

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ YÖNTEMİ İLE DİJİTAL
MAMOGRAMLARDA MEME KANSERİ TANISI**

YÜKSEK LİSANS TEZİ

Kübra KARACAN

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

HAZİRAN 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ YÖNTEMİ İLE DİJİTAL
MAMOGRAMLARDA MEME KANSERİ TANISI**

YÜKSEK LİSANS TEZİ

Kübra KARACAN

(509161208)

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Burcu TUNGA

HAZİRAN 2019

İTÜ, Fen Bilimleri Enstitüsü'nün 509161208 numaralı Yüksek Lisans Öğrencisi KÜBRA KARACAN, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ YÖNTEMİ İLE DİJİTAL MAMOGRAMLARDA MEME KANSERİ TANISI” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Burcu TUNGA**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Doç. Dr. Behçet Uğur TÖREYİN**
İstanbul Teknik Üniversitesi

Doç. Dr. Devrim ÜNAY
İzmir Ekonomi Üniversitesi

Teslim Tarihi : 2 Nisan 2019
Savunma Tarihi : 13 Haziran 2019





Aileme,



ÖNSÖZ

Bu çalışmada benden bilgisini, zamanını esirgemeyen ve çalışmanın her aşamasında emek veren danışman hocam Doç.Dr.Burcu Tunga'ya çok teşekkür ederim.

Son olarak bana her zaman destek veren, moral ve motivasyonum için elinden gelen her şeyi yapan aileme ve Dr.Tevfik Uyar'a minnettarım.

Nisan 2019

Kübra KARACAN



İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET.....	xvii
SUMMARY	xix
1. GİRİŞ	1
2. YÖNTEM	5
2.1 Yüksek Boyutlu Model Gösterilimi Yöntemi (YBMG)	5
2.2 Yüksek Boyutlu Model Gösterilimi Yöntemi İle Görüntü İşleme	9
2.3 Gözetimli Makine Öğrenmesi Teknikleri	11
2.3.1 K-en yakın komşu algoritması	12
2.3.2 Destek vektör makineleri	13
2.3.3 Lojistik bağlanım	16
2.3.4 Doğrusal ayırtaç çözümlemesi	18
2.3.5 Karar ağaçları	19
2.3.6 Sınıflandırma ve bağlanım ağaçları	20
2.3.7 Rastgele orman.....	23
2.4 Görüntü İyileştirme	23
2.4.1 Kontrast sınırlı uyarlamalı histogram eşitleme	24
2.4.2 Kontrast iyileştirme yöntemi olarak ybmg	27
3. YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ İLE MEME KANSERİ	29
4. BULGULAR VE KARŞILAŞTIRMA	35
4.1 Görüntü iyileştirme işlemi yapılmamış veriler için skorlar	35
4.1.1 Normal, iyi huylu ve kanser durumları için skorlar	35
4.1.2 Kanser ve kanser olmayan durumlar için skorlar.....	37
4.1.3 Kitle olan ve olmayan durumlar için skorlar	38
4.2 Görüntü iyileştirme işlemi yapılmış veriler için skorlar	40
4.2.1 Clahe uygulanmış mamogramlar	40
4.2.1.1 Normal, iyi huylu ve kanser durumları için skorlar	40
4.2.1.2 Kanser ve kanser olmayan durumlar için skorlar.....	42
4.2.1.3 Kitle olan ve olmayan durumlar için skorlar	44
4.2.2 Ybmg uygulanmış mamogramlar.....	46
4.2.2.1 Normal, iyi huylu ve kanser durumları için skorlar	46
4.2.2.2 Kanser ve kanser olmayan durumlar için skorlar.....	48
4.2.2.3 Kitle olan ve olmayan durumlar için skorlar	49
4.2.3 Clahe ve ybmg skorlarının karşılaştırması.....	51
5. SONUÇ VE ÖNERİLER.....	53
KAYNAKLAR	55



KISALTMALAR

ATD	: Alzheimer Tipi Demans
BDT	: Bilgisayar Destekli Tanı
CAD	: Computer-aided Diagnosis
CART	: Classification and Regression Trees
CLAHE	: Contrast Limited Adaptive Histogram Equalization
DDSM	: The Digital Database for Screening Mammography
IRMA	: Image Retrieval in Medical Applications
k-NN	: k-Nearst Neighbour
KKT	: Kalıntı Kareler Toplamı
LDA	: Linear Discreminant Analysis
LLNL	: The Lawrence Livermore National Laboratory
LogReg	: Logistic Regression
MATLAB	: Matrix Laboratory
MIAS	: The Mammographic Image Analysis Society Digital Mammogram Database
RF	: Random Forest
RWTH	: Rheinisch-Westfälische Technische Hochschule
SVM	: Support Vector Machine
UCI	: Univercity of California-Irvine
YBMG	: Yüksek Boyutlu Model Gösterilimi
WEKA	: Waikato Environment for Knowledge Analysis



ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1 : k-En yakın komşu algoritması.....	13
Çizelge 2.2 : CART algoritması.	21
Çizelge 2.3 : Rastgele orman algoritması.	23
Çizelge 2.4 : CLAHE algoritması.....	26
Çizelge 2.5 : YBMG kontrast iyileştirme algoritması.	28
Çizelge 3.1 : Mamogram etiketleri.	29
Çizelge 3.2 : Altı biçimde öznitelik matrisi.....	31
Çizelge 3.3 : Mamogram etiketleri.	32
Çizelge 3.4 : Mamogram etiketleri.	32
Çizelge 3.5 : Öznitelik matrislerinin boyutları	32
Çizelge 4.2 : Normal, iyi huylu ve kanser durumları için validasyon skoru	36
Çizelge 4.3 : Kanser ve kanser olmayan durumlar için test skoru.....	37
Çizelge 4.4 : Kanser ve kanser olmayan durumlar için validasyon skoru.	38
Çizelge 4.5 : Kitle olan ve olmayan durumlar için test skoru.	39
Çizelge 4.6 : Kitle olan ve olmayan durumlar için validasyon skoru.....	39
Çizelge 4.7 : Normal, iyi huylu ve kanser durumları için test skoru.	40
Çizelge 4.8 : Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırılması	41
Çizelge 4.9 : Normal, iyi huylu ve kanser durumları için validasyon skorları.....	41
Çizelge 4.10 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.	41
Çizelge 4.11 : Kanser/kanser olmayan veri kümesi CLAHE test skorları.	42
Çizelge 4.12 : Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırması	43
Çizelge 4.13 : Kanser ve kanser olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.....	43
Çizelge 4.14 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	44
Çizelge 4.15 : Kitle olan ve olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş test skorları.	44
Çizelge 4.16 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	45

Çizelge 4.17 : Kitle olan ve olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.....	45
Çizelge 4.18 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	45
Çizelge 4.19 : Normal, iyi huylu ve kanser durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş test skorları	46
Çizelge 4.20 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	46
Çizelge4.21 : Normal, iyi huylu ve kanser durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.....	47
Çizelge 4.22 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	47
Çizelge 4.23 : Kanser ve kanser olmayan durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş test skorları	48
Çizelge 4.24 : Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırması	48
Çizelge 4.26 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	49
Çizelge 4.27 : Kitle olan ve kitle olmayan durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş test skorları	50
Çizelge 4.28 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	50
Çizelge 4.29 : Kitle olan ve kitle olmayan durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.....	51
Çizelge 4.30 : Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması	51
Çizelge 4.31 : CLAHE ve YBMG test skorlarının karşılaştırılması.....	51
Çizelge 4.32 : CLAHE ve YBMG validasyon skorlarının karşılaştırılması.....	52

ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 2.1 : YBMG bileşenlerine karşılık gelen görüntüler	11
Şekil 2.2 : Orijinal ve süperpoze edilmiş görüntüler	11
Şekil 2.3 : $k=3$ ve 5 için en yakın komşu.....	12
Şekil 2.4 : Destek vektörler yardımıyla hesaplanan en uygun düzlem.....	14
Şekil 2.5 : Destek vektör makineleri yöntemi	14
Şekil 2.6 : Tanım aralığı farkı.....	17
Şekil 2.7 : Lojistik bağlanım adımları	18
Şekil 2.8 : Karar ağacı basit gösterimi.....	20
Şekil 2.9 : Bir meme görüntüsünün histogramı	25
Şekil 2.10 : CLAHE yöntemiyle kontrastı iyileştirmiş mamogram görüntüsü	26
Şekil 2.11 : YBMG ile kontrastı iyileştirilmiş mamogram görüntüsü	28
Şekil 3.1 : Orijinal IRMA veritabanı görüntüleri.	30
Şekil 3.2 : Önışlemeden geçirilmiş görüntüler.	30
Şekil 3.3 : Öznitelik matrisi oluşturulurken izlenen yol.	31



YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ İLE DİJİTAL MAMOGRAMLARDA MEME KANSERİ TANISI

ÖZET

Giderek gelişen teknolojik gelişmelerle birlikte yapay zekanın kullanımı günlük yaşantımızda yaygınlaşmaya başlamıştır. Sağlık sektöründe de gün geçtikçe artan kullanımıyla yapay öğrenme algoritmaları hayati önem taşımaktadır. Bilgisayar Destekli Tanı sistemlerinin gelişmesiyle birlikte yanlış tanı riski azaltılmaya ve erken tanı imkanı artmaya başlamıştır. BDT sistemleri, erken ve doğru teşhisin hayati önem taşıdığı kanser gibi ölümcül hastalıkların seyrini değiştirecek niteliktedir.

Bu tez çalışmasında Yüksek Boyutlu Model Gösterilimi yöntemi ve çeşitli gözetimli makine öğrenmesi yöntemleri kullanılarak meme kanseri teşhisi otomatikleştirilmiştir. Yeni bir algoritma geliştirilmiş ve sadece mamogram görüntüleri kullanılarak meme kanseri tanısı üzerine çalışılmıştır. Bu çalışmamın özgün tarafı, Yüksek Boyutlu Model Gösterilimi yönteminin ilk kez meme kanseri tanısı için kullanılmış olmasıdır. IRMA veri tabanından elde edilmiş mamogramlara, YBMG yöntemi uygulanmış ve YBMG bileşenleri hesaplanmıştır. Bu bileşenler, meme kanseri tanısında öznitelik olarak kullanılmıştır. Bu çalışmada YBMG aynı zamanda bir kontrast iyileştirme yöntemi olarak ele alınmış ve durumlara göre ayrılmış veri kümelerine uygulanmıştır.

Yüksek Boyutlu Model Gösterilimi yöntemi bir böl yönet algoritmasıdır. Günümüzde çok çeşitli uygulamaları olan bu yöntemin meme kanseri tanısı için kullanılması uygulaması da bu tez çalışmada yapılmıştır. Yüksek Boyutlu Model Gösterilimi Yöntemi, karesi integrallenebilen çok değişkenli bir fonksiyonu, daha az bağımsız değişkenli fonksiyonların toplamı şeklinde yeniden temsil etmeye yarayan bir yöntemdir. Daha az işlem yükü gerektiren problemlerde kesme yaklaşım fonksiyonları kullanılabilir. Bu çalışmada da ikinci dereceye kadar olan terimlerden oluşan YBMG denklemi esas alınmıştır.

Bu çalışmada meme görüntülerinden elde edilmiş YBMG bileşenleri ve bu bileşenlerin istatistiksel varyansları birer öznitelik olarak kabul edilmiştir. Öznitelik matrisleri YBMG bileşenlerinin çeşitli durumlarına göre oluşturulmuştur. Bu öznitelik matrisleri kullanılarak mamogramlar 3 farklı durumda incelenmiştir. Sırasıyla normal, iyi huylu ve kanser, kanser olan ve olmayan ve kitle olan ve olmayan durumlarını içeren veri kümeleri oluşturulmuş ve öznitelik matrisleri bu durumlarda elde edilmiştir. Gözetimli makine öğrenmesi yöntemleri kullanılarak yöntemin başarısı sınanmıştır. Meme kanseri tanısı için geliştirilen algoritma sınanmış ve çalışmada tablolar halinde sunulmuştur. IRMA veri kümesinden elde edilen görüntüler herhangi bir görüntü iyileştirme işlemi uygulanmadan bu sonuçlar elde edilmiştir. Çıkan sonuçlar karşılaştırılmış ve farklı arayışlara yönelinmiştir. Elde edilen skorları arttırılabileceği düşüncesiyle her bir mamaograma kontrast iyileştirme işlemi uygulanmıştır. Meme görüntülerine kontrast iyileştirme tekniklerinden olan Kontrast Sınırlı Uyarlamalı Histogram Eşitleme yöntemi ve Yüksek Boyutlu Model Gösterilimi

yönteminin kontrast iyileştirme için geliştirilmiş algoritması uygulanmıştır. Bu yöntemler uygulandıktan sonra başarıları sınanmış ve karşılaştırılmıştır. Bu çalışmada kontrast iyileştirme teknikleri arasındaki skor farklarının anlamlı olup olmadığına t testi uygulanarak bakılmıştır. Normal iyi huylu ve kanser veri kümesi için YBMG ve CLAHE yöntemleri arasında istatistiksel bir fark gözlemlenirken kanser ve kanser değil, kitle var ve kitle yok veri kümeleri için anlamlı bir fark gözlemlenmemiştir. Tüm veri kümeleri ve skorlara bakıldığında en yüksek test başarıları %89 olarak elde edilmiştir.

Geliştirilen özgün yöntemin performansı, mamogramların durumlara göre ayrılmış veri kümelerine ve oluşturulan öznelik matrislerine göre sınanmış ve karşılaştırma yapılmıştır. YBMG bileşenlerinin öznelik olarak değerlendirilebileceği düşüncesine, elde edilen başarılı sonuçlar neticesinde ulaşılmıştır.



BREAST CANCER DIAGNOSIS IN DIGITAL MAMMOGRAMS USING HIGH DIMENSIONAL MODEL REPRESENTATION

SUMMARY

With the advances in computer technologies, the use of artificial intelligence has started to become widespread in not only our daily life but also in the health sector. Computer Aided Diagnosis (CAD) systems have reduced the risk of misdiagnosis and increased the possibility of early diagnosis. CAD systems have the potential to positively affect the course of deadly diseases such as cancer.

In this study, we developed a computer-aided diagnosis algorithm based on the High Dimensional Model Representation (HDMR) method and various supervised machine learning methods. That is, we try to design an automated diagnosis algorithm for breast cancer diseases using mammogram images. This algorithm has two main steps. The first step is that the HDMR components which are features for the mammograms are obtained using HDMR decompositions method and the second one is to cluster the mammograms using machine learning algorithms such as k-nearest neighborhood algorithm, support vector machine, logistic regression, linear discriminant analysis, decision trees algorithms, random forest.

HDMR is a method used for function and data decomposition and image enhancement. In other words, HDMR is a divide-and-conquer method which is used for a low dimensional representation of a multivariate function.

The most general of an N -dimensional function is the HDMR expansion as follows:

$$f(x_1, \dots, x_N) = f_0 + \sum_{i=1}^N f_i(x_i) + \sum_{\substack{i,j=1 \\ i < j}}^N f_{ij}(x_i, x_j) + \dots + f_{12\dots N}(x_1, \dots, x_N)$$

In this multivariate function, the constant term f_0 , the univariate terms $f_i(x_i)$ and the bivariate terms are represented in the form of $f_{ij}(x_i, x_j)$.

HDMR can be applied to N -dimensional multivariate data set and as a result 2^N number of less variate data sets are obtained. Since images in RGB format have three-dimensional array structure. 2^3 data sets are gained. These are named as constant term, univariate terms and bivariate terms. The general structure of constant, univariate and bivariate terms are as follows:

$$f_0 = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \sum_{k_3=1}^{n_3} \alpha_{k_1}^{(1)} \alpha_{k_2}^{(2)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)})$$

$$f_1(\xi_1^{(k_1)}) = \sum_{k_2=1}^{n_2} \sum_{k_3=1}^3 \alpha_{k_2}^{(2)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0$$

$$f_2(\xi_2^{(k_2)}) = \sum_{k_2=1}^{n_2} \sum_{k_3=1}^3 \alpha_{k_1}^{(1)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0$$

$$f_3(\xi_3^{(k_3)}) = \sum_{k_2=1}^{n_2} \sum_{k_3=1}^3 \alpha_{k_1}^{(1)} \alpha_{k_2}^{(2)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0$$

$$f_{12}(\xi_1^{(k_{m1})}, \xi_2^{(k_{m2})}) = \sum_{k_3=1}^3 \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_1(\xi_1^{(k_1)}) - f_2(\xi_2^{(k_2)}) - f_0$$

$$f_{13}(\xi_1^{(k_{m1})}, \xi_3^{(k_{m3})}) = \sum_{k_2=1}^{n_2} \alpha_{k_2}^{(2)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_1(\xi_1^{(k_1)}) - f_3(\xi_3^{(k_3)}) - f_0$$

$$f_{23}(\xi_2^{(k_{m2})}, \xi_3^{(k_{m3})}) = \sum_{k_1=1}^{n_1} \alpha_{k_1}^{(1)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_2(\xi_2^{(k_2)}) - f_3(\xi_3^{(k_3)}) - f_0$$

Here we assume an image in RGB format as $n_1 \times n_2 \times n_3$ dimension. Since RGB has 3 channels for pictures, n_3 is always equal to 3. As it seen that there are seven components of HDMR are given here. The last remaining term is just calculated that the given above components are subtracted from the original image. Since HDMR method represents the original image exactly, the remaining term is out of scope.

HDMR method has many applications in various fields. In this study, we use this method to decomposed the mammograms with HDMR in this study. The original aspect of this study is that the HDMR method was used for the first time to diagnose breast cancer. Mammogram images were drawn from the IRMA database. We used the different combinations of main HDMR components and statistical properties to build six different feature matrices.

We trained various supervised machine learning algorithms to classify the patients. We used three different classification modes, with labeling the set as normal-benign-malign, normal-abnormal or cancer-not cancer. We reported and compared the accuracy scores of each method. First of all, a data set which was not used any contrast enhancements methods take for processing. Afterwards, image enhancement techniques were applied and a new data set was created. For the two datasets, three different paths were applied similarly.

First way; For the data set containing normal, benign and cancer states, the YBMG components of each mammogram were calculated. The calculated components were transformed into an attribute matrix in 6 formats, each with the same path.

Second way; HDMR components were calculated for a data set containing cancer and non-cancer states.

The third way; HDMR components of mamaograms containing masses and mass disappearances were calculated. Similarly, attribute matrices were formed and achievement scores were obtained.

In summary, the attributes to be measured with success were determined and a vector was created by adding to the tip. We also tested whether contrast enhancement

methods increase the accuracy or not and found that Contrast Limited Adaptive Histogram Equalization (CLAHE) does not affect the prediction accuracy, while HDMR significantly increased the success up to 89% when used as a contrast enhancement method.

In this study we found that HDMR is a successful method to extract features from mammograms for using in artificial learning algorithms and we confirmed that HDMR is an effective contrast enhancement method in medical image processing.





1. GİRİŞ

Tıbbi tanılamada bilgisayar desteği ve yapay öğrenme tekniklerinin kullanımı gün geçtikçe artmakta ve hayati önem taşımaktadır. Erken ve doğru teşhisin önemli olduğu kanser gibi hastalıklarda bilgisayar destekli tanı sistemleri yaygınlaşmaktadır. Kanser, insan hayatı için önemli bir tehdittir. Dünya Sağlık Örgütü istatistiklerine göre, kanserden kaynaklanan ölümler 2030 yılında yaklaşık 12 milyon kişiye ulaşacaktır. Bu nedenle hem tıbbi uygulamalarda hem de bilimsel araştırmalarda kanser teşhisine yönelik yapılan her çalışma önem kazanmıştır (Tang ve Aghayan, 2014).

Bilgisayar Destekli Tanı (Computer-aided diagnosis , CAD, BDT) tıbbi görüntüleme ve tanısal radyolojide oldukça önemli bir araştırma alanıdır. BDT ile radyologlar bilgisayar çıktısını ikinci görüş olarak kullanır ve gözden kaçabilecek bulguların yeniden değerlendirilmesi sağlanır. Burada BDT ile bilgisayarların performansının doktorlar ile karşılaştırılabilir ya da daha iyi olması gerekmez, ancak doktorlar tarafından tamamlayıcı olması gerekmektedir. Dolayısıyla BDT, doğrudan tanıyı koyan araç olarak değil ancak tanıyı koymada doktorlara yardımcı olacak tamamlayıcı bir sistemdir (Doi, 2007).

Bilgisayar Destekli Tanı, sağlık sektöründe verimliliği arttırmada oldukça önemli bir yere sahiptir. BDT, çeşitli kanser türlerinde gözden kaçabilecek bulguların tespitinde kullanılmış ve olumlu sonuçlar vermiştir. Örneğin meme, akciğer, kolon, prostat kanserleri, Alzheimer hastalığı ve diyabetik retinopati gibi birçok kanser türü ve hastalığın teşhisinde bilgisayar destekli tanıdan yararlanılmaktadır.

Daha önce yapılan çalışmalarda birinde kolonik poliplerin tanısında bilgisayar destekli tanı sistemi ve optik kolonoskopi performansı karşılaştırılmıştır. Bu karşılaştırmada 1186 hasta içeren bir veri kümesinden rastgele seçilen 394 veri eğitim, 792 veri test kümesi olarak ayrılmıştır. Burada BDT ile 8mm ve daha büyük boyuttaki polipler için optik kolonoskopiyle kıyasla benzer sonuçlar elde edilmiştir (Tang ve Aghayan, 2014).

Radyologların akciğer grafilerindeki iyi huylu ve kötü huylu pulmoner noldülleri belirlemede bilgisayar destekli tanı şeması kullanılıp kullanılmaması karşılaştırılarak radyologların performansı ölçülmüştür. 31 primer akciğer kanseri ve 22 iyi huylu nodül içeren toplam 53 akciğer grafisi içeren veri kümesinde her nodül için malignite olma olasılığı ölçüsü, otomatik bir bilgisayar destekli şema kullanılarak belirlenmiştir. Radyologların şema varken ve şema olmadan görüntüleri yorumlaması beklenmiş ve radyologların performansının şema varken konan tanıyı önemli ölçüde iyileştirdiği gösterilmiştir (Shiraishi ve diğ, 2003).

Alzheimer tipi demansın (ATD) tanısında erken teşhis için bir bilgisayar destekli tanı tekniği sunulmuştur. Önerilen metodoloji, Welch'in her iki sınıf arasındaki t-testini sunan normal ve Alzheimer görüntüleri arasındaki belirli bir eşikten büyük voksellerin seçimine dayanmaktadır. Seçilen vokseller için yoğunluk değerlerinin ortalama ve standart sapması hesaplanmış ve iki farklı sınıflandırma için öznelik vektörleri seçilmiştir. Önerilen metodoloji ile %95 üzerinde doğruluk seviyesine ulaşılmıştır (Salas-Gonzalez ve diğ, 2010).

Erken teşhis, diğer kanser türlerinde olduğu gibi meme kanserinin seyrini belirlemede de çok önemli bir faktördür. Bilgisayar destekli tanı, meme kanserinin erken teşhisi için önemli bir araçtır. Çift okuma ile karşılaştırıldığında BDT radyologların iş yükünü azaltmasının yanında etkili bir BDT sistemi ile erken teşhiste etkili olmuştur (Jinshantang ve diğ, 2009) .

Makine öğrenmesi verilerden bilgi çıkarmakla ilgilidir. İstatistik, yapay zeka ve bilgisayar bilimlerinin kesişimindeki bir araştırma alanıdır ve aynı zamanda tahmine dayalı analitik veya istatistiksel öğrenme olarak da bilinir. Makine öğrenmesini temel alan uygulamalar günlük yaşamımızda oldukça geniş yer tutmaktadır. Bu uygulamalardan en hayati olanları şüphesiz sağlık sektöründe yer alanlardır.

En başarılı makine öğrenme algoritmaları, bilinen örneklerden genelleştirerek karar verme süreçlerini otomatikleştiren algoritmalarıdır. Gözetimli öğrenme olarak bilinen bu teknikte, girdiler ve istenen çıktılar kullanıcı tarafından sağlanır ve algoritma, yeni veri için girdileri kullanarak istenen çıktıyı üretmenin yollarını bulur (Müller ve Guido, 2017).

Son yirmi yılda çeşitli makine öğrenmesi teknikleri ve öznelik seçme algoritmaları hastalık teşhisinde, prognozu ve öngörüsünde yaygın olarak kullanılmaktadır (Kourou

ve diğ., 2015). Özellikle kanser teşhisinde yardımcı yöntem olarak sıkça kullanılan makine öğrenmesi teknikleri, erken teşhis ve tedavi imkanı sağlayabilmesi açısından önemlidir. Literatürde yapılan bir çalışmada 286 meme kanseri teşhisi konmuş hastaya ait UCI (University of California-Irvine) Machine Learning Laboratuvarından alınmış veriler kullanılmıştır. Dokuz adet öznitelik içeren bu veriler ön işlemden geçirilmeden makine öğrenmesi teknikleri kullanılarak sınıflandırılmıştır. Bu sınıflandırma Weka Veri Madenciliği yazılımı kullanılarak yapılmıştır. Sınıflandırma başarısını %45 ila %79 arasında yöntemlere göre değişiklik göstermiştir (Kolayve ERdoğmuş, 2016). Başka bir çalışmada meme kanseri teşhisi için kullanılan makine öğrenmesi tekniklerinin sınıflandırma performansları karşılaştırılmıştır. Ayrıca meme kanserinde rol oynayan baskın genlerin belirlenmesi üzerine çalışılmış ve neticede 139 öznitelik ile elde edilen başarı %90,72 olmuştur. Bu çalışma etkin genlerin seçimi WEKA programı kullanılarak yapılmıştır (Bektas ve Babur, 2016).

Bu çalışmada yapay öğrenmede kullanılacak özniteliklerin eldesinde Yüksek Boyutlu Model Gösterilimi (YBMG) yöntemi kullanılmıştır. Bu yöntem, bir boyut azaltma tekniği olarak da ele alınır. Çok değişkenli bir fonksiyonu daha az değişkenli fonksiyonların toplamı olarak yazmayı sağlayan YBMG yöntemi, bir böl-yönet algoritmasıdır (M. A. Tunga ve Demiralp, 2008).

YBMG yöntemi ilk olarak Rus matematikçi I. M. Sobol tarafından ortaya atılmıştır (Sobol, 1993). Prof. Metin Demiralp ve grubu ile birlikte bu yöntem geliştirilmiş ve farklı bilimsel alanlarda çok çeşitli problemlerin çözümü için kullanılmaya başlanmıştır (Özay & Demiralp, 2014; Korkmaz & Metin, 2012; M.A.Tunga ve Demiralp, 2009; B. Tunga & Demiralp, 2012; B. Tunga & Demiralp, 2010). Örneğin renkli görüntüleri yeniden inşaa etmede kullanılmış ve başarılı sonuçlar vermiştir (B.Tunga ve Altın, 2014). Başka bir çalışmada hem YBMG görüntü kümeleme yöntemi hem de boyut azaltma tekniği olarak kullanılmıştı (Karcılı ve B.Tunga, 2017). Diğer bir çalışmada ise bu yöntem görüntü iyileştirme tekniği olarak sunulmuştur (B. Tunga ve Koçanaoğulları, 2017). Ayrıca YBMG yöntemi, yüksek boyutlu veriyi düşük boyutlarda temsil ederek imgedeki eksik kısımların tamamlanması yada bozuk kısımların düzeltilmesinde de kullanılmıştır (Karaca ve M.A.Tunga, 2018).

Bu çalışmada ise YMBG yöntemi ilk defa meme kanseri tanısı için kullanılmış ve yöntemin performansı ölçülmeye çalışılmıştır. Bu amaçla IRMA (Image Retrieval in Medical Applications) (http://irma-project.org/index_en.php) veri tabanında bulunan

mamogramlar kullanılmış ve YBMG yöntemi bu mamogramlar üzerine uygulanarak veri ayrıştırımı yapılmıştır. Bu sayede çok büyük boyutlu mamogramlar yerine daha az boyutta veri kümeleri ile elde edilen öznitelik matrisleri oluşturulmuştur. Bu matrisler gözetimli yapay öğrenmesi teknikleri kullanılarak normal, iyi huylu ve kanser, kanser olan ve olmayan ve kitle olan ve olmayan durumlar için tahmin başarıları ölçülmüştür.

Bu tez çalışmasının ikinci bölümünde Yüksek Boyutlu Model Gösterilimi yönteminin genel anlatımı, verilere ve görüntülere uygulanışı, makine öğrenmesi teknikleri ve görüntü iyileştirme algoritmaları verilmiştir. Bu çalışmanın özgün kısmı olan Yüksek Boyutlu Model Gösterilimi yöntemiyle meme kanseri tanısı işlemleri üçüncü bölümde gösterilmiştir. Dördüncü bölümde elde edilen sonuçlar tablolar halinde karşılaştırmalı olarak sunulmuştur. Beşinci bölümde ise elde edilen sonuçlardan ve önerilerden bahsedilmiştir.

2. YÖNTEM

Bu bölümde fonksiyon ve veri ayrıştırımı ile görüntü işleme için kullanılan Yüksek Boyutlu Model Gösterilimi (YBMG) yöntemi, gözetimli makine öğrenmesi teknikleri ve görüntü iyileştirme algoritmaları yer almaktadır.

2.1 Yüksek Boyutlu Model Gösterilimi Yöntemi (YBMG)

N -boyutlu bir $f(x_1, \dots, x_N)$ fonksiyonunun en genel haliyle YBMG açılımı aşağıdaki gibi sonlu bir toplama karşılık gelir:

$$f(x_1, \dots, x_N) = f_0 + \sum_{i=1}^N f_i(x_i) + \sum_{\substack{i,j=1 \\ i < j}}^N f_{ij}(x_i, x_j) + \dots + f_{12\dots N}(x_1, \dots, x_N) \quad (2.1)$$

Çok değişkenli bu fonksiyonda sabit terim f_0 , tek değişkenli terimler $f_i(x_i)$ ve iki değişkenli terimler $f_{ij}(x_i, x_j)$ şeklinde gösterilir. YBMG açılımına ait terim sayısı denklem 2.2'deki gibi hesaplanır.

$$\binom{N}{0} + \binom{N}{1} + \dots + \binom{N}{N} = \sum_{i=0}^N \binom{N}{i} - 2^N \quad (2.2)$$

Sonuç olarak toplam terim sayısı 2^N şeklinde bulunur.

YBMG terimleri olarak adlandırılan bu fonksiyonların eşsiz bir biçimde belirlenmesinde Sobol tarafından ortaya atılan ve daha sonra H. Rabitz tarafından genelleştirilen aşağıdaki yok etme koşulları geçerlidir.

$$\int_{a_1}^{b_1} dx_1 \dots \int_{a_N}^{b_N} dx_N W_1(x_1) \dots W_N(x_N) f_i(x_i) = 0, \quad i = 1, \dots, N \quad (2.3)$$

Bu koşulda yer alan $W_i(x_i)$, ağırlık fonksiyonuna karşılık gelmektedir. Daha sonraki adımlarda YBMG terimlerini hesaplarken kullanacağımız ağırlık fonksiyonlarının çarpımından elde edilen ifade denklem 2.4'teki gibidir.

$$W(x_1, \dots, x_N) = \prod_{i=1}^N W_i(x_i), \quad x_i \in [a_i, b_i], \quad 1 \leq i \leq N \quad (2.4)$$

Uygun bir iç çarpım tanımı altında YBMG açılımından elde edilen herhangi iki terimin birbirine dik olması koşulu, Prof. Demiralp tarafından kanıtlanan denklem 2.5'teki ifadeye karşılık gelmektedir.

$$\langle f_{i_1 i_2 \dots i_k}, f_{i_1 i_2 \dots i_l} \rangle = 0, \quad i_1 i_2 \dots i_k \neq i_1 i_2 \dots i_l, \quad k, l = 1, \dots, N \quad (2.5)$$

Buradan YBMG açılımının bir dik ayrıştırım olduğu sonucuna ulaşılabilir.

Dik olma koşulu için kullanılan uygun bir iç çarpım, her biri x bağımsız değişkenine bağlı, karesi integrallenebilir iki N -boyutlu fonksiyon, $u(x_1, \dots, x_N)$ ve $v(x_1, \dots, x_N)$, alınarak denklem 2.6'teki gibi tanımlanır.

$$\langle u, v \rangle = \int_{a_1}^{b_1} dx_1 \dots \int_{a_N}^{b_N} dx_N W(x_1, \dots, x_N) u(x_1, \dots, x_N) v(x_1, \dots, x_N) \quad (2.6)$$

YBMG terimlerinin ağırlık faktörleri hesaplanmasında işlem kolaylığı sağlamak adına belirlenen normalizasyon kriteri, ilgili a_i, b_i aralıklarda integraller 1 olacak şekilde seçilir.

$$\int_{a_i}^{b_i} dx_i W_i(x_i) = 1, \quad i = 1, \dots, N \quad (2.7)$$

Yukarıda anlatılan ağırlık fonksiyonları ve diklik koşulları kullanılarak YBMG sabit, tek değişkenli ve iki değişkenli bileşenleri sırasıyla hesaplanır.

YBMG sabit bileşenini hesaplamak için denklem 2.1'de verilen ifade ile denklem 2.4 çarpma işlemine tabii tutulur ve tüm değişkenler ilgili aralıklarda integrallenir. Elde edilen eşitliğin sağ tarafında yok etme koşulundan dolayı sadece f_0 kalır ve istenen ifadeye ulaşılır.

$$f_0 = \int_{a_1}^{b_1} dx_1 \dots \int_{a_N}^{b_N} dx_N W_1(x_1) \dots W_N(x_N) f(x_1, \dots, x_N) \quad (2.8)$$

Tek değişkenli terimler hesaplanırken de benzer bir yol izlenir. Herhangi tek değişkenli $f_i(x_i)$ fonksiyonu için denklem 2.1'deki eşitliğin her iki tarafı ağırlık fonksiyonları ile çarpılır ve ilgili bağımsız değişkenine göre yok etme koşulu altında

ağırlık fonksiyonu ve integralleri atılır. Sonuç olarak elimizde sadece YBMG'nin sabit f_0 ve tek değişkenli $f_i(x_i)$ terimleri kalır. Böylece bulmak istediğimiz ifade elde edilir.

$$f_i(x_i) = \int_{a_1}^{b_1} dx_1 \dots \int_{a_{i-1}}^{b_{i-1}} dx_{i-1} \int_{a_{i+1}}^{b_{i+1}} dx_{i+1} \dots \int_{a_N}^{b_N} dx_N W_1(x_1) \dots$$

$$W_{i-1}(x_{i-1}) \cdot W_{i+1}(x_{i+1}) \dots W_N(x_N) f(x_1, \dots, x_N) - f_0, \quad i = 1, \dots, N \quad (2.9)$$

İki değişkenli terimler hesaplanırken de sabit ve tek değişkenli terimlerdeki yol izlenir. Burada da iki değişkenli $f_{ij}(x_i, x_j)$ fonksiyonu için denklem 2.1'deki eşitliğin iki tarafı da $W(x_i), W(x_j)$ dışındaki ağırlık fonksiyonları ile çarpılır ve x_i, x_j dışındaki bağımsız değişkenlere göre ilgili aralıkta integrellenir. Yok etme ve normalizasyon koşulları uygulandıktan sonra istenen ifadeye ulaşılır.

$$f_{ij}(x_i, x_j) = \int_{a_1}^{b_1} dx_1 \dots \int_{a_{i-1}}^{b_{i-1}} dx_{i-1} \int_{a_{i+1}}^{b_{i+1}} dx_{i+1} \dots \int_{a_{j-1}}^{b_{j-1}} dx_{j-1} \int_{a_{j+1}}^{b_{j+1}} dx_{j+1} \dots \int_{a_N}^{b_N} dx_N$$

$$\cdot W_1(x_1) \dots W_{i-1}(x_{i-1}) W_{i+1}(x_{i+1}) \dots W_{j-1}(x_{j-1}) W_{j+1}(x_{j+1}) \dots W_N(x_N) f(x_1, \dots, x_N)$$

$$- f_i(x_i) - f_j(x_j) - f_0, \quad i, j = 1, \dots, N \quad (2.10)$$

Aynı yol izlenerek üç ve daha fazla değişkenli YBMG bileşenlerinin hesaplanması da mümkündür. Fakat literatür incelendiğinde ikili terimlere kadar YMBG açılımının yeterli olduğu ve yüksek başarı oranlarıyla karşılaşıldığı görülmüştür. Dolayısıyla daha az değişkenli fonksiyonlar yardımıyla orijinal fonksiyona yaklaştırım mümkündür. Sabit, tek değişkenli ve iki değişkenli yaklaşımlar Prof. Demiralp tarafından tanımlanmış ve kesme yaklaşımları adı verilmiştir.

$$s_0(x_1, \dots, x_N) = f_0$$

$$s_1(x_1, \dots, x_N) = s_0(x_1, \dots, x_N) + \sum_{i=1}^N f_i(x_i)$$

$$s_2(x_1, \dots, x_N) = s_1(x_1, \dots, x_N) + \sum_{\substack{i,j=1 \\ i < j}}^N f_{ij}(x_i, x_j)$$

⋮

$$s_k(x_1, \dots, x_N) = s_{k-1}(x_1, \dots, x_N) + \sum_{\substack{i_1, \dots, i_k=1 \\ i_1 < \dots < i_k}}^N f_{i_1 \dots i_k}(x_{i_1}, \dots, x_{i_k})$$

$$1 \leq k \leq N \quad (2.11)$$

Denklem 2.11'deki bağlantılar yardımıyla orijinal fonksiyona k. dereceden yaklaşıtırm denklem 2.12'de gösterilmiştir.

$$f(x_1, \dots, x_N) \approx s_k(x_1, \dots, x_N) \quad (2.12)$$

Bu çalışmada YBMG'ye açılacak fonksiyonlar, görüntüler kullanılarak elde edilecektir. Orijinal fonksiyonumuz üç bağımsız değişken içereceğinden ikinci dereceye kadar olan terimler kullanacağımız yaklaşıtırm için yeterlidir.

$$f(x_1, \dots, x_N) \approx s_2(x_1, \dots, x_N) \quad (2.13)$$

Sonuç olarak bu çalışmada kullanacağımız YBMG yaklaşımı üç bağımsız değişken içeren ikinci dereceye kadar olan terimlerden oluşan denklem 2.14'teki gibi olacaktır.

$$f(x_1, x_2, x_3) = f_0 + \sum_{i_1=1}^3 f_{i_1}(x_{i_1}) + \sum_{\substack{i_1, i_2=1 \\ i_1 < i_2}}^3 f_{i_1 i_2}(x_{i_1}, x_{i_2}) \quad (2.14)$$

Analitik yapısı bilinen fonksiyonlara uygulanan bu yöntem, veriler için de uygulanabilir. Her x_i bağımsız değişkeni için, ayrık bir D_i kümesi tanımlanır:

$$D_i = \{\varepsilon_i^1, \dots, \varepsilon_i^{k_i}, \dots, \varepsilon_i^{n_i}\}, i = 1, \dots, N \quad (2.15)$$

Burada orijinal verinin tanım kümesi, her bir x_1, \dots, x_N bağımsız değişkeni için tanımlanan D_1, \dots, D_N kümelerinin kartezyen çarpımına karşılık gelir. Dolayısıyla analitik yapısı bilinmeyen bir fonksiyon üzerinde tanımlanan sonlu sayıda ayrık nokta ile oluşturulmuş tanım kümesi denklem 2.16'daki gibi ifade edilir.

$$D = D_1 \times D_2 \times \dots \times D_N \quad (2.16)$$

Ağırlık fonksiyonu da bu yapıya uygun düzenlenir. Dirac delta fonksiyonlarının lineer kombinasyonları kullanılarak denklem ağırlık fonksiyonu 2.17'deki gibi elde edilir.

$$W_i(x_i) = \sum_{k_i}^{\alpha_i} \alpha_{k_i}^i \delta(x_i - \varepsilon_i^{k_i}), \quad x_i \in [a_i, b_i], \quad i = 1, \dots, N \quad (2.17)$$

Veriler için bu ağırlık fonksiyonu yardımıyla, analitik yapısı bilinen fonksiyonlar için uygulandığı şekliyle uygulanarak YBMG bileşenleri hesaplanır.

2.2 Yüksek Boyutlu Model Gösterilimi Yöntemi İle Görüntü İşleme

RGB formatındaki renkli görüntü verileri, yatay ve düşey piksel değerlerinden oluşan üç boyutlu dizi yapısına sahiptir. Bu yüzden dijital görüntülere uygulanan YBMG yöntemi, 3 değişkene bağlı bir veri kümesine uygulandığı şekliyle kullanılmıştır. Burada her bağımsız değişkenin aldığı değerler n_1, n_2, n_3 ile gösterilir ve sırasıyla düşey piksel sayısı, yatay piksel sayısı, renk sayısına karşılık gelir. Renkli görüntü verileri RGB formatında 3 kanala sahip olduğu için n_3 her zaman 3'e eşittir. Dolayısıyla YBMG ayrıştırım yöntemi dijital görüntüler üzerine uygulandığında her bir görüntü için sabit, tek değişkenli ve iki değişkenli bileşenleri içeren terimler elde edilir (B.Tunga ve Koçanaoğulları, 2017).

Aşağıda YBMG'ye açılan görüntülerden elde edilen sabit terim verilmiştir (B.Tunga ve Altın, 2014):

$$f_0 = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \sum_{k_3=1}^{n_3} \alpha_{k_1}^{(1)} \alpha_{k_2}^{(2)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) \quad (2.18)$$

Burada $f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)})$ RGB formatındaki üç boyutlu resmin piksel değerlerine, $\alpha_{k_j}^{(j)}$ parametresi ise Dirac delta fonksiyonu yardımıyla hesaplanmış her bir piksele ait ağırlık faktörüne karşılık gelir (M. A. Tunga ve Demiralp, 2008). Bu ağırlık faktörleri literatürde genellikle aşağıdaki şekilde seçilir:

$$\alpha_{k_j}^{(j)} = \frac{1}{n_j} \quad j = 1,2,3, \quad 1 \leq k_1 \leq n_1, \quad 1 \leq k_2 \leq n_2, \quad 1 \leq k_3 \leq n_3 \quad (2.19)$$

Böylece aynı boyuttaki tüm ağırlık faktörlerinin değerleri eşit olarak alınmıştır.

Ağırlık faktörlerinin seçimi değişiklik gösterse de aşağıdaki normalizasyon kriterini sağlamalıdır:

$$\sum_{k_j=1}^{n_j} \alpha_{k_j}^j = 1, \quad j = 1,2,3 \quad (2.20)$$

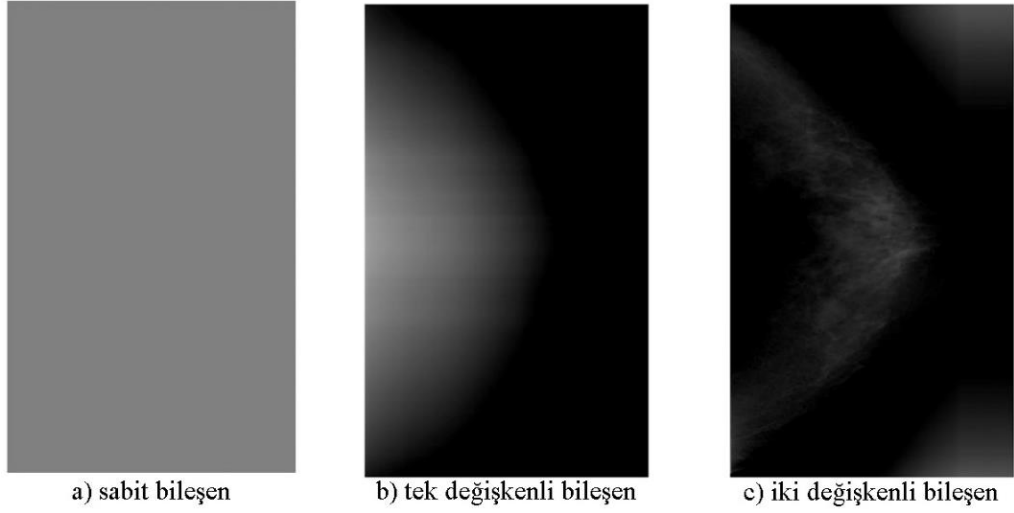
Verilen ağırlıklar altında görüntüler için YBMG veri ayrıştırımının tek değişkenli terimleri aşağıdaki gibi hesaplanır:

$$\begin{aligned}
f_1(\xi_1^{(k_1)}) &= \sum_{k_2=1}^{n_2} \sum_{k_3=1}^3 \alpha_{k_2}^{(2)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0 \\
f_2(\xi_2^{(k_2)}) &= \sum_{k_1=1}^{n_1} \sum_{k_3=1}^3 \alpha_{k_1}^{(1)} \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0 \\
f_3(\xi_3^{(k_3)}) &= \sum_{k_2=1}^{n_2} \sum_{k_1=1}^{n_1} \alpha_{k_1}^{(1)} \alpha_{k_2}^{(2)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_0
\end{aligned} \tag{2.21}$$

YBMG'ye açılan görüntü verilerinden elde edilen sabit ve tek değişkenli bileşenler yardımıyla iki değişkenli bileşen de aşağıdaki gibi hesaplanır:

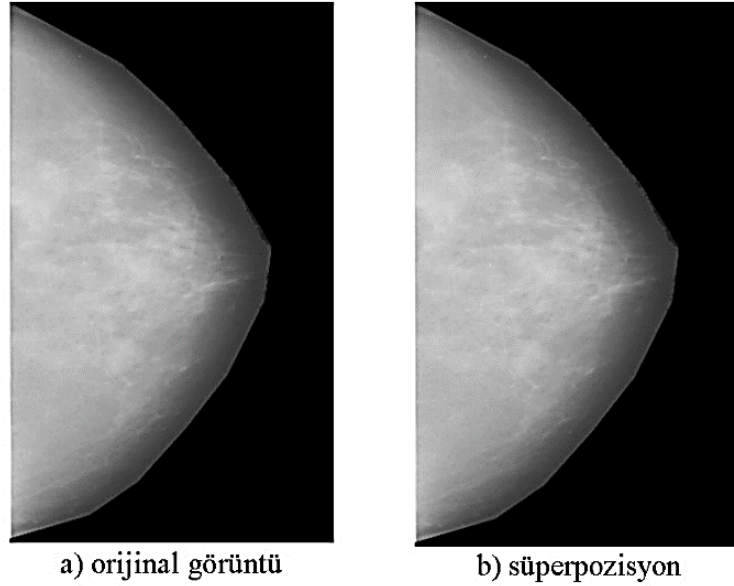
$$\begin{aligned}
f_{12}(\xi_1^{(k_{m1})}, \xi_2^{(k_{m2})}) &= \sum_{k_3=1}^3 \alpha_{k_3}^{(3)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_1(\xi_1^{(k_1)}) - f_2(\xi_2^{(k_2)}) - f_0 \\
f_{13}(\xi_1^{(k_{m1})}, \xi_3^{(k_{m3})}) &= \sum_{k_2=1}^{n_2} \alpha_{k_2}^{(2)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_1(\xi_1^{(k_1)}) - f_3(\xi_3^{(k_3)}) - f_0 \\
f_{23}(\xi_2^{(k_{m2})}, \xi_3^{(k_{m3})}) &= \sum_{k_1=1}^{n_1} \alpha_{k_1}^{(1)} f(\xi_1^{(k_1)} \xi_2^{(k_2)} \xi_3^{(k_3)}) - f_2(\xi_2^{(k_2)}) - f_3(\xi_3^{(k_3)}) - f_0
\end{aligned} \tag{2.22}$$

YBMG'ye açılan dijital görüntülerin hesaplanan tek değişkenli bileşenleri birer vektör, iki değişkenli bileşenleri birer matris yapısındadır. Şekil 2.1'de YBMG'ye açılan dijital mamogramların sabit, tek değişkenli ve iki değişkenli bileşenlerinin karşılığı olan görüntüler yer almaktadır.



Şekil 2.1 : YBMG bileşenlerine karşılık gelen görüntüler.

Tüm bu terimler yeniden yerleştirme yapılarak toplandığında, yani terimlerin süperpozisyonu ile, orijinal görüntü Şekil 2.2'deki gibi tam olarak ifade edilebilir.



Şekil 2.2 : Orijinal ve süperpoze edilmiş görüntüler.

Bu çalışmada YBMG'nin sabit, tek değişkenli ve iki değişkenli bileşenleri kullanılarak yeni algoritma geliştirilmiş ve bu algoritma kullanılarak meme kanseri teşhisi konulmaya çalışılmıştır.

2.3 Gözetimli Makine Öğrenmesi Teknikleri

Bu bölümde k-En Yakın Komşu (k-Nearest Neighbour, kNN), Destek Vektör Makineleri (Support Vector Machine, SVM), Lojistik Bağlanım (Logistic Regression, LogReg),

Doğrusal Ayırtaç Çözümlemesi(Linear Discriminant Analysis, LDA), Karar Ağaçları(Decision Trees) ve Rastgele Orman(Random Forest, RF) gözetimli makine öğrenmesi teknikleri anlatılmaktadır. Bu yöntemler YBMG yönteminin sabit, tek ve iki değişkenli bileşenlerinin kullanılmasıyla elde edilen ve bir sonraki bölümde nasıl bulunduğu anlatılacak olan öznelik matrisine uygulanmıştır.

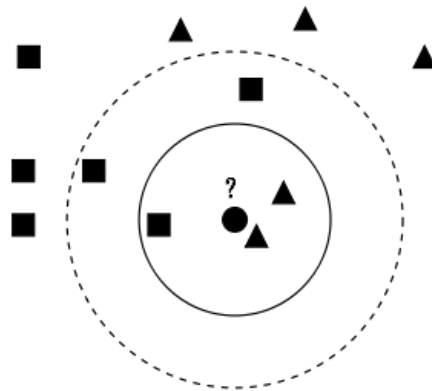
2.3.1 K-en yakın komşu algoritması

k-En Yakın Komşu algortitması, 1967 yılında Cover ve Hart tarafından ortaya atılmış, istatistiksel öğrenme teorisine dayalı örnek tabanlı bir gözetimli makine öğrenmesi yöntemidir (Cover ve Hart, 1967) . Bu yöntem bir veri kümesindeki örneklerin, genellikle yakın özelliklere sahip diğer örneklerle benzerlik göstereceği prensibine dayanır (Kotsiantis, 2007). Eğer örnekler, bir sınıflandırma etiketi ile belirlendiyse, sınıflandırılmamış bir örneğin etiketinin değeri, en yakın komşularının sınıfını gözlemleyerek belirlenir. Böylece kNN algoritması yardımıyla, işaretli örneğe en yakın diğer örnek bulunur ve en sık kullanılan etiketi tanımlayarak örneğin sınıfı belirlenir .

Pozitif bir k tamsayısı ve x_0 test örneği olarak verilsin. kNN sınıflandırıcısı, eğitim verilerini göz önünde bulundurarak x_0 'a en yakın k noktalarını gösterir ve N_0 olarak temsil edilir. Daha sonra, j sınıfına ait koşullu olasılığı N_0 'daki noktaların, j değerine eşit olduğu noktaların kesri olarak tahminde bulunur:

$$\Pr(Y = j | X = x_0) = \frac{1}{k} \sum_{x_i \in N_0} I(y_i = j) \quad (2.23)$$

Son olarak Bayes kuralı uygulanır ve x_0 test örneğini en yüksek olasılıklı sınıfa atar (Casella ve diğ, 2017).



Şekil 2.3 : k=3 ve 5 için en yakın komşu.

Şekil 2.3'te yuvarlak şekil, sınıfı belirlenecek veriyi, kareler ve üçgenler ise iki farklı sınıfa ait veri kümelerine ait örnekleri temsil eder. Ayrıca kesiksiz çember $k = 3$ için, kesikli çember ise $k = 5$ için en yakın komşuları çevreleyen sınırlardır.

Burada eğitim verisiyle sınıfı belirlenecek veri arasındaki “yakınlık” için çeşitli metrikler kullanılabilir. Bu Öklid metriği olabileceği gibi Minkowski ve Manhattan metriği gibi çeşitli uzaklık fonksiyonları tercih edilebilir (Hastie, 2017). Çizelge 2.1’de k-En Yakın Komşu algoritması verilmiştir (Hyun, 2014).

Çizelge 2.1 : k-En yakın komşu algoritması.

k-NN Algoritması

- 1: X : eğitim verisi, Y : X 'in belirlenmiş sınıfı, x : bilinmeyen
 - 2: **for** $i=1$ **to** m **do**
 - 3: Mesafeyi hesapla $d(X_i, x)$
 - 4: **end for**
 - 5: En yakın k mesafesi için $d(X_i, x)$ I 'daki indeksler hesaplanır.
 - 6: **return** $\{Y_i, i \in I\}$ için etiketlenir.
-

2.3.2 Destek vektör makineleri

İstatistiksel öğrenme teorisine dayanan gözetimli makine öğrenme tekniklerinden bir diğeri de Destek Vektör Makineleri yöntemidir. Bu yöntem bir çeşit ikili ayrıştırma tekniği olarak düşünülebilmektedir. Vapnik ve ark. tarafından geliştirilen bu yöntem ile ikili sınıflandırmada tam ayrışabilen ve tam ayrışamayan veriler üzerinde etkili sonuçlar alınabilmektedir (Cortes ve Vapnik, 1995).

Elimizde tam ayrışabilir bir veri kümesinden elde edilmiş p -boyutlu bir öznitelik matrisi ve bu veri kümesine ait sınıf bilgileri olsun. Burada öznitelik matrisi ve sınıflar sırasıyla girdiler ve çıktılar olarak düşünülebilir.

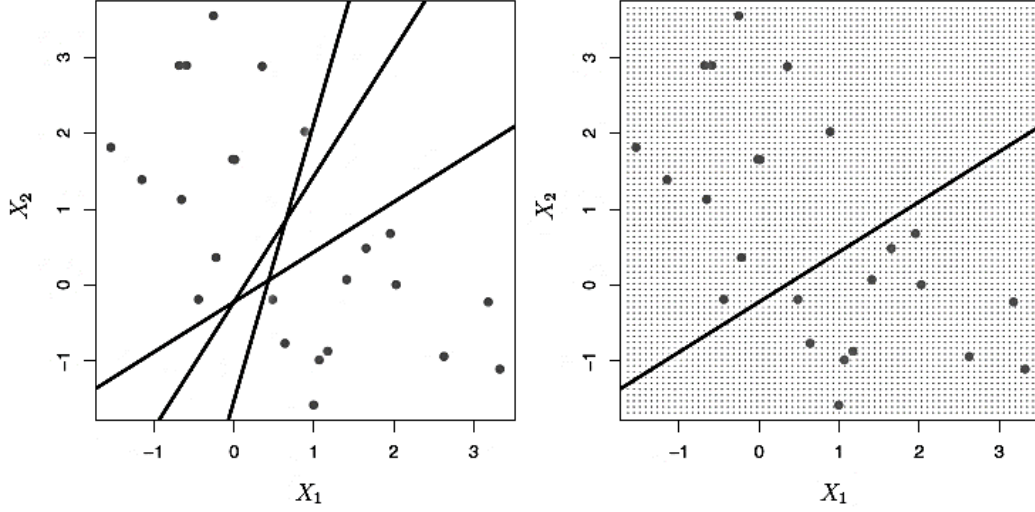
$$(x_i, y_i), \quad i = 1, 2, \dots, n \quad (2.24)$$

$$\text{Girdiler: } x_i \in \mathbb{R}^p, \quad \text{Çıktılar: } y_i \in \{-1, 1\}$$

Sınıflar için genellikle $\{-1, 1\}$ etiketi kullanılmakla birlikte herhangi iki etiket de seçilebilir. Yöntemin temel amacı, iki veya daha fazla sınıfa ait verileri bir düzlem (hyperplane) yardımıyla en uygun şekilde iki sınıfa ayırmaktır.

$$\text{Düzlem: } \{X \in \mathbb{R}^p: \beta^T X + \beta_0 = 0, \beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}\} \quad (2.25)$$

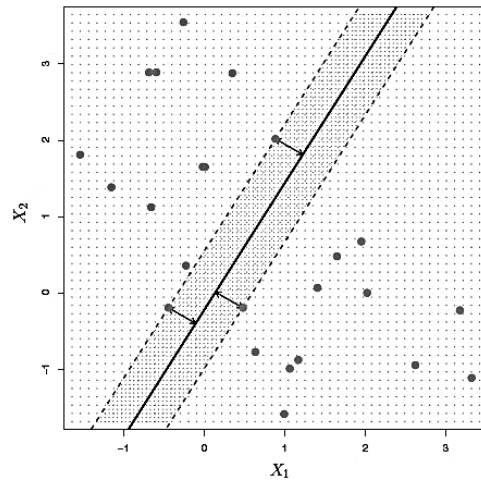
Denklem (2.25)'de β^T ve β_0 sırasıyla eğim ve kesme noktası olarak düşünülebilir. Bundan sonra ise amacımız, ayırmak istediğimiz iki sınıf arasından geçecek en uygun düzlemi belirleyerek sınıflandırma yapmaktır.



Şekil 2.4 : Destek vektörler yardımıyla hesaplanan en uygun düzlem.

Şekil 2.5'te veri noktalarını birbirinden ayırmak için belirlenen (iki boyutta doğrulara karşılık gelen) düzlemler gösterilmektedir. Şekil 2.5'te ise tam ayrışabilir veri kümesi için veriyi ikiye ayıracak şekilde seçilen en uygun düzlem gösterilmiştir.

Veri kümesinin elemanlarını birbirinden tam ayırıştıran düzlemin nereden geçtiğini, $\mp\Delta$ komşuluğunda kaydırılarak belirlenen düzlemler üzerindeki vektörler belirler. Yani Şekil 2.6'daki kesikli çizgiler üzerinde noktayla gösterilen veriler, sınıflandırmayı belirleyen verilerdir ve yöntemin ismini aldığı destek vektörleri oluşturur.



Şekil 2.5 : Destek vektör makineleri yöntemi.

Denklem (2.26), $\mp\Delta$ komşuluğunda kaydırılarak belirlenen düzlemlere aittir. Buradaki amaç, β ve Δ parametrelerini

$$\beta^T X + \beta_0 - \Delta = 0 \quad \text{ve} \quad \beta^T X + \beta_0 + \Delta = 0 \quad (2.26)$$

ile gösterilen iki düzlem arasındaki mesafeyi maksimum olacak şekilde belirlemektir. O halde x_i öznitelikleri ve y_i etiket değerleri için

$$\text{eğer } y_i = 1 \text{ ise } \beta^T x_i + \beta_0 - 1 \geq 0$$

$$\text{eğer } y_i = -1 \text{ ise } \beta^T x_i + \beta_0 + 1 \leq 0$$

$$y_i(\beta^T x_i + \beta_0) \geq 1 \quad (2.27)$$

eşitsizliği elde edilir.

Denklem 2.27'teki eşitliklerin her iki tarafı Δ ile bölünür ve Δ parametresi işlem kolaylığı açısından (genellikle) 1 olarak alınır.

Çeşitli geometrik işlemlerin ardından iki düzlem arasındaki mesafe aşağıdaki gibi bulunur:

$$\frac{2}{\|\beta\|} \quad (2.28)$$

İki sınıfın karşılaştırıldığı doğrusal ayırışan bir veri kümesinde aşağıdaki optimallik şartları geçerlidir:

$$\text{minimal tut: } \frac{2}{\|\beta\|} = \frac{1}{2} \beta^T \beta$$

$$\text{öyle ki: } y_i(\beta^T x_i + \beta_0) \geq 1, \quad i = 1, \dots, n \quad (2.29)$$

Bu yöntem tam ayırışmayan veri kümelerinde ve çoklu sınıflarda da benzer şekilde çalışmaktadır. Ayrıca doğrusal ayırışmayan veri kümelerinde sınıflandırma çeşitli çekirdekler(kernels) yardımıyla yapılabilir. Bu çekirdeklere örnek olarak polinomal ve radyal bazlı çekirdekler verilebilir.

Bu sınıflandırma algoritmasının temel problemi, çözümün veri kümesine bağlı, verideki ufak değişikliklerden kolay etkileniyor olması yani varyansının yüksek olmasıdır (Casella ve diğ, 2017).

2.3.3 Lojistik bağlanım

Sınıflandırma problemlerinde sıkça kullanılan gözetimli makine öğrenmesi yöntemlerinden biri Lojistik Bağlanım'dır. Bu yöntem genellikle iki etiketli durumlarda yaygın bir şekilde kullanılırken daha fazla etiket içeren problemlerde de kullanıldığı durumları bulunmaktadır.

$$\text{eğitim verisi: } (x_i, y_i), \quad i = 1, 2, \dots, n \quad (2.30)$$

Elimizde öznitelikleri temsil eden p -boyutlu X öznitelik kümesi ve bunlara karşılık gelen sınıf/etiket bilgilerini içeren Y etiketleri bulunsun. Amacımız eğitim verisi yardımıyla sınıfını sorguladığımız bir veriyi en uygun şekilde etiketlemek olacaktır. Y etiketlerinin olma olmasılığı ve veri kümesindeki X öznitelikleri arasındaki ilişkiyi veren olasılık ifadesi aşağıdaki gibidir:

$$p(x) = P(Y = 1|X) \quad (2.31)$$

Bir etiketin bu sınıfa ait olma olasılığını denklem (2.31) yardımıyla hesaplamış oluruz. $p = 1$, yani tek bir öznitelik için doğrusal bağlanım ile olasılık hesapladığımızda

$$p(x) = \beta_0 + \beta_1 X, \quad (-\infty, \infty) \quad (2.32)$$

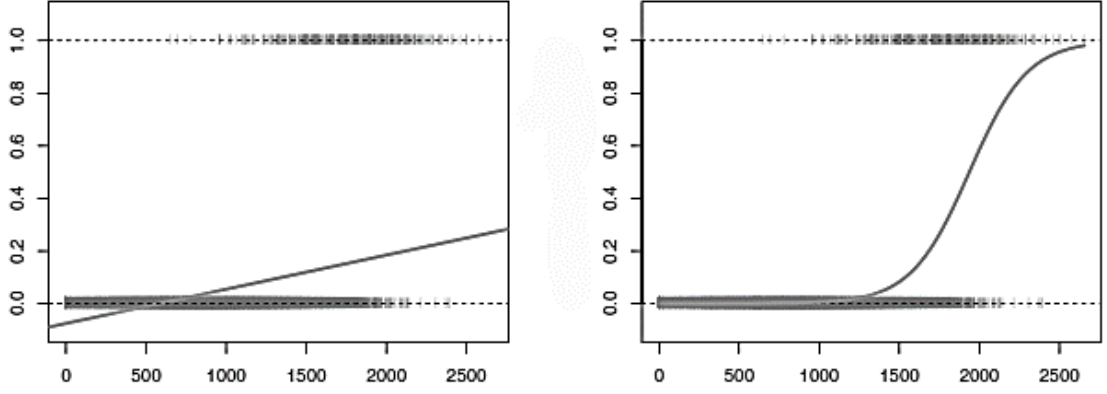
elde edilir.

Fakat olasılık hesabı için bu tanım aralığı uygun değildir. Bu yüzden bir dönüşüm fonksiyonu yardımıyla (Sigmoid fonksiyonu) tanım kümesini $[0,1]$ aralığına çekmek mümkün olacaktır.

$$p(x) = \sigma(\beta_0 + \beta_1 X), \quad [0,1]$$

$$\text{sigmoid fonksiyonu: } \sigma(y) = \frac{e^y}{1 + e^y} \quad (2.33)$$

Şekil 2.6'da tanım aralığı sırasıyla $(-\infty, \infty)$ ve $[0,1]$ olan iki durumun olasılık değerleri gösterilmektedir.



Şekil 2.6 : Tanım aralığı farkı.

İki etiketin tek bir öznitelikle olan ilişkisine ait sınıflandırma başarısını gösteren Lojistik fonksiyonu aşağıdaki gibidir:

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (2.34)$$

Bu fonksiyonu genelleştirmek mümkündür. Birden fazla öznitelik içeren problemlerde Lojistik fonksiyonu aşağıdaki gibi hesaplanır:

$$p(x) = \frac{e^{\beta_0 + \beta_1 x + \beta_2 x + \dots + \beta_p x}}{1 + e^{\beta_0 + \beta_1 x + \beta_2 x + \dots + \beta_p x}} \quad (2.35)$$

Bundan sonraki adımlarda yapmamız gereken eğitim verisi yardımıyla $\beta_0, \beta_1, \dots, \beta_p$ değerlerini bulmaktır.

Bir örnek yardımıyla hesaplamalar daha kolay anlaşılacaktır. Elimizde $p = 3$ öznitelik ve iki etiket durumunu içeren bir problem olsun ve sırasıyla x_1 özniteliği için 1 etiketinin, x_2 özniteliği için 0 etiketinin ve x_3 özniteliği için 1 etiketinin olma olasılığı denklem 2.36'teki gibi ifade edilir:

$$(x_1, 1), \quad (x_2, 0), \quad (x_3, 1)$$

$$P(Y = 1|X = x_1), \quad 1 - P(Y = 1|X = x_2), \quad P(Y = 1|X = x_3) \quad (2.36)$$

Her bir öznitelik için verinin istenilen etiketlerde olma olasılıklarını bulurken, tüm olasılıklar çarpılır.

$$p(x_1) * (1 - p(x_2)) * p(x_3) \quad (2.37)$$

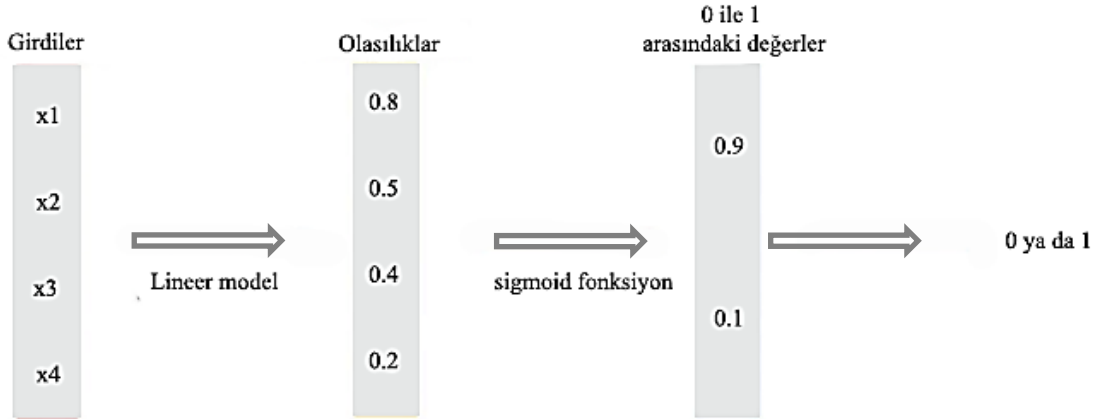
Bu ifade Denklem 2.38’da gösterilen Olabilirlik(Likelihood) Fonksiyonu’na karşılık gelir.

$$l(\beta_0, \dots, \beta_p) = \prod_{i:y_i=1} p(x_i) \prod_{j:y_j=0} (1 - p(x_j)) \quad (2.38)$$

Amacımız, bu fonksiyonun aldığı değeri maksimize etmektir.

$$\max_{\beta_0, \dots, \beta_p} l(\beta_0, \dots, \beta_p) \quad (2.39)$$

İşlem kolaylığı için maksimize edilecek ifadenin logaritması alınır ve toplamlar şeklinde yeniden ifade edilir. Böylece aranan $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_p$ değeleri bulunmuş olur (Casella ve diğ.,2017).



Şekil 2.7 : Lojistik bağlanım adımları.

2.3.4 Doğrusal Ayırtaç Çözümlemesi

Yüksek boyutlu veri kümelerinde analiz yaparken yaygın olarak boyut azaltma yöntemlerine başvururuz. Bir veri kümesi için sınıflandırma yaparken o kümenin karmaşıklığı öznelik sayısıyla bağlantılıdır. Sınıflandırma yöntemleri kullanılmadan önce veriye bir boyut azaltma yöntemi uygulayarak başarıya olumlu bir etkisinin olabileceği bilinmektedir. Gözetimli sınıflandırma teknikleriyle kullanılabilen Doğrusal Ayırtaç Çözümlemesi uygulanarak başarıya olan etkisi incelenmiştir (Alpaydın, 2012).

İlk olarak 1936 yılında R.Fisher'in ve 1939 yılında Welch'in ortaya attığı bu yöntemde, özniteliklerin lineer birleşimleri kullanılmıştır. Farklı sınıflardaki veri noktalarını birbirinden en iyi biçimde ayırabilmek için bir altküme seçimi yapılır. Bu seçimdeki amaç, var olan özniteliklerden en iyi altküme seçilerek en az sayıda boyut içeren yüksek başarılı öznitelikleri elde etmektir (Kuhn ve Johanson, 2016).

Bu işlem temelde üç adımdan oluşmaktadır. Öncelikle sınıflar arası varyans hesaplanır. Diğer bir deyişle farklı sınıfların ortalaması arasındaki uzaklık bulunur.

$$S_b = \sum_{i=1}^g N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (2.40)$$

İkinci adım olarak sınıf içi varyans olarak adlandırılan her bir sınıfın ortalaması ile örnek arasındaki uzaklığın hesaplanmasıdır.

$$S_w = \sum_{i=1}^g (N_i - 1)S_i = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T \quad (2.41)$$

Üçüncü adımda ise sınıflar arası ayrışımın en iyi olması için sınıf varyansının maksimum ve sınıf içi varyansının minimum olması beklenir. Fisher doğrusal ayırtacı, bu durumun maksimum olduğu değerle tanımlıdır (Wang ve diğ, 2010).

$$P_{lda} = \arg \max_p \frac{|P^T S_b P|}{|P^T S_w P|} \quad (2.42)$$

2.3.5 Karar Ağaçları

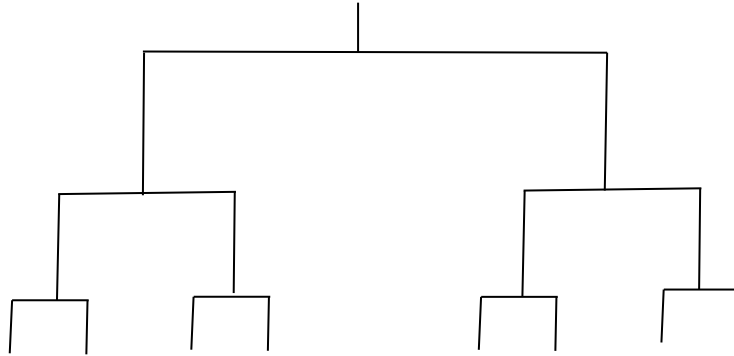
Ağaç tabanlı algoritmalar, sınıflandırma ve bağlanım problemlerinde kullanılan gözetimli makine öğrenmesi yöntemlerindedir. Bu yöntemle tahmin yapılırken eğitim verisi yardımıyla veri noktalarının ortalaması yada modu kullanılır. Tahmin uzayını bölmek için kullanılan bölme kuralları kümesi bir ağaçta özetlendiğinden, bu tür yaklaşımlara karar ağaçları yöntemi denir.

Ağaç tabanlı algoritmalar yorumlanırken basit ve kullanışlı olmaları bakımından avantajlıdır. Karar ağacı algoritmalarından bazıları ID3, CART, CHAID ve C4.5 olarak sıralanabilir (Adeyemo & Adeyeye, 2015). Bu çalışmada CART; sınıflandırma ve regresyon ağaçları kullanılmıştır.

Eğer birden fazla karar ağacı üreteceksek, Torbalama (Bagging) ve Takviye (Boosting) algoritmaları kullanılabilir. Bir karar ağacının varyansı azaltmak istendiğinde Torbalama algoritmasına başvurulabilir. Rastgele Orman algoritması başvuru bu yönteme örnek gösterilebilir. Rassal olarak bir ağacı üretmektense, bir ağacı üretip bu ağaçtan elde ettiğimiz bilgiyi diğer ağaca aktardığımız Takviye (Boosting) yöntemi ile tahmin başarısını iyileştirilebilir. Takviye, sadece karar ağaçlarında kullanılan bir yöntem değildir. Diğer makine öğrenmesi algoritmalarında da kullanılır. Torbalama, Rastgele Orman ve Takviye yöntemlerine, topluluk (ensemble) yöntemleri denir (Casella ve diğ, 2017).

2.3.6 Sınıflandırma ve Bağlanım Ağaçları

Sınıflandırma ve bağlanım ağaçları algoritması (Classification and Regression Trees, CART), 1984 yılında Leo Breiman ve arkadaşları tarafından ortaya atılmıştır (Breiman ve diğ, 1984). Her bir karar düğümünden itibaren ağacın iki dala ayrılması prensibine dayanan CART yönteminde temel amaç, x_1, \dots, x_p 'den oluşan p -boyutlu değişkenler uzayını daha az boyutlu alt bölgelere ayırmaktır. Eğer tahmin etmeye çalıştığımız çıktı sürekli bir değişkense bağlanım, sürekli olmayan değişkene ait veriler ise sınıflandırma ağaçları kullanılır.



Şekil 2.8 : Karar ağacı basit gösterimi.

Oluşturulan ağaç modelinde ara düğümlerde en iyi ayrımı yapan bağımlı değişkenler yer alır. Bu bağımlı değişkenlerin kritik değerleri, düğümlerin dallarında gösterilir. İlk düğüm ve son düğüm noktası arasında yer alan yol boyunca, sınıflar arası ayırım maksimize ve sınıf içi varyasyonun minimize edildiği kurallar yer alır.

Çizelge 2.2’de CART algoritması gösterilmiştir (Suknović ve diğ, 2005).

Çizelge 2.2 : CART algoritması.

CART Algoritması

```
1:  $d = 0$ , endtree= 0
2: Node(0) = 1, Node(1) = 0, Node(2) = 0
3: while endtree < 1
4:   if
      Node( $2^d - 1$ ) + Node( $2^d$ ) + ... + Node( $2^d - 2$ ) =  $2 - 2^{d+1}$ 
5:     endtree=1.
6:   else
7:     do  $i = 2^d - 1, 2^d, \dots, 2^{d+1} - 2$ .
8:       if Node( $i$ ) > -1
9:         Split Tree
10:        else
11:          Node( $2i + 1$ ) = -1
12:          Node( $2i + 2$ ) = -1
13:        end if
14:      end do
15:    end if
16:   $d = d + 1$ 
17: end while
```

Bağlanım Ağaçları: Bu yöntemde tahmin etmeye çalıştığımız veri, sürekli bir değişken olacaktır. Amacımız, x_1, \dots, x_p , p -boyutlu değişkenler uzayından çakışmayan, R_1, \dots, R_j , j -boyutlu bölgeler elde etmektir. Burada R_j bölgesindeki eğitim verisinin çıktı değerlerini (y_i) ortalaması, tahmin etmeye çalıştığımız veridir. Eğitim verisi kullanılarak bulunacak bu bölgeler Kalıntı Kareler Toplamı (KKT) minimize edilerek hesaplanır.

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (2.43)$$

Denklem 2.43'te y_i çıktı değerlerini ve \hat{y}_{R_j} , R_j bölgesindeki çıktıların ortalamasını gösterir. Tahmin değeri ile gerçek değer arasındaki farkın karesini bulduğumuz KKT işleminde amacımız bu değer en küçük olmasıdır. Böylece yapılacak hata minimize edilir.

Burada değişkenler uzayıyla çakışmayan R_j bölgelerini oluşturmaya çalışırken özyinelemeli ikili ayırma (recursive binary splitting) yöntemini kullanırız.

$$R_1(j, s) = \{x | x_j < s\}$$

$$R_2(j, s) = \{x | x_j \geq s\} \quad (2.44)$$

Burada s kısıtını gerçekleyen j ve s değerleri denklem 2.44'ün minimum olmasıyla bulunur.

$$\sum_{i: x_i \in R_1(j, s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j, s)} (y_i - \hat{y}_{R_2})^2 \quad (2.45)$$

Terminal düğümde çok az veri kalana kadar bu işlem yapılır (Casella ve diğ, 2017).

Sınıflandırma Ağaçları: Sınıflandırma Ağaçları, Bağlanım Ağaçlarına çok benzer şekilde çalışır. Fakat bu yöntemi kullanılırken sürekli verilerle işlem yapılmaz. Bu yüzden bölerek oluşturulan alt bölgelerin saflık derecesine bakılır. Gerçekleşen her bölünme, sınıflandırma hata oranını en aza indirmek için yapılır.

Sınıflandırma hata oranı, yalnızca en yaygın sınıfa ait olmayan bir bölgedeki eğitim gözlemlerinin oranıdır. O halde \hat{p}_{mk} , m. bölgedeki eğitim verisindeki k. sınıftan olanların oranı olarak gösterilsin.

$$E = 1 - \max(\hat{p}_{mk}) \quad (2.46)$$

Denklem 2.45, sınıflandırma hata oranı formülüdür. Fakat sınıflandırma ağacını büyütürken yeterince hassas değildir. Entropi ve Gini indeksi ile saflık ölçülebilir.

$$G = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk}) \quad (2.47)$$

Denklem 2.47 ile Gini indeksi formülü verilmiştir. Bu, tüm sınıflardaki toplam varyansın ölçüsüdür. Görüldüğü gibi eğer oran 0 veya 1'e yakında, Gini indeksi küçük olacaktır. Bu nedenle iyi bir düğüm sağlığı ölçüsüdür.

$$E = - \sum_{k=1}^K \hat{p}_{mk} \log(\hat{p}_{mk}) \quad (2.48)$$

Benzer bir mantık, Denklem 2.48'de çapraz entropi adı verilen diğer yonteme uygulanır (Casella ve diğ, 2017).

2.3.7 Rastgele Orman

Ağaç tabanlı gözetimli makine öğrenmesi tekniklerinden biri rastgele orman algoritmasıdır. Temel amaç, ağaçlar arasındaki korelasyonu azaltmak için dalları ayırırken tüm değişkenler yerine rassal sayıda değişkeni kullanmaktır. Rastgele orman, alt ağaçların öğrenilme biçiminin algoritmasını değiştirir, böylece tüm alt ağaçlardan elde edilen tahminler daha az korelasyon gösterir. Bu yöntem, Torbalama yönteminin geliştirilmiş bir halidir. CART’da bir bölünme noktası seçerken, en uygun ayrılma noktasını seçmek için öğrenme algoritmasının tüm değişkenlerine bakmasına izin verilir. Fakat Rastgele Orman algoritmasında bu prosedür değişir, öğrenme algoritması aranacak özelliklerin rastgele bir örneği ile sınırlıdır.

Her bir düğümde kullanılacak değişkenlerin sayısı m ve giriş değişkenlerinin sayısı p olsun. Bu parameter, en iyi bölünmeyi belirlemek için kullanılacaktır.

Sınıflandırma için varsayılan en iyi eşitlik:

$$m = \sqrt{p} \quad (2.49)$$

İken bağlanım için varsayılan en iyi eşitlik:

$$m = \frac{p}{3} \quad (2.50)$$

şeklinde olacaktır (Brownlee, 2017).

Bu sayede aşırı öğrenme probleminin önüne geçilebilir.

Aşağıda rastgele orman algoritması gösterilmiştir (Kuhn ve Johanson, 2016).

Çizelge 2.3 : Rastgele orman algoritması.

Rastgele Orman Algoritması

- 1: m , her düğümde kullanılacak değişkenlerin sayısı seçilir
 - 2: **for** $i = 1$ to m **do**
 - 3: Orijinal verinin önyükleme örneği oluşturulur
 - 4: Bu örnek üzerinden ağaç modeli eğitilir
 - 5: **for** her bölünme **do**
 - 6: Orijinal belireyicilerin rastgele $k (< P)$ ‘sı seçilir
 - 7: k belirleyicileri arasında en iyi tahminci seçilir ve veri bölünür
 - 8: **end**
 - 9: Bir ağacın ne zaman tamamlandığını belirlemek için tipik ağaç modeli durdurma ölçütleri kullanılır(ancak budama yapılmaz)
 - 10: **end**
-

2.4 Görüntü İyileştirme

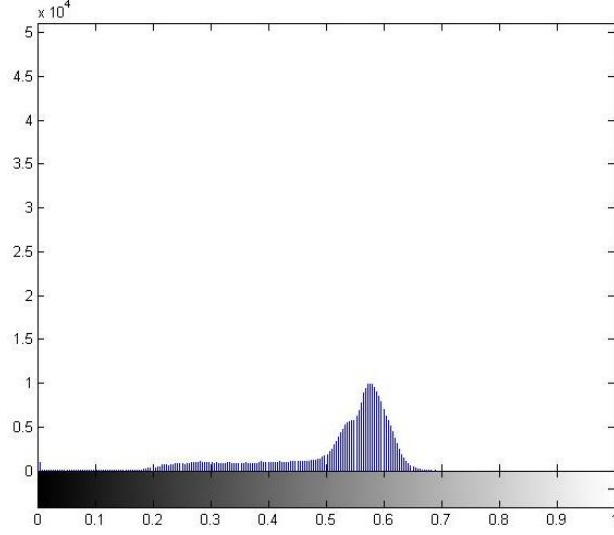
Temel olarak gürültüsüz bir görüntünün kalitesi; bulanıklık, yüksek çözünürlük ve iyi bir kontrast ile ilişkilidir. İyi bir kontrast için görüntünün siyah-beyaz dengesindeki mevcut gri değerlerin, insan görüşü hassasiyetindeki parlaklık seviyesinde olması gerekir (M.Petrou ve C.Petrou, 2010). Dolayısıyla görüntü ne kadar kaliteliyse o kadar az önışlemeden geçerek amaca uygun şekilde kullanılabilir.

Dijital görüntülerde kaliteyi arttırmak için genelde kontrast iyileştirme tekniklerine başvurulur. Örneğin, uyarlamalı bulanık mantık kontrast iyileştirme yöntemi kullanılarak mamografik konturların ve ince detayların etkili bir şekilde geliştirilebileceği gösterilmiştir (Cheng ve Xu, 2002). Başka bir çalışmada bölge bazlı kontrast iyileştirme yöntemi kullanılarak farklı boyut ve şekillerdeki mamografik özelliklerin kontrastını arttırmak için uyarlamalı bir yöntem uygulanmıştır. Bu yöntemle mikrokalsifikasyon kümelerinin ve anatomik detayların geçerliliğini önemli ölçüde arttırdığı gösterilmiştir (Morrow ve diğ, 1992). Diğer bir çalışmada ise histogramla modifiye edilmiş yerel kontrast iyileştirmesi yapılarak orijinal mamogram görüntülerinin yerel bilgisi korunmuş ve sağlıklı bir kontrast iyileştirmesi yapılmıştır. Bu yöntem, mikrokalsifikasyon kümelerinin saptanabilirliğini arttırarak optimum sonuçlar sağlamıştır (Lewin ve diğ, 2003).

Bu çalışmada kontrast iyileştirme yöntemleri olarak mamogramlara Kontrast Sınırlı Uyarlamalı Histogram Eşitleme (Zuiderveld, 1994) ve kontrast iyileştirme yöntemi olarak Yüksek Boyutlu Model Gösterilimi (B. Tunga ve Koçanaoğulları, 2017) uygulanmıştır.

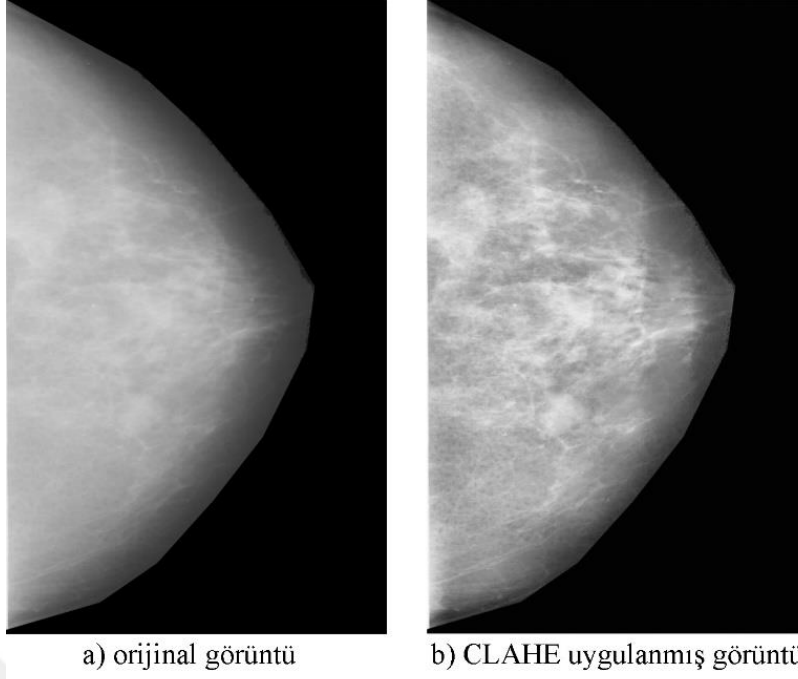
2.4.1 Kontrast sınırlı uyarlamalı histogram eşitleme

Bir görüntünün histogramı, ayrık fonksiyonlar olarak adlandırılır. Görüntüdeki piksel değerlerinin ağırlığını belirler ve Şekil 2.9'daki gibi bir grafikte temsil edilir.



Şekil 2.9 : Bir meme görüntüsünün histogramı.

Uyarlamalı histogram eşitleme (AHE), görüntülerde kontrastı arttırmak için kullanılan bir görüntü işleme ve iyileştirme tekniğidir. Uyarlamalı yöntemin, her biri görüntünün ayrı bir bölümüne karşılık gelen birkaç histogramı hesapladığı ve görüntünün açıklık değerlerini yeniden dağıtmak için kullandığı yönündeki sıradan histogram eşitlemesinden farklıdır. Bu nedenle, yerel kontrastı arttırmak ve görüntünün her bölgesindeki kenar tanımlarını geliştirmek için uygundur. Bununla birlikte, AHE, bir görüntünün nispeten homojen bölgelerindeki gürültüyü uzatma eğilimine sahiptir. Kontrast sınırlı uyarlamalı histogram eşitleme (CLAHE) olarak adlandırılan uyarlamalı histogram eşitleme çeşidi, amplifikasyonu sınırlayarak bunu önler (Zuiderveld, 1994).



Şekil 2.10 : CLAHE yöntemiyle kontrastı iyileştirmiş mamogram görüntüsü.

Şekil 2.10’de orijinal meme görüntüsüyle iyileştirme işlemi yapılmış mamogramlar gösterilmektedir. Görüldüğü üzere CLAHE uygulanmış meme görüntüsünde damarlanmalar belirgin hale gelmiştir.

Çizelge 2.4 : CLAHE algoritması.

Kontrast Sınırlı Uyarlamalı Histogram Eşitleme

```

1: for  $i=1 \rightarrow n$  do
2:   Resimler okunur. Resimler image ile gösterilir.
3:   if renkli resim then
4:     Görüntüler gri tonlamalı resme dönüştür.
5:   else if gri tonlamalı resim then.
6:     adaphisteq(image, 'clipLimit', 0.01, 'Distribution', 'exponential',
7:     'Alpha', 0.2, 'NBins', 350)
8:     Oluşturulan yeni resim image2'ye atanır.
9:     Elde edilen yeni görüntüler kaydedilir.
10:  end if
11: end for

```

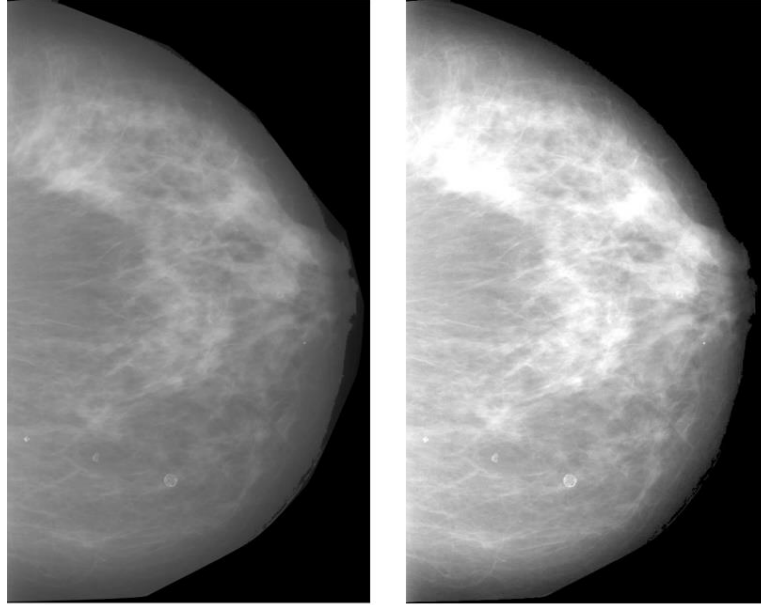
Çizelge 2.4’te Kontrast Sınırlı Uyarlamalı Histogram Eşitleme yöntemine ait algoritma verilmiştir. MATLAB’da *adaphisteq* fonksiyonu kullanılarak dijital mamogramların histogramları eşitlenmiştir. Bu fonksiyonda ‘clipLimit’, ‘Distribution’, ‘Alpha’ ve ‘NBins’ parametleri kullanılmıştır. ClipLimit, kontrast iyileştirme limitidir ve varsayılan değer olan -0.01- kullanılmıştır. Bu parametre [0, 1]

arasındaki deęerleri almaktadır. Görüntünün özellikle homojen alanlarında aşırı doymayı önleyen bir kontrast faktörüdür. Distribution, istenilen histogram şeklidir ve varsayılan deęeri -uniform-'dur. Buna ek olarak 'rayleigh' ve 'exponential' daęılımlar da histogramı şekillendirmek için kullanılmaktadır. Bu daęılım, kontrast dönüşümü işlevini oluşturmak için kullanılan daęılımdır. Seçilen daęılım, giriş görüntüsüne baęlıdır. Örneęin rayleigh daęılımı, sualtı görüntülerde daha iyi sonuç vermektedir. Bu çalışmada daha iyi sonuç alınan exponential daęılım kullanılmıştır. Alpha, daęılım parametresini belirleyen negatif olmayan gerçel sayı deęerlerini alır. Varsayılan deęeri -0.04-'tür fakat bu çalışmada 0.2 alınmıştır. Alpha deęeri, yalnızca rayleigh ve exponential daęılımlarda kullanılır. NBins, kontrast iyileştiren dönüşümü oluşturur ve varsayılan deęeri 256'dır. Pozitif sayı deęerlerini alır ve kontrastı arttıran bir dönüşümü oluşturmak için kullanılan histogram kutularının sayısına karşılık gelir. Daha yüksek deęerler, daha düşük işlem hızına sebep olsa da daha iyi sonuç verir. Bu çalışmada NBins deęeri 350 olarak seçilmiştir. Bu parametreler, en iyi sonuç alacak şekilde belirlenmiş ve algoritma oluşturulmuştur.

2.4.2 Kontrast iyileştirme yöntemi olarak YBMG

Yüksek Boyutlu Model Gösterilimi yöntemi bir kontrast iyileştirme algoritması olarak öncelikle renkli ve gri seviyeli resimlere uygulanmış ve başarılı sonuçlar vermiştir.

Dijital görüntülerden elde edilen iki ve üç deęişkenli YBMG bileşenleri sırasıyla gri seviyeli ve renkli görüntüleri temsil ederken kullanılır. YBMG yaklaşımlarının yapısı yeniden oluşturularak bir kontrast iyileştirme teknięi haline getirilir. Bir gri tonlamaları resmin tek deęişkenli bileşenleri, görüntünün baskın renk bilgisini içerir. Ana desen, görüntünün iki deęişkenli YBMG bileşenlerinde tutulduğundan gri tonlamalı görüntülerde kontrastı arttırmak için tek deęişkenli ve iki deęişkenli bileşenleri üst üste eklenir. Eęer kontrastı daha fazla arttırmak istenirse tek deęişkenli ve iki deęişkenli bileşenler birkez daha eklenir. Bu işlem kontrast iyileştirme seviyesini yönetmek için tekrarlanabilir. Gri seviyeli görüntülerde olduęu gibi renkli görüntülerde de kontrast benzer şekilde iyileştirilir.



a) orijinal görüntü

b) YBMG uygulanmış görüntü

Şekil 2.11 : YBMG ile kontrastı iyileştirilmiş mamogram görüntüsü.

Şekil 2.11’de original görüntü ve YBMG ile kontrastı iyileştirilmiş mamogram resmi verilmiştir. Çizelge 2.4’te ise kontrast iyileştirme yöntemi olarak kullanılan YBMG algoritması özetlenmiştir (B. Tunga ve Koçanaoğulları, 2017).

Çizelge 2.5 : YBMG kontrast iyileştirme algoritması.

Kontrast İyileştirme için YBMG Yöntemi

- 1: Resmin matris formunu oluşturun (Gri tonlama resimler için $N = 2$, renkli resimler için $N = 3$).
 - 2: $N = N + 1$ (Resmin çoğaltılması. Ek boyut 1 ve 2 değerlerini alabilir.)
 - 3: **if gri tonlamalı resim then**
 - 4: Resmin sabit, tek değişkenli, iki değişkenli YBMG bileşenleri hesaplanır.
 - 5: Tek değişkenli bileşenlerine kadar YBMG yaklaşımı hesaplanır.
 - 6: **for $i=1 \rightarrow l$ do** *l: Kontrast İyileştirme seviyelerinin sırası*
 - 7: YBMG yaklaşımına tek değişkenli ve iki değişkenli bileşenler eklenir.
 - 8: **end for**
 - 9: **else if renkli resim then**
 - 10: Resmin sabit, tek değişkenli ve iki değişkenli YBMG bileşenleri bulunur.
 - 11: Üç değişkenli YBMG yaklaşımı oluşturulur.
 - 12: **for $i=1 \rightarrow l$ do** *l: Kontrast İyileştirme seviyelerinin sırası*
 - 13: Yaklaşımına iki ve üç değişkenli bileşenler eklenir.
 - 14: **end for**
 - 15: **end if**
-

3. YÜKSEK BOYUTLU MODEL GÖSTERİLİMİ İLE MEME KANSERİ TANISI

Bu çalışmada meme kanseri tanısı için IRMA (Image Retrieval in Medical Applications) veri tabanından mamogramlar seçilmiştir. Bu proje MIAS (The Mammographic Image Analysis Society Digital Mammogram Database) , DDSM (The Digital Database for Screening Mammography) , LLNL (the Lawrence Livermore National Laboratory) ve RWTH (Rheinisch-Westfälische Technische Hochschule) verilerini içermektedir. Bu projenin amacı, mamografileri tanı için kullanılabilir hale getirmek ve meme kanseri teşhisi için kullanılacak dijital mamografi görüntülerinden bir veritabanı oluşturmaktır.

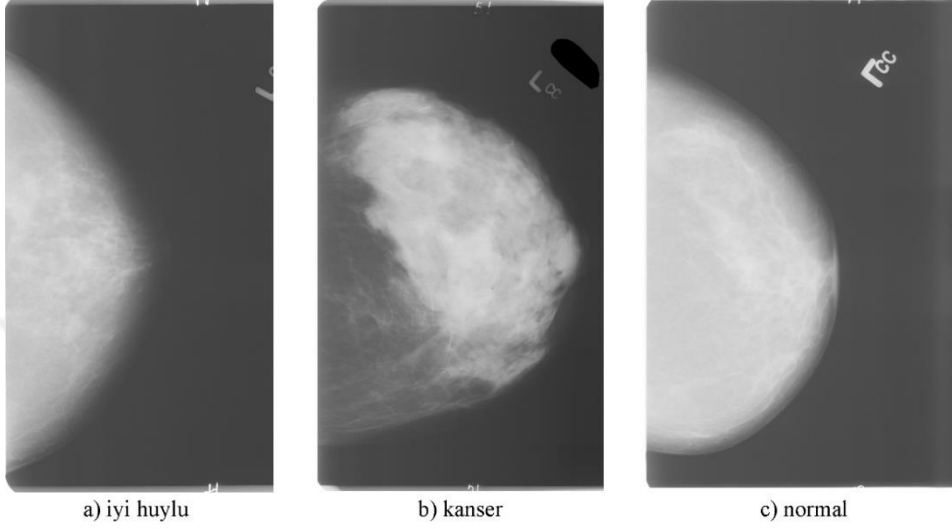
Meme kanseri tanısı, 225 tane sol meme mamogramı kullanılarak yapılmıştır. Sol memeye ait mamogramlar IRMA veri tabanında LEFT_CC.LJPEG etiketiyle isimlendirilmektedir. Bu etiket altında iyi huylu, normal ve kanser durumlarına ait mamogramlar bulunmaktadır. Burada gerçekleştirilecek olan tanı için her durumdan 75 tane mamogram seçilmiştir. Mamografi görüntüleri Çizelge 3.1'deki gibi gruplandırılmış ve etiketlenmiştir.

Çizelge 3.1 : Mamogram etiketleri.

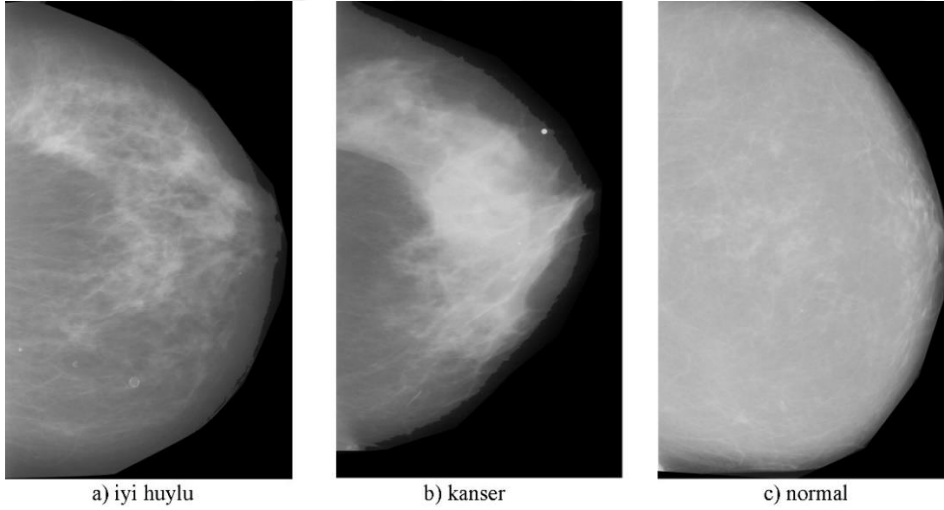
Normal	İyi Huylu	Kanser
0	1	2

Ele alınan 225 mamogramın %80'i eğitim kümesine ayrılmış kalan görüntüler ise, oluşturulan modelin test edilebilmesi için test kümesi olarak kullanılmıştır. Bu ayırım sırasında eğitim ve test kümeleri, görüntülerin iyi huylu, kanser ya da normal olmasına bakılmaksızın rastgele oluşturulmuştur. IRMA veri tabanından seçilen farklı durumlara ait işlenmemiş mamogramlar Şekil 2.12'de verilmektedir. Bu görüntüler, temizlenmemiş ve arka planında etiket bulunan, 10-30MB büyüklüğündeki görüntülerdir ve geliştirilen algoritmanın daha yüksek performans ile çalışabilmesi

için görüntülerin temizlenerek arka planının siyaha boyanması ön işleme gereksinim duyulmuştur. Bu işlem MATLAB (R2013a) programı yardımıyla yapılmış ve görüntüler 100-200KB civarına indirilmiştir. Ayrıca her biri farklı piksel sayısına sahip olan mamogramlar, 600x1000 piksel boyutuna getirilmiştir. Ön işlem yapıldıktan sonra oluşan farklı durumlara ait görüntüler Şekil 2.13'te verilmiştir.



Şekil 3.1 : Orijinal IRMA veritabanı görüntüleri.

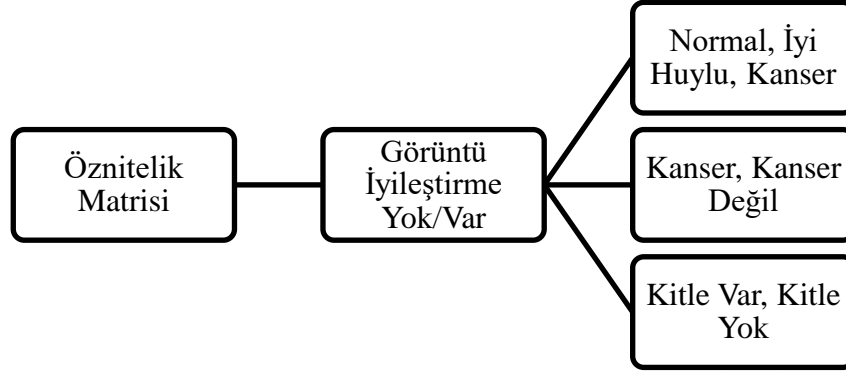


Şekil 3.2 : Önışlemeden geçirilmiş görüntüler.

Ön işlemden geçirilen ve eğitim kümesine atanan her bir mamogram öznitelik matrisi oluşturulmak üzere öncelikle YBMG bileşenlerine ayrılmalıdır. Bu amaçla denklem 2.15'te verilen f_0 , denklem 2.18'de verilen f_1, f_2 tek değişkenli ve denklem 2.19'da verilen f_{12}, f_{13}, f_{23} iki değişkenli YBMG bileşenleri hesaplanır. Tek değişkenli f_3

bileşenine mamogramların gri seviyeli olması nedeniyle ayırt edici bir özellik taşımadığı düşünüldüğünden algoritma içinde yer verilmemiştir.

Öznitelik matrisi oluşturulurken aşağıdaki gibi özetlenen üç farklı yol izlenmiştir.



Şekil 3.3 : Öznitelik matrisi oluşturulurken izlenen yol.

Öncelikle herhangi bir görüntü iyileştirme işlemi uygulanmamış veri kümesi kullanılmıştır. Daha sonra başarı skorlarını arttırabileceği düşüncesiyle Yöntem bölümünde anlatılan görüntü iyileştirme teknikleri uygulanmış ve yeni bir veri kümesi oluşturulmuştur. İki veri kümesi için de aşağıda anlatılan 3 farklı yol benzer şekilde uygulanmıştır.

Birinci yol; normal, iyi huylu ve kanser durumlarını içeren veri kümesi için her bir mamogramın YBMG bileşenleri hesaplanmıştır. Hesaplanan bileşenler, her bir yolda aynı olmak üzere 6 biçimde öznitelik matrisine dönüştürülmüştür.

Çizelge 3.2 : Altı biçimde öznitelik matrisi.

Birinci Biçim	İkinci Biçim	Üçüncü Biçim
Etiketler, f_0, f_1	Etiketler, f_0, f_1, f_2	Etiketler, f_0 , Frobenius Norm
Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
Etiketler, f_0 , Ortalama	Etiketler, f_0 , Medyan	Etiketler, f_0 , Frobenius Norm, Ortalama, Medyan

İkinci yol; kanser ve kanser olmayan durumları içeren Çizelge 3.3'teki gibi etiketlenen bir veri kümesi için YBMG bileşenleri hesaplanmıştır.

Çizelge 3.3 : Mamogram etiketleri.

Kanser	Kanser Değil
0	1

Bu bileşenlerden Çizelge 3.2'teki gibi öznitelik matrisleri elde edilmiş ve sonuçları incelenmiştir.

Üçüncü yol; kitle var ve kitle yok durumlarını içeren Çizelge 3.4'teki gibi etiketli mamaogramların YBMG bileşenleri hesaplanmıştır. Benzer şekilde Çizelge 3.2'deki öznitelik matrisleri oluşturulmuş ve başarı skorları elde edilmiştir.

Çizelge 3.4 : Mamogram etiketleri.

Kitle Var	Kitle Yok
0	1

Özetle, başarısı ölçülecek öznitelikler belirlenmiş ve uç uca eklenerek bir vektör oluşturulmuştur. Sonrasında her bir vektör satırlar haline getirilerek mamogram sayısı kadar satıra sahip bir matris elde edilmiştir. Bu matrisin ilk sütununa her biçim için mamogramaların etiket değerleri de eklenerek bir öznitelik matrisi oluşturulmuştur.

Burada mamogramaların durumlarını gösteren etiket değerleri Çizelge 3.2, Çizelge 3.3 ve Çizelge 3.4'de gösterildiği gibi sırasıyla normal durum için 0, iyi huylu durum için 1 ve kanser durumu için 2; kanser durumu için 0, kanser değil için 1 ve kitle var durumu için 0, kitle yok durumu için 1 olarak belirlenmiştir. Ön işlemde geçirilen mamaogramlar için, piksel boyutu, 600x1000 olduğundan öznitelik matrislerinin boyutu Çizelge 3.5'te özetlenmiştir. Burada YBMG bileşenleri f_0, f_1, f_2 sırasıyla 1, 1000, 600 boyutludur.

Çizelge 3.5 : Öznitelik matrislerinin boyutları.

Birinci Biçim	İkinci Biçim	Üçüncü Biçim
225x1002	225x1602	225x7
Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
225x7	225x7	225x17

Eđitim kümesine ait öznitelik matrisinin elde edilmesinden sonra aynı yollar izlenerek test kümesine ait mamogramların özniteliklerinin de çıkarılması gerekmektedir. Bu sayede test kümesindeki mamogramaların öznitelikleri eğitim kümesindeki mamogramların öznitelikleriyle gözetimli makine öğrenmesi teknikleri kullanılarak meme kanseri tanısı için tahmin başarısı ölçülmüştür. Bu hesaplama için MATLAB programı ile elde edilen öznitelikler Python çalışma ortamına aktarılmış, scikit-learn isimli özel bir kütüphane kullanılmıştır. Her bir yola ve her bir biçime ait sonuçlar Bulgular ve Karşılaştırma bölümde yer almaktadır.





4. BULGULAR VE KARŞILAŞTIRMA

Bu bölümde görüntü iyileştirme işlemi uygulanmış ve uygulanmamış dijital mamogramlara ait veriler için, üçüncü bölümde anlatılan öznelik elde etme işlemlerinden geçirilerek ve ikinci bölümde matematiksel arkaplanı verilmiş makine öğrenmesi yöntemleri kullanılarak test ve validasyon skorları elde edilmiştir. Test ve validasyon skorları karşılaştırılarak en yüksek sonuç alınan öznelik biçimi belirlenmiştir.

4.1 Görüntü iyileştirme işlemi yapılmamış veriler için skorlar

Mamogram etiketlerindeki üç durum için test ve validasyon skorlarını incelediğimiz dijital mamogram verisi, görüntü iyileştirme işlemi uygulanmış ve uygulanmamış şekilde ikiye ayrılmıştır. Bu etiketlerdeki üç durum sırasıyla; normal, iyi huylu ve kanser durumu, kanser ve kanser olmayan durumu ve kitle olan ve olmayan durumudur. Ayrı ayrı incelenen ve doğruluk skorları elde edilen veriler, aşağıda verilen tablolarda gösterilmektedir.

4.1.1 Normal, iyi huylu ve kanser durumları için skorlar

Üç farklı durumu içeren bu veri kümesi, 75 adet normal, 75 adet iyi huylu ve 75 adet kanser olarak etiketli toplam 225 mamogramdan oluşmaktadır. Bu veri kümesinin %20'si test %80'i eğitim kümesi olarak ayrılmıştır. IRMA projesi kapsamında toplanan ve bu çalışmada da kullanılan dijital mamogram görüntüleri, yeniden boyutlandırma, meme sınırlarından keserek görüntüde memeye odaklanma, arka planı siyaha boyama gibi işlemlerden geçmiş ve veri kümesi oluşturulmuştur. Öncelikle YBMG bileşenleri teker teker hesaplanmış ve üçüncü bölümde gösterilen 6 biçimdeki öznelik matrisleri elde edilmiştir. Bu öznelikler yardımıyla hesaplanmış test ve validasyon skorları Çizelge 4.1 ve Çizelge 4.2'de elde gösterilmektedir. Çizelgelerdeki değerlerin rahat görülebilmesi için en yüksek değerler kalın font ile gösterilmiştir.

Çizelge 4.1 : Normal, iyi huylu ve kanser durumları için test skoru.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.38	0.62	0.44	0.40	0.31	0.38
LDA	0.40	0.49	0.49	0.62	0.44	0.62
k-NN	0.44	0.47	0.31	0.42	0.51	0.31
CART	0.47	0.40	0.38	0.49	0.42	0.42
SVM	0.22	0.73	0.36	0.29	0.36	0.27
RF	0.46	0.51	0.42	0.51	0.62	0.51

Test skorları için makine öğrenmesi yöntemleri karşılaştırıldığında, en yüksek başarı ikinci öznitelik biçiminde destek vektör makineleri yönteminde %73 olarak elde edilmiştir.

Diğer veri kümelerinde olduğu gibi bu veri kümesi için de k katlamalı çapraz doğrulama yöntemi kullanılmıştır. 6 farklı öznitelik matrisi için ayrı ayrı k=10 katlı çapraz doğrulama yöntemiyle veri kümesi eğitilip test edilmiştir. Bu testte eğitim, her kat bir kez test kümesi olacak şekilde yapılır. Elde edilen sonuçların ortalaması, Çizelge 3.7’de gösterilen validasyon skoruna karşılık gelir.

Çizelge 4.2: Normal, iyi huylu ve kanser durumları için validasyon skoru.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.42±0.18	0.48±0.20	0.43±0.20	0.39±0.19	0.44±0.21	0.41±0.23
LDA	0.40±0.19	0.39±0.20	0.43±0.20	0.49±0.19	0.50±0.24	0.59±0.21
k-NN	0.43±0.19	0.42±0.21	0.32±0.21	0.44±0.23	0.51±0.20	0.29±0.15
CART	0.38±0.26	0.49±0.23	0.32±0.19	0.38±0.22	0.39±0.18	0.42±0.23
SVM	0.22±0.16	0.49±0.18	0.25±0.17	0.29±0.18	0.36±0.20	0.36±0.21
RF	0.46±0.18	0.43±0.20	0.38±0.18	0.39±0.20	0.44±0.20	0.45±0.20

Validasyon ve test skorları incelendiğinde bazı yöntemlerde aşırı öğrenme meydana geldiği görülmektedir. Birinci ve ikinci biçim özniteliklerin sayısı, veri kümesi boyutuna göre daha büyük olduğu için aşırı öğrenme, karşılaşılabilecek muhtemel problemlerdendir.

Birinci, beşinci ve altıncı biçim öznitelite LogReg, ikinci biçim öznitelik CART, üçüncü ve dördüncü biçim öznitelikte k-NN, beşinci biçim öznitelikte LDA, ve altıncı biçim öznitelikte SVM yönteminde, aşırı öğrenme problemi yaşandığından bu skorlar ihmal edilmiştir. Sonuç olarak en yüksek validasyon başarı altıncı biçim öznitelik LDA yöntemiyle 0.59 ± 0.21 olduğu görülmektedir.

4.1.2 Kanser ve kanser olmayan durumlar için skorlar

Herhangi bir görüntü iyileştirme işlemi uygulanmamış ve kanser, kanser değil şeklinde etiketlenmiş veri kümesi kullanılmıştır. Bu veri kümesi 75 adet kanser, 150 adet kanser olmayan şekilde etiketlenmiş toplam 225 adet mamogram içermektedir. Python çalışma ortamında veri kümesi rastgele olarak %20'si test, %80'i eğitim kümesi şeklinde ayrılmıştır. İkinci bölümde ayrıntılı bir şekilde nasıl elde edildiği anlatılan YBMG bileşenleri ve makine öğrenmesi yöntemleri kullanılarak Çizelge 4.3'deki test kümesine ait doğruluk skorları elde edilmiştir.

Çizelge 4.3: Kanser ve kanser olmayan durumlar için test skoru.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.69	0.69	0.67	0.69	0.76	0.69
LDA	0.62	0.60	0.67	0.71	0.80	0.80
k-NN	0.60	0.67	0.58	0.71	0.71	0.58
CART	0.69	0.71	0.67	0.62	0.67	0.58
SVM	0.71	0.71	0.67	0.69	0.76	0.69
RF	0.71	0.73	0.64	0.64	0.78	0.60

Karşılaştırılan yöntemler arasından en yüksek başarıya beşinci ve altıncı biçimdeki özniteliklerle LDA yönteminde %80 test başarıları olarak ulaşılmıştır.

Çizelge 4.4: Kanser ve kanser olmayan durumlar için validasyon skoru.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.67±0.21	0.68±0.19	0.66±0.17	0.66±0.20	0.64±0.18	0.66±0.17
LDA	0.62±0.18	0.59±0.22	0.64±0.17	0.77±0.18	0.72±0.15	0.72±0.17
k-NN	0.70±0.15	0.62±0.18	0.59±0.18	0.63±0.18	0.66±0.15	0.55±0.21
CART	0.62±0.21	0.69±0.19	0.52±0.22	0.52±0.20	0.61±0.15	0.53±0.18
SVM	0.66±0.16	0.66±0.20	0.64±0.18	0.66±0.19	0.64±0.15	0.66±0.20
RF	0.64±0.19	0.68±0.19	0.62±0.22	0.53±0.17	0.63±0.17	0.62±0.20

Bu karşılaştırmada görüyoruz ki en yüksek başarı skoru 0.77 ± 0.18 ile dördüncü biçim öznitelik kullanılarak LDA yönteminden elde edilmiştir. Fakat test ve validasyon skorları karşılaştırıldığında bazı makine öğrenmesi yöntemlerinde aşırı öğrenme (overfitting) problemi yaşanmıştır. Öznitelik sayısı ile veri kümesi büyüklüğünün dengesizliğinden yaşanabilen bu problem yüzünden birinci ve üçüncü biçim öznitelikteki k-NN ve dördüncü biçim öznitelikteki LDA skoru ihmal edilecektir. Bu durumda en yüksek validasyon skoru altıncı biçim öznitelikle LDA yönteminde 0.72 ± 0.17 olarak elde edilmiştir.

4.1.3 Kitle olan ve olmayan durumlar için skorlar

Elimizdeki veri kümesi, herhangi görüntü iyileştirme işlemi uygulanmamış, kitle olan ve olmayan şeklinde ayrılmış biçimdedir. Bu veri kümesi 150 adet kitle olan, 75 adet kitle olmayan dijital mamogramlardan oluşmaktadır. Burada da veri kümesi rastgele biçimde %20'ye %80 olacak şekilde test ve eğitim kümesi olarak ayrılmıştır. Çizelge 4.5'de bu çalışma için kullandığımız makine öğrenmesi teknikleri ile elde edilmiş test skorları yer almaktadır.

Çizelge 4.5: Kitle olan ve olmayan durumlar için test skoru.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.69	0.64	0.67	0.73	0.69	0.73
LDA	0.69	0.68	0.60	0.76	0.82	0.80
k-NN	0.67	0.73	0.56	0.62	0.60	0.62
CART	0.67	0.71	0.67	0.67	0.64	0.62
SVM	0.73	0.71	0.67	0.73	0.69	0.78
RF	0.69	0.73	0.67	0.69	0.71	0.69

Dijital mamogramlarda kitle olup olmadığını test ederken elde ettiğimiz en yüksek skor, beşinci biçim öznelik ile LDA yönteminde %82'dir.

Çizelge 4.6: Kitle olan ve olmayan durumlar için validasyon skoru.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.63±0.16	0.69±0.20	0.66±0.20	0.65±0.18	0.66±0.17	0.66±0.19
LDA	0.61±0.20	0.63±0.20	0.66±0.17	0.61±0.20	0.68±0.15	0.73±0.17
k-NN	0.62±0.18	0.64±0.18	0.55±0.21	0.62±0.14	0.63±0.18	0.58±0.21
CART	0.63±0.19	0.61±0.17	0.57±0.17	0.51±0.21	0.59±0.23	0.60±0.19
SVM	0.65±0.19	0.66±0.20	0.65±0.18	0.65±0.21	0.66±0.17	0.62±0.20
RF	0.64±0.18	0.64±0.18	0.64±0.16	0.51±0.19	0.64±0.17	0.69±0.21

Bu karşılaştırmada en yüksek validasyon skoru altıncı biçim ile elde edilen öznelik kullanılarak LDA yönteminde 0.73 ± 0.17 'dir. Test ve validasyon skorları karşılaştırıldığında ise validasyon skorunun, test skorundan büyük olması aşırı öğrenmeye işaret etmektedir. Bu yüzden üçüncü ve beşinci biçim k-NN, ikinci biçim LogReg sonuçları ihmal edilecektir. Bu koşullar altında en yüksek validasyon başarısı değişmemiştir.

4.2 Görüntü iyileştirme işlemi yapılmış veriler için skorlar

Dijital mamogramlara kontrast iyileştirme işlemleri olarak sırasıyla Kontrast Sınırlı Uyarlamalı Histogram Eşitleme (CLAHE) ve Yüksek Boyutlu Model Gösterilim (YBMG) yöntemleri uygulanmıştır. Bu işlemlerin matematiksel arkaplanı ikinci bölümünde anlatılmıştır. Bu bölümde görüntü iyileştirme işlemi uygulandıktan sonra elde edilen skorlar verilmiştir.

4.2.1 CLAHE uygulanmış mamogramlar

Dijital mamogram görüntülerine kontrast iyileştirme tekniklerinden biri olan CLAHE yöntemi uygulanmış ve üç durum için skorlar altı farklı makine öğrenmesi yöntemi kullanılarak elde edilmiştir. Test ve validasyon skorlarını içeren tablolarda altı biçim öznitelikte elde edilen skorlar verilmiştir.

4.2.1.1 Normal, iyi huylu ve kanser durumları için skorlar

Görüntü iyileştirme işlemi uygulanmış normal, iyi huylu ve kanser durumlarını içeren veri kümesi kullanılarak altı biçimde öznitelik matrisleri elde edilmiştir. Python çalışma ortamında altı farklı gözetimli makine öğrenmesi yöntemi kullanılarak sınıflandırma başarısı ölçülmüştür.

Çizelge 4.7: Normal, iyi huylu ve kanser durumları için test skoru.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.56	0.51	0.33	0.40	0.49	0.42
LDA	0.42	0.40	0.60	0.58	0.64	0.64
k-NN	0.38	0.49	0.36	0.42	0.58	0.36
CART	0.44	0.38	0.37	0.47	0.53	0.42
SVM	0.31	0.27	0.33	0.29	0.42	0.31
RF	0.62	0.42	0.44	0.42	0.42	0.49

En yüksek test başarısı, beşinci ve altıncı biçim öznitelik matrisleri kullanılarak LDA yöntemiyle %64 başarı elde edilmiştir.

Çizelge 4.8: Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırılması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
SVM - %73	LDA - %64

Görüntü iyileştirme işlemi uygulamadan aldığımız en yüksek skorla görüntü iyileştirme işlemi yaptığımız mamogram skorları karşılaştırmış ve Çizelge 4.8’de gösterilmiştir. Görüldüğü üzere görüntü iyileştirme işlemi yapılmış dijital mamogramlardan elde edilen en yüksek başarı, görüntü iyileştirme uygulanmamış görüntüler kullanılarak elde edilen skorun altında kalmıştır.

Çizelge 4.9: Normal, iyi huylu ve kanser durumları için validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.40±0.15	0.48±0.21	0.37±0.17	0.39±0.22	0.41±0.21	0.34±0.20
LDA	0.40±0.19	0.46±0.17	0.50±0.21	0.57±0.21	0.58±0.18	0.62±0.20
k-NN	0.36±0.16	0.40±0.22	0.41±0.18	0.36±0.17	0.42±0.25	0.36±0.19
CART	0.46±0.16	0.43±0.19	0.37±0.19	0.43±0.19	0.49±0.19	0.44±0.22
SVM	0.17±0.15	0.31±0.19	0.32±0.17	0.29±0.17	0.27±0.17	0.27±0.17
RF	0.39±0.18	0.42±0.22	0.43±0.20	0.41±0.21	0.46±0.20	0.46±0.20

Validasyon ve test skorları incelendiğinde birinci biçim öznitelik için CART, ikinci biçim öznitelik için CART, SVM, üçüncü biçim öznitelik için LogReg, k-NN, dördüncü biçim öznitelik için LogReg, beşinci biçim öznitelik için RF ve altıncı biçim öznitelik için CART yönteminde aşırı öğrenme sorunu yaşanmıştır. Bu yüzden bu skorlar karşılaştırma yapılırken ihmal edilmiştir. Bu ihmal göz önünde bulundurularak en yüksek validasyon başarısı altıncı biçim öznitelik matrisi kullanılarak LDA yöntemiyle 0.62 ± 0.20 olarak elde edilmiştir.

Çizelge 4.10: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
LDA - 0.59 ± 0.21	LDA - 0.62 ± 0.20

Görüntü iyileştirme işlemi yapılmış ve yapılmamış verilerden elde edilen skorlar karşılaştırıldığında görüldüğü üzere altıncı biçim öznelik matrisi ile LDA yöntemi kullanılarak en yüksek başarı elde edilmiştir. Validasyon skorlarına bakıldığında CLAHE yönteminin başarıyı anlamlı bir farkla olumlu bir şekilde etkilediği görülememiştir.

4.2.1.2 Kanser ve kanser olmayan durumlar için skorlar

İki sınıfa ayrılmış mamogram görüntülerinin altı farklı biçimdeki öznelik matrisleri kullanılarak hesaplanmış test ve validasyon skorları aşağıdaki gibidir. Çizelge 4.11’de görüldüğü gibi en yüksek test skoru üçüncü ve dördüncü biçim öznelik kullanılarak LDA yöntemiyle elde edilmiştir.

Çizelge 4.11: Kanser/kanser olmayan veri kümesi Clahe test skorları.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.73	0.73	0.64	0.67	0.67	0.67
LDA	0.64	0.60	0.78	0.78	0.67	0.76
k-NN	0.58	0.60	0.62	0.58	0.71	0.53
CART	0.58	0.64	0.62	0.58	0.62	0.64
SVM	0.67	0.67	0.67	0.67	0.67	0.69
RF	0.64	0.60	0.62	0.62	0.71	0.60

Çizelge 4.11 ve Çizelge 4.12 karşılaştırıldığında bazı yöntemlerde aşırı öğrenme sorunuyla karşılaşıldığı görülmüştür. Birinci biçim öznelikte k-NN ve CART, ikinci biçim öznelikte k-NN ve RF, üçüncü biçim öznelikte LogReg, dördüncü biçim öznelikte RF ve altıncı biçim öznelikte CART ve RF yöntemleri kullanıldığında validasyon skorunun test skorunun altında kaldığı açıktır. Bu değerler karşılaştırma yapılırken ihmal edilmiştir.

Çizelge 4.12: Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
LDA - 0.80	LDA - 0.78

Kanser ve kanser olmayan durumlarının görüntü iyileştirme işlemi uygulanmış ve uygulanmamış hallerdeki en yüksek başarı skorları Çizelge 4.12’de gösterilmiştir. Görüntü iyileştirme uygulanmamış verilerin beşinci ve altıncı biçim özniteliklerle LDA yöntemi kullanılarak elde edilmiş skor %80’dir. CLAHE yöntemiyle kontrastı iyileştirilmiş görüntülerden elde edilen veriler kullanılarak elde edilen en yüksek skor üçüncü ve dördüncü biçim öznitelikte LDA yöntemi kullanılarak %78 şeklinde elde edilmiştir. CLAHE yönteminin başarıya olumlu bir etkisinin olmadığı görülmüştür.

Çizelge 4.13: Kanser ve kanser olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.68±0.15	0.66±0.22	0.67±0.16	0.67±0.20	0.66±0.14	0.67±0.17
LDA	0.59±0.17	0.58±0.20	0.72±0.17	0.72±0.19	0.66±0.19	0.75±0.17
k-NN	0.62±0.20	0.61±0.21	0.54±0.18	0.47±0.22	0.54±0.17	0.53±0.21
CART	0.59±0.19	0.58±0.17	0.57±0.18	0.56±0.20	0.54±0.22	0.53±0.19
SVM	0.67±0.16	0.67±0.16	0.67±0.16	0.67±0.19	0.67±0.17	0.66±0.19
RF	0.63±0.20	0.64±0.21	0.61±0.23	0.63±0.17	0.58±0.20	0.64±0.21

Çizelge 4.13’de CLAHE yöntemiyle kontrastı iyileştirilmiş mammogram görüntülerinden elde edilmiş özniteliklerle hesaplanmış validasyon skorları karşılaştırılmıştır. En yüksek validasyon başarısı altıncı biçim öznitelikle LDA yöntemi kullanılarak elde edilmiştir.

Çizelge 4.14: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
LDA - 0.72±0.17	LDA - 0.75±0.17

CLAHE yönteminin validasyon skorları üzerindeki etkisine bakıldığında görüntü iyileştirme işlemi uygulanmış ve uygulanmamış verilerde altıncı biçim öznitelikle LDA yöntemiyle en yüksek başarı skoru elde edildiği görülmüştür. CLAHE yönteminin skoru anlamlı bir biçimde etkilediği görülmemiştir.

4.2.1.3 Kitle olan ve olmayan durumlar için skorlar

Mamogram görüntülerinde herhangi bir kitle olup olmadığı sorgulanan bu durumda CLAHE yöntemi uygulanmış meme görüntüleri kullanılmıştır. Diğer tüm durumlarda olduğu gibi altı çeşit öznitelik kullanılmıştır. Bu öznitelikler kullanılarak 6 çeşit gözetimli makine öğrenmesi tekniği ile test ve validasyon skorları elde edilmiştir.

Çizelge 4.15: Kitle olan ve olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş test skorları.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.80	0.67	0.69	0.78	0.71	0.80
LDA	0.53	0.53	0.73	0.82	0.71	0.76
k-NN	0.64	0.69	0.67	0.53	0.78	0.67
CART	0.62	0.76	0.62	0.60	0.64	0.62
SVM	0.76	0.67	0.71	0.78	0.71	0.73
RF	0.64	0.64	0.60	0.64	0.67	0.80

Çizelge 4.15 incelendiğinde en yüksek test skorunun dördüncü biçim öznitelikle LDA yöntemi kullanılarak elde edildiği görülmüştür.

Çizelge 4.16: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
LDA - %82	LDA - %82

Görüntü iyileştirme yapılmış ve yapılmamış mamogramlardan elde edilmiş özniteliklerle hesaplanmış başarı skorları Çizelge 4.17’de verilmiştir. CLAHE’siz mamogramlar beşinci öznitelikle LDA yöntemi kullanılarak %82 başarıya ulaşılmıştır. Aynı başarı skoru CLAHE’li mamogramlar dördüncü öznitelikle LDA yöntemiyle elde edilmiştir.

Çizelge 4.17: Kitle olan ve olmayan durumlar için CLAHE uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.66±0.15	0.68±0.22	0.66±0.19	0.64±0.17	0.66±0.20	0.66±0.14
LDA	0.57±0.19	0.63±0.20	0.62±0.21	0.71±0.19	0.63±0.19	0.76±0.17
k-NN	0.62±0.17	0.61±0.16	0.60±0.20	0.49±0.19	0.60±0.22	0.66±0.19
CART	0.52±0.20	0.62±0.17	0.60±0.23	0.58±0.19	0.56±0.20	0.60±0.16
SVM	0.64±0.15	0.67±0.16	0.64±0.17	0.64±0.19	0.66±0.15	0.66±0.19
RF	0.60±0.21	0.62±0.20	0.69±0.17	0.57±0.20	0.67±0.19	0.66±0.22

CLAHE’li mamogramların validasyon skorları incelendiğinde en yüksek başarımın 0.76 ± 0.11 olduğu görülür. Fakat Çizelge 4.17 ve Çizelge 4.18 incelendiğinde birinci öznitelikle LDA, ikinci öznitelikle LR ve LDA ve üçüncü öznitelikle RF yöntemlerinde aşırı öğrenme meydana gelmiştir. Bu skorlar başarı karşılaştırmada dikkate alınmamıştır.

Çizelge 4.18: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- CLAHE
LDA - 0.73 ± 0.17	LDA - 0.76 ± 0.17

Çizelge 4.18’te CLAHE’nin validasyon skorları üstündeki etki karşılaştırılmıştır. Görüntü iyileştirmenin başarıya anlamlı bir fark katmadığı görülmüştür.

4.2.2 YBMG uygulanmış mamogramlar

Bu bölümde normal, iyi huylu ve kanser, kanser olan ve kanser olmayan, kitle olan ve kitle olmayan durumları içeren altı biçim öznitelik matrisi kullanarak ve altı adet gözetimli makine öğrenmesi yöntemiyle mamogramların test ve validasyon skorları incelenmiştir.

4.2.2.1 Normal, iyi huylu ve kanser durumları için skorlar

Mamogramların normal, iyi huylu ve kanser durumlarını içeren veri kümesinden, YBMG ile kontrast iyileştirilip altı farklı öznitelikle altı adet makine öğrenmesi yöntemi kullanılarak başarı skorları elde edilmiştir.

Çizelge 4.19: Normal, iyi huylu ve kanser durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş test skorları.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.40	0.49	0.31	0.38	0.31	0.53
LDA	0.51	0.53	0.47	0.31	0.51	0.60
k-NN	0.29	0.49	0.29	0.44	0.59	0.33
CART	0.47	0.33	0.42	0.47	0.44	0.44
SVM	0.22	0.29	0.26	0.31	0.27	0.33
RF	0.40	0.53	0.42	0.38	0.49	0.51

En yüksek test skoru, altıncı biçim öznitelikle LDA yöntemi kullanılarak %60 olarak elde edilmiştir.

Çizelge 4.20: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- YBMG
SVM - %73	LDA - %60

YBMG yöntemi uygulanmış ve uygulanmamış veriler karşılaştırıldığında kontrast iyileştirmenin olumlu bir etkisinin olmadığı görülmüştür.

Çizelge 4.21: Normal, iyi huylu ve kanser durumları için YBMG uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.44±0.18	0.48±0.22	0.42±0.20	0.27±0.19	0.38±0.20	0.43±0.16
LDA	0.45±0.20	0.44±0.21	0.52±0.20	0.45±0.16	0.45±0.18	0.48±0.18
k-NN	0.43±0.21	0.43±0.24	0.34±0.16	0.44±0.21	0.47±0.18	0.32±0.18
CART	0.43±0.20	0.42±0.24	0.37±0.16	0.41±0.21	0.49±0.17	0.44±0.19
SVM	0.36±0.18	0.28±0.16	0.25±0.19	0.22±0.12	0.42±0.18	0.31±0.14
RF	0.46±0.18	0.49±0.15	0.42±0.20	0.41±0.19	0.40±0.23	0.45±0.19

Çizelge 4.21’de normal, iyi huylu ve kanser durumları için YBMG uygulanmış veriler kullanılarak elde edilmiş altı çeşit öznelikle ilgili altı adet gözetimli makine öğrenmesi yönteminin performansı karşılaştırılmıştır. En yüksek validasyon skoru altıncı biçim öznelikle LDA kullanılarak 0.48 ± 0.18 olarak elde edilmiştir. Fakat Çizelge 4.19 ve Çizelge 4.21 incelendiğinde birinci öznelikle ilgili LogReg, k-NN, SVM yöntemlerinde, ikinci biçim öznelikle ilgili CART yönteminde, üçüncü biçim öznelikle ilgili LogReg, LDA, k-NN ve SVM yöntemlerinde, dördüncü biçim öznelikle ilgili LDA ve RF yöntemlerinde, beşinci biçim öznelikle ilgili LogReg, CART ve SVM yöntemlerinde aşırı öğrenme sorunuyla karşılaşmıştır. Bu skorlar karşılaştırma yapılırken ihmal edilmiştir.

Çizelge 4.22: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var-YBMG
LDA - 0.59 ± 0.21	LDA - 0.48 ± 0.18

Çizelge 4.22’de YBMG yönteminin uygulanıp uygulanmamasının etkisi gösterilmiştir. Validasyon skorları incelendiğinde YBMG ile kontrast iyileştirme işleminin validasyon skoruna olumlu bir katkısının olmadığı gözlemlenmiştir.

4.2.2.2 Kanser ve kanser olmayan durumlar için skorlar

YBMG yöntemi uygulanmış kanser ve kanser olmayan durumları içeren mamogramların altı biçim öznitelik matrisleri oluşturulmuştur. Gözetimli makine öğrenmesi teknikleri uygulandıktan sonra test ve validasyon skorlarına ait sonuçlar Çizelge 4.23 ve Çizelge 4.25’de verilmiştir.

Çizelge 4.23: Kanser ve kanser olmayan durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş test skorları.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.68	0.69	0.71	0.71	0.76	0.76
LDA	0.67	0.71	0.80	0.71	0.77	0.89
k-NN	0.67	0.67	0.62	0.60	0.71	0.69
CART	0.75	0.69	0.67	0.59	0.71	0.78
SVM	0.73	0.76	0.71	0.71	0.73	0.82
RF	0.71	0.71	0.64	0.62	0.67	0.71

YBMG uygulanmış veri için en yüksek test skoruna altıncı biçim öznitelikle LDA yöntemiyle ulaşılmıştır.

Çizelge 4.24: Görüntü iyileştirme etkisinin test sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- YBMG
LDA - %80	LDA - %89

YBMG ile kontrast iyileştirme işlemi uygulandığında ve uygulanmadığında kanser ve kanser olmayan durumlar için elde edilen test skorları Çizelge 4.24’de özetlenmiştir. Görüntü iyileştirme işlemi uygulanmamış mamogramlarda en yüksek başarı beşinci ve altıncı biçim öznitelikle LDA yönteminde %80 iken YBMG ile görüntü iyileştirme işlemi uygulanmış mamogramlar için altıncı biçim öznitelikle yine LDA yönteminde %89 başarı elde edilmiştir. YBMG yönteminin test skorunu iyileştirdiği görülmüştür.

Çizelge 4.25: Kanser ve kanser olmayan durumlar için YBMG uygulanmış mamogram görüntülerinden elde edilmiş validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.68±0.19	0.67±0.19	0.64±0.20	0.66±0.19	0.69±0.17	0.64±0.20
LDA	0.57±0.17	0.62±0.16	0.67±0.22	0.65±0.19	0.73±0.18	0.64±0.20
k-NN	0.64±0.19	0.59±0.23	0.56±0.17	0.64±0.20	0.64±0.16	0.57±0.23
CART	0.60±0.20	0.61±0.20	0.53±0.19	0.59±0.19	0.62±0.20	0.57±0.18
SVM	0.65±0.18	0.64±0.15	0.59±0.19	0.66±0.17	0.66±0.21	0.59±0.23
RF	0.67±0.18	0.63±0.17	0.60±0.24	0.61±0.20	0.66±0.17	0.62±0.19

Kanser ve kanser olmayan durumları içeren veri kümesinde validasyon skorları incelendiğinde beşinci öznitelik için LDA yönteminde en yüksek skorun elde edildiği görülmüştür.

Çizelge 4.26: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- YBMG
LDA - 0.72±0.17	LDA - 0.73±0.18

Görüntü iyileştirme işleminin etkisinin özetlendiği Çizelge 4.26'da, YBMG'nin başarıyı anlamlı bir oranda arttırdığı gözlenmemiştir.

4.2.2.3 Kitle olan ve olmayan durumlar için skorlar

Kontrast iyileştirme işlemi olarak YBMG uygulanmış mamogram görüntü verileri kullanılarak hesaplanmış altı biçim öznitelik matrisleri, çeşitli gözetimli makine öğrenmesi teknikleri kullanılmış ve mamogramlarda kitle olup olmadığı sonuçlarına ait doğru tahmin etme skorları Çizelge 4.27 ve Çizelge 4.29'da verilmiştir.

Çizelge 4.27: Kitle olan ve kitle olmayan durumlar için YBMG uygulanmış mamogram görüntülerinden elde edilmiş test skorları.

Yöntem	Test Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.73	0.76	0.67	0.67	0.80	0.80
LDA	0.73	0.58	0.71	0.69	0.87	0.80
k-NN	0.76	0.64	0.62	0.67	0.69	0.58
CART	0.76	0.69	0.69	0.64	0.49	0.76
SVM	0.69	0.71	0.62	0.67	0.71	0.73
RF	0.73	0.78	0.66	0.67	0.67	0.73

Mamogramlarda kitle olan ve olmayan durumları için elde edilen test skorunda en yüksek başarı beşinci biçim öznitelikle LDA yöntemi kullanılarak %87 olarak elde edilmiştir.

Çizelge 4.28: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var- YBMG
LDA - %82	LDA - %87

YBMG işlemi uygulanmış ve uygulanmamış mamogramlardan elde edilen özniteliklerle yapılan karşılaştırmada YBMG'nin skoru olumlu yönde etkilediği görülmüştür.

Çizelge 4.29: Kitle olan ve kitle olmayan durumlar için YBMG uygulanmış mammogram görüntülerinden elde edilmiş validasyon skorları.

Yöntem	Validasyon Skoru					
	Birinci Biçim	İkinci Biçim	Üçüncü Biçim	Dördüncü Biçim	Beşinci Biçim	Altıncı Biçim
LogReg	0.66±0.21	0.64±0.17	0.67±0.20	0.67±0.11	0.69±0.17	0.67±0.19
LDA	0.63±0.16	0.58±0.23	0.66±0.17	0.66±0.21	0.69±0.15	0.69±0.19
k-NN	0.57±0.20	0.64±0.20	0.61±0.18	0.63±0.13	0.64±0.16	0.58±0.21
CART	0.55±0.22	0.65±0.18	0.56±0.20	0.62±0.28	0.64±0.17	0.60±0.20
SVM	0.66±0.20	0.66±0.19	0.62±0.19	0.67±0.16	0.66±0.14	0.61±0.19
RF	0.66±0.22	0.66±0.18	0.61±0.17	0.63±0.18	0.67±0.19	0.61±0.22

Aynı durum için validasyon skorları incelendiğinde beşinci biçim özellikle LDA yöntemiyle elde edilen en yüksek skor 0.69 ± 0.15 olmuştur.

Çizelge 4.30: Görüntü iyileştirme etkisinin validasyon sonuçlarıyla karşılaştırması.

Görüntü iyileştirme yok	Görüntü iyileştirme var-YBMG
LDA - 0.73 ± 0.17	LDA - 0.69 ± 0.15

Çizelge 4.30 incelendiğinde YBMG işlemi uygulanmış mamogramlarda validasyon başarısının düştüğü gözlenmiştir.

4.2.3 CLAHE ve YBMG skorlarının karşılaştırması

Mamogramlara kontrast iyileştirmek için uygulanan Kontrast Sınırlı Uyarlamalı Histogram Eşitleme ve Yüksek Boyutlu Model Gösterilimi yöntemlerine ait test ve validasyon skorları karşılaştırılmıştır.

Çizelge 4.31: CLAHE ve YBMG test skorlarının karşılaştırılması.

	Iyi huylu, kanser, normal durumu	Kanser var, kanser yok durumu	Kitle var, kitle yok durumu
CLAHE	LDA - %64	LDA - %78	LDA - %82
YBMG	LDA - %60	LDA - %89	LDA - %87

Çizelge 4.31’de gösterildiği üzere tüm durumlarda en yüksek test başarısı LDA yöntemi kullanılarak elde edilmiştir. Ayrıca normal, iyi huylu ve kanser durumları için sırasıyla en yüksek CLAHE başarısı 5.ve 6. biçim öznitelikle kullanılarak %64, YBMG başarısı 6. biçim öznitelikle %60 olarak elde edilmiştir. Kanser ve kanser olmayan durum için en yüksek CLAHE başarısı 3 ve 4. biçim öznitelikle %78 şeklinde elde edilmiştir. Buna karşılık 6.biçim öznitelikle %89 olarak YBMG skoruna ulaşılmıştır. Son olarak kitle olan ve olmayan durum için CLAHE 4.biçim öznitelikle en yüksek %82 skoruna ulaşırken YBMG 5.biçim öznitelikle %87’dir. Sonuç olarak kanser ve kanser olmayan durum ve kitle olan ve olmayan durum için CLAHE yöntemine kıyasla YBMG yöntemi daha yüksek sonuç vermiştir.

Çizelge 4.32: CLAHE ve YBMG validasyon skorlarının karşılaştırılması.

	Iyi huylu, kanser, normal durumu	Kanser var, kanser yok durumu	Kitle var, kitle yok durumu
CLAHE	LDA - %62±0.20	LDA - %75±0.17	LDA - %76±0.17
YBMG	LDA - %48±0.18	LDA - %73±0.18	LDA - %69±0.15

Çizelge 4.32’de ise en yüksek validasyon skorları verilmiştir. Bu skorlar karşılaştırıldığında ise CLAHE yönteminin, YBMG yönteminden daha yüksek validasyon skorları sağladığı görülmüştür. Aradaki farkın anlamlılığını ölçmek için t testi kullanılmıştır.

Normal, iyi huylu, kanser veri kümesi için CLAHE ve YBMG yöntemlerinin arasındaki farkın anlamlılığı için t testi uygulandığında $p = 0.001 < 0.05$ olduğu görülmüş ve iki yöntem arasında anlamlı istatistiksel bir fark elde edilmiştir.

Kanser olan ve olmayan veri kümesi için $p = 0.6598 > 0.05$, kitle olan ve olmayan veri kümesi için $p = 0.0962 > 0.05$ olarak elde edilmiş ve iki yöntem arasında istatistiksel olarak anlamlı bir farklılık saptanmamıştır.

5. SONUÇ VE ÖNERİLER

Bu tez çalışmasında, Yüksek Boyutlu Model Gösterilimi yöntemi dijital mamogramlara uygulanmış ve çeşitli öznitelikler elde edilerek ilk kez meme kanseri tanısı için kullanılmıştır. IRMA veri kümesinden alınan mamogramlar, çeşitli ön hazırlık süreçleriyle öznitelik çıkarılacak meme görüntüleri haline getirilmiştir. Bu çalışmada, meme kanseri tanısına yönelik 3 farklı durum incelenmiştir. Birinci durum normal, iyi huylu, kanser, ikinci durum kanser olan ve olmayan ve üçüncü durum ise kitle olan ve olmayan meme görüntülerinden veri kümeleri oluşturulmuştur. Bu veri kümeleri için toplam 6 farklı öznitelik matrisi oluşturulmuştur ve Çizelge 2.2’de gösterilmiştir. Oluşturulan öznitelik matrisleri ile Python çalışma ortamında gözetimli makine öğrenmesi teknikleri kullanılarak sonuçlar elde edilmiştir.

Meme görüntülerine, performansı olumlu etkileyebileceği düşüncesiyle kontrast iyileştirme işlemleri uygulanmıştır. Kontrast Sınırlı Uyarlamalı Histogram Eşitlemesi ve Yüksek Boyutlu Model Gösterilimi yöntemi kullanılarak mamogramlarda görüntü iyileştirme işlemi uygulanmış ve sonuçlar karşılaştırılmıştır.

Herhangi bir görüntü iyileştirme işlemi uygulanmamış veri kümesinde normal, iyi huylu ve kanser durumlarını ve YBMG’nin f_0, f_1, f_2 bileşenlerini içeren öznitelik matrisi ile en fazla %73 test başarısı elde edilmiştir. Kanser olan ve olmayan durumlar kullanılarak YBMG bileşenlerinin ortalaması ve frebenius norm, ortalama ve medyanına karşılık gelen öznitelik matrisi yardımıyla en yüksek test skoru %80’dir. Kitle olan ve olmayan durumlara bakıldığında YBMG bileşenlerinin medyanına karşılık gelen 5.bıçım öznitelik ile en yüksek %82 test başarısı elde edilmiştir.

Kontrast Sınırlı Uyarlamalı Histogram Eşitleme yöntemi uygulanmış normal, iyi huylu ve kanser durumlarını içeren veri kümesi kullanılarak YBMG bileşenlerinin medyanına ve frebenius norm, mean ve medyanına karşılık gelen 5. ve 6.bıçım öznitelik matrisi ile en yüksek test skoru %64 olarak elde edilmiştir. Kanser ve kanser

olmayan veri kümesi için YBMG bileşenlerinin frobenius normuna ve ortalamasına karşılık gelen 3. ve 4. biçim öznitelik matrisi ile en yüksek %78 test başarısı elde edilmiştir. Kile olan ve olmayan veri kümesi için YBMG bileşenlerinin ortalamasına karşılık gelen 4. biçim öznitelik matrisi ile en yüksek %82 test başarısı elde edilmiştir.

Yüksek Boyutlu Model Gösterilimi kontrast iyileştirme işlemi uygulanmış normal, iyi huylu ve kanser durumlarını içeren veri kümesi kullanılarak YBMG bileşenlerinin frebenius norm, mean ve medyanına karşılık gelen 6.b biçim öznitelik matrisi ile en yüksek test skoru %60 olarak elde edilmiştir. Kanser ve kanser olmayan veri kümesi için YBMG bileşenlerinin frobenius normuna ve ortalamasına karşılık gelen 6. biçim öznitelik matrisi ile en yüksek %89 test başarısı elde edilmiştir. Kile olan ve olmayan veri kümesi için YBMG bileşenlerinin medyanına karşılık gelen 5. biçim öznitelik matrisi ile en yüksek %87 test başarısı elde edilmiştir.

Kontrast Sınırlı Uyarlamalı Histogram Eşitleme ve Yüksek Boyutlu Model Gösterilimi kontrast iyileştirme yöntemleri validasyon skorları karşılaştırılmış ve anlamlı bir fark olup olmadığı t testiyle sınanmıştır. Normal, iyi huylu ve kanser veri kümesi için iki yöntem arasında anlamlı bir istatistiki fark elde edilmişken kanser olan ve olmayan, kitle olan ve olmayan veri kümeleri için istatistiki olarak anlamlı bir fark elde edilememiştir.

Yapılan karşılaştırmalarda sıklıkla YBMG bileşen değerlerinin medyanına ve frebenius norm, ortalama ve medyanına karşılık gelen 5. ve 6.b biçim öznitelik matrisi kullanılarak yüksek sonuçlar elde edildiği görülmüştür.

Bundan sonra yapılacak çalışmalarda, ortaya koyduğumuz 6 farklı biçimdeki öznitelik matrisleri geliştirilebilir, farklı görüntü iyileştirme teknikleri uygulanabilir. Ayrıca IRMA veri kümesinden alınan mamogramların dışında mini-MIAS gibi farklı veri kümeleri de kullanılıp sonuçlar incelenebilir.

KAYNAKLAR

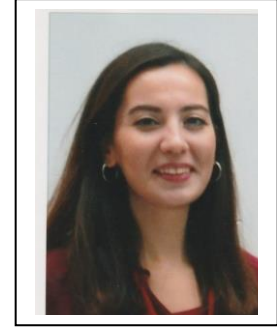
- Adeyemo, O. O., & Adeyeye, T. O.** (2015). Comparative Study of ID3 / C4 . 5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever, 8(1).
- Babur, S.** (2016). Makine Öğrenmesi Teknikleri Kullanılarak Meme Kanseri Teşhisinin Performans Değerlendirmesi Machine Learning Based Performance Development for Diagnosis of Breast Cancer, 1–4.
- Casella, G., Fienberg, S., & Olkin, I.** (2017). *An Introduction to Statistical Learning*. <https://doi.org/10.1007/978-1-4614-7138-7>
- Cheng, H. D., & Xu, H.** (2002). A novel fuzzy logic approach to mammogram contrast enhancement, 148, 167–184.
- Cortes, C., & Vapnik, V.** (1995). Support-Vector Networks, 297, 273–297.
- Doi, K.** (2007). Computer-aided diagnosis in medical imaging : Historical review , current status and future potential, 31, 198–211. <https://doi.org/10.1016/j.compmedimag.2007.02.002>
- Hastie, T. T.** (2017). The Elements of Statistical Learning Second Edition. *Math. Intell.*, 27(2), 83–85. <https://doi.org/111>
- Karaca, E., & Tunga, M. A.** (2018). An interpolation-based texture and pattern preserving algorithm for inpainting color images. *Expert Systems with Applications*, 91, 223–234. <https://doi.org/10.1016/j.eswa.2017.09.001>
- Karçılı, A., & Tunga, B.** (2017). High dimensional model representation (HDMR) with clustering for image retrieval. *AIP Conference Proceedings*, 1798. <https://doi.org/10.1063/1.4972664>
- Kolay, N., Erdoğan, P.** (2016). Makine Öğrenmesi Teknikleri ile Göğüs Kanserinin Sınıflandırılması. Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT).<https://doi.org/10.1109/EBBT.2016.7483683>
- Korkmaz, E., & Metin, Ö.** (2012). Combined small scale high dimensional model representation. <https://doi.org/10.1007/s10910-012-0018-6>
- Kotsiantis, S. B.** (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249–268. <https://doi.org/10.1007/s10751-016-1232-6>
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I.** (2015). Machine learning applications in cancer prognosis and prediction. *CSBJ*, 13, 8–17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- Kuhn, Max Johanson, K.** (2016). *Applied Predictive Model*. <https://doi.org/10.1007/978-1-4614-6849-3>
- Lewin, J. M., Isaacs, P. K., Vance, V., & Larke, F. J.** (2003). Radiology enhanced Digital Subtraction Mammography :, (January 2002), 261–268.

- Master Machine Learning. (1385), 302.
- Morrow, W. M., Paranjape, R. B., Rangayyan, R. M., Member, S., Edward, J., & Desautels, L.** (1992). Region-Based Contrast Enhancement of Mammograms, *11*(3).
- Müller, Andreas C. & Guido, S.** (2017). *Introduction to Machine Learning with Python*.
- Özay, E. K., & Demiralp, M.** (2014). Reductive enhanced multivariate product representation for multi-way arrays. *Journal of Mathematical Chemistry*, *52*(10), 2546–2558. <https://doi.org/10.1007/s10910-014-0396-z>
- Petrou, M. & Petrou, C.** (2010). *Image Processing: The Fundamentals Image Processing: The Fundamentals*.
- Search, H., Journals, C., Contact, A., Iopscience, M., & Address, I. P. (n.d.)**. Computer-aided diagnosis of Alzheimer ' s disease, 2807. <https://doi.org/10.1088/0031-9155/55/10/002>
- Shiraishi, J., Abe, H., Engelmann, R., & Aoyama, M.** (2003). Computer-aided Diagnosis to Distinguish Benign from Malignant Solitary Pulmonary Nodules on Radiographs : ROC Analysis of Radiologists ' Performance — Initial, (6), 469–474.
- Sobol, M.** (1993). Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modeling and Computational Experiment*, *1*(4), 407–414. <https://doi.org/1061-7590/93/04407-008>
- Stevens, K. N., Cover, T. M., & Hart, P. E.** (1967). Nearest Neighbor Pattern Classification, *1*. <https://doi.org/10.1109/TIT.1967.1053964>
- Suknović, M., Čupić, M., Martić, M., & Krulj, D.** (2005). Data warehousing and data mining - A case study. *Yugoslav Journal of Operations Research*, *15*(1), 125–145. <https://doi.org/10.2298/YJOR0501125S>
- Tang, Jinshan Agaian, S.** (2014). *Computer-Aided Cancer Detection And Diagnosis*.
- Tang, J., Member, S., Rangayyan, R. M., Xu, J., & Naqa, I. El.** (2009). Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography : Recent Advances, *13*(2), 236–251.
- Tay, B., Hyun, J. K., & Oh, S.** (2014). A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained, (March). <https://doi.org/10.1155/2014/276589>
- Tunga, Burcu; Altın, E. M.** (2014). High Dimensional Model Representation in Image Processing. *Proceedings of the 13th International Conference on Computational and Mathematical Methods in Science and Engineering*, 1120–1126. <https://doi.org/10.1063/1.4904687>
- Tunga, M. A., & Demiralp, M.** (2009). *Advances in Numerical Methods. Electrical Engineering* (Vol. 11). <https://doi.org/10.1023/B:TICU.0000025642.20704.51>
- Tunga, B., & Demiralp, M.** (2010). The influence of the support functions on the quality of enhanced multivariate product representation, 827–840. <https://doi.org/10.1007/s10910-010-9714-2>

- Tunga, B., & Demiralp, M.** (2012). Hybrid HDMR method with an optimized hybridity parameter in multivariate function representation. *Journal of Mathematical Chemistry*, 50(8), 2223–2238. <https://doi.org/10.1007/s10910-012-0026-6>
- Tunga, B., & Koçanaoğulları, A.** (2017). Digital image decomposition and contrast enhancement using high-dimensional model representation. *Signal, Image and Video Processing*. <https://doi.org/10.1007/s11760-017-1158-8>
- Tunga, M. A., & Demiralp, M.** (2008). A new approach for data partitioning through high dimensional model representation. *International Journal of Computer Mathematics*, 85(12), 1779–1792. <https://doi.org/10.1080/00207160701576095>
- Wang, H., Ding, C., & Huang, H.** (2010). Multi-label linear discriminant analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6316 LNCS(PART 6), 126–139. https://doi.org/10.1007/978-3-642-15567-3_10
- Zuiderveld, K.** (1994). *Contrast Limited Adaptive Histogram Equalization*. *Graphics Gems*. Academic Press, Inc. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- Url-1** <<http://www.mathworks.com/help/images/ref/adaphisteq.html>>, erişim tarihi 14.03.2019.
- Url-2** <http://irma-project.org/index_en.php />, erişim tarihi 10.01.2013.



ÖZGEÇMİŞ



Ad-Soyad : Kübra Karacan
Doğum Tarihi ve Yeri : 15.01.1994 / İstanbul
E-posta : kubrakrcn.kk@gmail.com

ÖĞRENİM DURUMU:

- **Lise** :2008, Oğuzkaan Koleji
- **Lisans** :2012, Yıldız Teknik Üniversitesi, Matematik Bölümü, Matematik

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **Karacan K., Tunga B.**, 2018. Yüksek Boyutlu Model Gösterilimi ile Dijital Mamogramlarda Meme Kanseri Tanısı, 12. İstanbul Bilişim Kongresi, 6 Aralık, 2018 İstanbul, Türkiye.

DİĞER YAYINLAR, SUNUMLAR VE PATENTLER:

- **Karacan K., Uyar T. ve Öztürk M.K.**, 2018. Görüntü İşleme ve Gözetimli Makine Öğrenmesi Teknikleriyle Ticari Hava Aracı Sınıflandırma, 7. Ulusal Havacılık ve Uzay Konferansı, 12-14 Eylül, 2018 Samsun, Türkiye.