılmış ve kesitler Mayer hemotoksilin ile karşıt renklendirilmiştir.

Östrojen alıcısı IHC boyamasından sonra, ER pozitif hücre çekirdekleri kahve tonlarda ve ER negatif hücre çekirdekleri mavi tonlarda boyanır. Kimyasal boyama tümörlü hücre çekirdekleri üzerinde değerlendirildi ve bu işlem Allred puanlama sistemine göre yapıldı [7]. Toplanan 40 adet örnekten 7 tanesi çalışma kapsamından çıkarıldı. Çıkarılan bu örnekler üzerinde sitoplazmik boyama, yapay boyama kusurları bulunmuştur ya da gözlemci doktorların tanı kararları farklı olmuştur. Sitoplazmik boyama içeren örnek bir resim kesiti Şekil 5.1'de gösterilmektedir. Diğer örneklerden farklı olarak burada hücre çekirdekleri değil sitoplazma kahve tonlarda boyanmış haldedir.

Patolog F.A., her bir örneği mikroskop (Lecia DMD 108 Micro-imaging Device) altında inceleyerek durumu en iyi temsil eden bölgeye seçmiş ve bu bölgeyi 40x büyütme ölçeğinde sayısal resim dosyası olarak kayıt etmiştir. Her bir resim dosyası 2048x1536x24 bit (8 bit her bir RGB kanalı için) boyutlu JPEG sıkıştırma türünde kayıt

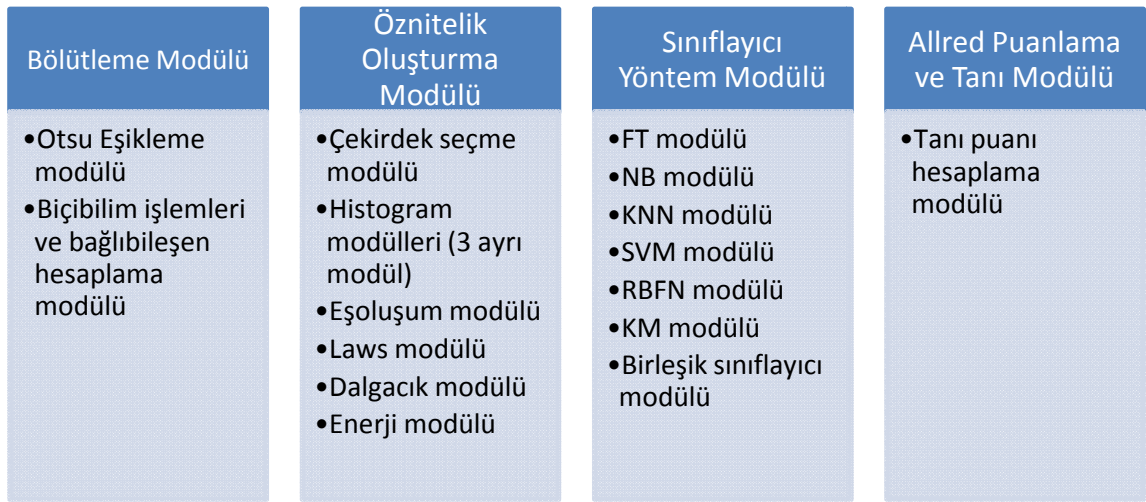
altına alınmıştır. Gözlemciler arası farklılık dikkate alınarak, iki uzman patolog H.A. ve M.K. kayıtlı örnekler ait sayısal resim dosyalarını inceleyerek Allred protokolüne göre puanlama yapmış ve tanı kararını vermişlerdir.



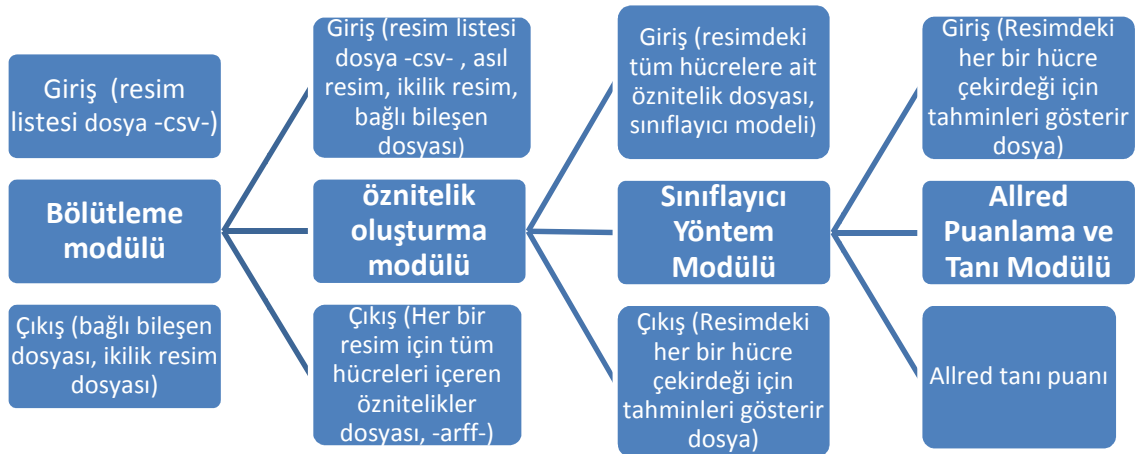
Şekil 5.1. Sitoplazmik boyama

Başlı başına ayrı bir eğitim kümesi oluşturmak için, geriye kalan 33 resim dosyasından 6 tanesi seçilmiş ve bu resimler deneysel incelemenin yapıldığı test veri kümesinden çıkarılmıştır. Geriye kalan 27 adet resim test veri kümesi olarak ayrılmış ve iki uzman doktorun bu resimler üzerindeki puanlamaları ve tanı kararları not edilmiştir. Pozitif veya negatif boyamaya ve boyama şiddetine bağlı olarak iki uzman doktor Microsoft Paint Brush yazılımı arcılığıyla eğitim veri kümesinde bulunan 6 adet resim üzerinde temsil kabiliyeti olan her bir hücre çekirdeğini işaretlemişlerdir. Bu resimler üzerinde Allred protokolüne göre her iki doktorun aynı boyanma türü ve aydınlık değeri işaretlediği 384 adet hücre çekirdeği ortaya çıkmıştır. Bu hücre çekirdekleri danışmanlı yapay öğrenme yöntemlerinin eğitim veri kümesini oluşturmaktadır. Ancak, yüzlerce hücre çekirdeğinin farklı uzmanlarca incelenmesi ve işaretlenmesi emek ve zaman yoğun bir işlem olduğundan, ayrıca bir geçerleme veri kümesi oluşturulamamıştır. Bunun yerine tıbbi uygulamalarda sıklıkla kullanılan 10-kümeli çapraz geçerleme yordamı uygulanarak sınıflayıcı başarımlar değerlendirilmesi gerçekleştirilmiştir.

Bu tez çalışması kapsamında kullanılan otomatik tanı sistemi, Matlab [55] geliştirme ortamında yazılım kodlanarak oluşturulmuştur. Ayrıca, sınıflayıcı modelin icrasında Java kodları kullanılmaktadır. Tanı sistemi meydana getiren ana yazılım modülleri ve ana yazılım modüllerinin çağırdıkları alt modüller aşağıdaki diyagramda özetlenmiştir (Şekil 5.2). Bu ana modüllerin aldıkları giriş verileri ve ürettikleri çıkışlar Şekil 5.3'te görülebilir.



Şekil 5.2. Tanı sistemini oluşturan ana ve alt yazılım modülleri



Şekil 5.3. Tanı sisteminde ana modüllere ait giriş ve çıkış verileri

Tanı sistemini oluşturan tüm modüllerin 20 defa çalıştırması durumunda ana yazılım parçacıkların ortalama çalışma süresi ve bu ortalamadan elde edilen standart sapma değerleri şu şekilde ölçülmüştür. Denemeler için en çok sayıda hücre çekirdeği barındıran bir test resmi seçilmiştir. Bu resimde 298 adet hücre çekirdeği tespit edilmiş ve tüm bu çekirdekler, tanı kararına varabilmek için sistem tarafından otomatik olarak incelenmiştir. Test veri kümesinde bulunan resimlerde ortalama tespit edilen hücre çekirdeği sayısının 192 olması dikkate alındığında yapılan çalışma zamanı ölçümleri ortalama başarıma göre karamsar olacaktır.

Buna göre, bölütleme modülü Otsu eşikleme yöntemini kullanarak yapılan bölütle işlemini test resmi üzerinde $3.75 (\pm 0.25)$ saniyede tamamlamıştır. Öznitelik oluşturma modülü, test resmi üzerinde yapılan 20 denemenin sonunda $25.62 (\pm 0.21)$ saniye zaman almıştır. Sınıflayıcı modelin test resminin tüm hücre çekirdekleri ortalama sınıflama süresi $0.014 (\pm 0.011)$ saniye olmuştur. Son olarak, Allred tanı sistemine göre test resmi üzerine östrojen alıcı tanı puanının hesaplanması $0.009 (\pm 0.009)$ saniye zaman tüketmiştir. Özetle, işlem yükü fazla olan bir test resmi için tanı sisteminin tüm modüllerinin çalıştırılması ve nihai tanı kararının elde edilmesi yaklaşık 30 saniye sürmektedir. Bu denemeler, 32 bit komut setli Intel Pentium M 1.73 GHz İşlemcili ve 1.25 GB DDR2 RAM özelliklerinde bir PC üzerinde gerçekleştirilmiştir. Bugün çok daha güçlü kişisel bilgisayarların kullanımda olduğu dikkate alınırsa sistemin çalışma zamanı belirgin biçimde azaltılabilir.

Sonuç itibarıyla, tanı kararına varmak için talep edilen zaman gereksinimi bakımından bu tez çalışmasında kullanılan tanı sistemi kullanım değeri taşıyabilir.

5.2.1. Bölütleme ve Biçimbilim İşlemleri

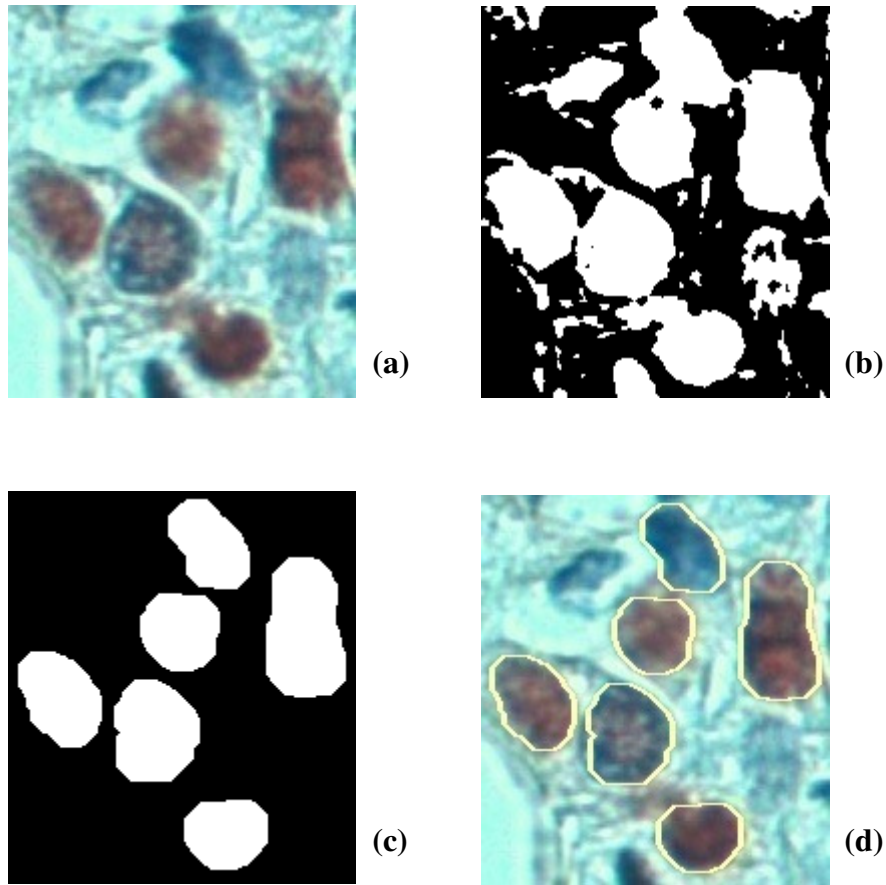
Allred puanlama protokolüne göre, incelenen örnek üzerinde boyanmış kesimin alan olarak yüzdeliğine veya boyanmış piksel sayısına bakılmamaktadır. Bu yordam pozitif ve negatif hücre çekirdeklerinin sayısına göre puanlama sonucu verir. Bu nedenle, öncelikle hücre çekirdeklerinin tespiti ve tasnifi bu çalışmada hedef olarak benimsenmiştir.

Böylece, ilk aşama olarak bölümlleme aşaması öngörölmüş ve hücre çekirdeklerinin geri kalan diğler doku yapılarından ayırt edilmesi gerçekleştirilmiştir. Bu doğrultuda, temel bölümlleme yöntemi olarak Otsu otomatik eşikleme yöntemi kullanılmıştır [52]. Eşikleme safhasında IHC resminde hücre çekirdekleri ile onları kaplayan doku kesimi arasındaki aydınlık şiddeti farkından yararlanılmıştır. Çekirdekler geri kalan doku yapılarına göre daha koyu renklidir. Buradan hareketle, önce resim 8 bitlik gri ton bir resim dosyasına dönüştürölmüş ve sonra bu dosya üzerinde Otsu eşikleme yöntemi çalıştırılarak koyu renkli çekirdek yapıları aydınlık değeri yüksek doku kesimlerinden otomatik olarak ayırt edilmiştir. Kullanılan eşikleme yöntemi giriş parametresi veya önceden ayarlama gerektirmeden otomatik olarak resmi siyah-beyaz ikilik bir renk düzenine dönüştürür. Elde edilen çıkış resminde çekirdekler siyah, doku yapısı ise beyazdır. Daha sonra bu resim renk düzeni bakımından terslenerek hücre çekirdekleri beyaz ve doku kesimi siyah hale getirilir. Bu tersine dönüştürme işlemi hücre çekirdeklerinin birer bağılı bileşen olarak algılanması için gereklidir.

Daha sonra biçimbilim işlemcileri görev alır. 12 piksel yarıçaplı disk görünlümlü biçimbilim açma işlemcisi ikilik resme uygulanır ve resim açılır. Bu biçimsel düzeltme safhasından sonra her biri bir hücre çekirdeğini temsil eden bağılı bileşenler resim üzerinde hesaplanır ve etiketlenirilir. Bölümlleme ve biçim bilim işlemlerini gösteren örnek bir resim kesiti Şekil 5.4’de sunulmuştur.

Bölümlleme aşamasında bazı istisnai haller önceden dikkate alınmıştır. Örneğin doku içerisinde bulunan stroma hücreleri de koyu renklidir ve bunların hücre çekirdeği olarak seçilmemesi gerekmektedir. Bir diğler olasılık, görsel olarak birbiri içerisinde geçmiş ve kaynaşmış halde görölen birleşik hücrelerin varlığıdır. Bu kaynaşma, açma işlemcisinin bir sonucu olabileceği gibi inceleme yapılan örnekte hücrelerin üç boyutlu doğal konumlamasının iki boyutlu mikroskop görüntüsünde çakışmasından da oluşabilir. Ayrıca, kayıt edilen resim çerçevesinin kenarlarında kalan ve biyopsi kesitinde derinde olup ancak bir kısmı görönlü hücre çekirdekleri de bulunabilir. Bu olasılıklar dikkate alınarak, sadece 300 pikselden büyük ve 3000 pikselden küçük boyutlu çekirdeklerin seçilmesine tecrübeye dayalı olarak karar verildi.

Böylesi bir biçimsel süzgeç ile küçük boyutlu stroma hücrelerinin çekirdek olarak algılanmasının önüne geçilmiştir. Çünkü stromalar hücre çekirdeklerine göre belirgin biçimde daha küçük boyutludurlar. Bu yolla kaynaşmış hücreleri de saf dışı bırakmak mümkün olmaktadır. Sonuçta, boyut süzgecinin yaptığı eleme bir kısım verinin kayıp edilmesine sebep olur. Ancak, bu bilgi kaybı pozitif veya negatif hücre çekirdek sayısında bir ağırlık etkisi oluşturmamaktadır. Çünkü eleme işlemi pozitif veya negatif ayrımı yapmadan sadece boyuta bağlı olarak gerçekleştirilmektedir. ER durum incelemesinde esas olan her resim üzerinde yeter sayıda hücre çekirdeğinin elde edilmesidir. Bölümleme ve biçimbilim işlemlerinden sonra yeter sayıda hücre çekirdeği elde edilmektedir. Bu aşamada son olarak her hücre çekirdeğinin biçimsel ve konumsal özellikleri ayrı bir dosyaya kayıt edilerek saklanır.



Şekil 5.4. Bölümleme aşaması alt işlemleri; (a) asıl resimden bir kesit, (b) kesit üzerinde Otsu eşikleme işlemi sonucu, (c) biçimbilim işlemlerinin sonucu, (d) tespit edilen hücre çeperinin asıl resim kesiti üzerine giydirilmesi işlemi sonucu

5.2.2. Öznitelik Çıkarımı ve Elemesi

Bu tezin önceki bölümünde açıklanan tüm öznitelikler hesaplanarak her bir hücre çekirdeğinden toplam 144 adet öznitelik elde edilmiştir. Bu kadar çok miktarda giriş verisinin doğrudan sınıflayıcıya verilmesi halinde hem sınıflayıcı başarımlar değerleri düşmekte hem de işlem yükü ve karmaşası yükselmektedir. Bu nedenle yine 5. Bölümde açıklanan ilinti tabanlı boyut azaltma yöntemi kullanılmış ve sınıflayıcı girişine verilecek olan öznitelik miktarının sınıflayıcı başarımlarını düşürmemek kaydıyla en az sayıda olması hedeflenmiştir. Boyut azaltma sonrası seçilen ve sınıflayıcı modelinin tanı işleminde kullanılan öznitelikler ve onları temsil eden değişken adları Tablo 5.1’te gösterilmiştir.

Tablo 5.1. Azaltılmış renkli öznitelikler ve kısaltma adları

Ortalama aydınlık kırmızı renk kanalında (R-AvInt)
Ortalama aydınlık mavi renk kanalında (B-AvInt)
Birliktelik matrisi tekdüzeliği kırmızı renk kanalında (R-Hom)
Birliktelik matrisi iç-ilinti mavi renk kanalında (B-Auto)
Dalgacık dönüşümü birinci ölçekte resim yaklaştırmasının enerjisi kırmızı kanalda (R-EnAp1)
Dalgacık dönüşümü ikinci ölçekte resim yaklaştırmasının enerjisi kırmızı kanalda (R-EnAp2)
Dalgacık dönüşümü birinci ölçekte resim yaklaştırmasının enerjisi mavi kanalda (B-EnAp1)
Dalgacık dönüşümü ikinci ölçekte resim yaklaştırmasının enerjisi mavi kanalda (B-EnAp2)
Dalgacık dönüşümü ikinci ölçekte köşegensel detay resim enerjisi mavi kanalda (B-EnDg2)
R5L5 Laws desen maskesi sarmaşım enerjisi kırmızı renk kanalında (R-R5L5)
E5L5 Laws desen maskesi sarmaşım enerjisi mavi renk kanalında (B-E5L5)

5.2.3. Sınıflayıcı

Sınıflayıcı seçiminde yapılan denemelere dayanarak FT sınıflayıcısının kullanılmasına karar verilmiştir. Bu modelin başarımlar ve yapısı C4.5 karar ağacına [78] ve destek vektör makinesine [79] kıyasla karşılaştırılarak incelenmiştir. Bu incelemede destek vektörlerinin doğrusal, ikinci dereceden ve yarıçap fonksiyonlu yaklaşımları da başarımlar bakımından ele alınmıştır.

Bilindiği üzere, tek değişkenli bir karar ağacı, bir düğümdeki tek bir değişken üzerinde basit bir olasılık hesaplaması yapar. Öte yandan, FT karar ağacının hem düğüm hem de yaprak noktasında birden çok değişkenin doğrusal bileşimlerinden faydalanarak karar vermektedir. Böylece, FT ağaç yapısı hem kolay anlaşılır hem de doğrusal yaklaştırma yöntemlerinin genelleyebilme kabiliyetini ortaya koyar.

Bu farklılığın etkisi, bu bölümde yapılan ER tanı uygulaması üzerinde gösterilebilir. FT sınıflayıcısı mevcut eğitim veri kümesi üzerinde sadece tek bir düğümden oluşan bir ağaç yapısı tanımlar. Bu tek düğüm katsayıları ve sabit değeri Tablo 5.2’de tanımlanan yaklaştırma fonksiyonlarını içerir. Sınıfların sayılardan değil de metne dayalı etiketlerle temsil ediliyor olması FT’nin kullanımını kısıtlamaz. Oysa aynı problemi doğrusal yaklaştırma fonksiyonları ile doğrudan çözmek istediğimizde sınıfları ifade eden metne dayalı etiketlerin var ve yok biçiminde yeniden tanzim edilmesi ve mevcut sınıf sayısının iki katına çıkarılarak sınıfa aidiyetin 1 ve 0 rakamlarından oluşan sayısal değerlerle gösterilmesi zorunludur.

Tablo 5.2. Doğrusal fonksiyonlarla FT modeli

Sınıf-1 ağırlıklar ve katsayılar	Sınıf-2 ağırlıklar ve katsayılar	Sınıf-3 ağırlıklar ve katsayılar
0.07 B-EnDg2 -11.3 sabit değer	0.01 R-AvInt 7.8 R-Hom 0.03 B-Autoc 0.01 B-EnDg2 -8.24 sabit değer	0.02 R-AvInt -0.11 B-AvInt 34.66 R-Hom -0.5 B-EnDg2 -12.6 sabit değer

Ağaç yapısında sınıflayıcılara ek olarak, destek vektör makineleri (SVM) aynı veri kümesi üzerinde uygulanmıştır. SVM modelinin eğitilmesinde WEKA tarafından sağlanan sıralı en küçük değerle iyileştirme algoritması, “sequential minimal optimization algorithm” [79] kullanılmıştır. Çeşitli SVM sınıflayıcılarının başarımları incelendiğinde (doğrusal, ikinci dereceden, yarıçap fonksiyonlu) küçük bir miktar sınıflama başarımında artış sağlandığı ikinci dereceden yaklaştırma kullanan SVM sınıflayıcısında gözlemlenmiştir. Fakat sınırlı miktarda başarımın iyileştirilmesinin karşılığında daha çok özneliğe giriş verisi olarak ihtiyaç duyan çok büyük bir

sınıflayıcı model ortaya çıkmıştır. SVM sınıflayıcılarına bakıldığında en küçük boyutlu olan model doğrusal yaklaştırma fonksiyonları kullanan modeldir ve bu modelin ürettiği sınıflama fonksiyonları Tablo 5.3’te verilmiştir.

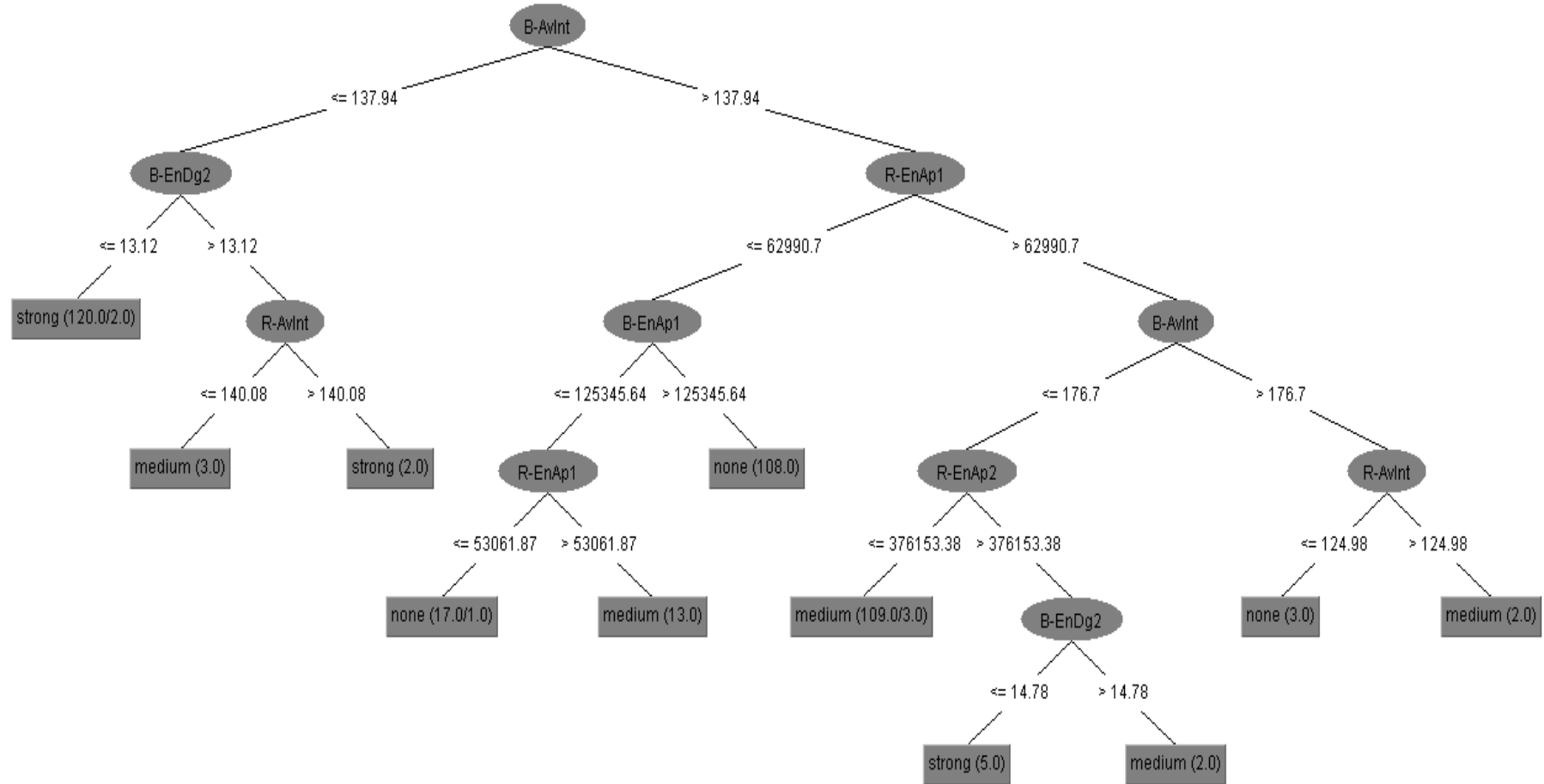
Tablo 5.3. Doğrusal SVM modeli ve katsayıları (destek vektörleri değil)

sınıf-1 ve sınıf-2 arasında sınıflayıcı katsayıları		sınıf-1 ve sınıf-3 arasında sınıflayıcı katsayıları		sınıf-2 ve sınıf-3 arasında sınıflayıcı katsayıları	
2.6555	R-AvInt	0.9832	R-AvInt	0.2458	R-AvInt
-1.7299	B-AvInt	-1.6367	B-AvInt	-3.1246	B-AvInt
1.5779	R-Hom	0.8466	R-Hom	1.863	R-Hom
-0.0356	B-Autoc	-0.1566	B-Autoc	-1.2469	B-Autoc
2.6781	R-EnAp1	0.9358	R-EnAp1	0.8085	R-EnAp1
2.3948	R-EnAp2	0.861	R-EnAp2	1.1131	R-EnAp2
-2.0498	B-EnAp1	-1.4463	B-EnAp1	-1.79	B-EnAp1
-2.0421	B-EnAp2	-1.2969	B-EnAp2	-1.087	B-EnAp2
-0.4596	B-EnDg2	-1.1557	B-EnDg2	-2.8398	B-EnDg2
1.4834	R-R5L5	0.7278	R-R5L5	0.6264	R-R5L5
-0.445	B-E5L5	-0.8425	B-E5L5	-0.4056	B-E5L5
-1.027	sabit değer	0.7128	sabit değer	0.9939	sabit değer

Diğer taraftan, aynı eğitim veri kümesi üzerinde, C4.5 karar ağacına göre bir sınıflayıcı model tanımlandığında Şekil 5.5’te gösterilen ve nispeten daha karmaşık olan ağaç yapısı elde edilir. Bu ağacın oluşturulmasında “confidence factor” katsayısı WEKA’da 0.25 olarak sabitlenmiş ve budama yapılmasına izin verilmiştir.

Her iki modele bakıldığında FT modelinin daha az yer kapladığı görülmektedir. Bilindiği üzere, ne kadar az bileşenli ve küçük boyutlu bir sınıflayıcı model olursa, o kadar aşırı öğrenmeye karşı bağışık ve genelleme kabiliyeti fazla bir öğrenme elde edilir.

Tablo 5.3 ve Şekil 5.5 belirtilen öznitelikler, özniteliklerin hesaplanması ve seçimine ilişkin tez bölümünde açıklanmıştır. Özniteliklerin anlamlarını bilmeksizin dahi FT sınıflayıcısının diğer sınıflayıcılardan daha basit bir model sunduğu görülmektedir. Bu nedenle, FT sınıflayıcısının, Allred puanlama yöntemine göre test resimlerinin ER durum değerlendirmesinde uygulanması kararlaştırılmıştır. Ancak, yine de tüm modellerin doğru sınıflama yüzdelikleri sonuçlar alt başlığında karşılaştırılmıştır.



Şekil 5.5. C4.5 karar ağacı

5.2.4. Bilgisayar Tabanlı Puanlama Yöntemi

Bilgisayar tabanlı yüzdeler puanlaması Allred puanlama tablosunun puan aralıkları dikkate alınarak tıp doktorun elle yaptığı puanlamada olduğu gibi gerçekleştirilir. Bilgisayar tabanlı aydınlık puanlamasında ise şu yol izlenmiştir. Resme genel olarak bakılmaz, tespit edilen hücre çekirdeklerinin aydınlık değerine bakılır. Eğer hiç bir pozitif hücre çekirdeği tespit edilmemişse resmin aydınlık değeri 0 olur. Eğer tespit edilen orta şiddette boyanmış hücre çekirdeğinin sayısı kuvvetli boyanan çekirdek sayısından büyükse resmin aydınlık puanlaması 2, değilse 3 olarak belirlenmektedir.

5.3. Sonuçlar ve Tartışma

Bilgisayar tabanlı otomatik tanı sisteminde sınıflayıcı modelinin eğitilmesinden sonra, kullanıcının resimler üzerinde etkileşimine gerek duyulmamaktadır. Tüm test resimleri otomatik olarak değerlendirilir. Bu otomatik değerlendirme sürecinde, 27 test resminden 5150 adet hücre çekirdeği bölümlenerek tespit edilmiş ve sınıflandırılmıştır. Her resimde ortalama 191 adet hücre çekirdeği okunmuştur. Bölümleme aşamasında kullanılan bilgisayar kodları Matlab'te (The Mathwork Increment) yazılmıştır. FT sınıflayıcısı ve boyut azaltma yönteminin kodu WEKA, (Waikato Environment for Knowledge Analysis library, WEKA) kütüphanesinin Java'da yazılmış kodları çağrılarak çalıştırılmıştır [80]. İstatiksel testler ve grafikler R programlama dilinde gerçekleştirilmiştir. Tüm yazılımlar 32 bit komut setli Intel Pentium M 1.73 GHz İşlemcili ve 1.25 GB DDR2 RAM özelliklerinde bir PC üzerinde koşturulmuştur. Diğer araştırmacıların burada verilen yöntemleri kendi veri kümeleri üzerinde denemelerine imkân vermek için kullanılan yöntemlere ait WEKA kodları ve bunların başlangıç parametreleri bir sonraki bölümde tüm sınıflayıcı modelleri için gösterilmiştir.

5.3.1. Öznitelik Çıkarım ve Eleme Sonuçları

Her bir hücre çekirdeğinden üretilen 144 adet öznitelik şu başlıklar altında gruplandırılabilir; birinci dereceden gri ton, ikinci dereceden gri ton, Laws gri ton, dalgacık gri ton, birinci dereceden renkli, ikinci dereceden renkli, Laws renkli, dalgacık renkli öznitelikler. Ayrıca, üç yeni öznitelik grubu sınıflayıcı başarımını ölçen deneysel

çalışmaya dahil edilmiştir. Bu yeni öznitelik grupları; gri ton öznitelikleri arasından eleme yöntemiyle seçilmiş gri ton öznitelikler, aynı biçimde eleme yoluyla seçilmiş renkli öznitelikler ve tüm öznitelikler arasından eleme yöntemi ile seçilmiş öznitelikler olarak adlandırılabilir.

Sınıflayıcı başarımı içim yapılan deneysel çalışma “WEKA Experiment Utility” kullanılarak gerçekleştirilmiştir [80]. FT sınıflayıcısının tüm öznitelik grupları üzerinde doğru sınıflama yüzdesi ve parantez içinde standart sapma değerleri Tablo 5.4 verilmiştir.

Tablo 5.4. Farklı öznitelik grupları üzerinde FT başarımı

Öznitelik veri grupları	Öznitelik sayısı	FT doğru sınıflama yüzdesi (%)
Birinci derece gri ton öznitelikler	6	59.33 (7.05)
İkinci derece gri ton öznitelikler	8	54.17 (7.57)
Dalgacık gri ton enerji öznitelikleri	8	61.45 (7.70)
Laws gri ton enerji öznitelikleri	14	51.77 (7.22)
Birinci derece renkli öznitelikler	18	94.93 (3.52)
İkinci derece renkli öznitelikler	24	90.45 (4.77)
Dalgacık renkli enerji öznitelikleri	24	97.34 (3.10)
Laws renkli enerji öznitelikleri	42	92.81 (4.24)

Tablo 5.4’ün ilk sütununda çeşitli öznitelik grupları listelenmiştir. İkinci sütunda FT yönteminin sınıflama başarımı gösterilmiştir. Yansıtılan sonuçlar, 10 defa 10-parçalı çapraz geçерleme yordamıyla gerçekleştirilen toplam 100 adet yürütmenin ortalamasıdır. Sonuçlar, renkli özniteliklerin gri ton özniteliklere göre istatistiksel olarak daha başarılı olduğunu göstermektedir. Ayrıca, renkli öznitelikler arasından dalgacık öznitelikleri 97.34% doğru sınıflama yüzdesi ile diğerlerine nazaran daha başarılıdır.

İlinti tabanlı öznitelik eleme yöntemine göre, öznitelik sayısı 144’ten 11’e indirilmiştir (Tablo 5.1). Öznitelik sayısı azaltılmış veri kümesi üzerinde FT sınıflayıcısının doğru sınıflama yüzdesi 96.67% olarak elde edilmiştir. Belirgin boyut azaltmaya rağmen, elde edilen doğru sınıflama yüzdesi dalgacık dönüşümü ile elde edilen en yüksek sınıflama yüzdesine çok yakındır.

5.3.2. Sınıflama Sonuçları

Sınıflayıcı başarımını ölçen deneysel çalışmada boyut azaltma işleminden sonra seçilen 11 adet öznitelik kullanılmıştır. Başarım sonuçları bu öznitelik kümesi üzerinde 10 defa 10-parçalı çapraz geçerleme yordamının uygulanmasıyla elde edilen 100 adet sonucun ortalaması alınarak kayıt edilmiştir. C4.5 ağacının (WEKA’da J48 ağacı olarak gösterilmektedir.) doğru sınıflama yüzdesi 93.73%, SVM’nin yarıçap fonksiyonlu yaklaştırma ile elde ettiği oran 84.05% olarak not edilmiştir. Bu sonuçlar, istatistiksel olarak elde edilen diğer sonuçlardan; FT 96.67%, SVM doğrusal 97.50% ve SVM ikinci dereceden 96.80% daha kötü olmuştur. FT, SVM doğrusal ve SVM ikinci derece arasında istatistiksel anlamda belirgin (eşleşmeli t-testine göre) bir fark bulunmamaktadır. Ancak, model karmaşıklığına bakıldığında FT sınıflayıcısının klasik C4.5 sınıflayıcısına göre daha iyi olduğu görülebilir. Model karmaşası ve boyutu büyüdükçe öğrenme azalmakta ve sistem öğrenmek yerine giriş çıkış verilerini ezberlemektedir. Bu bakışla, FT modeli SVM modelinden daha sade ve küçük boyutlu bir sınıflayıcı ortaya koymaktadır.

Karşılaştırmalarda, eşleşmeli t-testi kullanılmış ve belirgin farkın karar eşik düzeyi olarak 0.05 ($P < 0.05$) olasılığı benimsenmiştir. Her ne kadar, FT diğer seçeneklere nazaran daha basit bir sınıflayıcı model verse de elde edilen sonuçlar bakımından diğer yöntemlerin kayda değer bir başarımlar getirisi yoktur.

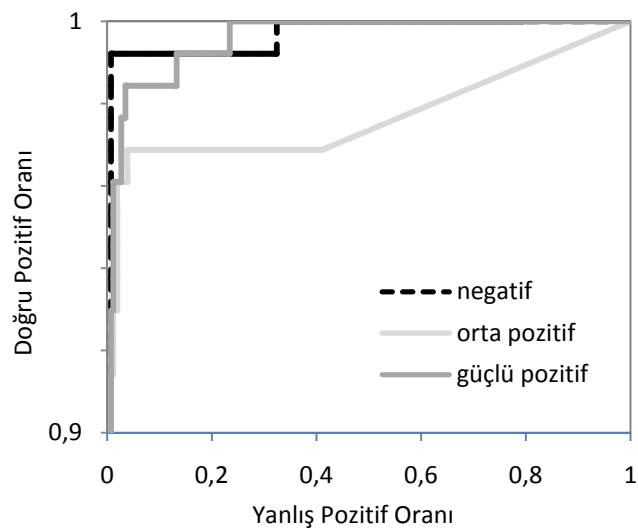
FT modelini oluştururken ayrıca duyarlılık ve özgüllük değerleri 10-parçalı çapraz geçerleme yordamına göre tüm koşulların ortalamaları alınarak belirlenmiştir. Nihai sınıflayıcı modeli, tüm eğitim kümesi kullanılarak elde edilen tek modeldir. Çapraz geçerleme yoluyla üretilen ve başarımlar sonuçları dikkate alınan ara modeller ise ortalama başarımlar sonuçları elde edildikten sonra silinmiştir. Bu yordama göre, FT sınıflayıcısının azaltılmış öznitelik veri kümesi üzerinde gösterdiği ortalama duyarlılık değeri 0.969 ve ortalama özgüllük değeri 0.984 olmuştur. Bu değerlere göre, sınıflayıcının her iki boyanma türünü yüksek bir olasılıkla doğru sınıflayabildiği anlaşılmaktadır.

Hata matrisi (Tablo 5.5) incelendiğinde ise eğitim kümesinde bulunana 384 hücre çekirdeğinden sadece 12 tanesinin yanlış sınıflandırıldığı görülmektedir.

Tablo 5.5. FT hata matrisi

Gerçek\Tahmin	Sınıf-1	Sınıf-2	Sınıf-3
Sınıf-1 (negatif)	127	1	0
Sınıf-2 (orta pozitif)	2	123	3
Sınıf-3 (güçlü pozitif)	0	6	122

Alıcı işletim karakteristik eğrileri (ROC curves) sınıflayıcı başarımını sınıflama kararı belirleyen çeşitli eşik düzeylerini temel alarak ölçer. Eşik değeri değiştikçe, sınıflayıcının hedef sınıfı ve kalan diğer sınıfları ayırt etme kabiliyeti aşağıda gösterilmiştir (Şekil 5.6).



Şekil 5.6. FT sınıflayıcısı alıcı işletim karakteristik (ROC) eğrileri

Bu grafikten anlaşılmaktadır ki, negatif, orta şiddetli ve kuvvetli boyanmış olan hücre çekirdekleri sınıflayıcıda karar eşik düzeyinin değişimine karşı bağımsız yapıdadır. Başka bir deyişle değişik veriler nedeniyle karar eşik düzeyi değişse bile sınıflayıcı tüm sınıf türlerinde istikrarlı bir ayırt etme kabiliyeti sergilemektedir. Güçlü pozitif ve negatif hücre çekirdeklerinin sınıflandırılmasında, sınıflayıcı modelin alıcı işletim karakteristik

eğrisi orta pozitif hücre çekirdeklerine nazaran daha dinç bir ayır etme davranışı göstermektedir.

5.3.3. Tanı Kararları ve Puanlama Sonuçları

Test resimleri üzerinde puanlama sonuçlarını uzman doktorlar ve bilgisayar tabanlı sistem arasında karşılaştırmak için ağırlıklandırılmış kappa istatistiksel örtüşme katsayısı kullanılmıştır. Puanlamaların karşılaştırılması nedeniyle, örtüşmeme durumunun uzaklığını ve yakınlığını da dikkate almak amacıyla ikinci dereceden ağırlık katsayıları kappa istatistiksel örtüşme hesaplamalarına dahil edilmiştir.

Kappa istatistiksel örtüşme katsayısı $[-1 \ 1]$ aralığında değerler alır. Bu değerler içerisinde negatif değerler örtüşmezliği, 0 değeri tesadüfî örtüşmeyi ve pozitif değerler ise örtüşmeyi ifade eder. Bu aralıkta, -1 değeri hiç örtüşmeme durumunu ve 1 değeri ise tam bir örtüşmeyi gösterir. Kappa katsayılarına göre olası sonuçlar ve bunların yorumlanması Tablo 5.6’de verilmiştir [81]. Ayrıca, Allred puanlama protokolüne göre gözlemcilerin ve bilgisayar tabanlı sistemin puanlamaları Tablo 5.7’de gösterilmiştir.

Tablo 5.6. Kappa katsayılarının yorumlanması

Kappa değeri	Örtüşme Derecesi
< 0.20	Kötü
0.21 - 0.40	Az
0.41 - 0.60	Orta
0.61 - 0.80	İyi
0.81 - 1.00	Çok iyi

Örtüşmeyi ölçmenin yanı sıra, ayrıca gözlemciler arası ilinti de Spearman’ın sıralı ilinti katsayısına göre incelenmiştir. Bu yöntem yaygın kullanılan bir gözlemciler arası uyumu ölçme yöntemidir. Ancak, gözlemci puanları arasında yüksek düzeyde ilinti bulunması puan dizilerinin birbirine yeterince yakın olduğunu göstermeye yetmez. Çünkü ortalamaları çok farklı iki puan dizisi arasında yüksek ilinti çıkabilir. Bu nedenle, örtüşme ve ilinti ölçmenin yanı sıra puan dizileri arasında istatistiksel bakımdan anlamlı

bir fark olup olmadığını da incelemek zorunludur. Bu amaçla, Wilcoxon eşleştirilmiş işaretli diziler testi kullanılarak, işaretli birer dizi olarak kabul edilen gözlemci puanlamaları arasında istatistiksel bakımdan anlamlı bir fark olup olmadığı araştırılmıştır.

Tablo 5.7. Test resimlerinde puanlamalar

Örnekler	FT	Gözl.- 1	Gözl.-2
1	0	0	0
2	6	7	6
3	0	0	0
4	5	2	5
5	0	0	0
6	4	0	2
7	0	0	0
8	0	0	0
9	0	0	2
10	6	6	7
11	5	4	5
12	0	0	0
13	7	6	8
14	7	7	5
15	0	0	0
16	6	4	4
17	8	8	8
18	0	0	0
19	4	0	0
20	8	8	8
21	6	5	6
22	7	8	7
23	7	8	8
24	8	8	8
25	6	5	6
26	5	3	3
27	7	8	8

Fark sınavında, kullanılan varsayım (sıfır hipotezi) puan dizilerinin ortalama değerlerinin birbirlerinden farksız olduğudur. Karar eşik düzeyi olasılığı 0.05 olarak benimsenmiştir. Eğer p değeri 0.05'ten büyükse varsayım doğru, küçükse yanlış olarak kabul edilmiştir.

Gözlemci doktor ve otomatik tanı sistemi puanlamaları (Tablo 5.7) üzerinde yapılan örtüşme, ilinti ve farksızlık sınamaların sonuçları aşağıda verilmiştir (Tablo 5.8).

Tablo 5.8. Puanlar arasında örtüşme, ilinti ve farksızlık

Değerlendirme	Ağırlıklı Kappa Örtüşmesi	Spearman İlinti	Wilcoxon Farksızlık testi
Gözl.-1-Gözl.-2	0.944*	0.948*	$p = 0.148$
FT-Gözl.-1	0.899*	0.963*	$p = 0.051$
FT-Gözl.-2	0.927*	0.943*	$p = 0.316$

Tablo 5.8’da verilen Kappa örtüşme katsayıları Tablo 5.6’ye başvuru yaparak incelenmiştir. Bu tabloda, ‘*’ sembolü ($p < 0.001$) olasılık bağıntısının geçerli olduğunu göstermektedir. Buna göre, bilgisayar tabanlı sistem ve uzman doktor gözlemcilerin puanlamaları arasında çok iyi düzeyde bir örtüşmenin bulunduğu görülmektedir. Ayrıca, hesaplama sonrasında, eşleştirilmiş puanlama dizileri arasında ilinti katsayısının yüksek olduğu belirlenmiştir. Son olarak, farksızlık sınaması, bilgisayar tabanlı sistem ve uzman doktor puanlamaları arasında istatistiksel bakımdan anlamlı bir fark bulunmadığını göstermiştir.

5.3.5. Tartışma ve Sonuçlar

Bu çalışmada, bilgisayar tabanlı otomatik sistem ve gözlemcilerin verdiği Allred puanları arasında örtüşme katsayısı ($\kappa = 0.899$ ve $\kappa = 0.927$, $P < 0.001$) olarak, Spearman ilinti katsayısı ($\rho = 0.963$ ve $\rho = 0.943$, $P < 0.001$) olarak ve Wilcoxon farksızlık sınaması sonucu ($P = 0.051$ ve $P = 0.316$) olarak bulunmuştur. Test resimlerinden 25 tanesi (25/27) doğru olarak pozitif veya negatif tanıyla tasnif edilmiştir. Fakat elde edilen bu sonuçları literatürde yapılan diğer çalışmalarla doğrudan karşılaştırmak mümkün olmamaktadır. Çünkü ER tanısının doğruluğunun bildirilmesinde çok çeşitli başarımlar ölçütleri rapor edilmiştir. Dahası, literatürde yayınlanan öncü çalışmalarda temel alınmış puanlama protokolleri, karar eşik yüzdeleri çeşitlidir. Benimsenmiş genel bir puanlama protokolü bulunmamaktadır. Yine de, benzer çalışmalara burada atıf yapılarak

literatürün ilgili kısmı genel olarak yorumlanacaktır. Bu sayede, bilgisayar tabanlı nesnel ER tanı sistemlerinin geliştirilmesinde tekdüze bir puanlama başarımlı ölçütünün benimsenmesine katkı sağlamak umulmaktadır. Tek düze bir puanlama ve başarımlı ölçütü arayışı, klinik çalışmanın kimyasal süreçlerinde standart yordamların uygulanmasını da teşvik edecektir.

Malzeme ve Yöntemler bölümünde Allred Puanlama tablosunda verildiği üzere, Allred puanlama yordamı düşük düzeyde boyama yüzdeliklerine karşı çok hassastır (0-1%, 1-10%, 10-33%, 33-66%, 66-100%). %2'den daha az boyanma yüzdesinde bile üç farklı tanı puanlaması ortaya çıkabilir. Bu hassasiyet dikkate alındığında, yanlış bir ya da birkaç hücre çekirdeği sınıflaması kolayca tüm tanı sonucunu değiştirebilmektedir. Oysa, literatürde yayınlanan diğer puanlama yordamlarında, daha geniş ve eşit aralıklarla bölünmüş çekirdek sayısı yüzdelikleri (0-25%, 25-50% etc. or 0-33%, 33-66%, 66-100% vb.) dikkate alınmaktadır [34, 35, 37, 82]. Ayrıca, 8 farklı ve doğrusal olmayan bölümlere sahip bir puanlama protokolünde gözlemciler arası örtüşmenin başarılı olması 2 veya 3 adet ve doğrusal bölümlere sahip yordamlara göre daha zordur. Sonuç olarak, Allred puanlama yordamı düşük boyanma yüzdeliklerine daha duyarlı ve gözlemciler arası örtüşmeyi literatürde yayınlanan diğer puanlama yordamlarına nazaran daha zor sağlayan bir puanlama yordamıdır. Bu zorluğa karşın, Allred puanlama yordamı gün geçtikçe laboratuvarlarda yaygın biçimde kullanılmaktadır [7,9]. Geniş çaplı klinik çalışmalar bu yordamın kullanılmasını tavsiye etmektedir [6].

ER durum değerlendirmesinde düşük boyanma yüzdeliklerine duyarlı bir puanlama yordamının önemi literatürde rapor edilmiştir [8]. Bu etkenleri dikkate alarak, bu çalışmada, bilgisayar tabanlı sistemin Allred puanlama yordamına göre tanı sonucu vermesi sağlanmıştır. Bu protokol, literatürde hücre çekirdeğinin bölümlenmesine dayalı önceki çalışmalarda kullanılmamıştır [34, 35, 37, 82-84].

Şu anki bilgimize göre literatürde Allred puanlama yöntemini baz alan ve hücre çekirdeği tespiti yaparak sayan az çalışma [32, 33] bulunmaktadır. Benzer bir çalışmada, Krecksak ve ark. Allred protokolüne göre ER durum değerlendirmesi için bilgisayar tabanlı otomatik bir sistem [32] önermiştir. Her ne kadar, daha yüksek bir örtüşme katsayısı ($\kappa = 0.981$) duyurulmuş olsa da, bu tez çalışmasından farklı olarak

önerilen NQ algoritması her bir resim için kullanıcı etkileşimi ve hassasiyet ayarlama işlemleri gerektirmektedir. Diğer taraftan, Krecksak ve ark. hücre çekirdeklerinin renk, aydınlık ve biçim özniteliklerinden faydalanırken bizim çalışmamız ayrıca dalgacık dönüşümü gibi birinci derece özniteliklerden daha başarılı özniteliklerden de (Tablo 5.1, Tablo 5.4) faydalanmaktadır. Bir başka çalışmada, yine hücre çekirdeği tespiti ve sayımı Allred yordamına göre gerçekleştirilmiştir [33]. Bu çalışmada, gözlemci ile bilgisayar algoritması arasında yüksek düzeyde ilinti bulunduğu duyurulmuştur (ER durum değerlendirmesi için ilinti 0.77; güven aralığı %95, Progesteron durum değerlendirmesi için ilinti 0.59, güven aralığı %0.95). Ancak, yapılan bu tez çalışmasında, bilgisayar tabanlı otomatik sistemle ile gözlemci doktorlar arasında ER durum değerlendirmesinde erişilen ilinti değeri çok daha yüksektir (Tablo 5.8).

Bazı çalışmalarda, ER tanısı için hücre çekirdeği tespiti ve sayımı yapmakla birlikte farklı puanlama protokolü kullanılmıştır. Kostopoulos ve ark. danışmansız bir öğrenme yöntemi kullanarak gözlemci ve bilgisayar tabanlı sistem arasında orta düzeyde bir örtüşme (Kendall örtüşme katsayısı $W = 0.79$) elde etmiştir [37]. Aynı yazar ve ark. renk ve desen tabanlı bir görüntü analiz yöntemi önermiş [35] ve bu yöntemle örtüşme düzeyini geliştirmişlerdir (Kendall's $W = 0.875$, $P < 0.001$). Aynı yazarın bu konudaki bir başka çalışmasında desen enerji özniteliklerinden faydalanılmış ve böylece sistem ve gözlemci puanlamaları arasında yüksek ilinti değerine erişildiği bildirilmiştir ($\rho = 0.89$, $P < 0.001$) [34]. Schnorrenberg ve ark. bilgisayar destekli bir hücre çekirdeği tespit sistemi (BASS) sunmuşlardır. Bu sistemde güçlü boyanma değeri olan hücre çekirdekleri üzerinde gözlemciler ve sistem arasında yüksek ilinti ($0.78 < \rho < 0.86$, $P < 0.001$) bulunduğu bildirilmiştir. Ancak, zayıf boyanma olan hücre çekirdekleri üzerinde sistem ve gözlemci doktorlara arasında ilinti değeri düşmüştür ($\rho = 0.51$ ve $\rho = 0.38$, $P < 0.001$) [36].

Literatürde global eşikleme ve piksel tabanlı ölçütlere dayalı bir çok çalışma vardır. Bu çalışmalar içerisinden gözlemci ve sistem arasında örtüşme ve ilinti başarımını verenler üzerinde durulacaktır. Bu nedenle, ER tanısında biyokimyasal değerlendirme ile immünohistokimyasal değerlendirme başarımını karşılaştıran şu çalışmaların sonuçları burada ele alınmayacaktır: Charpin ve ark.'nın SAMBA 200 sistemi [39], Furukawa ve

ark. CAS 200 sistemi [45], Hatanaka ve ark.'nın WinROOF sistemi [43], Lehre ve ark.'nın Photoshop of Adobe Systems'e dayalı sistemi [41].

Öte yandan gözlemciler ile immünohistokimyasal yöntemde ER tanı başarımını karşılaştıran çalışmalar şunlardır. Diaz ve ark. QCA görüntü analiz sistemi ile gözlemciler arasında örtüşmeyi $\kappa=0.84$ olarak bildirmişlerdir [85]. Turbin ve ark. Ariol adını verdikleri sistemi kullanarak elde ettikleri en yüksek örtüşme oranının ($\kappa=0.9021$, 95% CI: 0.8854–0.9180) olduğunu duyurmuşlardır [40]. Gokhale ve ark. iki ticari sistemi, “The ChromaVision Automated Cellular Imaging System (ACIS)” ve “Applied Imaging Ariol SL-50” i karşılaştıran bir çalışma yayınlamıştır. İstatiksel Gamma sınavına göre gözlemcilerin ve bilgisayar tabanlı sistemlerin puanlamaları arasında, en yüksek ilinti değeri ACIS ile ($\gamma=0.91$) elde edilmiştir [47]. Mofidi ve ark. “Adobe Photoshop” kullanarak optik yoğunluğa dayalı görüntü maskeleme ve gözlemci yüzdelikleri arasında yakın bir ilinti ($r^2=0.844$, $P<0.0001$) bulunduğunu bildirmiştir [42]. McClelland ve ark. görüntü üzerinde boyanan alanın yüzdelik oranına CAS 100 sistemi ile bakarak gözlemci ile sistemin ürettiği ER pozitif yüzdelik arasında ($r = 0.919$, $P<0.01$, $n = 94$) ilinti değerini, ER negatif yüzdelik arasında daha düşük ($r = 0.821$, $P<0.01$, $n = 68$) ilinti değerini elde etmişlerdir [44]. Chung ve ark. 11 örnekten alınan 29 ayrı görüntü üzerinde yapılan incelemede AQUA sistem ve altın standart yüzdelikler arasında logaritmik ölçekte doğrusal bir yaklaşıklığın ($R=0.8903$) bulunduğunu göstermişlerdir [46]. Camp ve ark. patoloğ değerlendirmesi ve AQUA tabanlı sistem arasında yüksek ilinti bulunduğunu belirtmişlerdir [48]. Ayrıca, onlar patoloğ gözlemci değerlendirmelerinin farklılaşması üzerinde bir inceleme yapmış ve bilgisayar tabanlı sistemin patoloğ gözlemciye nazaran daha az farklılaşan ve daha iyi yeniden üretilebilir ($R = 0.824$ karşı $R = 0.732$) sonuçlar verdiğini rapor etmişlerdir. Loyd ve ark. iki görüntü inceleme sistemini karşılaştırmış, “Definiens (Munich, Germany)” ve “Aperio Technologies (Vista, CA, USA)” ve istatistiksel sonuç vermeksizin her iki yöntemin de örnekler üzerinde patoloğ değerlendirmelerine göre 10/10 başarı sağladığını ifade etmişlerdir [49]. Bolton ve ark., gözlemci ve bilgisayar tabanlı sistem arasında ER negatif/pozitif ve tasnifi bakımından çok iyi örtüşme değerleri elde ettiklerini (ER- α için $\kappa=0.86$ -0.91 aralığında ve ER- β için $\kappa=0.80$ -0.86 aralığında) duyurmuşlar ve bu çalışmada ayrıca üç ayrı ticari sistemin, “TMAX (Beecher Instruments, Sun Prairie, WI)”, “Ariol (Applied Imaging, Grand Rapids, MI)” ve

“TMALab II (Aperio, Vista, CA)”, başarımlı karşılaştırmasını yapmışlardır [50]. Ters sarmaşım yöntemiyle renk ayrıştırma analizi yapabilen “ImmunoRatio” isimli web tabanlı bir IHC analiz yazılımını, Tuominen ve ark., ücretsiz olarak kamu kullanımına sunmuşlardır. Bu çalışmada, doğrulama ayarlaması yapılmış ImmunoRatio yazılımının gözlemci tespitleri ile pozitif çekirdek yüzdeliği bakımından çok yüksek ilinti değerini ($r=0.98$) verdiği ve bunun çeşitli biyolojik renklendiricilerde, östrojen, Progesteron ve Ki-67 alıcı, gözlemlendiği bildirilmiştir [83]. Rexhepaj ve ark. pozitif boyanma yüzdeliği bakımından gözlemci ve sistem arasında yüksek ilinti bağıntısı elde ettiklerini (Spearman's $\rho = 0.9$, $P < 0.001$) bildirmişlerdir [84]. Aynı çalışmada, ER puanlamaları arasında ise güçlü bir ilinti bulunduğu (Spearman's $\rho = 0.74$, $P < 0.001$) belirtilmiştir. Sharangpani ve ark. kendi önerdikleri yöntemin yanı sıra ticari yazılımları listeleyen ve bir çalışma yayınlamışlardır [33]. Bu konuda ticari yazılımlar üzerinde özelleşen bir başka yayın Cregger ve ark. tarafından yapılmıştır [86]. Tüm bu çalışmaların ortak özelliği, bunların piksel tabanlı ve alan tabanlı ölçümlere (örneğin boyanmış piksel yüzdelik, alan yüzdelik) bağlı olması ve küresel eşikleme yöntemlerini kullanıyor olmalarıdır.

Bunlardan farklı olarak, bu tez çalışması hücre çekirdeği tespitine, tasnifine ve sayılmasına dayanmaktadır. Bu yaklaşım gerçekte uygulanan tıbbi işlem ile aynıdır ve elde edilen çıkış verisi doğrudan özel bir yazılım gerektirmeksizin tıp hekimi tarafından yorumlanabilir. Ayrıca, renk ve aydınlık niteliklerinin yanı sıra kullanılan desen öznitelikleri ve bunların içersinde bu çalışmada önerilen dalgacık enerji özniteliği sınıflayıcının ayırım yapabilme kabiliyetini arttırmıştır. Nitekim bu sonuç, öznitelik eleme aşaması sonrasında seçilen 11 öznitelikten 10 tanesinin desen tabanlı olmasıyla doğrulanmaktadır (Tablo 5.1).

Hücre çekirdeği sayımına ve tasnifine dayalı yaklaşımla elde edilen sonuçlar nesnel ve yeniden üretilebilir sonuçlardır. Yeniden üretilebilir olmasının önemi; her incelemede gözlemcilerin farklı olmasından veya aynı gözlemcinin farklı zamanlarda yaptığı incelemelerden kaynaklanan farklılaşmanın, başka bir deyişle öznelliğin, ortadan kalkmasıdır. Ayrıca, kullanılan puanlama protokolü ve ER tanısında karar eşik değeri ne olursa olsun, tanı için başvurulacak temel veri hazır ve değişmez olduğundan örnek üzerinde tekrar inceleme yapmaya gerek kalmayacaktır. Öte yandan, başvurulacak temel

veriyi bildirmeksizin sadece tanı kararını bildiren ve öznel olan geleneksel değerlendirmelerde bu esneklik olanaksızdır. Bu nedenlerle, bilgisayar tabanlı nesnel ve hücre çekirdeği tespitine dayalı yaklaşım, doktorlar için tıbbi kararların kayıt altına alınarak belgelendirilmesinde ve sonradan yeniden kullanılmasında fayda sağlayacaktır. En azından, farklı laboratuvarlar sonuçlarını karara esas temel veriler üzerinden kolayca paylaşabilirler.

Piyasada bulunan ticari yazılımlar, pahalı ve özel kullanım bilgisi gerektiren uygulamalardır. Bu tez çalışmasında kullanılan yöntemler herkesin ücretsiz olarak erişebileceği kütüphane dosyalarından oluşmaktadır [80]. Aynı yaklaşımların diğer araştırmacılar tarafından denenmesini sağlamak amacıyla açık kodlu Java kütüphanelerinin adresi bir sonraki bölümde verilmiştir. Kullanıma açık birçok yapay öğrenme yönteminin bu ve benzeri uygulamalarda açık kodlu kütüphane dosyaları ile sınanması, yapay öğrenme yöntemleri ile elde edilen başarımın yükselmesine aracı olacaktır. Diğer taraftan yapay öğrenme yöntemlerinin kıyaslanmasında kullanılan başarım ölçütlerinin de bu yolla ortak bir kabule yakınsayacağı umulmaktadır.

Bu bölümde sergilenen ve karar ağacı tabanlı olan öğrenme yöntemin bir diğer özelliği, diğer karar ağacı yöntemlerinde olduğu gibi insanlar tarafından yapay sinir ağları vb. yapısı karmaşık olan diğer öğrenme yöntemlerine kıyasla daha anlaşılır olmalarıdır [63, 72]. Diğer taraftan, karar ağacı tabanlı yöntemler bazı kusurları beraberinde getirir. Bu kusurlardan birisi en az düğümü içeren ve çözüme en uygun ağaç yapısının garanti edilememesi, diğeri ise ikilik karar ağaç yapısında araştırma uzayının dik düzlemli kareler biçiminde bölünmesi ve bu nedenle eğrisel bağıntıların karmaşık yapılarla tarif edilmesidir. Bu kusurların üstesinden gelmek amacıyla önerilmiş olan FT yöntemi [77] özellikle tercih edilmiştir.

FT sınıflayıcısını uyguladıktan sonra, gözlemci puanlamalarıyla çok iyi düzeyde örtüşen puanlama sonuçları elde edilmiştir (Tablo 5.8). Ayrıca elde edilen puanlamaları arasında yüksek düzeyde ilinti bulunmuş ve sonuçların istatistiksel anlamda farksız olduğu Wilcoxon fark sınamasıyla anlaşılmıştır. Ancak, şu dikkat çekmektedir ki; FT ve gözlemci-1 arasında örtüşme diğer gözlemciye nazaran azdır ve gözlemci-1 ile sistem arasındaki farksızlık olasılığı istatistiksel karar eşiği olan %5 değerine çok yakındır. Yine

de, gözlemciler ve önerilen sistem arasında ortaya çıkan bağıntı çalışmanın başarı olduğunu kanıtlamaya yeterlidir. Başarımı daha da fazla arttırmak mümkündür. Bu amaçla, daha geniş bir eğitim veri kümesi kullanılabilir, çeşitli boyut azaltma yöntemlerinin başarımı incelenerek elde edilen sınıflama başarımı yükseltilebilir veya daha başka sınıflayıcılar aynı sistem içerisinde kullanılabilir. Her ne kadar resim ön işleme teknikleri sistem içerisinde kullanılmamış olsa da, resim ön işleme tekniklerinin sistem başarımına etkisi incelenebilir ve Otsu eşikleme yönteminden farklı bölümlendirme yöntemleri, farklı biçimbilim işlemleri kullanılarak genel sistem başarımının yükseltilmesi araştırılabilir.

Önerilen bilgisayar tabanlı ER durum değerlendirme sisteminin uygulanmasında bazı kısıtlar bulunmaktadır. Birincisi, Mikroskop görüntüsünde (x40) örnek üzerinde hangi kesimin durumu temsil kabiliyetine sahip olduğu doktor tarafından seçilmeli ve resim dosyası olarak kaydedilmelidir. İkincisi, tüm örnekler aynı kimyasal boyama süreçlerinden geçirilmelidir. Üçüncüsü, üzerinde elle yapılan hatalar, bulanık kesimler veya sitoplazmik boyanma bulunan resimler çıkarılmalıdır. Çünkü otomatik yöntem sitoplazmik boyanmayı ayırt edecek biçimde tasarlanmamıştır, bu nedenle sitoplazmik boyanan kesimler hücre çekirdeği olarak tanımlanabilir. Son kısıt, Yapay öğrenme yöntemleri kullanıldığından bu yöntemlere dayalı modellerin eğitilmesi için bir defaya mahsus ek bir zaman ve emek harcanmalıdır.

Özetle, yapılan deneysel incelemenin sonunda, fonksiyonel karar ağacı modeline dayanan bilgisayar tabanlı otomatik ER durum değerlendirme sistemi, patoloğlar için destekleyici bir karar sistemi olarak kullanılabilir. Bu yolla, Allred puanlama protokolüne göre yeniden elde edilebilir ve nesnel puanlama ve ER tanı kararları verilebilir.

6. BÖLÜM

ÇEŞİTLİ ÖZNETELİK GURUPLARI VE SINIFLAYICILARIN OTOMATİK TANI BAŞARIMINDA BULGULAR

6.1. Özet

ER durum değerlendirmesinde nesnel ve yeniden üretilebilir sonuçlar alabilmek amacıyla, bu bölümde daha önce literatürde yayınlanan çeşitli yapay öğrenme modellerinin sınıflama ve tanı başarımları ölçülerek karşılaştırmalı olarak incelenmiştir.

Yarıçap fonksiyonlu yapay sinir ağları, en yakın- k komşuluk metodu, destek vektör makineleri, fonksiyonel karar ağaçları, sade bayes sınıflayıcı ve k -orta nokta öbekleme yöntemlerinin sınıflama başarımları karşılaştırılmıştır. Ayrıca, çeşitli öznetelik gurupları, sağladıkları doğru sınıflama yüzdeleri bakımından karşılaştırılmıştır. Son olarak, sınıflayıcıların tanı ve puanlama başarımları test resimleri üzerinde deneysel olarak incelenmiştir.

Elde edilen sonuçlara göre, fonksiyonel karar ağaçları ve sade bayes sınıflayıcı bileşiminden oluşan tümleşik bir sınıflayıcı ER durum değerlendirmesinde hem puanlama hem de negatif veya pozitif tanı konması bakımından en başarılı kararları göstermiştir. Bu iki metot diğer yapay öğrenme yöntemlerine nazaran daha iyi genelleme kabiliyeti ortaya koymuştur. Öte yandan öznetelik gurupları arasında renkli dalgacık desen özneteliklerinin, Laws desen enerji ve eş oluşum matrisi özneteliklerinden daha başarılı sonuçlar ortaya koyduğu tüm yapay öğrenme yöntemleri üzerinde gözlenmiştir.

6.2. Malzeme ve Yöntemler

Bu bölümde kullanılan eğitim ve test verisi, Bölüm-5'te kullanılan veri ile tamamen aynıdır. Eğitim ve test veri kümesinin oluşturulmasında da aynı usuller takip edilmiştir. Sınıflayıcıların eğitilmesinde ve başarımlarının ölçülmesinde Bölüm-5'te açıklanan yaklaşıma sadık kalınmıştır. Böylece, elde edilen sınıflayıcı başarımların değerleri, 10 defa tekrarlanan 10-parçalı çapraz geçişleme yordamıyla elde edilen 100 adet koşmanın ortalaması alınarak hesaplanmıştır. Bir önceki bölümde açıklandığı üzere, 11 elemandan oluşan ve eleme yapılarak seçilmiş olan renkli öznitelik grubu sınıflayıcı giriş verisi olarak kullanılmıştır. Tanı kararının verilmesinde Allred puanlama yordamına esas alınmıştır. Yapılan denemelerde, yine "WEKA Experiment Tool" kullanılmış, istatistiksel analizde R programlama dili ve sistemin kodlanmasında ve icrasında Tablo 6.1'de verilen WEKA Java programlama dili kütüphaneleri ve Matlab, The Mathwork Inc. yazılım araçları kullanılmıştır. Bahse konu Java kütüphaneleri hakkında ayrıntılı bilgi [80] numaralı kaynakta bulunabilir.

Tablo 6.1. WEKA kütüphanesinden çağrılan sınıflayıcılar ve ilk değer atamaları

Sınıflayıcı	WEKA kütüphanesi sınıfları ve ilk değer atamaları
Voting Scheme	weka.classifiers.meta.Vote -S 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.trees.FT -I 15 -F 0 -M 15 -W 0.0" -R MAX
Functional Trees	trees.FT '-I 15 -F 0 -M 15 -W 0.0'
K-nearest neighborhood search	lazy.IBk '-K 3 -W 0 -I -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\"\"
Support vector machines	functions.SMO '-C 1.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1 - K \"functions.supportVector.PolyKernel -C 250007 -E 1.0\"
Radial basis functions network	functions.RBFNetwork '-B 3 -S 1 -R 1.0E-8 -M -1 -W 0.1'
Naive Bayes	bayes.NaiveBayes
Simple K-means	weka.classifiers.meta.ClassificationViaClustering -W weka.clusterers.SimpleKMeans -- -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

6.3. Sonuçlar

6.3.1. Öznitelik Sonuçları

Başlangıçta üretilen öznitelik grupları ve sonradan bunlardan boyut azaltma yöntemi ile seçilerek çıkarılan yeni öznitelik gruplarının incelemeye alınan yapay öğrenme yöntemlerine göre ortaya koydukları sınıflama başarımları Tablo 6.2’de gösterilmektedir.

Tablo 6.2. Özniteliklere göre öğrenme yöntemlerinin doğru sınıflama yüzdeleri

Veri kümesi ve öznitelik sayısı	VOTE	FT	KNN	SVM	RBFN	NB	KM
Birinci derece gri ton öznitelikler (6)	59,45 (5,88)	59,33 (7,05)	57,44 (7,56)	51,30 (6,56)*	51,10 (8,22)*	50,59 (6,36)*	41,32 (6,76)*
İkinci derece gri ton öznitelikler (8)	53,76 (7,43)	54,17 (7,57)	44,47 (6,97) *	55,56 (7,97)	48,06 (7,75)	49,85 (6,87)	41,17 (7,90) *
Laws gri ton enerji öznitelikleri (14)	45,74 (7,71)	51,77 (7,22) v	42,06 (6,75)	46,39 (6,78)	42,95 (6,14)	43,01 (6,11)	42,42 (7,66)
Dalgacık gri ton enerji öznitelikleri (8)	57,78 (6,44)	61,45 (7,70)	56,47 (7,71)	57,92 (6,37)	57,46 (7,14)	54,98 (6,41)	41,89 (8,27)*
Seçilen gri ton öznitelikler (4)	63,73 (6,09)	63,65 (7,38)	60,53 (6,48)	62,69 (6,43)	64,75 (6,56)	58,68 (6,44)*	56,49 (6,81)*
Birinci derece renkli öznitelikler (18)	93,29 (4,74)	94,93 (3,52)	88,47 (4,84)*	93,62 (4,20)	87,00 (5,49)*	80,98 (7,16)*	53,13 (7,26)*
İkinci derece renkli öznitelikler (24)	84,69 (5,48)	90,45 (4,77) v	84,06 (6,04)	92,09 (4,77) v	82,65 (5,55)	75,23 (6,64)*	55,79 (6,45)*
Laws renkli enerji öznitelikleri (42)	60,81 (8,38)	92,81 (4,24) v	80,00 (5,81) v	90,45 (4,61) v	63,16 (6,93)	49,83 (6,66)*	43,30 (7,59)*
Dalgacık renkli enerji öznitelikleri (24)	94,51 (4,62)	97,34 (3,10) v	91,56 (5,24)	97,39 (2,65)	88,51 (5,55)*	80,32 (6,42)*	63,00 (4,84) *
Seçilen renkli öznitelikler (11)	97,39 (2,88)	96,67 (3,10)	95,89 (3,03)	97,50 (2,67)	97,37 (2,83)	97,58 (2,74)	87,53 (5,70)*
İstatiksel olarak (iyi ‘v’/farksız/kötü, ‘*’)	(v / /*)	(4/6/0)	(1/7/2)	(2/7/1)	(0/7/3)	(0/4/6)	(0/1/9)

İlk sütunda (Tablo 6.2), çeşitli öznitelik grupları isimleriyle birlikte belirtilmiş ve kaç adet öznitelik barındırdığı parantez içindeki sayı ile gösterilmiştir. Diğer sütunlar, incelemeye alınan yapay öğrenme yöntemlerinin doğru sınıflama yüzdelerini 10 defa tekrarlanan 10-parçalı çaprazlamanın ortalama değerleri olarak göstermektedir.

Her bir öğrenme yöntemi için aynı öznitelik türü üzerinde 100 adet başarımlar sonucu elde edilerek, öğrenme yöntemlerinin ikili karşılaştırmalarla istatistiksel bakımdan anlamlı bir farka sahip olup olmadıkları tespit edilmiştir. Bu yolla iki öğrenme yöntemini kıyaslamak için eşleştirilmiş t-testi kullanılmıştır. Bu testte diğer tüm öğrenme

yöntemleri ile ikili kıyaslamaya alınan baz öğrenme yöntemi VOTE ile temsil edilen sınıflayıcı birleşimi olmuştur. Çünkü bu birleşik öğrenme modeli diğer öğrenme yöntemlerine kıyasla başarımlar bakımından öne çıkmaktadır. Kullanılan diğer yapay öğrenme yöntemleri için kullanılan kısaltmalar şunlardır; fonksiyonel karar ağacı (FT), k -en yakın komşuluk yöntemi (KNN), destek vektör makineleri (SVM), yarıçap fonksiyonlu ağlar (RBFN), sade bayes sınıflayıcı (NB) ve k -orta nokta öbekleme yoluyla sınıflama (KM).

Deneyisel incelemede ayrıca, öğrenme yöntemleri üzerinde farklı öznitelik gruplarının sınıflama başarımlar değerleri de ikili olarak karşılaştırılmıştır. Bu karşılaştırma sonuçları Tablo 6.3’de gösterilmiştir. Yine burada, 100 adet koşturmanın sonuçları eşleştirilmiş t-testi ile iki öznitelik grubu arasında kıyaslanmıştır. İkili kıyaslamalarda, dalgacık desen öznitelik grubu karşılaştırma tabanı olarak alınmıştır.

Tablo 6.3. Öznitelik gruplarının öğrenme yöntemlerinde doğru sınıflama yüzdeleri

Sınıflayıcı	Dalgacık renkli desen	Birinci derece gri ton	İkinci derece gri ton	Laws gri ton desen	Dalgacık gri ton desen	Seçilen gri ton	Birinci derece renkli	İkinci derece renkli	Laws renkli desen	Seçilen renkli
VOTE	94,51 (4,62)	59,45 (5,88)*	53,76 (7,43)*	45,74 (7,71)*	57,78 (6,44)*	63,73 (6,09)*	93,29 (4,74)	84,69 (5,48)*	60,81 (8,38)*	97,39 (2,88)
FT	97,34 (3,10)	59,33 (7,05)*	54,17 (7,57)*	51,77 (7,22)*	61,45 (7,70)*	63,65 (7,38)*	94,93 (3,52)	90,45 (4,77)*	92,81 (4,24)*	96,67 (2,75)
KNN	91,56 (5,24)	57,44 (7,56)*	44,47 (6,97)*	42,06 (6,75)*	56,47 (7,71)*	60,53 (6,48)*	88,47 (4,84)	84,06 (6,04)*	80,00 (5,81)*	95,89 (3,03) _v
SVM	97,39 (2,65)	51,30 (6,56)*	55,56 (7,97)*	46,39 (6,78)*	57,92 (6,37)*	62,69 (6,43)*	93,62 (4,20)*	92,09 (4,77)*	90,45 (4,61)*	97,50 (2,67)
RBFN	88,51 (5,55)	51,10 (8,22)*	48,06 (7,75)*	42,95 (6,14)*	57,46 (7,14)*	64,75 (6,56)*	87,00 (5,49)	82,65 (5,55)*	63,16 (6,93)*	97,37 (2,83) _v
NB	80,32 (6,42)	50,59 (6,36)*	49,85 (6,87)*	43,01 (6,11)*	54,98 (6,41)*	58,68 (6,44)*	80,98 (7,16)	75,23 (6,64)	49,83 (6,66)*	97,58 (2,74) _v
KM	63,00 (4,84)	41,32 (6,76)*	41,17 (7,90)*	42,42 (7,66)*	41,89 (8,27)*	56,49 (6,81)	53,13 (7,26)*	55,79 (6,45)*	43,30 (7,59)*	87,53 (5,70) _v
	(_v / /*)	(0/ 0/7)	(0/ 0/7)	(0/ 0/7)	(0/ 0/7)	(0/10/6)	(0/ 5/2)	(0/1 /6)	(0/0 /7)	(4/3 /0)

Her iki tablonun (Tablo 6.2 ve Tablo 6.3) son satırında, yeniden örneklemeyle düzeltilmiş iki ölçüm dizisinin karşılaştırılmasında kullanılan t-testi sonuçları özetlenmiştir. ($x/y/z$) gösterimi kıyaslanan dizinin taban alınan diziye göre istatistiksel bakımdan anlamlı bir fark taşıyıp taşımadığını gösterir. Başka bir deyişle, iki dizisinin eşit ortalamaları olan birer topluluktan örneklenmiş olma olasılığın bakılır. Eğer bu olasılık %5’ten küçükse bu iki dizinin aynı topluluktan örneklenmiş olma olasılığının

çok küçük olduğuna ve istatistiksel bakımdan anlamlı bir fark bulunduğuna kanaat getirilir. (x) taban alınan ölçüm dizisinin diğerin göre kaç defa istatistiksel bakımdan anlamlı olarak daha iyi sonuçlar verdiğini gösterir. Benzer biçimde, (y) kaç defa farksız sonuçlar alındığını ve (z) ise taban ölçüm dizisinin istatistiksel bakımdan anlamlı olarak kaç defa daha başarılı sonuçlar verdiğini ifa eder. Sonuçların yanında gösterilen semboller; taban ölçüm dizisine göre (v) istatistiksel bakımdan anlamlı olarak daha iyi, (*) daha kötü sonuçları işaret eder. Ayrıca, her iki tabloda elde edilen ortalama doğru sınıflama yüzdelerinin yanına parantez içerisinde ortalamadan standart sapma değerleri de verilmiştir.

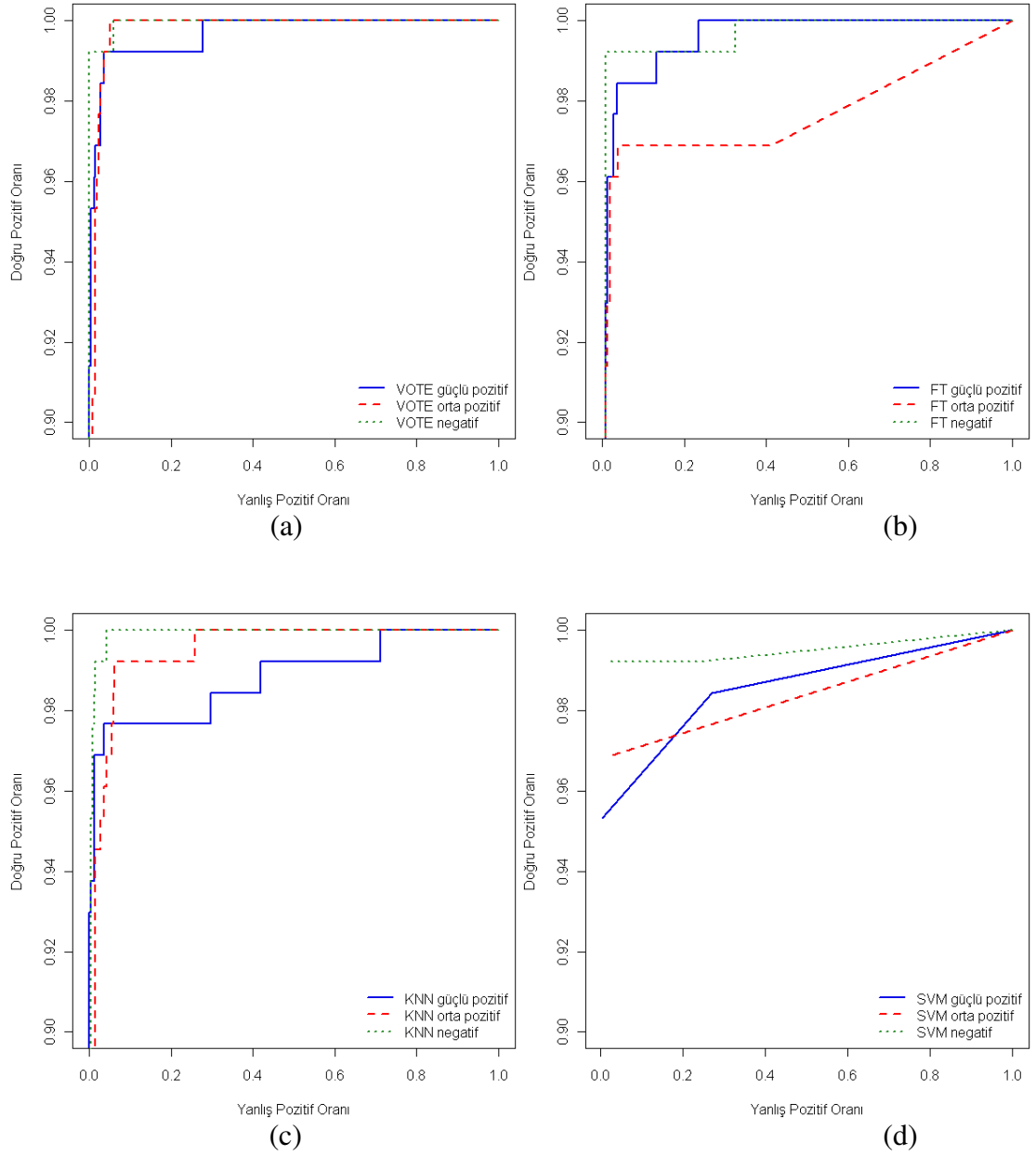
6.3.2. Sınıflayıcı Sonuçları

Sınıflama başarımını karşılaştırmak amacıyla tüm öğrenme yöntemlerinin, duyarlılık, özgüllük, doğru sınıflama yüzdesi ve alıcı işletim karakteristik eğrisi (ROC) özellikleri çapraz geçerleme yordamının yürütülmesi esnasında kayıt altına alınmıştır. Öğrenme yöntemlerinin doğru sınıflama yüzdeleri zaten Tablo 6.2’de verilmişti. Burada ayrıca, öğrenme yöntemlerinin 100 defa koşturulması sonrasında elde edilen duyarlılık ve özgüllük değerleri Tablo 6.4’te verilmiştir. Burada verilen duyarlık ve özgüllük değerleri sınıf-1, 2 ve 3 için elde edilmiş tüm sonuçların ortalamasıdır.

Tablo 6.4. Öğrenme yöntemlerinin duyarlılık ve özgüllük değerleri

	VOTE	FT	KNN	SVM	RBFN	NB	KM
Duyarlılık	0.994	0.996	0.962	0.992	0.994	0.992	0.706
Özgüllük	0.994	0.994	0.992	0.994	0.993	0.991	1.000

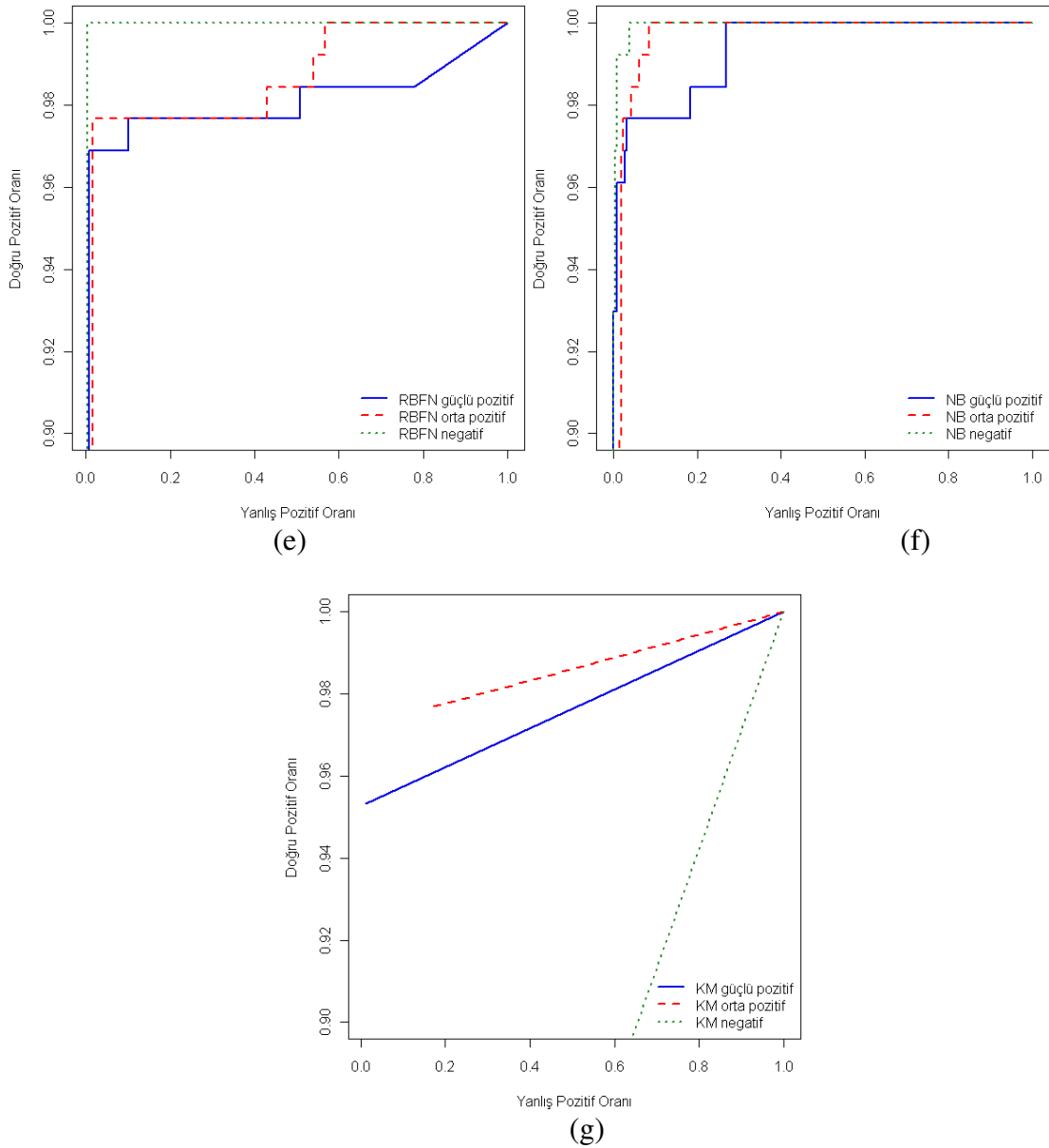
Sınıflama başarımının ölçümünde kullanılan bir başka ölçü, alıcı işletim karakteristik eğrilerinin grafik çıktılarıdır. Her bir sınıf için tüm öğrenme yöntemlerinin sergilediği alıcı işletim karakteristik eğrileri Şekil 6.1 ve Şekil 6.2’de gösterilmektedir. Burada negatif doğru ve pozitif doğru oranlarının hedef sınıflayıcı için farklı karar eşik düzeylerine göre değişimi gösterilmiştir. Bu eğriler, “WEKA Experiment Tool” yazılım aracı ve R programlama dili ile elde edilmiştir.



Şekil 6.1. Sınıflayıcıların ROC eğrileri; (a) tümleşik sınıflayıcı (b) fonksiyonel karar ağacı, (c) en yakın-k komşuluk arama yöntemi, (d) destek vektor makineleri

Bilindiği üzere, bu grafiklerde karakteristik eğri sol üst köşeye yaklaştıkça sınıflama modelinin hedef sınıfı başarıyla ayırt etme kabiliyeti artar. Bu eğrilere bakıldığında dikkat çeken şudur; sınıflayıcı modelin farklı karar eşik düzeylerinde doğru pozitif oranında görülen 0.90 ve üstü yüksek değerlere karşın yanlış pozitif oranı artış 0-0.10 aralığında kalmaktadır. K -orta nokta öbekleme yöntemi istisna olmak üzere gözlenen bu durum, incelenen modellerin dinç bir genelleme ve ayırt etme kabiliyeti sergilediğine yorumlanabilir. Fonksiyonel karar ağaçları ve birleşik sınıflayıcı dışında burada

incelenen tüm sınıflayıcılar geçmişte literatürde yayınlanan ve östrojen alıcı varlığının tespitinde önerilen yöntemlerdir ve bu danışmanlı yöntemler yarışmacı karakteristik eğrileri sergilemektedir. *K*-orta nokta öbekleme yönteminin özellikle östrojen alıcı negatif hücre çekirdeklerinde ortaya koyduğu yüksek yanlış pozitif oranı önemli bir zafiyet olarak dikkat çekmiştir (Şekil 6.2(g)). Eğrilerde, 0.96 doğru pozitif oranından sonra sınıflayıcı modellerinde davranışında ayrışma belirgin hale gelmektedir. Sade bayes sınıflayıcı ve özellikle birleşik sınıflayıcı en başarılı ayırt etme davranışı göstermiştir (Şekil 6.1(a)) ve (Şekil 6.2(f)).

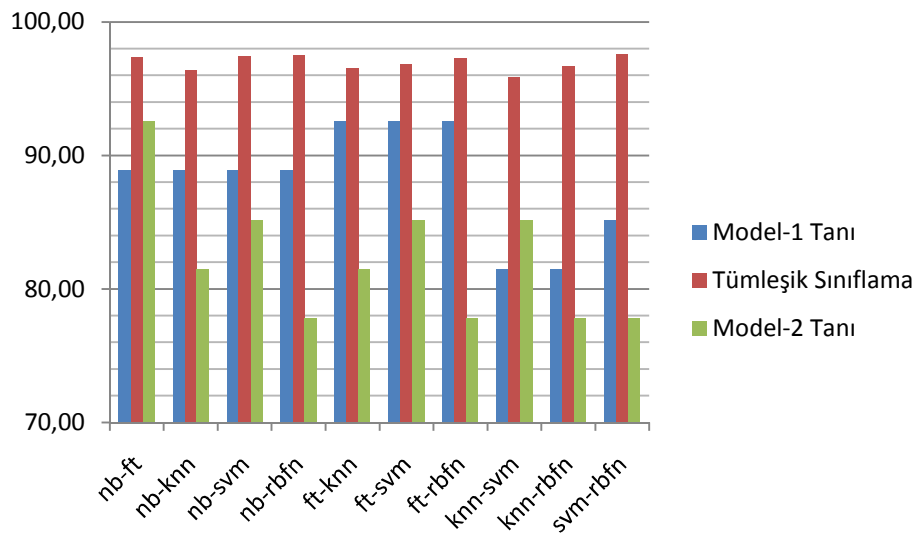


Şekil 6.2. Sınıflayıcıların ROC eğrileri; (e) yarıçap fonksiyonlu ağlar, (f) sade bayes sınıflayıcı, (g) *k*-orta nokta öbekleme

Alıcı işletim karakteristik eğrileri ayrıca hangi sınıf türünde hangi sınıflayıcının ne kadar istikrarla doğru sınıflama yapabildiğinin bir ölçüsüdür. Örneğin fonksiyonel karar ağaçlarına bakıldığında, östrojen alıcı durumu orta pozitif hücre çekirdeklerinin sınıflandırılması diğer iki tür sınıfa göre belirgin bir zayıflık göze çarpmaktadır (Şekil 6.1(b)). Öte yandan östrojen alıcı durumu güçlü pozitif çekirdeklerin sınıflandırılmasında fonksiyonel karar ağaçları sade bayes sınıflayıcıdan (Şekil 6.2(f)) daha başarılıdır. Bu iki tür sınıflayıcının birlikte kullanımı birleşik modelde (Şekil 6.1(a)) etkisini göstermiştir ve sonuçta birbirlerinin zayıflıklarını kapatmıştır. Yukarıdaki eğriler tasarımcıya hangi tür sınıfta hangi tür modelin diğerlerine göre üstünlük sağlayabileceği hakkında fikir vermektedir.

6.3.3. Puanlama Sonuçları

Çeşitli sınıflayıcı birleşimleri ve bu birleşimlerden alınan başarımların yüzdeleri Şekil 6.3’de grafikte verilmektedir. Bu grafikte birinci seri, birleşik sınıflayıcıyı oluşturan birinci modelin negatif pozitif ayrımında doğru östrojen alıcı tanısı verme yüzdesini, ikinci seri birleşik sınıflayıcının doğru sınıflama yüzdesini ve üçüncü seri birleşik sınıflayıcıyı oluşturan diğer modelin negatif pozitif ayrımında doğru tanı yüzdesini göstermektedir.



Şekil 6.3. Birleşik sınıflayıcı modellerinin doğru sınıflama yüzdeleri

Allred puanlama protokolüne göre gözlemcilerin ve bilgisayar tabanlı sistemde yer alan yapay öğrenme yöntemleriyle elde edilen puanlamaların Kappa örtüşme katsayısına, Spearman sıralı ilinti katsayısına ve Wilcoxon fark sınavına göre başarımları Tablo 6.5'te verilmiştir.

Tablo 6.5. Öğrenme yöntemleri ve gözlemciler arasında örtüşme, ilinti ve farklılık

Değerlendirme	Cohen Kappa Örtüşme	Spearman İlinti katsayısı	Wilcoxon farklılık testi
gözl.-1-gözl.-2	0.944*	0.948*	$p = 0.148$
VOTE-gözl.-1	0.899*	0.961*	$p = 0.051$
VOTE-gözl.-2	0.949*	0.972*	$p = 0.422$
FT-gözl.-1	0.899*	0.963*	$p = 0.051$
FT-gözl.-2	0.927*	0.943*	$p = 0.316$
KNN-gözl.-1	0.802*	0.944*	$p = 0.002$
KNN-gözl.-2	0.823*	0.904*	$p = 0.022$
RBFN-gözl.-1	0.782*	0.901*	$p = 0.003$
RBFN-gözl.-2	0.803*	0.874*	$p = 0.024$
NB-gözl.-1	0.880*	0.953*	$p = 0.008$
NB-gözl.-2	0.925*	0.953*	$p = 0.095$
SVM- gözl.-1	0.861*	0.952*	$p = 0.005$
SVM- gözl.-2	0.886*	0.924*	$p = 0.065$
KM-gözl.-1	0.454*	0.780*	$p < 0.001$
KM-gözl.-2	0.482*	0.726*	$p < 0.001$

Bir önceki bölümde olduğu gibi, kapa örtüşme katsayısında ikinci derece ağırlık katsayıları kullanılmıştır. Wilcoxon fark sınamasında, kullanılan sıfır hipotezi puan dizilerinin ortalama değerlerini aynı olan iki topluluktan örneklenmiş olmasıdır. Karar eşik düzeyi olan örneklenme olasılığı 0.05 olarak benimsenmiştir. Eğer p değeri 0.05'ten büyükse varsayım doğru, küçükse yanlış olarak kabul edilmiştir. '*' sembolü ($p < 0.001$) küçüklük bağıntısını gösterir.

6.3.4. Tartışma ve Sonuç

Bu çalışmada, ER durum değerlendirmesini otomatik ve nesnel bir yaklaşımla gerçekleştirmek amacıyla, bazı yapay öğrenme yöntemlerinin başarımı invaziv duktal karsinom teşhisli meme kanseri biyopsi örnekleri üzerinde incelenmiştir. Bu deneysel çalışmada, tümleşik bir sınıflayıcı modelinin hem eğitim veri kümesi üzerinde geçerleme yordamıyla sınıflamada hem de test resimleri üzerinde tanı puanlamasında en başarılı sonuçları verdiği gözlenmiştir. Bu tümleşik sınıflama modeli, fonksiyonel karar ağacı ve sade bayes sınıflayıcısının birleşimidir, bu birleşim geri kalan bölümde VOTE kısa adlandırmasıyla gösterilmiştir. Tüm yapay öğrenme yöntemlerini icra eden bilgisayar tabanlı otomatik tanı yaklaşımı daha önce yayına kabul edilen [87] ve bir önceki bölümde takdim edilen sistem ile aynıdır. Bu sisteme göre, VOTE, FT, KNN, SVM, RBFN, NB ve KM yöntemleri aynı veri kümesi üzerinde eğitildiler ve aynı test resimler üzerinde (veri kümesinden bağımsız) test edildiler.

Bu yöntemlerin seçiminde daha önce literatürde yayınlanan çalışmalar temel alınmıştır. Bu çalışmalar şunlardır; RBFN [36], KNN ağırlık katsayıları ile birlikte [35], KM öbekleme yöntemi [37], olasılık tabanlı ağlar ve SVM [34] ve FT [87]. Böylece, çeşitli yapay öğrenme yöntemlerinin gerçek bir uygulamada ve bağımsız test verisi üzerinde genelleme kabiliyetlerini kıyaslamak mümkün olmuştur. Tanı amaçlı puanlamaya ilaveten, doğru sınıflama yüzdeleri, gözlemci doktorlarla örtüşme ve ilinti özellikleri ortaya konmuştur. Bu çalışmada yapay öğrenme yöntemleri WEKA kütüphaneleri Matlab içerisinden çağrılarak icra edilmiştir. Burada uygulanan sınıflayıcılar standart yöntemlerdir. Ancak, taban alınan iki çalışmada burada uygulanan standart yöntemlerden farklı olarak, RBFN'ye dinamik nöronlar eklenmiştir [36] ve KNN'de farklı bir ağırlıklandırma fonksiyonu kullanılmıştır [35].

Sınıflayıcı ve öznitelik türlerinin sınıflama başarımlarına bakıldığında, geçmişte yapılan diğer çalışmalarla tutarlı sonuçlar tespit edilmiştir. ER durum değerlendirmesinde Laws desen enerji özniteliklerini kullanan iki çalışma yayınlanmıştır [34, 35]. Bu çalışmaların sonuçlarına bakıldığında KNN'in doğru sınıflama yüzdeliğinin %82.73 ve %86.71 arasında gezindiği ve Laws desen özniteliklerin farklı bileşimlerine göre bu aralıkta farklı sonuçlar gözlemlendiği belirtilmiştir [35]. Bizim çalışmamızda KNN sınıflayıcısının doğru sınıflama yüzdeliği Laws desen öznitelikleri üzerinde %80 olarak bulunmuştur (Tablo 6.2).

Ayrıca, bir başka makalede Laws ve Gabor desen öznitelikleri üzerinde olasılık tabanlı sinir ağlarının ve SVM'nin daha iyi sınıflama sonuçları gösterdiği yayınlamıştır [34]. Bu çalışmada belirtilen en yüksek sınıflama yüzdesi %94.23 olarak SVM kullanımı ile elde edilmiştir. Bizim çalışmamızda renkli Laws desen öznitelikleri üzerinde SVM ile elde edilen doğru sınıflama yüzdesi % 90.45 olmuştur. Ancak, yaptığımız çalışmada SVM ile en yüksek doğru sınıflama yüzdesi renkli dalgacık öznitelikleri üzerinde %97.39 olarak kayıt edilmiştir. Atıf yapılan çalışmalardan ve bizim bu çalışmamızdan çıkan ortak sonuç şudur: ER durum değerlendirmesinde SVM, KNN'den daha başarılı sınıflama yapmaktadır.

Hücre çekirdeği bölümlenmesine ve sayımına dayalı olarak daha önce yapılmış olan ER durum değerlendirmesi çalışmalarında dalgacık öznitelikleri kullanılmamıştır. Bu tez çalışmasında, ayrıca dalgacık özniteliklerinin başarımları araştırılmıştır. Dalgacık özniteliklerinin özellikle literatürde atıf yapılan ikinci dereceden diğer özniteliklerden istatistiksel anlamda belirgin olarak daha başarılı sonuçlar verdiği gözlemlenmiştir (Tablo 6.3). Özellikle, öznitelik grupları içerisinde kullanılan öznitelik sayısına bakıldığında dalgacık desen özniteliklerinin Laws desen özniteliklerine kıyasla ne kadar daha iyi sonuçlar verdiği açıkça görülebilir. Dalgacık desen özniteliklerinin sınıflama kabiliyeti bakımından geçmişte yayınlanan diğer özniteliklere karşı gösterdiği üstünlük tüm yapay öğrenme yöntemleri üzerinde gözlemlenmiştir (Tablo 6.3).

ER durum değerlendirmesinde incelemeye alınan sınıflayıcılar içerisinde danışmanlı öğrenmeye başvurmeyen tek yöntem KM yöntemidir [37]. Bu çalışma, bir örnek üzerinde doğru sınıflama yüzdesini olarak çekirdeğin türüne göre %81.5 ve %87.5

olarak bildirmiştir. Bizim çalışmamızda, KM yönteminin doğru sınıflama yüzdesi azaltılmış öznitelik kümesi üzerinde %87.53 olarak ölçülmüştür (Tablo 6.2). Bu tablonun sonuçlarına bakıldığında, KM sınıflayıcısı en kötü sınıflama başarımını, içerisinde en fazla öznitelik içeren Laws desen öznitelikleri kümesinde göstermiştir.

Doğru sınıflama yüzdelikleri (Tablo 6.2 ve Tablo 6.3) genel olarak incelendiğinde, gri ton özniteliklerinin sınıflama başarımının renkli özniteliklere göre daha az olduğu açıktır. KM, RBFN ve NB, sınıflayıcı girişine verilen öznitelik sayısı arttıkça sınıflama başarımını düşüren bir eğilim gösterir. NB'nin bu eğilimine bağlı olarak, VOTE öznitelik sayısı artan öznitelik kümelerinde düşen bir başarı yüzdesi gösterir. Örneğin, yine Laws desen özniteliklerinde VOTE düşen bir doğru sınıflama yüzdeliği ortaya koymuştur. Öte yandan, FT, SVM ve KNN artan öznitelik sayısından olumsuz etkilenmemiştir. Tam aksine, bu üç sınıflayıcı büyük veri kümelerinde de iyi bir sınıflama başarımı gösterirler. Danışmansız tek sınıflayıcı olan KM danışmanlı diğer yöntemlere kıyasla belirgin biçimde geride kalan sonuçlar vermiştir.

NB'nin azaltılmış veri kümesi üzerinde en yüksek doğru sınıflama yüzdesini, %97.58, verdiği dikkat çekmektedir (Tablo 6.2). Her ne kadar tüm danışmanlı öğrenen sınıflayıcılar sınıflamada yakın sonuçlar verse de, test resimleri üzerinde yapılan tanı puanlamasında yöntemlerin genelleme kabiliyetlerine bağlı olarak farklılaşan başarımlar elde edilmiştir.

Duyarlılık ve özgüllük değerlerine gelince (Tablo 6.4), KM dışındaki tüm sınıflayıcıların birbirine yakın ve yarışmacı düzeyde başarılı sonuçlar verdiği görülmüştür. KM öbekleme yöntemi diğerlerine göre düşük duyarlılık davranışı sergilemiştir. Düşük duyarlılık şu anlama gelir; KM hücre çekirdeğini tanımlamaya çalışırken birçok yanlış negatif, *YN*, üretir. Boyanmamış hücre çekirdekleri üzerinde *YN* verilmesi, bu hücre çekirdeklerinin orta şiddetli boyanma veya kuvvetli boyanma ile tanımlanmasına sebep olur. Sonuç olarak olması gerekenden daha yüksek Allred puanları ortaya çıkacaktır. Bu durum aynı zamanda genel olarak ER durum değerlendirmesinde ER negatif bir örneğin ER pozitif olarak yanlış tanımlanmasını da meydana getirebilir. Diğer taraftan, *YN* hatasının orta ya da kuvvetli boyanmış hücre

çekirdekleri üzerinde olması halinde ER pozitif tanısı değişmeyecektir ancak pozitiflik derecesi başka bir değişle tanı puanı değişebilir.

Alıcı işletim karakteristik eğrileri incelendiğinde (Şekil 6.1 ve Şekil 6.2), tüm danışmanlı öğrenme yöntemlerinin yanlış pozitif oranındaki değişime karşı bağımsız olduğu gözlenmiştir. Duyarlılık sonuçlarına benzer biçimde, KM boyanmayan hücre çekirdeklerinde en kötü sınıflama eğrisini ortaya çıkarmıştır. Tümlşik sınıflayıcı VOTE'un alıcı işletim karakteristik eğrilerinde FT'ye nazaran küçük ama gözle görülebilir bir iyileşme varlığı dikkat çekmektedir. Eğrilerdeki bu iyileşme, VOTE sınıflayıcısının doğru sınıflama yüzdesinde gözlenen ve FT'ye kıyasla elde edilen başarının tesadüfî değil kalıcı olduğunu ve genellemeyi olumlu etkilediğini göstermektedir.

Gözlemcilerin ve bilgisayar tabanlı sistemin tanı sonuçları incelendiğinde (Tablo 6.5), RBFN ve KM hariç olmak üzere diğer tüm öğrenme yöntemleriyle gözlemci patolog hekimlerin ER durum değerlendirmeleri arasında çok iyi düzeyde örtüşme bulunmaktadır. RBFN ile gözlemci-1 arasında örtüşme iyi düzeydedir ancak bu örtüşme değeri diğer danışmanlı sınıflayıcılara karşı geridedir. KM öbeklemesine dayalı sınıflayıcı vasat örtüşme değeri ortaya koymuştur. Ancak bu düzeyde bir örtüşme tıbbi uygulamalar için uygulamada bir kullanım değeri taşımaz. Burada patolog hekimlerle bilgisayar tabanlı tanı sistemi arasında en iyi örtüşme değeri tümlşik sınıflayıcı, VOTE, ile elde edilmiştir ($\kappa=0.899$ ve $\kappa=0.949$, $p < 0.001$).

İlinti sonuçlarında (Tablo 6.5), tüm yapay öğrenmeye dayalı yaklaşımlar gözlemci hekimlerin tanı puanlamalarıyla güçlü düzeyde bir ilinti sergilemiştir. Ne var ki, RBFN'nin gözlemci-2 ile ilinti diğerlerinin gerisinde kalmaktadır ($\rho=0.874$, $p < 0.001$). RBFN için doğru sınıflama yüzdeliğinde elde edilen başarı test resimlerinin puanlanmasında gözlenmemiştir. Diğer taraftan, gözlemci-1 ile en yüksek ilinti katsayısını FT ($\rho=0.963$, $p < 0.001$) ve gözlemci-2 ile en yüksek ilinti katsayısını ise VOTE ($\rho=0.972$, $p < 0.001$) göstermiştir.

İlinti katsayısının yanı sıra, gözlemci doktorlar ve bilgisayar tabanlı sistemlerin puanlama dizilerinin farksızlığını incelemek için Wilcoxon eşleştirmeli işaretli ve sıralı

(iki yönlü) fark testi uygulanmıştır (Tablo 6.5). Bu sına sonuçlarına göre, VOTE ve FT gözlemci-1 puanlamalarından istatistiksel anlamda farklı olmayan puanlamalar yapmıştır (her ikisi için de $p=0.051$). Gözlemci-2'ye gelince, şu yöntemler VOTE ($p=0.422$), FT ($p=0.316$), NB ($p=0.095$) ve SVM ($p=0.065$) gözlemci-2 ile istatistiksel anlamda farklı olmayan puanlama sonuçları vermektedir. Tüm sınıflayıcılar içerisinde yalnızca VOTE ve FT her iki gözlemciden de istatistiksel anlamda farklı olmayan puanlama sonuçları göstermiştir.

Test resimleri üzerinde Allred puanlarının elde edilmesinden sonra, puanlamaların yanı sıra ER pozitif veya ER negatif olarak pozitiflik derecesine bakmayan genel bir tanı değerlendirmesi elde etmek ve buna göre sonuçları karşılaştırmak da mümkündür. Bu yaklaşımla sonuçlara bakıldığında, en iyi tanı başarımı 27 test resmi içerisinde 26 tanesini doğru tanı ile ayırt eden (26/27) tümleşik sınıflayıcı, VOTE, tarafından elde edilmiştir. Böylelikle, ER negatif veya ER pozitif temelinde doğru tanı yüzdesi % 96.3 değerinde gerçekleşmiştir. Bu netice ile bir önceki bölümde FT'ye dayalı çalışmamızda elde edilen doğru tanı yüzdeliği, % 92.59, daha da geliştirilmiştir [87]. Diğer yöntemlerin doğru tanı oranları, FT (25/27), NB (24/27), SVM (23/27), KNN (22/27), RBFN (21/27) ve KM (19/27) olarak elde edilmiştir. Doğru tanı oranı, tanı puanlamasının bir sonucudur ve bu bakımdan yapay öğrenme yöntemleri arasında ortaya çıkan sıralama Tablo 6.5'da verilen örtüşme, ilinti ve farksızlık sınınamalarıyla tutarlıdır.

Bağımsız test resimleri üzerinde gerçekleştirilen deneysel çalışmaları sonuçlarına (Tablo 6.2 ve Tablo 6.3) dayalı olarak, şu genel çıkarımları yapılabilir. Yöntemlerin başarımları, özneteliklerin türüne bağlı olarak değişmektedir. Renkli öznetelikler, gri ton özneteliklere kıyasla daha iyi sınıflama yapabilmektedir. Öznetelik kümesinde kullanılan öznetelik sayısı arttıkça, KM, NB ve RBFN öğrenme yöntemlerinin sınıflama başarımı düşüş göstermiştir. Öte yandan, SVM ve FT öğrenme yöntemleri sınıflayıcı girişinde artan öznetelik sayısından olumsuz olarak etkilenmemiştir. Özetle, FT ve VOTE diğer öğrenme yöntemlerinden daha iyi genelleme yeteneği göstermiştir. VOTE, sade bayes ve karar ağacı sınıflayıcılarının birleşimi, invaziv duktal meme kanseri örnekleri üzerinde ER tanısında en iyi Allred puanlama değerlerini ve en iyi tanı sonuçlarını ortaya koymuştur.

7. BÖLÜM

TARTIŞMA - SONUÇ VE ÖNERİLER

7.1. Tartışma Sonuçlar ve Öneriler

Bu çalışmada, süt kanalı meme kanseri biyopsi örnekleri üzerinde otomatik östrojen alıcı tanısının elde edilmesi amaçlandı. Belirlenen amaç doğrultusunda ERÜ Tıp Fakültesi Patoloji Bölümü arşivinden alınan östrojen alıcısı boyalı biyopsi örnekleri ışın mikroskobu altında incelendi, tümörü temsil eden bölge uzman patolog tarafından seçilerek sayısal resim olarak kaydedildi. Bu sayısal resimler üzerinde sırasıyla bölümlendirme, hücre tespiti ve sayımı, öznelik üretme, hücre tanıma ve biyopsi örneğine tanı konulması adımlarını uygulayan bir otomatik tanı sistemi oluşturuldu.

Bu sistemin tasarımında aşağıda verilen üç ayrı savın geçerliliği ve uygulanabilirliği hususunda sonuçlar elde edilmiştir.

Sav-1: Allred puanlama yordamına uygun otomatik bir östrojen alıcı tanı sistemi bir karar ağacı modeli ile gerçekleştirilebilir.

Sav-2: Dalgacık dönüşümü ile elde edilen öznelikler, hem uzay hem de frekans boyutunda bilgi içerdiği için hücre çekirdeği tanıma başarımını arttırmaktadır.

Sav-3: Sınıflayıcıların birleşimi, otomatik östrojen alıcı tanı sisteminin genelleme kabiliyetini - tanı başarımını- yükseltmektedir.

Birinci savın geçerliliği Bölüm-5'te incelenmiştir. Elde edilen sonuçlar ve bu sonuçlara ilişkin tartışma ayrıntılı olarak Bölüm-5'in tartışma başlığı altında ele alınmıştır. Geleneksel puanlama yordamlarına göre daha karmaşık olan Allred puanlama

yordamına dayalı olarak otomatik östrojen alıcı tanısının verilebileceğini göstermektedir. Fonksiyonel karar ağaçları ile elde edilen otomatik östrojen alıcı tanı modeli, klasik C4.5 karar ağacına göre daha başarılı ve basit bir ağaç yapısı sergiler. Ayrıca, bu yapı literatürde sıklıkla başvurulmuş destek vektör makinelerine göre çok daha sadedir. Önerilen bu modelin östrojen alıcı tanısı Allred puanları, gözlemci doktor puanları ile yüksek ilinti düzeyi sağlar ($\rho=0.963$ ve $\rho=0.943$, $P<0.001$) ve kappa örtüşme katsayısına göre gözlemci doktorlarla tanı puanlaması bakımından örtüşme düzeyi çok yüksektir ($\kappa=0.899$ ve $\kappa=0.927$, $P<0.001$). Wilcoxon farklılık sınaması, gözlemci doktorlar ve fonksiyonel karar ağacına dayalı otomatik tanı sistemi arasında istatistiksel anlamda belirgin bir fark olmadığını sonucu çıkarmıştır ($P=0.051$ ve $P=0.316$).

Bölüm-5'te ve Bölüm-6'da uygulanan tanı sistemini, literatürde yayınlanan diğer çalışmalardan ayıran iki önemli özellik bulunmaktadır. Birincisi, bu tez çalışmasında kullanılan tanı sistemi görüntü üzerinde hücre çekirdeği tespiti ve sayımı yapmaktadır. Literatürde yayınlanan birçok çalışma görüntü üzerinde piksel yüzdeliğine veya alanına bağlı analiz ya da resim üzerinde küresel eşikleme esasına dayalı yöntemler kullanılmaktadır [38-50]. Oysa gerçek tıbbî uygulamada çekirdek sayımı esastır. İkincisi, östrojen alıcı varlığının değerlendirilmesinin Allred puanlama protokolüne göre yapılmış olmasıdır. Bu yöntem alternatiflerine göre yenidir ve geniş çaplı klinik çalışmalara dayanarak kliniklerde bu yöntemin kullanımı özellikle tavsiye edilmektedir [6]. Hücre çekirdeği tespiti ve sayımına dayalı olarak östrojen alıcı varlığının değerlendirmesini yapan diğer çalışmalarda [34-37, 82] Allred puanlama protokolü kullanılmamıştır.

İkinci savın geçerliliği hem Bölüm-5 ve hem de Bölüm-6'da yapılan deneysel karşılaştırmalarla incelenmiştir. Buna göre, dalgalık dönüşümü enerji öznitelikleri, seçilen birinci dereceden istatistiksel öznitelikler, eş-oluşum matrisi öznitelikleri ve Laws desen enerji özniteliklerine kıyasla en başarılı ayırt etme kabiliyetine sahiptir (Tablo-5.4 ve Tablo-6.3). Dalgalık dönüşümünden elde edilen öznitelikler östrojen alıcı tanısının elde edilmesi amacıyla ilk defa bu tez çalışmasında kullanıldı. Bölüm-5'in ve özellikle Bölüm-6'nın tartışma kısmında, bu öznitelik kümesinin sağladığı sınıflama başarımı daha önce yayınlanan çalışmalarda önerilen öznitelik kümeleri ile karşılaştırılmıştır (Tablo-6.3). Dalgalık özniteliklerinin diğer öznitelik türlerine göre tüm sınıflayıcılarda

sınıflama başarımını arttırdığı gözlenmiştir. Öte yandan, dalgacık öznitelikleri dışında kalan diğer özniteliklerin sınıflayıcı türüne göre başarımlar oranı geçmişte yayınlanan çalışmalarla uyumlu sonuçlar göstermiştir.

Üçüncü savın doğrulanması için yapılan deneysel inceleme Bölüm-6'da açıklanmıştır. Çeşitli sınıflayıcıların çeşitli öznitelik kümeleri üzerinde başarımları bu bölümde karşılaştırmalı olarak incelendi. Bu incelemelerde sınıflama başarımı yüksek olan bazı sınıflayıcıların puanlama başarımının test kümesinde düşmekte olduğu gözlenmiştir. Örneğin yarıçap fonksiyonlu yapay sinir ağları yüksek doğru sınıflama yüzdesine karşılık nispeten düşük doğru puanlama başarısı göstermekte ve böylece aşırı öğrenme davranışı sergiledi. Diğer taraftan fonksiyonel karar ağaçları, k -en yakın komşuluk algoritması, destek vektör makineleri ve sade bayes sınıflayıcısı hem sınıflama hem de puanlama başarımı bakımından istikrarlı sonuçlar vermiş ve iyi bir genelleme kabiliyeti ortaya koymuştur. Bölüm-6'da açıklanan deneysel sonuçlar tanı bakımından en iyi sonucu (doğru tanı yüzdesi % 96.3) birleşik sınıflayıcı modelin verdiğini ortaya koydu. Bu birleşik sınıflayıcı modeli oylama esasına dayalı olan ve sade bayes sınıflayıcısı ile fonksiyonel karar ağaçlarının birleşiminden oluşan bir modeldir. Önerilen birleşik sınıflayıcı modelin östrojen alıcı tanısı Allred puanları ve gözlemci doktor puanları arasında sağlandığı ilinti düzeyi ($\rho=0.961$ ve $\rho=0.972$, $P<0.001$) ve kappa örtüşme katsayısı bakımından örtüşme düzeyi ($\kappa=0.899$ ve $\kappa=0.949$, $P<0.001$) değerlerindedir (Tablo-6.5). Wilcoxon farksızlık sınaması, gözlemci doktorlar ve birleşik sınıflayıcı modeline dayalı otomatik tanı sistemi arasında istatistiksel anlamda belirgin bir fark olmadığını ($P=0.051$ ve $P=0.422$) göstermiştir.

Yukarıda verilen savlar ve bu savların geçerliliğini sınamak için yapılan deneysel karşılaştırmalar, 10 defa tekrarlanan 10-parçalı çapraz geçerleme yordamının verdiği istatistiksel ortalamalardır. Başka bir deyişle, her bir modelin sınıflama başarımı aynı modelin farklı eğitim ve test kümelerinde 100 defa oluşturulmasından elde edilen sonuçların ortalamasıdır. Modellerin diğerlerinden daha iyi ya da daha kötü olması, 100 adet elemandan oluşan bu başarımlar örneklemelerinin eşleştirmeli t-testinde istatistiksel bakımdan anlamlı fark taşıyıp taşımadıklarına göre saptanmıştır. Buna göre, tek sınıflayıcılar içerisinde tanı kararlarında, birinci sırada fonksiyonel karar ağaçlarının (doğru tanı yüzdesi % 92.59) ve ikinci sırada sade bayes sınıflayıcısının genelleme

kabiliyeti (doğru tanı yüzdesi %88.89) yüksektir. Öte yandan, destek vektör makinelerinin ve k -en yakın komşuluk algoritmasının artan giriş veri miktarına rağmen sınıflama başarımını düşürmedikleri ve dinç bir sınıflama başarımı gösterdikleri tespit edilmiştir. Çalışmada incelenen tek danışmansız öğrenme yöntemi olan k -orta nokta öbekleme yöntemi ise artan öznitelik sayısından belirgin biçimde olumsuz etkilenmektedir. Bu yöntem danışmalı rakiplerine göre hem sınıflama hem de tanı başarımında istatistiksel bakımdan belirgin bir farkla geride kalmıştır.

Önerilen otomatik tanı sistemi, kullanılan öznitelik veri kümesi ve sınıflayıcı modeli bakımından en uygun olanı seçme arayışının bir ürünüdür. Ancak bu ürünü geliştirmek için gelecekte şu çalışmaların yapılması öngörülmektedir. Sistemin ham verisi olan resimler üzerinde beyaz dengelemesi, aydınlık dengelemesi gibi ön işlemlerin genel başarıma etkisi incelenebilir. Öte yandan, bölütleme yönteminin bir başarımlı ölçütüne dayandırılması ve bu ölçüte göre en uygun bölütleme yönteminin belirlenmesi araştırılabilir. Benzer biçimde belirlenen bölütleme ölçütüne göre biçimbilim işlemcilerinin katkısı matematiksel kesinlikte araştırılabilir. Bu tespitler aynı zamanda gelecekte yapılabilecek devam niteliğinde çalışmaların basamaklarını oluşturmaktadır. Ayrıca bölütleme, hücre (nesne) tespiti, öznitelik üretme, hücre sınıflama ve doku değerlendirme işlem adımları sadece tanı sisteminde değil aynı zamanda kanser teşhis sistemlerinde de kullanılabilecek genel işlemlerdir. Bu yönüyle, önerilen genel tanı sisteminin uygulamaya göre özelleştirilmesi ve kanser teşhisi amacıyla kullanılması da mümkündür.

Önerilen tanı sisteminin dört önemli uygulama kısıtı bulunmaktadır. Biricisi, bu tanı sistemi sitoplazmik boyamada kullanılamaz. İkincisi, danışmanlı öğrenme nedeniyle kullanılan sınıflayıcının en az bir kez eğitilmesi gerekir. Üçüncüsü, sayısal resimler üzerinde bulanıklık, odaklanma hataları vb. resmin bir kısmını diğerlerinden farklı kılan yapay hatalar bulanmamalıdır. Dördüncüsü, elde edilen sayısal resimler aynı kimyasal süreçlere, optik okuma ve kayıt cihazlarına bağlı olarak doktor tarafından elde edilmelidir.

Özetle; önerilen tanı sistemi, hücre tanıma ve sayma yapabilir ve bu yetenekle Allred puanlama yordamına göre östrojen alıcı tanısı verebilmektedir. Üretilen tanı kararları,

öznellik içermez, matematiksel kesinlikte nesneldir. Sınıflayıcı modeli değiştirilmedikçe, elde edilen tanı karaları gerektiğinde sayısal resim üzerinden yeniden üretilebilir. Dalgacık dönüşümü öznitelikleri geçmişte literatürde tanıtılan ve östrojen alıcı tanısı amacıyla kullanılan diğer öznitelik türlerine kıyasla sınıflama başarımını arttırmıştır. Fonksiyonel karar ağaçları ve sade bayes sınıflayıcının birleşiminden oluşan tümleşik sınıflayıcı yaklaşımı, otomatik östrojen alıcı tanısının başarımını yükseltmektedir.

KAYNAKLAR

1. Department of Biology, Davidson College, 2002. Pathology/Histology Slides. (Web page: <http://www.bio.davidson.edu/courses/genomics/method/Histology.html>), (Date accessed: February 2012).
2. Donegan, W. L. ,1997. Tumor-related prognostic factors for breast cancer. **CA Cancer J. Clin.**, **47**: 28–51.
3. Sommer, S., & Fuqua, S. A., 2001. Estrogen receptor and breast cancer. **Semin Cancer Biol.**, **11**: 339–352.
4. Layfield, L. J., Gupta, D., & Mooney, E. E., 2000. Assessment of tissue estrogen and progesterone receptor levels: a survey of current practice, techniques, and quantitation methods. **The Breast Journal**, **6**: 189–196.
5. Walker, R. A., 2006. Quantification of immunohistochemistry issues concerning methods, utility and semiquantitative assessment. **I. Histopathology**, **49**: 406–410.
6. Umemura, S., Itoh, J., Itoh, H., Serizawa, A., Saito, Y., Suzuki, Y., et al., 2004. Immunohistochemical evaluation of hormone receptors in breast cancer: which scoring system is suitable for highly sensitive procedures. **Appl. Immunohistochem. Mol. Morphol.**, **12**: 8-13.
7. Harvey, J., Clark, G., Osborne, C., & Allred, D. C.,1999. Estrogen receptor status by im-munohistochemistry is superior to the ligand-binding assay for predicting response to ad-juvant endocrine therapy in breast cancer. **J. Clin. Oncol.**, **17**: 1474–1481.
8. Walker, R. A., 2008. Immunohistochemical markers as predictive tools for breast cancer. **J Clin. Pathol.**, **61**, 689-696.
9. Qureshi, A., & Pervez, S., 2010. Allred scoring for ER reporting and its impact in clearly distinguishing ER negative from ER positive breast cancers. **JPMA**, **60**: 350-353.
10. Dhawan, A. P., D'Alessandro, B., & Fu, X., 2010. Optical Imaging Modalities for Biomedical Applications. **IEEE Reviews in Biomedical Engineering**, **3**: 69-92.
11. Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B., 2009. Histopathological Image Analysis: A Review. **IEEE Reviews in Biomedical Engineering**, **2**: 147-171.

12. Rudiger, T., Hofler, H., Kreipe, H., Nizze, H., Pfeifer, U., Stein, H., et al., 2002. Quality assurance in immunohistochemistry: results of an interlaboratory trial involving 172 pathologists. **Am J Surg Pathol**, **26**: 873-882.
13. Demir, C., & Yener, B. Automated cancer diagnosis based on histopathological images: a systematic survey. Technical Report, Rensselaer Polytecnic Institute, Department of Computer Science, TR-05-09.
14. Symptoms & Diagnosis., 2012, (Web page: <http://www.breastcancer.org/symptoms/>), (Date accessed: June 2012).
15. Afzan, A., & Khairuddin, O., 2012. Computerized breast cancer diagnosis with genetic algorithms and neural network. (Web page: <http://fit.mmu.edu.my/caic/papers/afzanicaiet.pdf>), (Date accessed: July 2012).
16. Koçer, H., 2012. Kanser tanısı.(Web sayfası: <http://www.memekanseri.com.tr/kansernedir/kansertanisi.html>), (Erişim tarihi: Mayıs 2012).
17. Mayo Clinic., 2012. Breast cancer types: What your type means. (Web page: <http://www.mayoclinic.com/health/breast-cancer/HQ00348>), (Date accessed: April 2012).
18. Zingman, I., Meir, R., & El-Yaniv, R. , 2005. Breast cancer image data. (Web page: <ftp://ftp.cs.technion.ac.il/pub/projects/medic-image/>) (Date accessed: May 2012).
19. Susan G. Komen for the Cure, 2012. Diagnosis. (Web page: <http://ww5.komen.org/BreastCancer/Diagnosis.html>), (Date accessed: April 2012).
20. Wikipedia, the free encyclopedia., 2012. Breast cancer classification. (Web page: http://en.wikipedia.org/wiki/Breast_cancer), (Date accessed: April 2012)
21. Ellis, I., Schnitt, S., & Sastre-Garau, X., 2003. Invasive breast carcinoma, pp. 13-59. *In*, Pathology and Genetics of Tumours of the Breast and Female Genital Organs. (Eds. F. Tavassoli, & P. Devilee). IARC Press.
22. Elston, C., & Ellis, I., 1991. Pathological prognostic factors in breast cancer I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. **Histopathology**, **19**: 403–410.
23. *Stages of Breast Cancer*. 2012, (Web page: <http://www.breastcancer.org/symptoms/diagnosis/staging.jsp>). (Date accessed: April 2012)

24. Vikipedi, özgür ansiklopedi., 2012. Meme kanseri. (Web sayfası: http://tr.wikipedia.org/wiki/Meme_kanseri). (Erişim tarihi: Mayıs 2012).
25. Kanserleyasamak, 2012. Genel Bilgiler. (Web sayfası: http://www.kanserleyasamak.org/genel_bilgiler.php) (Erişim tarihi: Mayıs 2012)
26. HER2 Status. 2012. (Web page: <http://www.breastcancer.org/symptoms/diagnosis/her2.jsp>). (Date accessed: July 2012)
27. Hormone Receptor Status. 2012. (Web page: http://www.breastcancer.org/symptoms/diagnosis/hormone_status/) (Date accessed: March 2012)
28. Wu, H. S., Barbara, J., & Gil, J., 1998a. Optimal segmentation of cell images. **IEEE Proc., Vision, Image, Signal Process.** (145): 50–56.
29. Wu, H. S., Barbara, J., & Gil, J., 1998b. Parametric fitting algorithm for segmentation of cell images. **IEEE Trans. Biomed. Eng.**: 400–407.
30. Wu, H. S., Barbara, J., & Gil, J., 1996. Region growing segmentation of textured cell images. **Electron. Lett.** (32): 1084–1085.
31. Wu, H. S., Barbara, J., & Gil, J., 2000. Iterative thresholding for segmentation of cell images. **J. Microsc.** (197): 296–304.
32. Krecsák, L., Micsik, T., Kiszler, G., Krenács, T., Szabó, D., Jónás, V., et al., 2011. Technical note on the validation of a semi-automated image analysis software application for estrogen and progesterone receptor detection in breast cancer. **Diagnostic Pathology**, 6: 6.
33. Sharangpani, G. M., Joshi, A. S., Porter, K., Deshpande, A. S., Keyhani, S., Naik, G. A., et al., 2007. Semi-automated imaging system to quantitate estrogen and progesterone receptor immunoreactivity in human breast cancer. **Journal of Microscopy** (226): 244–255.
34. Kostopoulos, S., Cavouras, D., Daskalakis, A., Kalatzis, I., Bougioukos, P., Kagadis, G., et al., 2007. Assessing estrogen receptors' status by texture analysis of breast tissue specimens and pattern recognition methods. **CAIP 2007, LNCS 4673**: 221–228.
35. Kostopoulos, S., Cavouras, D., Daskalakis, A., Bougioukos, P., Georgiadis, P., Kagadis, G., et al., 2007. Colour-texture based image analysis method for assessing the hormone receptors status in breast tissue sections. pp. 4985-88. *Conf Proc IEEE Eng Med Biol Soc.*

36. Schnorrenberg, F., Tsapatsoulis, N., Pattichis, C., Schizas, C., Kollias, S., Vassiliou, M., et al., 2000. Improved detection of breast cancer nuclei using modular neural networks. **IEEE Eng. Med. Biol. Mag.**, **19**: 48–63.
37. Kostopoulos, S., Cavouras, D., Daskalakis, A., Ravazoula, P., & Nikiforidis, G., 2006. Image analysis system for assessing the estrogen receptor's positive status in breast tissue carcinomas . *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine*.
38. Diaz, L., & Sneige, N., 2005. Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy. **Advances in Anatomic Pathology**, **12**: 10-19.
39. Charpin, C., Andrac, L., Habib, M., Vacheret, H., Xerri, L., Devictor, B., et al., 1989. Immunodetection in fine-needle aspirates and multiparametric (SAMBAs) image analysis. Receptors (monoclonal antiestrogen and antiprogesterone) and growth fraction (monoclonal Ki67) evaluation in breast carcinomas. **Cancer**, **63**: 863-872.
40. Turbin, D., Leung, S., Cheang, M., Kennecke, H. A., Montgomery, K., McKinney, S., et al., 2008. Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. **Breast Cancer Res. Treat**, **110**: 417–426.
41. Lehr, H., Mankoff, D., Corwin, D., Santeusano, G., & Gown, A., 1997. Application of photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. **J. Histochem. Cytochem**, **45**: 1559-1566.
42. Mofidi, R., Walsh, R., Ridgway, P., Crotty, T., McDermott, E., Keaveny, T., et al., 2003. Objective measurement of breast cancer oestrogen receptor status through digital image analysis. **Eur. J. Surg. Oncol**, **29**: 20–24.
43. Hatanaka, Y., Hashizume, K., Nitta, K., Kato, T., Itoh, I., & Tani, Y., 2003. Cytometrical image analysis for immunohistochemical hormone receptor status in breast carcinomas. **Pathology International**, **53**: 693–699.
44. McClelland, R. A., Finlay, P., Walker, K. J., Nicholson, D., Robertson, J. F., Blamey, R. W., et al., 1990. Automated quantitation of immunocytochemically localized estrogen receptors in human breast cancer. **Cancer Res**, **50** (12): 3545-50.
45. Furukawa, Y., Kimijima, I., & Abe, R., 1998. Immunohistochemical image analysis of estrogen and progesterone receptors in breast cancer. **Breast Cancer**, **5** (4): 375-380.

46. Chung, G., Zerkowski, M., Ghosh, S., Camp, R., & Rimm, D., 2007. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. **Laboratory Investigation**, **87**: 662–669.
47. Gokhale, S., Rosen, D., Sneige, N., Diaz, L., Resetkova, E., Sahin, A., et al., 2007. Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma. **Applied Immunohistochemistry & Molecular Morphology**, **15** (4): 451–455.
48. Camp, R., Chung, G., & Rimm, D., 2002. Automated subcellular localization and quantification of protein expression in tissue microarrays. **Nature Medicine**, **8** (11): 1323–1328.
49. Lloyd, M. C., Nandyala, P. A., Purohit, C. N., Burke, N., Coppola, D., & Bui, M. M., 2010. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer, how reliable is it. **J. Pathol. Inform.**, **1**, 29.
50. Bolton, K., Closas, M., Pfeiffer, R., Duggan, M. A., Howat, W., Hewitt, S., et al., 2010. Assessment of automated image analysis of breast cancer tissue microarrays for epidemiologic studies. **Cancer Epidemiol. Biomarkers Prev.**, **19** (4): 992–999.
51. Gonzales, R. C., & Woods, R. E., 2002. Digital Image Processing 2nd Ed. New Jersey: Prentice-Hall Inc., 793 pp.
52. Otsu, N., 1979. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, **9** (1): 62–66.
53. Otsu's method, 2012.(Web page: http://en.wikipedia.org/wiki/Otsu's_method), (Date accessed: May 2012).
54. Sonka, M., Hlavac, V., & Boyle, R., 2008. Image Processing, Analysis and Machine Vision . Toronto: Thomson Learning. 829 pp.
55. MATLAB., 2010. version 7.10.0 (R2010a). Natick, Massachusetts: The MathWorks Inc.
56. Raimondo, F., Gavrielides, M. A., Karayannopoulou, G., Lyroudia, K., & Pitas, I., 2005. Automated evaluation of her-2/neu status in breast tissue from fluorescent in situ hybridization images. **IEEE Transactions on Image Processing**, **14** (9): 1288–99.
57. Weaver, J. R., & Au, J. L.-S., 1997. Application of Automatic Thresholding in Image Analysis Scoring of Cells in Human Solid Tumors Labeled for Proliferation Markers. **Cytometry**, **29**: 128–135.

58. Haralick, R. M., Shanmugam, K., & Dinstein, I. H., 1973. Textural features for image classification. **IEEE Trans. Sys. Man. Cyb.**, **3**: 610–621.
59. Soh, L., & Tsatsoulis, C., 1999. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. **IEEE Transactions on Geoscience and Remote Sensing**, **37** (2).
60. Clausi, D. A., 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization. **Can. J. Remote Sensing**, **28** (1): 45-62.
61. Uppuluri, A., 2008. GLCM texture feature. (Web page: <http://www.mathworks.com/matlabcentral/fileexchange/22187>),(Date accessed: May 2012).
62. Kelly, P. M., Laws' texture measures, (Web page: <http://www.ccs3.lanl.gov/~kelly/ZTRANSITION/notebook/laws.shtml>), (Date accessed: May 2012).
63. Alpaydin, E., 2004. Introduction to Machine Learning. The MIT Press, 537 pp.
64. Theodoridis, S., Koutroumbas, K., 2009. Pattern Recognition. London: Elsevier, 961 pp.
65. Hall, M. A., 1999. Correlation-based Feature Selection for Machine Learning, thesis for the degree of doctor of philosophy at The University of Waikato, 198 pp.
66. Ley, P., 1972. Quantitative Aspects of Psychological Assessment, 234 pp.
67. Wesolkowski, S. E., 1999. Global color image segmentation strategies: Euclidean distance vs. vector angle. *Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop*.
68. Haykin, S., 2009. Neural Networks and Learning Machines. New Jersey: Pearson International Edition, 934 pp.
69. MacQueen, J. B., 1967. Some Methods for classification and Analysis of Multivariate Observations. pp. 281-297. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.
70. Ian H. Witten, E. F., 2005. Data Mining Practical Machine Learning Tools and Techniques . San Francisco: Elsevier, 525 pp.
71. Richard O. Duda, P. E., 2000. Pattern Classification. New York: John Wiley & Sons Inc. 654 pp.
72. Webb, A., 2002. Statistical Pattern Recognition, 2nd Edition. John Wiley and Sons. 496 pp.

73. Thomo, A., 2012. Linear Separators. (Web page: <http://www.engr.uvic.ca/~seng474/>) (Date accessed: June 2012).
74. Naive Bayes classifier. 2012, (Web page: http://en.wikipedia.org/wiki/Naive_Bayes_classifier), (Date accessed: June 2012).
75. Rendell, M. L., 1989. Constructive induction on decision trees. **Proceedings of IJCAI, 89**: 645-650.
76. Gama, J., & Brazdil, P., 2000. Cascade generalization. **Machine Learning, 41**: 315–343.
77. Gama, J., 2004. Functional trees. **Machine Learning, 55**: 219–250.
78. Quinlan, R., 1993. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers.
79. Platt, J., 1999. Fast training of support vector machines using sequential minimal optimization. B. Schoelkopf, C. Burges, & A. Smola içinde, *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA: MIT Press.
80. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I., 2009. The WEKA data mining software: an update. **SIGKDD Explorations, 11**.
81. Altman, D., 1991. *Practical Statistics for Medical Research*. London: Chapman and Hall.
82. Schnorrenberg, F., Pattichis, C., Kyriacou, K., & Schizas, C., 1997. Computer-Aided Detection of Breast Cancer Nuclei. **IEEE Transactions on Information Technology in Biomedicine, 1** (2): 128-140.
83. Tuominen, V., Ruotoistenmäki, S., Viitanen, A., Jumppanen, M., & Isola, J., 2010. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. **Breast Cancer Research, 12** (4).
84. Rexhepaj, E., Brennan, D., Holloway, P., Kay, E., McCann, A., Landberg, G., et al., 2008. Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. **Breast Cancer Research, 10** (5): 89.
85. Diaz, L., Sahin, A., & Sneige, N., 2004. Interobserver agreement for estrogen receptor quantitation by manual and automated scoring. **Ann. Diagn. Pathol., 8**: 23–27.

86. Cregger, M., Berger, A., & Rimm, D., 2006. Immunohistochemistry and quantitative analysis of protein expression. **Archives of Pathology & Laboratory Medicine**, **130** (7): 1026-1030.
87. Sarikoc, F., Kalinli, A., Akgün, H., & Öztürk, F., 2012. An automated prognosis system for estrogen hormone status assessment in breast cancer tissue samples. **Turk. J. Elec. Eng. & Comp. Sci.**, (yayına kabul edildi).

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: Fatih SARIKOÇ

Uyruğu: Türkiye Cumhuriyeti (T.C.)

Doğum Tarihi ve Yeri: 11 Kasım 1978, Tomarza/KAYSERİ

Medeni Durumu: Evli

Tel: +90 505 468 4140

E-posta: fsarikoc@erciyes.edu.tr, fsarikoc@tcdd.gov.tr

Yazışma Adresi: TCDD Genel Müdürlüğü, Tesisler Dairesi Başkanlığı, Sinyalizasyon ve Kumanda Sistemleri Şube Müdürlüğü, Talatpaşa Bulvarı, 06330, ANKARA

EĞİTİM

Derece	Kurum	Mezuniyet Tarihi
Yüksek Lisans	ERÜ Bilgisayar Sistemleri	2004
Lisans	ERÜ Kontrol ve Bilgisayar. Mühendisliği, Kayseri	2002
Lise	TCDD Meslek Lisesi, Eskişehir	1995

İŞ DENEYİMLERİ

Yıl	Kurum	Görev
1997-2004	TCDD Yol ve Tesisler Dairelerinde	Teknisyen
2004-2005	TCDD Tesisler Dairesinde	Mühendis
2005-2006	TSK	OBİ Subayı
2006-2008	TCDD, Hızlı Tren Kontrol Sisteminde (ETCS) mantık fonksiyonlarının (interlock) incelenmesi, onaylanması ve test edilmesi (Ankara-Eskişehir)	Proje Mühendisi
2008-	TCDD, ETCS Seviye-1 Kontrol Sistemi fonksiyonlarının incelenmesi ve onaylanması (Kayseri-Mersin-Adana)	Proje Yönetim Görevlisi
2010-	TCDD, Yol Yenileme ve ETCS Seviye-1 Kontrol Sistemi Projesi (Irmak-Karabük-Zonguldak)	İhale Komisyon Üyesi ve Proje Yönetim Görevlisi

YABANCI DİL İngilizce (iyi, KPDS-89), Arapça (başlangıç)

YAYINLAR

(2008). A Parallel Ant Colony Optimization Algorithm Based on Crossover Operation. A. Kalinli, & F. Sarikoc içinde, *Advances in Metaheuristics for Hard Optimization* (s. 87-110). Springer.

Kalinli, A., Sagiroglu, S., & Sarikoc, F. (2010). Parallel Ant Colony Optimization Algorithm Based Neural Method for Determining Resonant Frequencies of Various Microstrip Antennas. *Electromagnetics*, 30 (5), 463-481.

Sarikoc, F., Kalinli, A., Akgün, H., & Öztürk, F. (2012). An automated prognosis system for estrogen hormone status assessment in breast cancer tissue samples. *Turk J Elec Eng & Comp Sci*, (yayına kabul edildi).