TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MEME KANSERİNİN GELİŞTİRİLMİŞ MAKİNE ÖĞRENME YÖNTEMLERİ İLE TESPİTİ

DOKTORA TEZİ

Erkan AKKUR

Biyomedikal Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Osman EROĞUL

OCAK 2023

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, alıntı yapılan kaynaklara eksiksiz atıf yapıldığını, referansların tam olarak belirtildiğini ve ayrıca bu tezin TOBB ETÜ Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlandığını bildiririm.

Erkan AKKUR

ÖZET

Doktora Tezi

MEME KANSERİNİN GELİŞTİRİLMİŞ MAKİNE ÖĞRENME YÖNTEMLERİ İLE TESPİTİ

Erkan AKKUR

TOBB Ekonomi ve Teknoloji Üniveritesi
Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Osman EROĞUL

Tarih: Ocak 2023

Meme kanseri dünya genelinde kadınlar arasında en sık görülen kanser türüdür. Meme kanseri erken evrede teşhis edilirse, tedavi edilmesi mümkündür. Bu çalışma meme kanserinin tanısı için geliştirilmiş makine öğrenme algoritmalarına dayalı yeni bir sınıflandırma sistemi önermektedir. Geliştirilmiş makine öğrenme algoritmaları oluşturmak amacıyla öznitelik seçim ve hiperparametre optimizasyon yöntemleri kullanılmıştır. Makine öğrenme algoritması olarak sırasıyla Karar Ağacı, Naive Bayes, Destek Vektör Makinesi, K-En Yakın Komşu ve Topluluk Öğrenme yöntemleri kullanılmıştır. Tüm deneyler Wisconsin Meme Kanseri Veri (WBCD) seti ve Mamografi Meme Kanseri Veri Seti (MBCD) olmak üzere iki farklı meme kanseri veri seti üzerinde test edilmiştir. Veri setlerinin en ayırt edici özniteliklerini belirlemek amacıyla sırasıyla Relief, En Küçük Mutlak Daralma ve Seçme Operatörü ((Least Absolute Deviation and Least Absolute Shrinkage and Selection Operator-LASSO) ve Ardışık İleri Yönde Seçim yöntemleri kullanılmıştır. Makine öğrenme algoritmalarındaki en uygun hiperparametreleri bulmak için Bayes optimizasyon (BO) yöntemi kullanılmıştır. Çalışma kapsamında en iyi sınıflandırma oranını elde etmek amacıyla farklı deneyler yapılmıştır. Önerilen öznitelik seçim-Bayes optimizasyon hibrit yöntemleri çalışmada kullanılan makine öğrenme algoritmalarının sınıflandırma

oranlarını artırmıştır. Yapılan deneyler sonucunda, LASSO-BO-DVM yöntemi her iki meme kanseri veri setinde de en yüksek doğruluk, kesinlik, duyarlılık ve F1-skorunu göstermiştir (WBCD için %98,95, %97,17, %100 ve %98,56; MBCD için %97,95, %98,28, %98,28 ve %98,28).

Anahtar Kelimeler: Meme kanseri, Makine öğrenmesi, Hiperparametre optimizasyonu, Öznitelik seçim yöntemleri

ABSTRACT

Doctor of Philosophy

DETECTION OF BREAST CANCER WITH IMPROVED MACHINE LEARNING ALGORITHMS

Erkan AKKUR

TOBB University of Economics and Technology
Institute of Natural and Applied Sciences
Biomedical Engineering Science Programme

Supervisor: Prof. Dr. Osman EROĞUL

Date: January 2023

Breast cancer is the most common cancer type among women worldwide. If breast cancer is detected at an early stage, it can be cured. This study proposes a novel classification model based improved machine learning algorithms for diagnosis of breast cancer. Feature selection and hyperparameter optimization methods are used to build improved the machine learning algorithms. Decision Tree, Naive Bayes, Support Vector Machine, K-Nearest Neighbor and Essemble Learning methods are used as machine learning algorithms, respectively. All experiments are tested on two different datasets, Wisconsin Breast Cancer Dataset (WBCD) and Mammographic Breast Cancer Dataset (MBCD). Relief, Least Absolute Deviation and Least Absolute Shrinkage and Selection Operator (LASSO) and Sequential Forward Selection methods are used to determine the most distinctive features of the datasets, respectively. Bayesian optimization (BO) method is used to find optimal hyperparameters in machine learning algorithms. Within the scope of this study, different experiments are conducted in order to obtain the best classification rate. The proposed feature selection-Bayes optimization hybrid methods have increased the classification rates of the machine learning algorithms used in the study. As a result of the experiments, LASSO-BO-SVM has showed the highest accuracy, precision, recall

and F1-score in both datasets (%98,95, %97,17, %100, %98,56 for WBCD; %97.95, %98,28, %98,28, %98,28 for MBCD).

Keywords: Breast cancer, Machine learning, Hyperparameter optimization, Feature selection methods

TEŞEKKÜR

Çalışmalarım boyunca kıymetli bilgi, birikim ve deneyimleri ile beni yönlendiren ve destek olan danışman hocam Sayın Prof. Dr. Osman EROĞUL'a sonsuz teşekkür ve saygılarımı sunarım.

Komite üyelerim Prof. Dr. Fatih BÜYÜKSERİN ve Doç. Dr. Mehmet Feyzi AKŞAHİN'e araştırmalarım gereken için geribildirimi sağladıkları için ve savunmamda yer alan Dr. Fuat TÜRK ve Dr. Aykut EKEN'e değerli fikirlerini benimle paylaştıkları için en içten dileklerimle teşekkür ederim.

Kıymetli tecrübelerinden faydalandığım TOBB Ekonomi ve Teknoloji Üniversitesi Biyomedikal Mühendisliği Bölümü öğretim üyelerine teşekkür ederim.

Çalışmamda kullanılan verilerin temini konusundaki yardımlarından dolayı Doç. Dr. Pelin Seher ÖZTEKİN ve Dr. Oğuz LAFCI'ya teşekkürlerimi sunarım.

Tez çalışmam sürecinde bana verdikleri desteklerden dolayı Dr. Ahmet İlker KESKİN'e, Nazlı Buket YAKIŞIR'a, Ayça YILDIRIM'a ve Yasin KINDAP'a çok teşekkür ederim.

Çalışmam boyunca bana inanmaya devam eden, sevgi ve desteklerini hiçbir zaman esirgemeyen annem Samiye YILDIZ AKKUR'a, babam Erhan AKKUR'a, kardeşim Ebru Ceren AKKUR BAYRAM'a ve eniştem Can BAYRAM'a sonsuz teşekkür ederim.

İÇİNDEKİLER

ŞEKİL LİSTESİ

Sayfa

ÇİZELGE LİSTESİ

Sayfa

KISALTMALAR

AYİS : Ardışık Yönde İleri Seçim
BDS : Bilgisayar Destekli Sistemler
BO : Bayes Optimizasyonu
BÖ : Bütün Öznitelikler
BI-RADS : Breast Imaging Reporting and Data System
DICOM :Tıpta Dijital Görüntüleme ve İletişim (Digital Imaging and Communications in Medicine)
DN : Doğru Negatif
DP : Doğru Pozitif
DVM : Destek Vektör Makineleri
EI : Beklenen İyileştirme (Expected Improvement)
GS : Gauss Süreci
GSEOM : Gri Seviye Eş Oluşum Matrisi
GSKUM : Gri Seviye Koşu Uzunluğu Matrisi
KA : Karar Ağacı
KK : Kraniokaudal
KKM : Kendi Kendine Muayene
KMM : Klinik Meme Muayenesi
K-NN : K-En Yakın Komşu
LASSO : En Küçük Mutlak Daralma ve Seçme Operatörü (Least Absolute Shrinkage and Selection Operator)
LR : Lojistik Regresyon
MBCD : Mamografi Meme Kanseri Veri Seti (Mammographic Breast Cancer Dataset)
MIAS : Mamographic Image Analysis Society
MLO : Mediolateral Oblik
MR : Manyetik Rezonans
NB : Naive Bayes
PACS : Görüntü Arşivleme ve İletişim Sistemleri (PACS- Picture Archiving Communication Systems)
RMSE : Karekök Ortalama Hatası (Root Mean Square Error)
RO : Rastgele Orman
ROI : İlgenilen Alan (Region of interest)
PI : İyileştirme Olasılığı (Probability of Improvement)
TÖ : Topluluk Öğrenme
WBCD : Wisconsin Breast Meme Kanseri Veri Seti (Wisconsin Breast Cancer Dataset)
VİGY : Veri İşleme Grup Yöntemi
YN : Yanlış Negatif
YP : Yanlış Pozitif
YSA : Yapay Sinir Ağları

SEMBOL LİSTESİ

Bu çalışmada kullanılmış olan simgeler açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklama**
A	Kümenin bölünmüş bir parçasını
argmax	Bir fonksiyonun maksimumu
As	Aynı sınıftaki en yakın öznitelik değerini
b	Sapma değeri
C_i	Her sınıfın başka bir olay olasılığı
D	Gözlemlenen eğitim seti
E	Beklenen değer
E_{test}	Validasyon verisindeki amaç fonksiyonu
f(x)	Bayes optimizasyonu amaç fonksiyonu
$f(x^+)$	Mevcut optimal değer
Fs	Farklı sınıftaki en yakın öznitelik değerini
f_{t+1}	Yeni örnek noktası
$f_{t+1}(x)$	Yeni nokta değeri
g(x)	Optimizasyon fonksiyonu
H	Hiperparametre kümesi
H_{xy}	Ortak entropi
I	İyileştirmenin derecesi
i	Merkez piksel
j	Komşu piksel
K	Destek vektör makinelerindeki çekirdek fonksiyonu
$k(x, x^{'})$	Kovaryans fonksiyonu
m(x)	Gauss sürecindeki ortalama fonksiyon
L	Herhangi bir hiperparametre uzayı
l	L uzayındaki herhangi bir değer
l^*	En düşük değeri veren hiperparametre seti
N	Gözlem sayısı
N_g	Gri seviye sayısı
N_r	Maksimum koşu uzunluğu
P	Eş oluşum matrisi
p	Olasılık değeri
p (i, j)	Uzunluk matrisinin (i, j) inci noktadaki değeri
p^+	En küçük değer veren hiperparametre seti
R	Relief özniteliğinin önem derecesi
S	Orijinal veri kümesini
Sv	Özniteliği A değeri olan S alt küme sayısı
v	Öznitelik değeri
v’	Yeni değer
w	Ağırlık vektörü
x	H uzayındaki herhangi bir hiperparametre

x^+	Yeni hiperparametre alt kümesi
x_i	i. dereceden örnek noktalarını
x_j	j. dereceden örnek noktalarını
y'	Çıkış vektörü
y_i	i. dereceden çıkış değeri
Δx	x yönündeki piksel değeri
Δy	y yönündeki piksel değeri
μ	Ortalama
σ	Standart sapma
β	LASSO katsayıları
λ	Ceza parametresi
φ_{tahmin}	Model tahmini
φ'_j	Test veri setinin referansını
ε	Epsilon değeri
α	Reel sayılarda herhangi bir değer

1. GİRİŞ

Meme kanseri küresel istatistiklere göre kadınlar arasında en sık görülen kanser türüdür [1]. Meme kanseri meme dokusundaki hücrelerin kontrolsüz olarak büyümeleri ile tümör adı verilen kistlerin oluşması sonucu meydana gelmektedir. Meme tümörleri iyi huylu ve kötü huylu olarak sınıflandırılmaktadır. İyi huylu tümörler kanser ile ilişkisi bulunmamaktadır. Bu tip tümörler vücudun diğer bölgelerine yayılmamaktadır. Kötü huylu tümörler ise kanser özelliğine sahiptir. Kötü huylu tümörler tedavi edilmedikleri taktirde kontrolsüz bir şekilde çoğalabilmektedir. Bu tür tümörlerin yakınlarındaki dokulara yayılma potansiyelleri bulunmaktadır [2]. Meme kanserinin erken teşhisi meme kanserine bağlı ölüm oranlarını önemli oranda azaltmaktadır [3]. Meme kanserinin erken evrede teşhisi için yapay zekâ yöntemleri kullanılarak yapılan çalışmalar oldukça önem arz etmektedir. Makine öğrenmesi bir sistemin geçmiş tecrübelerinden elde edilen öğrenmelerinden yararlanarak bir model oluşturan ve gelecekte meydana gelebilecek durumlar karşısında tahminler yapmasını sağlayan bir yapay zeka dalıdır. Bu yeteneklerinden dolayı makine öğrenme algoritmaları meme kanserinin teşhisinde yaygın bir şekilde kullanılmaktadır [4-6].

Öznitelik seçim yöntemleri ve hiperparametre optimizasyonu makine öğrenme algoritmalarının sınıflandırma performanslarını etkileyen iki önemli unsurdur. Öznitelik seçim yöntemleri veri setlerindeki en faydalı öznitelikleri seçme ve bulma süreci olarak tanımlanmaktadır. Öznitelik seçim yöntemleri veri setlerindeki öznitelik kümesinin boyutunu azaltmakta, makine öğrenme algoritmalarının hızını ve başarı oranlarını artırabilmektedir [6,7]. Makine öğrenme algoritmaları çok sayıda hiperparametre içermektedir. Hiperparametreler, modelin öğrenemediği ve eğitim sürecinden önce sağlanması gereken parametrelerdir. Uygun hiperparametrelerin ayarlanması makine öğrenme algoritmaların performanslarını artırmaktadır. Hiperparametre optimizasyonu makine öğrenme algoritmaları için en uygun hiperparametre kombinasyonun belirlenmesidir [8-9]. Sonuç olarak, öznitelik seçim yöntemlerinin ve hiperparametre optimizasyonun kullanılması makine öğrenme algoritmalarının başarı oranlarını artırmaktadır. Literatürde meme kanserinin makine öğrenmesi ile tespiti ile ilgili çok sayıda çalışma yapılmaktadır. Ancak konu ile ilgili

literatür çalışmaları incelendiğinde hiperparametre optimizasyonu ve öznitelik seçim yöntemlerinin makine öğrenmesi algoritmalarının sınıflandırma performanslarına etkisini inceleyen çalışmaların daha az olduğu görülmüştür.

1.1 Tezin Amacı

Bu çalışmada meme kanserinin tespiti için geliştirilmiş makine öğrenme algoritmalarına dayalı bir sınıflandırma sistemi önerilmiştir. Bu kapsamda; çalışmada farklı makine öğrenme algoritmaları kullanılmış ve bu algoritmaların sınıflandırma performanslarının artırılması hedeflenmiştir. Makine öğrenme algoritmalarının sınıflandırma oranlarını artırmak amacıyla öznitelik seçim yöntemleri ve hiperparametre optimizasyonu birleştirilerek farklı hibrit modeller oluşturulmuştur. Bununla birlikte öznitelik seçim yöntemlerinin ve hiperparametre optimizasyonun makine öğrenme algoritmalarına etkisi irdelenmiş ve en yüksek başarı oranına sahip hibrit modelin bulunması hedeflenmiştir.

Bu tez çalışmasında özet olarak;

- İki farklı meme kanseri veri seti 5 farklı makine öğrenme algoritması ile sınıflandırılmış ve algoritmaların başarı oranları karşılaştırılmıştır.
- Veri setlerindeki en seçici ve ayırt edici öznitelikler 3 farklı öznitelik seçim yöntemi kullanarak belirlenmiştir.
- Makine öğrenme algoritmalarının en uygun hiperparametrelerinin belirlenmesi maksadıyla Bayes optimizasyon yöntemi kullanılmıştır.
- En yüksek sınıflandırma oranlarını elde etmek amacıyla her bir veri seti için 25 adet farklı deney gerçekleştirilmiştir.

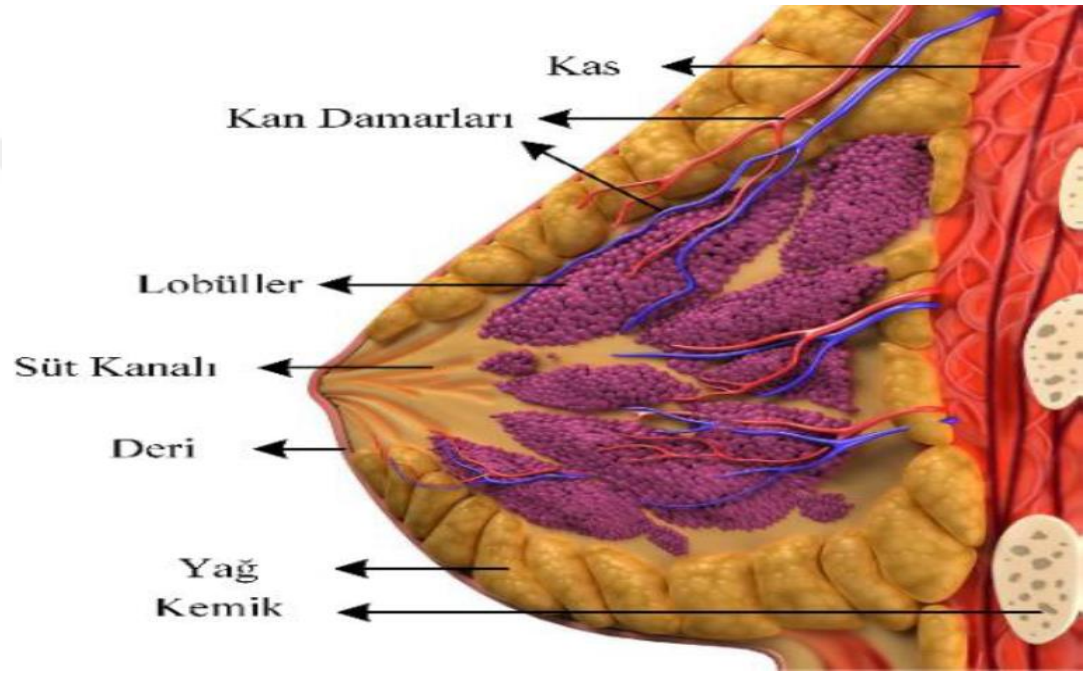
1.2 Tezin Kapsamı

Bu çalışma sekiz bölümden oluşmaktadır. Birinci bölümde tezin amacı ve kapsamı sunulmuştur. İkinci bölümde meme kanserinin tanımı yapılmış, meme kanserinin türlerinden, risk faktörlerinden ve tanı yöntemlerinden bahsedilmiştir. Üçüncü bölümde yapay zekanın tanımı yapılmış, makine öğrenme, yapay sinir ağları ve derin öğrenme ile ilgili bilgiler verilmiştir. Dördüncü bölümde son yıllarda meme kanserinin sınıflandırılması ile ilgili yapılan literatür çalışmaları özetlenmiştir.

Beşinci bölümde önerilen hibrit sınıflandırma sisteminin akış şeması sunulmuş ve çalışma kapsamında kullanılan veri setleri ile bilgiler verilmiş olup kullanılan öznitelik çıkarım, öznitelik seçim yöntemleri, makine öğrenme algoritmaları ve Bayes optimizasyon yöntemi ayrıntılı olarak anlatılmıştır. Son olarak makine öğrenme algoritmalarının sınıflandırma oranlarının değerlendirilmesinde kullanılan performans kriterleri ile ilgili bilgiler verilmiştir. Altıncı bölümde önerilen hibrit modellerin meme kanseri veri setleri üzerinde elde edilen sınıflandırma sonuçlarına yer verilmiştir. Yedinci bölümde önerilen modellerin performans sonuçlarının birbirleri ve literatürdeki benzer çalışmalar ile karşılaştırmaları sunulmuştur. Sekizinci bölümde çalışma ile ilgili genel değerlendirmeler yapılmıştır.

2. MEME KANSERİ

Meme göğüs kaslarının üzerinde bulunan ve orta hat göğüs kemiğinin dış bölgesi ve koltuk altının ön sınırından aşağı doğru uzanan süt bezlerinden oluşan damla şeklinde bir organdır [10]. Meme dokusunun içerisinde farklı anatomik yapılar bulunmaktadır [11]. Şekil 2.1'de bu anatomik yapılar gösterilmiştir.



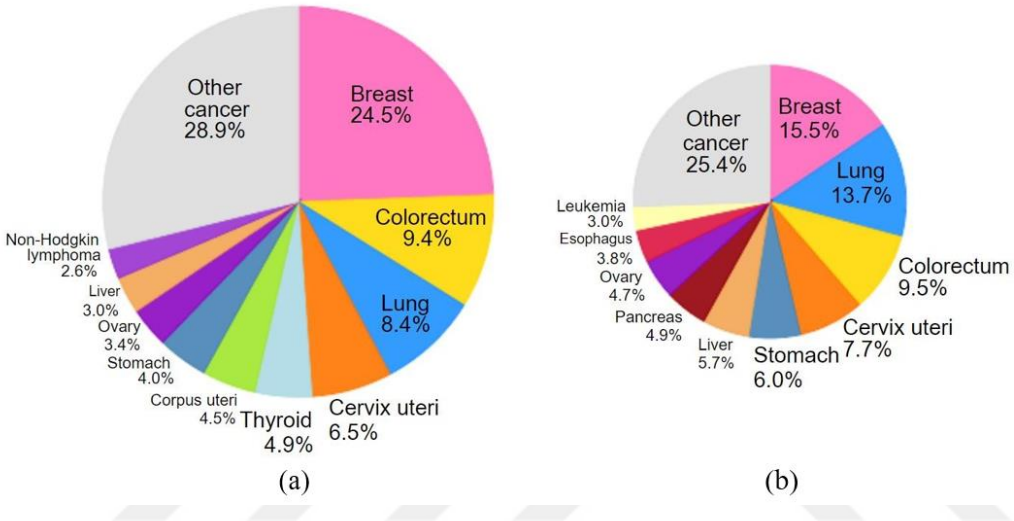
Şekil 2.1: Meme anatomisi [11].

Meme dokusu anatomik olarak süt üretimini sağlayan bezlerin oluşturduğu lobüllerden, sütün boşaltma işleminin gerçekleştiği kanallardan, kanallar ile lobüller arasını dolduran yağ ve bağ dokularından oluşmaktadır [10-11].

Meme kanseri meme dokusunda bulunan hücre büyümesini denetleyen genlerde oluşabilen mutasyon veya anormal değişiklikler sebebiyle oluşmaktadır. Sağlıklı bir bireyde bir hücre bölüneceği zamanı ve yeri bilme yeteneğine sahipken, bilinci kaybolmuş bir hücre kontrolsüz bir şekilde bölünerek kansere neden olabilmektedir [11].

Dünya Sağlık Örgütü'nün 2020 yılı için yayımlamış olduğu kanser istatistikleri raporlarına göre, meme kanseri dünya genelinde kadınlarda en yaygın kanser türüdür.

Şekil 2.2'de dünya genelinde kadınlar arasındaki görülen kanser türlerinin görülme ve ölüm oranları gösterilmiştir. 2020 yılı için yaklaşık olarak 2,3 milyon kişiye meme kanseri teşhisi konulmuştur. Bu kanser türü tüm kadınlarda görülen kanser vakalarının yaklaşık olarak %24,5'ini oluşturmaktadır. Aynı raporlarda 2020 yılında yaklaşık olarak 685.000 kadının meme kanseri nedeniyle yaşamını kaybettiği bildirilmektedir. Meme kanseri sebebiyle hayatını kaybeden kadın sayısı tüm kanser ölüm sayısının yaklaşık olarak %15,5'ini oluşturmaktadır [1].



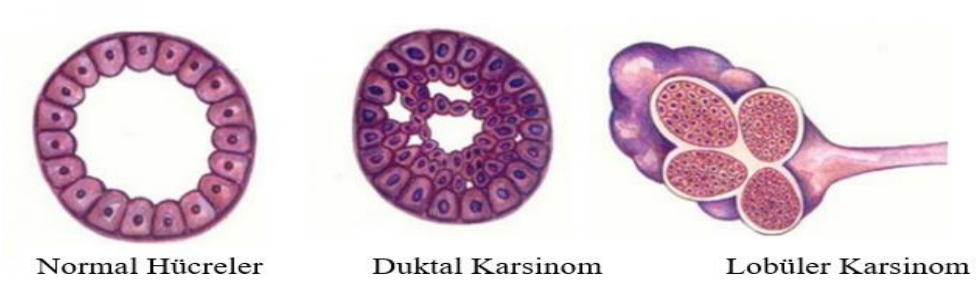
Şekil 2.2: Kanser tiplerinin dünya genelinde (a) görülme ve (b) ölüm oranı [1].

Ülkemizde de dünya genelinde olduğu gibi meme kanseri kadınlar arasında görülen en yaygın kanser türüdür. Kanser Erken Teşhis, Tarama ve Eğitim Merkezi (KETEM) tarafından ülkemiz genelinde ulusal meme kanseri taramaları çalışmaları yürütülmektedir. Toplum bazlı taramalar Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü bünyesinde bulunan Kanser Dairesi Başkanlığı tarafından yayımlanan "Ulusal Kanser Tarama Standartları" doğrultusunda yapılmaktadır. Bu standartlara göre 40-69 yaş arasındaki kadınlar hedef grubu oluşturmaktadır. Bu hedef grubuna mamografi ile meme kanseri taraması yapılmaktadır. Meme kanseri taramasının aralığı 2 yıldır [12].

2.1 Meme Kanseri Türleri

Meme kanseri meme dokusunda bulunan hücrelerde meydana gelmektedir. Meme dokusunda oluşan her meme tümörü kanser değildir. Meme tümörleri iyi huylu ve kötü

huylu olmak üzer iki sınıfta değerlendirilmektedir. İyi huylu meme tümörlerinin meme kanseri ile herhangi bir ilişkisi bulunmamaktadır. Fibradenom, adenozis, duktal ektazi, yağ nekrozu ve fibrokistik değişiklikler iyi huylu meme tümörü çeşitleridir [13]. Kötü huylu meme tümörleri ise kanser özelliğine ve vücutta başka dokulara yayılma potansiyeline sahiptir. Bu tür meme tümörlerinin kendi içerisinde bir sınıflandırılması bulunmaktadır. Eğer kanser süt kanallarında oluşuyorsa duktal karsinom, süt bezlerinde oluşuyorsa lobüler karsinom adını almaktadır. Bu iki kanser türü de kendi içlerinde iki grupta incelenmektedir. Lobüler karsinom süt bezinin içinde meydana gelirse lobüler karsinom in sutu, süt bezinin dışına çıkmış ise invaziv lobüler karsinom adını almaktadır. Duktal karsinom süt kanalının içindeyse duktal karsinom in sutu, dışındaysa invaziv duktal karsinom şeklinde tanımlanmaktadır [14-Url-1]. Şekil 2.3'de normal, duktal karsinom ve lobüler karsinom meme dokusu hücreleri gösterilmektedir.



Şekil 2.3: Meme dokusunun patolojisinin animatif olarak gösterimi [Url-2].

Duktal karsinom in sutu meme kanserinin en erken aşamasında oluşmaktadır. Bu kanser türünde anormal hücreler süt kanalının dışına çıkmayıp meme dokusuna yayılmamış durumdadır. Tedavi edilmedikleri zaman ise bazen meme dokusuna yayılma potansiyeli bulunmaktadır. Kanserli bölgenin çıkarılması ile tedavi edilebilmektedir. Elle muayene edilemeyecek kadar küçük olmaları sebebiyle mamografi ile teşhis edilmesi gerekmektedir [14, 15, Url-1, Url-2]. İnvaziv duktal karsinom kanser hücrelerinin süt kanallarının dışına çıktığı zaman oluşmaktadır. Bu kanser türünde, anormal hücre önce kanal içinde oluşmakta daha sonra duvarı geçerek meme dokusuna yayılmaktadır. Bazı durumlarda anormal hücre meme dokusu dışına yayılma potansiyeline sahiptir. Bu kanser türü eğer tedavi edilmezse ölümcül olabilmektedir[14-Url-1]. Lobüler karsinom in sutu, kanser hücrelerinin süt bezlerinde

(lobüller) oluşması ile meydana gelmektedir. Lobüllerde meydana gelmeye başlayan oluşum lobüllerin içindedir ve meme dokusuna yayılmamış durumdadır. [Url-1]. İnvaziv lobüler karsinom, kanser hücrelerinin süt bezlerinin dışına geçmeye başladığı formdur. Kanserli hücre meme içine ve meme dışına yayılmış durumdadır [14, Url-1].

2.2 Meme Kanseri Risk Faktörleri

Günümüzde, meme kanseri ile ilgili olarak nedeni kesin olarak bilinmemekle beraber yapılan bilimsel çalışmalar sayesinde, bu kansere neden olan bazı faktörler tespit edilmiştir. Meme kanserinin oluşmasında etkili olduğu varsayılan bu risk faktörlerini 5 ana grupta incelemek mümkündür.

1. Demografik özellikler
2. Reproduktif Öykü
3. Ailesel ve genetik faktörler
4. Çevresel etmenler
5. Diğer faktörler

Demografik özellikler: Kadın cinsiyeti meme kanserine yakalanma açısından en büyük risk faktörü olarak değerlendirilmektedir. Kadınların artan yaşı cinsiyet kadar önemli risk faktörlerinden birisidir. Bir kadının meme kanserine yakalanma ihtimali sekizde bir olarak kabul edilmektedir. Yaş oranı artıkça bir kadının meme kanserine yakalanma oranı da artmaktadır. Otuz yaşında bir kadının meme kanserine yakalanma oranı önündeki yıl boyunca 1/250 iken, yetmiş yaşında bir kadının meme kanserine yakalanma olasılığı 1/27'dir. Meme kanserine yakalanma açısından önemli paradokslardan birisi de meme kanserinin görülme oranın siyahi kadınlara göre beyaz kadınlarda %20 daha yüksek olmasına rağmen, ölüm oranın siyahi kadınlarda daha yüksek olmasıdır. Bu durumun sosyoekonomik koşullardan ve yaşam tarzından kaynaklandığı düşünülmektedir [16-17].

Reproduktif Öykü: Reproduktif öykü bir kadının ergenlik döneminde ilk adet kanamasından menopoz dönemine kadar olan süre olarak tanımlanmaktadır. Kadınların uzun süre östrojen hormonuna maruz kalması meme kanserinin gelişmesindeki artışla ilişkilidir. İlk doğumun ileri yaşlarda yapılması ya da hiç doğum yapılmaması da meme kanserinin gelişmesindeki artışla ilişkilidir.

Ailesel ve genetik faktörler: Aile öyküsü meme kanserine yakalanma açısından önemli bir risk faktörüdür. Meme kanserine yakalanan olguların yaklaşık olarak %5-10 oranında ailesel olduğu görülmektedir. Bir adet birinci derece yakınlarda meme kanseri olgusu varsa, risk 1.80 oranında artmaktadır. İki adet ikinci derecede yakınlarda meme kanseri olgusu varsa, risk 2.9 oranında artmaktadır. Bir akraba otuz yaşından önce meme kanserine yakalanmışsa risk oranı 2.9 kat, altmış yaşından sonra meme kanseri görülmüşse risk oranı 1.5 kat artmaktadır. Meme kanseri görülen kadınlarda kansere neden olan çeşitli kalıtsal genler tanımlanmıştır. Bu kalıtsal genler içinde en önemlileri BRCA1 ve BRCA2 genleridir. Bu genlere sahip kadınlarda yaşamları boyunca meme kanserine yakalanma riskinin %45-80 oranında olduğu kabul edilmektedir.

Çevresel etmenler: Çevresel etmenler sigara ve alkol kullanımı, beslenme alışkanlıkları, pasif yaşam, hormon replasman tedavisi, sosyoekonomik düzey, radyasyona maruz kalma gibi faktörlerdir.

Diğer faktörler: Vücut kitle indeksi, proliferatif meme hastalıkları, meme dokusunun dens özelliği ve kişisel meme kanseri öyküsü meme kanserini etkileyen diğer faktörler olarak değerlendirilebilmektedir [16-17].

2.3 Meme Kanserinde Tanı

Meme kanseri erken evrede saptanabilirse tedavi edilebilme ihtimali oldukça yüksektir. Erken teşhis ile meme kanserine bağlı ölüm oranlarını azaltmak mümkün olabilmektedir. Tanı yöntemleri olarak kendi kendine muayene, klinik meme muayenesi, görüntüleme yöntemleri ve biyopsi kullanılmaktadır [10].

Kendi kendine muayene (KKM): Yirmi yaşını geçen her kadının adet dönemi bittikten sonra ilk hafta içinde, adet görmeyenlerin ise ayın belli bir günü kendisini kontrol etmesi önerilmektedir. KKM yöntemi sırasıyla bir ayna karşısında vücuttaki görsel değişiklikler kontrol değerlendirildikten sonra yatarak elle yapılan muayene olarak tanımlanabilmektedir. Düzenli olarak yapılan KKM yöntemi ile her kadın belli bir süre sonra meme yapısını tanıyabilmektedir. Böylece meme dokusunda çıkan ya da çıkabilecek kitleleri erken evrede saptayabilecek seviyeye ulaşılabilmektedir. Meme kanserinin erken evrede belirtileri çok net olmamaktadır. Kanserin ilerleyen süreçlerinde ise meme dokusunda izlenmesi gereken birtakım değişikler meydana

gelebilmektedir. Meme dokusunda kitlelerin ele gelmesi, boyutlarında ya da şeklinde değişikler oluşması, meme başından akıntı gelmesi ve dokudaki renk değişimleri meme yapısındaki değişikliklere örnek olarak verilebilmektedir. [18].

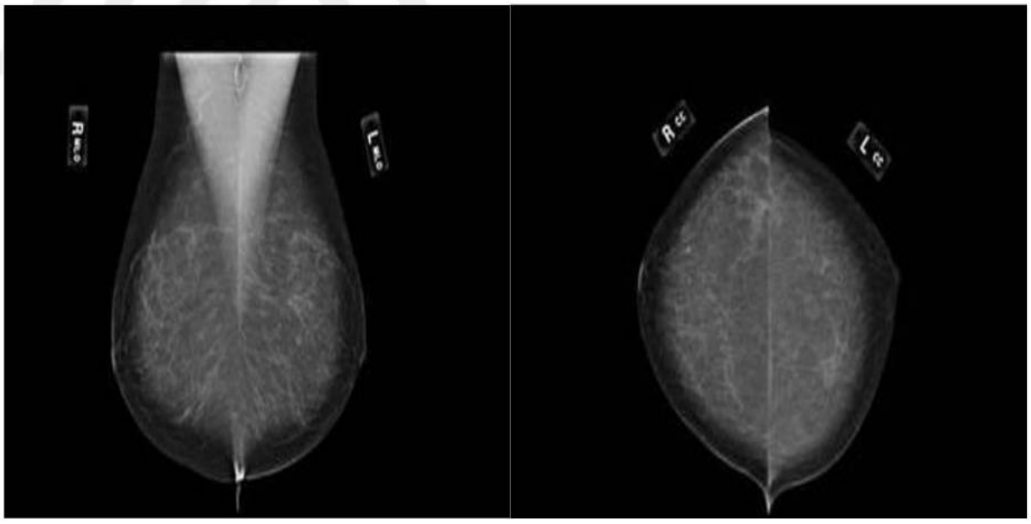
Klinik meme muayenesi (KMM): Bu yöntem bir doktor tarafından KKM ile KMM yöntemindeki benzer adımlar takip edilerek yapılmaktadır. Hastalık hikayesinin öğrenilmesi ve gerekirse bazı tetkiklerin yapılması bu yöntemin aşamalarından birisidir. Bu sayede sorun ile ilgili tanı konabilmekte ve tedavi süreci başlatılabilmektedir [19].

Görüntüleme Yöntemi: Meme kanserinin erken tanısı amacıyla Mamografi, Manyetik Rezonans (MR) Görüntüleme, Ultrasonografi ve Dijital Meme Tomosentez yöntemleri kullanılmaktadır [20]. Bu görüntüleme yöntemlerinden mamografi meme kanserinde en sık kullanılan ve ilk başvurulan yöntemdir. Kırk yaş üzerindeki kadınların meme dokusu ile ilgili şikâyeti olsun ya da olmasın yılda bir kez mamografi çektirmesi önerilmektedir [21].

Mamografi; memeyi görüntüleme amacıyla X-ışınları kullanan meme dokusunun özel bir radyografi türüdür. Mamografi meme kanserinin erken teşhisinde tarama ve tanı amaçlı kullanılmaktadır. Taramadaki temel amaç semptomu olmayan hastalarda erken teşhis koymaktır. Tanıdaki amaç ise semptomu olan hastalara teşhis koymaktır. Mamografi meme kanserinin yaklaşık olarak %25-30 oranında mortalite oranını azaltmaktadır. Mamografi meme dokusunun glandüler ve yağ oluşumlarını inceleyen yumuşak bir doku radyografi yöntemi olarak değerlendirilmektedir. Konvansiyonel ve dijital olmak üzere iki grupta incelenmektedir. Konvansiyonel olan mamografilerde, X-ışını meme dokusundan geçerek bir kaset içinde bulunan röntgen filmini etkilemektedir. Kaset içinde filmin banyo edilmesi ile mamografi görüntüsü oluşmaktadır. Dijital mamografi 2000'li yıllardan sonra geliştirilen bir teknolojidir. Bu teknoloji meme dokusunun bir görüntü dedektörü ve bir kompresyon plakası ile sıkıştırılarak röntgen filminin çekilmesi prensibine dayanmaktadır. Bu yöntemde meme iki yönden hafifçe sıkıştırılmaktadır. Kompresyona uğramış memedeki farklı dokuların X-ışınına karşı göstermiş olduğu zayıflama miktarları önce dedektörler yardımıyla elektriksel sinyale çevrilmektedir. Daha sonra da çeşitli bilgisayar algoritmaları ile bu sinyaller görüntü haline dönüştürülmektedir. Sonuç olarak; iki boyutlu ve parlaklık değişimini gösteren bir görüntü oluşmaktadır [22]. Tek bir açıdan

X-ışın kaynağının olması sebebiyle farklı dokuların üst üste çakışması ile görüntü üzerinde süperpozisyon meydana gelmektedir. Ortaya çıkan görüntünün tersinin alınması ile pektoral kas, damar ve şüpheli bölgeler gibi yoğun yapıların açık tonlarda, yağ dokusu gibi bölgeler ise koyu çıkmaktadır [23-25].

Görüntülerde meydana gelen süperpozisyon olayının etkisini ortadan kaldırmak amacıyla mamografi çekimi yapılırken meme dokusunun farklı açılarda görüntüsü alınmaktadır. Mamografi cihazlarında mediolateral oblik (MLO) ve kraniokaudal (KK) şeklinde iki farklı projeksiyonda görüntü tercih edilmektedir [23]. MLO projeksiyonunda uygun açıda meme dokusunun büyük bir bölümü görüntü alanına girmekte ve meme dokusunun üst kadran ve aksiyel kuyruk bölgesi daha iyi görüntülenmektedir. KK projeksiyonunda ise subareolar bölge, santral ve medial meme dokusu görüntülenmektedir. Bu iki farklı projeksiyondaki grafiler birbirini tamamlamakta ve birinin görüntü alanına girmeyen bir lezyon diğer projeksiyondaki görüntü alanına girebilmektedir [25]. Rutin olarak bir mamografi çekimi yapılırken her iki projeksiyonda da görüntü alınmaktadır. Şekil 2.4'de sağ ve sol meme dokusunun iki projeksiyonda da görüntüsü gösterilmiştir.



Şekil 2.4: Sağ ve sol memenin KK ve MLO projeksiyonundaki görüntüleri [26].

Mamografi görüntüleme yöntemi ile meme dokusundaki farklı anormallikler gözlemlenebilmektedir. Kitle, mikro-kalsifikasyon, asimetrik dansite, yapısal distorsiyon gibi radyolojik bulgular mamografi yöntemi ile tespit edilebilmekte ve

izlenebilmektedir. Mamografi görüntülerinde saptanan anormalliklerin klinisyen ve radyologlar tarafından değerlendirme sürecinde standart bir yöntemin kullanılması hastaların doğru bir şekilde yönlendirilmesini sağlamaktadır [27]. Bu amaçla 1992 yılında Amerikan Radyoloji Derneği tarafından "Meme görüntüleme raporlama ve veri sistemi (Breast Imaging Reporting and Data System"-(BI-RADS)) olarak adlandırılan ve mamografi yorumlanması için standart bir yöntem geliştirilmiştir. BI-RADS yönteminin son baskısı 2013 yılında yayımlanmıştır. BI-RADS sistemi,

- Tanımlama ve raporlamada kullanılacak terminolojide standardizasyon sağlama
- Malignite olasılığının değerlendirilmesi,
- Klinik ve radyoloji arasında iletişim kolaylığı,
- Raporların standardizasyonu,
- Tıbbi kayıt ve izlem kolaylığı,

gibi özellikleri sayesinde radyologlara rehberlik etmektedir. Benzer şekilde manyetik rezonans ve ultrasonografi için de meme dokusunda oluşan anormallikleri değerlendirmek amacıyla BI-RADS rehberi yayımlanmıştır. BI-RADS sistemi meme lezyonlarında oluşan başta mikro-kalsifikasyon ve kitle anormalliklerini 6 farklı kategoride yorumlamaktadır [27]. BI-RADS kategorileri:

Kategori 0: Ek tetkik gerektirdiğini ifade etmektedir. Bu kategori mevcut olan görüntüler ile net bir karar verilemediğini ve başka görüntülere ihtiyaç duyulduğunu ifade etmektedir.

Kategori 1: Görüntülerde herhangi bir lezyon olmadığını ifade etmektedir. Buna göre, mamografi görüntülerinde anormal bir oluşum gözlenmemektedir. Görüntüde herhangi bir kist, kalsifikasyon veya lezyon görülmemiştir. Mamografi normal özelliklere sahiptir.

Kategori 2: Görüntülerde tespit edilen lezyonların iyi huylu özellikte olduğunu varsaymaktadır. Bu kategoride görüntü üzerinde lezyona benzer yapılar görülmüştür ancak görülen oluşumlar çok büyük olasılıkla iyi huyludur.

Kategori 3: Görüntülerde tespit edilen anormalliklerin büyük çoğunlukla iyi huylu olduğunu varsaymaktadır. Bu kategoride kısa aralıklar ile lezyonun durumunun takip

edilmesi önerilmektedir. Lezyonlarda herhangi bir büyüme veya değişim olması halinde biyopsi önerilebilmektedir.

Kategori 4**:** Şüpheli anormallikler görüntüde açıkça görünmektedir. Anormalliklerin iyi huylu veya kötü huylu özelliklere sahip olma durumu görüntü üzerinde görülmemektedir. Genellikle görüntüde görülen bu anormallikler için biyopsi istenebilmektedir.

Kategori 5: Görüntüde görülen anormallikler büyük oranda kötü karakteristik özelliklere sahiptir. Yine de kesin tanı için biyopsi istenmektedir.

Kategori 6: Görüntüde tespit edilen anormalliklerin kesin bir şekilde kötü karakteristik özelliklere sahip olduğu anlaşılmaktadır [27].

Ultrasonografi; ses dalgalarını kullanarak görüntü oluşturma temeline dayanmaktadır. Bu yöntemle, meme dokusunun iç yapısı görüntülenebilmektedir. Meme yapısı yoğun olan hastalarda kullanılan etkili bir görüntüleme tekniğidir. Mamografi veya klinik muayenede şüpheli lezyonların teşhisinde ek bir görüş için tercih edilmektedir. Ultrasonografi maliyetinin düşük olması, taşınabilirliği, uygulama kolaylığı ve risk taşımaması nedeniyle tercih edilebilmektedir.

Manyetik Rezonans Görüntüleme; meme dokusu içerisindeki yapıları görüntülemek amacıyla büyük mıknatıslar vasıtasıyla oluşturulan güçlü manyetik alan içindeki radyo-frekans dalgalarından yararlanarak görüntü oluşturmaktadır. Vücudun doğal yapısında görünmeyen anormallikleri tespit etmede kullanılmaktadır. Genellikle, yumuşak dokuların görüntülenmesinde genellikle tercih edilmektedir. MR, diğer yöntemlerle tespit edilemeyen lezyonları görüntülemede yaklaşık olarak %20-%25 oranında etkilidir. Genellikle, meme taramasının ilk aşamasında kullanılmamaktadır. Bazı durumlarda ultrasonografiye destek amacıyla kullanılmaktadır. Meme kanseri açısından yüksek riskli hastalarda mamografiye destek amacıyla da kullanılabilmektedir [2].

Dijital Meme Tomosentezinin; mamografiye göre üstünlüğü üç boyutlu kesit görüntü sağlamasıdır. Bu teknik sayesinde dens glandüler doku süperpozisyonu giderilerek hem lezyonu saptama hem de var olan lezyonun karakterize edilmesi kolaylaşmaktadır. Meme Tomosentezi mamografi tekniğinin yalancı pozitiflik oranını

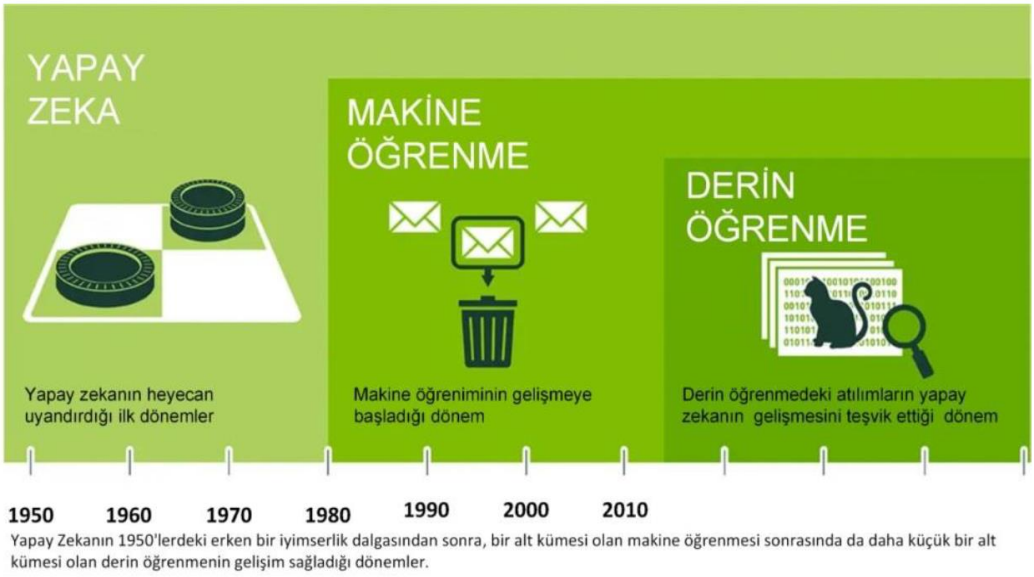
düşürmektedir. Bu yöntemde mamografiye göre ortalama glandüler doz oranı artmaktadır [24, 28].

Meme Biyopsisi, meme görüntüleme tekniklerinde şüpheli bir lezyon saptandığı durumlarda, lezyonun kanserli olup olmadığını anlamak amacıyla meme biyopsisi yapılmaktadır. Meme biyopsisi memeden alınan bir doku örneğinin hücre yapısı düzeyinde patoloji laboratuvarında incelenmesi temeline dayanmaktadır. Meme biyopsisinin birçok farklı türü bulunmaktadır. Alınacak örnek miktarına, meme yapısına ve lezyonun karakteristiğine göre biyopsi türüne karar verilmektedir. En çok kullanılan biyopsi çeşidi iğne türü sebebiyle tru-cut biyopsisidir. Bu teknikte, önce lezyon ultrason ile görüntülenerek iğnenin giriş noktası bulunmaktadır. Noktanın çevresi steril hale getirilmekte ve lokal anestezi yapılmaktadır. Daha sonra ultrason kılavuzluğunda tru-cut iğne ile ile biyopsi işlemi yapılmaktadır. İlgili bölgeden örnek alınmakta ve iğne çıkarılmaktadır. Giriş bölgesi bantlanarak biyopsi işlemi bitirilmektedir [29].

Radyologlar gün içinde çok sayıda mamografi görüntüsünü değerlendirme ve bu değerlendirme süreci radyologlar üzerinde ciddi bir iş yüküne neden olabilmektedir. Ayrıca, görüntüleme yöntemleri ile elde edilen görüntülerin yüksek boyutlu ve karmaşık yapısından dolayı içeriğinin değerlendirilmesi radyologlar için zor ve zaman alıcı bir işlemdir. Son yıllarda radyologların iş yüklerini azaltmak ve meme kanserinin erken teşhisinde radyologlara yardımcı olmak maksadıyla bilgisayar destekli sistemler (BDS) kullanılmaya başlanmıştır. BDS sistemleri gözden kaçan ya da yanlış değerlendirilen lezyon sayısının minimize edilmesi açısından oldukça önemlidir. Yapay zeka destekli BDS sistemleri meme kanseri tarama programlarının geliştirilmesine katkıda bulunmaktadır. Yapılan çalışmalar yapay zeka yöntemleri ile oluşturulmuş karar destek sistemlerinin meme kanseri teşhisinde yüksek bir tanı performansına sahip olduğunu göstermektedir [24].

3. YAPAY ZEKA

Yapay zeka bir veri örneği üzerinden karar alabilen ve tahmin işlemlerini yapılabilme yeteneğine sahip algoritmalar bütünüdür. Yapay zeka kavramı ilk kez 1956 yılında Amerika'da üretilmiştir. Yapay zeka verilerin depolanmasındaki iyileştirmeler ve algoritmaların hesaplama gücünün artırılması sonucunda günümüzde daha da popüler olmaya başlamıştır. Yapay zeka makinelerin deneyimlerinden öğrenmeyi, yeni girdilere uyum sağlamayı ve insana benzer şekilde görevler gerçekleştirmeyi hedefleyen algoritmalar oluşturmayı hedeflemektedir. 1980'li yıllarda yapay zekanın bir alt dalı olan makine öğrenmesi geliştirilmiştir. Makine öğrenmeye ek olarak derin öğrenme de 2010'lu yıllarda kullanılmaya başlanmıştır [30]. Şekil 3.1'de yapay zekanın makine öğrenme sürecinden derin öğrenme sürecine kadar gelişim evreleri gösterilmiştir.



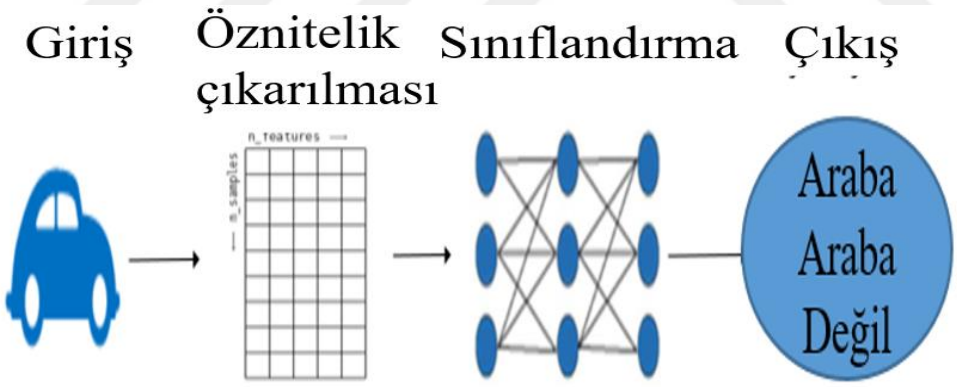
Şekil 3.1: Yapay zeka kavramının gelişim süreci [30].

Yapay zeka yöntemleri olarak sırasıyla makine öğrenme algoritması, yapay sinir ağları ve derin öğrenme algoritması kullanılmaktadır. Şekil 3.2'de yapay zeka türleri arasındaki ilişki gösterilmiştir.

Şekil 3.2: Yapay zeka türleri arasındaki ilişki.

3.1 Makine Öğrenmesi

Makine öğrenmesi; bir veri setinin giriş değerleri ile çıkış değerleri arasında doğrusal bir model oluşturarak problemlerin çözümü için geliştirilmiş hesaplama sistemi olarak tanımlanabilmektedir. Şekil 3.3'de bir makine öğrenme mimarisi gösterilmiştir.



Şekil 3.3: Makine öğrenme mimarisi.

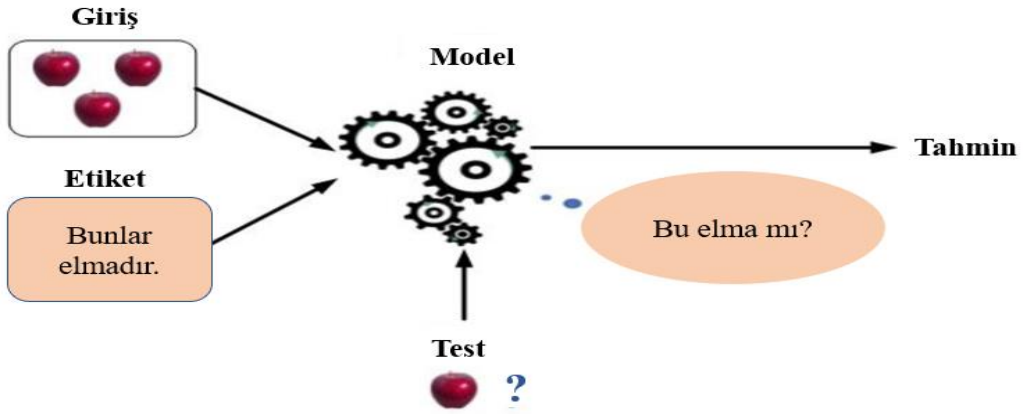
Bir makine öğrenme modeli giriş, öznitelik çıkarılması, sınıflandırma ve çıkış aşamalarından oluşmaktadır. Giriş bölümünde modele verilecek girdiler tanımlanmaktadır. Bu girdiler görüntü veya text verisi olabilmektedir. Bir sonraki aşamada girdilerin tanımlanabilmesi için öznitelik kümesi oluşturulmaktadır. Sınıflandırma aşamasında ise oluşturulan öznitelik kümesine göre hedeflenen sınıf çeşitli modeller kullanılarak tahmin edilmektedir. Makine öğrenmesi; denetimli ve

denetimsiz olmak üzere iki grupta incelenmektedir. Şekil 3.4'de denetimli ve denetimsiz bazı makine öğrenme algoritmaları gösterilmiştir.

Şekil 3.4: Makine öğrenmesi yöntemleri.

3.1.1 Denetimli öğrenme

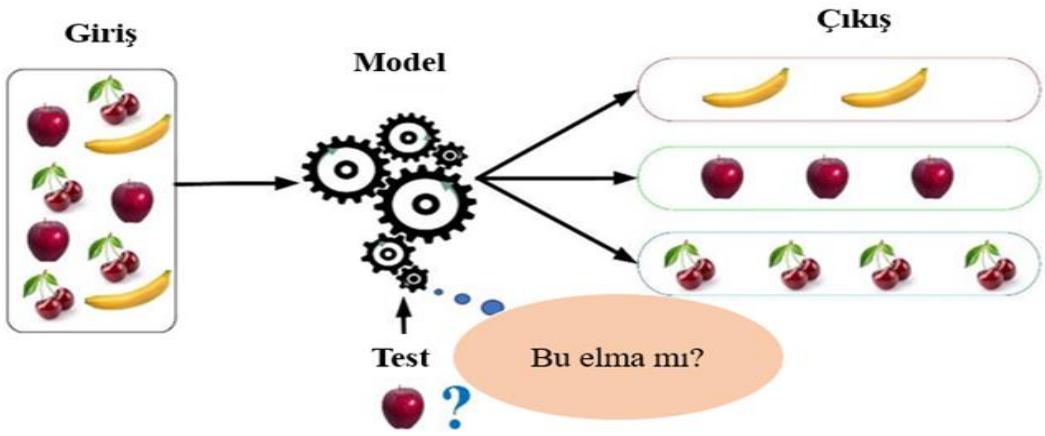
Denetimli öğrenme giriş ve çıkış arasında kesin bir ilişkinin bulunduğu durumlarda kullanılmaktadır. Veri setlerinin tamamının etiketli olması gerekmektedir. Bu etiketli veri setlerinin bir kısmı eğitim sürecinde öğrenme işlemi için, bir kısmı da test verisi için kullanılmaktadır. Eğitim sürecinde öğrenme işlemi etiketli bilgilerden yararlanılarak yapılmaktadır. Test verisi kullanarak tahmin ya da sınıflandırma işlemleri yapılmaktadır. Eğitim sürecinde kullanılan veriler ile eğitilen modelin gerçekte üretmesi gereken değer olup olmadığı denetlenmektedir. Şekil 3.5'de bir denetimli öğrenme modeli gösterilmektedir [30]. Genel olarak sınıflandırma işlemi; veri setlerinin daha önceden etiketleri belirlenmiş olan verilere uygun bir formata ayrılması olarak tanımlanmaktadır. Karar ağacı (KA), destek vektör makineleri (DVM) ve K-en yakın komşu (K-NN) algoritmaları en sık tercih edilen yöntemlerdir. Regresyon; bir veya birden fazla bağımsız değişken ile hedeflenen değişkenin birbirleri ile ilişkilerinin matematiksel olarak ifade edilmesidir [30]. Doğrusal regresyon ve temel bileşenler regresyonu literatürde kullanılan regresyon yöntemlerinden bazılarıdır.



Şekil 3.5: Denetimli öğrenme modeli [30].

3.1.2 Denetimsiz öğrenme

Denetimsiz öğrenme yönteminin denetimli öğrenme yöntemlerinden farkı, veri setlerinin etiketlenmemiş olmasıdır. Denetimsiz öğrenme mevcut olan veri setlerinde bulunan öznitelikler arasındaki ilişkiyi araştırarak veri setlerini kendi içinde gruplandırmaktadır. Şekil 3.6'da bir denetimsiz öğrenme modeli gösterilmektedir. Kümeleme işlemi ise benzer veri setlerinin bir araya getirme işlemidir. K-ortalamalar, hiyerarşik kümeleme ve temel bileşenler analizi en sık kullanılan kümeleme yöntemidir [30].

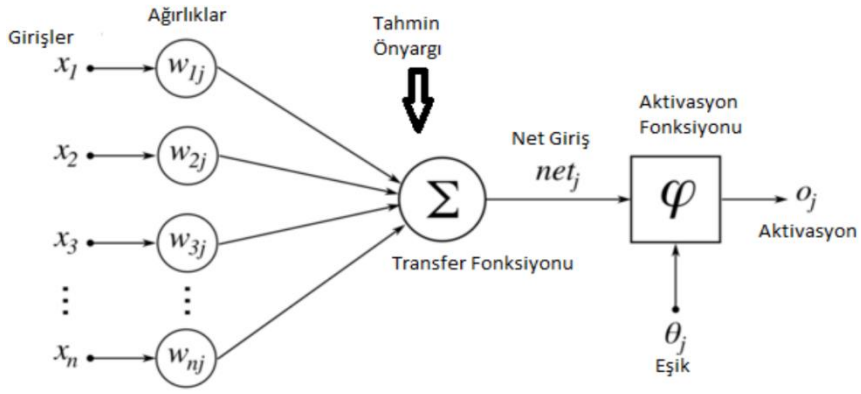


Şekil 3.6: Denetimli öğrenme modeli [30].

3.2 Yapay Sinir Ağları

Yapay sinir ağları (YSA) esas olarak insan beyninin öğrenme, karar verme, algılama gibi yetenekleri taklit ederek bu süreçleri otomatik olarak uygulamayı amaçlayan algoritmalar bütünüdür. Genel olarak YSA sistemi insan beynindeki biyolojik sinir

hücrelerinin yapılarından esinlenilerek oluşturulmaktadır. Sinir hücreleri, yeni bilgileri çözümleme ve başka hücrelere bu bilgiyi aktarabilme yeteneğine sahiptir [31]. Şekil 3.7'de bir YSA modeli yapısı gösterilmiştir.



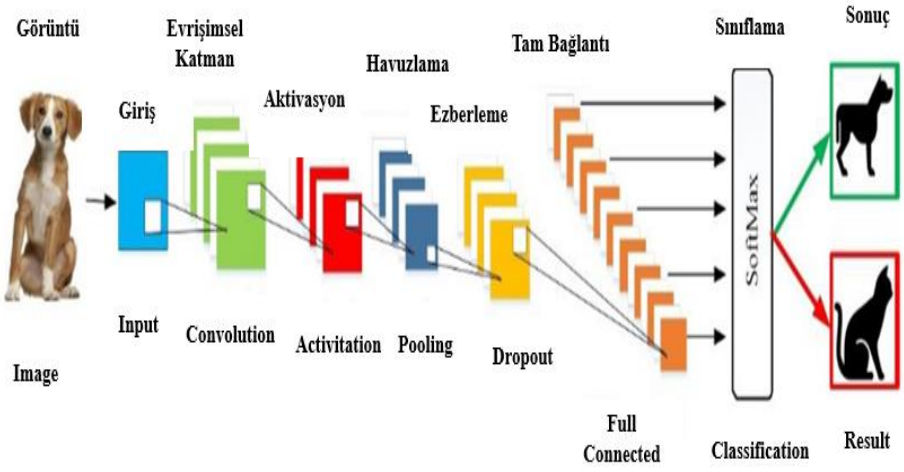
Şekil 3.7: Yapay sinir ağları yapısı [31].

Şekil 3.7'de görüleceği üzere bir hücreye n tane giriş verisi verilmektedir. Giriş verileri ağırlıklar ile çarpılarak tüm veriler toplanmaktadır. Bir sonraki aşamada önyargı eklenmektedir. Bunun sonucunda da net yargı elde edilmektedir. Daha sonraki aşamalarda ise net girdi bir aktivasyon fonksiyonundan geçirilmekte ve bir çıktı verisi oluşturulmaktadır

3.3 Derin Öğrenme

Derin öğrenme; çok katmanlı sinir ağlarından bir model oluşturan ve bu modele sunulan veri setlerinin hiyerarşisini otomatik olarak yapılandıran bir makine öğrenme algoritmasıdır. Makine öğrenme ve YSA'dan farklı olarak, derin öğrenme veri setlerinden otomatik olarak öznitelik çıkarabilmektedir. Öznitelik çıkarma ve dönüştürme işlemleri için doğrusal olmayan işlem birim katmanından yararlanılmaktadır. Derin öğrenmede her bir katman önceki katmanın çıkış bilgisini giriş bilgisi olarak kabul etmektedir. Derin öğrenme, denetimli ya da denetimsiz olarak modellenebilmektedir. Genel olarak derin öğrenme, veri setleri birçok öznitelik seviyesinin öğrenilmesine dayanan bir model yapısına sahiptir. Üst seviyedeki öznitelikler daha alt düzeydeki özniteliklerden üretilmekte ve bu şekilde hiyerarşik bir yapı oluşturulmaktadır. Derin öğrenme modellerine örnek olarak evrişimsel sinir ağları, uzun-kısa vadeli hafıza ağları, tekrarlayan ağlar, derin inanç sinir ağları, sınırlı

Boltzmann makineleri ve derin oto kodlayıcılar verilebilmektedir [32-33]. Şekil 3.8'de bir derin öğrenme modeli gösterilmektedir.



Şekil 3.8: Derin öğrenme modeli [32].

Genel olarak bir derin öğrenme modeli; giriş, ara ve çıkış katmanlarından oluşmaktadır. Giriş katmanı verilerin standart bir formatta hazırlanarak ağa sunulduğu bölümdür. Ara katmanlar sırasıyla evrişimsel, aktivasyon, havuzlama, ezberleme ve tam bağlı aşamalarından oluşmaktadır. Evrişimsel katmanda giriş katmanında belirlenmiş olan bir filtre veriler üzerinde gezdirilerek verilerden ayırt edici öznitelikler çıkarılmaktadır. Aktivasyon katmanında elde edilen öznitelik değerlerinin belirli bir aralığa indirme işlemi yapılmaktadır. Özniteliklerin daha küçük aralıklara indirgendiği katmana havuzlama denmektedir. Bazı durumlarda eğitim sürecinde sistem kullanılan verileri ezberleyebilmektedir. Bu nedenle tasarlanmış olan ağ yapısında bu verilerin ağa unutturulması gerekebilmektedir.Verilerin ağa unutturulma işlemi ezberleme olarak tanımlamaktadır. Çok boyutlu öznitelik kümesinin tek boyuta indirgendiği bölüm tam bağlı katman olarak isimlendirilmektedir. Son bölüm ise sınıflandırmanın yapıldığı çıkış katmanıdır. Bu katmanda kendisinden daha önce oluşturulan bilgiler değerlendirilerek ağ modelinin çıkış değerleri oluşturulmaktadır. Bu katmanda çıkış değerlerinin [0-1] arasına sıkıştırılmasını sağlayan ve olasılıksal bir hesaplamaya dayanan yumuşatma işlevi kullanılmaktadır [32].

4. LİTERATÜR ÇALIŞMALARI

Meme kanseri dünyada çok sayıda kadının ölümüne neden olan hastalıklardan biridir. Son yıllarda meme kanserinin teşhisi ile ilgili olarak çok sayıda çalışma yapılmıştır. Bu bölümde makine öğrenme algoritmalarını kullanarak meme kanserinin teşhisi ile ilgili olarak yapılmış güncel çalışmalar özetlenmiştir.

Asri ve diğ. [34] çalışmalarında Naive Bayes (NB), DVM, KA) ve K-NN gibi makine öğrenme yöntemlerinin performanslarını karşılaştırmıştır. Tüm algoritmalar Wisconsin Üniversitesi meme kanseri veri (WBCD) setinde uygulanmıştır. Yapılan deneyler sonucunda, %97,3 doğruluk oranı, %98 kesinlik oranı, %97 duyarlılık ve %97 F1-puanı ile DVM yöntemi diğer yöntemlerden daha iyi bir performans göstermiştir.

Naji ve diğ. [35] WBCD veri seti üzerinde DVM, Lojistik Regresyon (LR), KA, K-NN ve Rastgele Orman (RO) yöntemlerin performanslarını karşılaştırmıştır. DVM yöntemi %97,2 doğruluk oranı ile diğer yöntemlerden daha iyi bir performans göstermiştir.

Amrane ve diğ. [36] WBCD veri seti üzerinde NB ve K-NN algoritmalarını uygulamış ve algoritmaların performanslarını karşılaştırmıştır. K-NN algoritması %97,51 doğruluk oranı ile en iyi sonucu göstermiştir.

Ak [37] çalışmalarında WBCD veri seti üzerinde LR, DVM, K-NN, NB, DT, RO yöntemlerinin sınıflandırma performanslarını karşılaştırmıştır. LR yöntemi %98,1 doğruluk oranı ile en iyi sonucu göstermiştir.

Khan ve diğ. [38] WBCD veri seti üzerinde RO, KA, K-NN, LR yöntemlerinin performanslarını karşılaştırmıştır. Çalışmada, LR yöntemi %98 doğruluk oranı ile en iyi sonuca ulaşmıştır.

Omondiagbe ve diğ. [39] WBCD veri seti üzerinde DVM, YSA ve NB yöntemlerini karşılaştırmıştır. Öznitelik seçim yöntemi olarak Doğrusal Diskriminant Analiz yöntemi tercih edilmiştir. DVM yöntemi %98,82 doğruluk, %98,41 duyarlılık ve %99,07 özgüllük oranı ile en iyi sonuca ulaşmıştır.

Khandezamin ve diğ. [40] Wisconsin Meme Kanseri veri setleri üzerinde meme kanserinin erken teşhisi amacıyla LR-Veri İşleme Grup Yöntemi(VİGY) tabanlı bir model önermiştir. Çalışmada; LR yöntemi öznitelik seçim yöntemi, VİGY ise sınıflandırma işlemleri için kullanılmıştır. Önerilen hibrit yöntem sırasıyla %97,9, %99,1 ve %84,6 doğruluk oranına ulaşmıştır.

Haq ve diğ. [41] Wisconsin Meme Kanseri veri setleri üzerinde meme kanserinin erken teşhisi amacıyla farklı öznitelik seçim yöntemlerinin karşılaştırmıştır. Öznitelik seçim yöntemi olarak sırasıyla Relief, Temel Bileşen Analizi ve Otokodlayıcı metotlarını kullanmıştır. Sınıflandırma için DVM algoritması tercih edilmiştir. Çalışmanın sonucunda Relief-DVM hibrit yöntemi %99,91 doğruluk oranı ile en yüksek başarıyı göstermiştir.

Bacha ve diğ. [42] yeni bir yöntem olarak Radyal Tabanlı Kernel Aşırı Öğrenme Makineleri yöntemini geliştirmiş ve geliştirdikleri yöntemi Mamographic Image Analysis Society (MIAS) ve WBCD veri setleri üzerinde test etmiştir. Önerilen yöntem MIAS için %100 doğruluk ve WBCD için ise %91,13 doğruluk oranı göstermiştir.

Vadivel ve diğ. [43] çalışmalarında meme tümörlerini bulanık mantık kuralları ile karakterize etmiş ve mamografi görüntülerinden farklı geometrik öznitelikler üretmiştir. Çalışmada Tarama Mamografisi için Dijital Veri Tabanı veri seti kullanılmıştır. Sınıflandırma için C5.0 algoritması kullanılmış ve algoritma %100 doğruluk oranına ulaşmıştır.

Jadoon ve diğ. [44] çalışmalarında DVM ve Evrişimli Sinir Ağı yöntemlerini Mamografi için Dijital Veri Tabanı veri seti üzerinde uygulamıştır. Evrişimsel Sinir Ağı %83,74 doğruluk oranı ile en başarılı sonuca ulaşmıştır.

Punitha ve diğ. [45] iyi ve kötü huylu meme tümörlerini sınıflandırmak için YSA modelini kullanmıştır. Öznitelik çıkarım yöntemi olarak Gri Seviye Eş Oluşum Matrisi (GSEOM) ve Gri Seviye Koşu Uzunluğu Matrisi (GSKUM) metotlarını kullanılmıştır. Önerilen yöntem Mamografi için Dijital Veri Tabanı veri seti üzerinde test edilmiş ve %98 doğruluk oranına ulaşmıştır.

Bajcsi ve diğ. [46] meme kanserini sınıflandırmak amacıyla KA ve RO algoritmalarını kullanmıştır. Öznitelik çıkarım yöntemi olarak GSKUM metodu kullanılmıştır.

Önerilen yöntem MIAS veri seti üzerinde test edilmiş ve %100 doğruluk oranına ulaşmıştır.

Wang ve diğ. [47] farklı makine öğrenme yöntemlerini kullanarak meme kanserinin teşhisi için bir çalışma gerçekleştirmiştir. Çalışmanın amacı hasta kayıtlarının klinik verilerine dayanarak meme kanserinin etkin bir şekilde tespit edilmesini sağlamaktır. Deneyler WBCD veri seti üzerinde test edilmiştir. Sınıflandırma algoritması olarak sırasıyla DVM, YSA, NB ve Adaboost ağacı yöntemlerini kullanılmıştır. Çalışmalarında öznitelik seçim yöntemi olarak Temel Bileşenler Analizi yöntemini tercih etmişlerdir.

Kumar ve diğ. [48] meme kanserinin sınıflandırılması amacıyla Bayes optimizasyon yöntemi ve RO algoritmasını kullanarak hibrit bir metot önermişlerdir. Önerilen yöntem WBCD veri seti üzerinde test edilmiş ve %97,9 doğruluk oranına ulaşmıştır.

Bensaoucha [49] çalışmasında meme kanserinin tespiti için farklı makine öğrenme algoritmalarını WBCD veri seti üzerinde test eden bir model önermiştir. Makine öğrenme yöntemlerinin hiperparametreleri Bayes optimizasyon (BO) yöntemi ile belirlenmiştir. DVM yöntemi %96,52 doğruluk oranı ile en yüksek başarı oranına ulaşmıştır.

Mate ve Somai [50] meme kanserini sınıflandırmak amacıyla öznitelik seçim yöntemleri ile BO yöntemini birleştiren bir hibrit bir yöntem kullanmıştır. Öznitelik seçim yöntemi olarak Ki-kare, Pearson Katsayısı, RO, LR, Özyinelemeli Özellik Eliminasyonu ve Hafif Gradyan Artırma kullanılmıştır. Sınıflandırma için K-NN, LR, NB, RO, Ekstra Ağaçlar ve Quadratik Diskriminant Analizi gibi yöntemler kullanılmıştır. Deneyler WBCD veri seti üzerinde test edilmiştir. Ekstra Ağaçlar yöntemi %96,2 doğruluk oranı ile en yüksek başarı oranına ulaşmıştır.

Dhanya ve diğ. [51] meme kanserini sınıflandırmak amacıyla farklı makine öğrenme ve öznitelik seçim algoritmalarını kullanmıştır. Deneyler WBCD üzerinde test edilmiştir. Deneyler sonucunda, NB ve Ardışık İleri Yönde Seçim (AİYS) hibrit yöntemi %98,24 doğruluk oranı ile en yüksek başarı oranına ulaşmıştır.

Kumari ve diğ. [52] meme kanserini sınıflandırmak amacıyla Gradyan Artırma Algoritması, RF, K-NN, YSA ve DVM yöntemlerini kullanmıştır. Öznitelik çıkarım yöntemi olarak GSEOM metodunu kullanmışlardır. Deneyler MIAS veri seti üzerinde

test edilmiştir. Gradyan Artırma Algoritması %95,6 doğruluk oranı ile en yüksek başarı oranına ulaşmıştır.

Vijayarajeswari ve diğ. [53] meme kanserini sınıflandırmak amacıyla DVM yöntemini kullanmıştır. Öznitelik çıkarım yöntemi olarak Hough Dönüşüm yöntemini kullanılmıştır. Önerilen metot MIAS veri seti üzerinde test edilmiş ve %94 oranında başarı oranına ulaşmıştır.

Ancy ve Nair [54] meme kanserini sınıflandırmak amacıyla DVM yöntemini kullanmıştır. Segmentasyon için medyan filtre, gri seviye eşikleme ve morfolojik yöntemleri kullanılmıştır. Öznitelik çıkarım yöntemi olarak GSEOM metodunu kullanmıştır. Önerilen yöntem MIAS veri seti üzerinde test edilmiş ve %81 oranında başarı sağlamıştır.

Alshammari ve diğ. [55] meme kanserini sınıflandırmak amacıyla farklı makine öğrenme yöntemlerini kullanmıştır. Çalışma kapsamında 13 adet doku ve morfolojik öznitelik çıkarılmıştır. Önerilen yöntem Imam Abdulrahman Bin Faisal Universitesi'nden alınan 42 adet mamografi görüntüsü üzerinde test edilmiştir. Segmentasyon için MATLAB Segmenter Tool programını kullanılmıştır. Makine öğrenme yöntemlerinin hiperparametreleri BO yöntemi ile belirlenmiştir. Öznitelik seçim algoritması olarak AYİS yöntemi kullanılmıştır. Deneyler sonucunda, DVM ve NB yöntemleri %100 oranında başarı sağlamıştır.

Farid ve diğ. [56] meme kanserinin erken teşhisi amacıyla DVM ve Genetik Algoritma tabanlı bir model önermiştir. Deneyler WBCD veri seti üzerinde test edilmiştir. Önerilen hibrit yöntem %98,25 oranında başarı sağlamıştır.

Ergin ve diğ. [57] meme kanserinin erken teşhisi amacıyla GSEOM tabanlı bir model önermiştir. Segmentasyon için bölge büyütme yöntemini, öznitelik çıkarım yöntemi olarak GSEOM yöntemini, sınıflandırma için ise Fisher'in Doğrusal Diskriminant Analiz yöntemini tercih edilmiştir. Deneyler MIAS veri seti üzerinde test edilmiştir. Deneyler sonucunda, önerilen yöntem %82,48 doğruluk oranına ulaşmıştır.

Farhan ve diğ. [58] mamografi görüntülerinin iyi huylu veya kötü huylu olarak sınıflandırılması amacıyla farklı öznitelik çıkarım yöntemlerini karşılaştıran bir yöntem önermiştir. Öznitelik çıkarım yöntemi olarak Yerel İkili Örüntü, Yönlendirilmiş Gradyanların Histogramı ve GSEOM metotlarını

kullanılmıştır. Deneyler MIAS veri seti üzerinde test edilmiştir. Sınıflandırma için LR ve DVM algoritmalarını kullanılmıştır. Deneyler sonucunda, Yerel İkili Örüntü-LR yöntemi %92,5 doğruluk oranı ile en yüksek sonuca ulaşmıştır.

Wang ve diğ. [59] mamografi görüntülerinin iyi-kötü huylu sınıflandırılması amacıyla bir yöntem önermiştir. Çalışmada 288 adet mamografi görüntüsü kullanılmıştır. 188 adet doku ve morfolojik öznitelik çıkarılmıştır. Öznitelik seçim algoritması olarak LASSO yöntemi, sınıflandırma için DVM yöntemi kullanılmıştır. Eğitim verisi için 97,39%, test verisi için ise %98,7 duyarlılık oranı elde edilmiştir.

Li ve diğ. [60] mamografi görüntülerinin iyi-kötü huylu sınıflandırılması amacıyla bir yöntem önermiştir. Çalışmada 463 adet mamografi görüntüsü kullanılmıştır. 846 adet doku ve morfolojik öznitelik çıkarılmıştır. Öznitelik seçim algoritması olarak LASSO yöntemi, sınıflandırma için DVM ve LR algoritmaları kullanılmıştır. Deneyler sonucunda 10 adet öznitelik seçilmiş ve %80 doğruluk oranı ile DVM yöntemi en iyi oranına ulaşmıştır.

Stelzer ve diğ. [61] mamografi görüntülerinde saptanan mikrokalsifikasyonlarının sınıflandırılması amacıyla doku öznitelikleri ve makine öğrenme yöntemlerini birlikte kullanan bir yöntem önermiştir. Çalışmada 226 adet mamografi görüntüsü, 249 adet gri seviye histogram, GSEOM ve GSKUM doku öznitelik yöntemleri ve öznitelik yöntemi olarak da temel bileşen analizi, sınıflandırma algoritması olarak çok katmanlı algılayıcı kullanılmıştır. Çalışmanın sonucunda %98,4 doğruluk oranı elde edilmiştir.

Nugroho ve diğ. [62] mamografi görüntülerinin sınıflandırılması amacıyla histogram ve GSEOM yöntemlerini kullanarak toplam 12 adet öznitelik çıkarmıştır. Öznitelik seçim algoritması olarak Korelasyona dayalı öznitelik seçim yöntemini tercih edilmiştir. Makine öğrenme algoritması olarak çok katmanlı algılayıcı kullanılmıştır. 40 adet mamografi görüntüsü kullanılmış ve sunulan yöntem %91,66 doğruluk oranına ulaşmıştır.

Vijayarajeswari ve diğ. [63] mamografi görüntülerinin sınıflandırılması amacıyla etkin bir sınıflandırma yöntemi önermiştir. MIAS veri seti üzerinden 95 adet mamografi görüntüsü değerlendirilmiştir. Öznitelikler Hough dönüşümünden yararlanılarak çıkarılmıştır. Sınıflandırma işlemi için DVM algoritması kullanılmıştır. Çalışma sonucunda %94 doğruluk oranı elde edilmiştir.

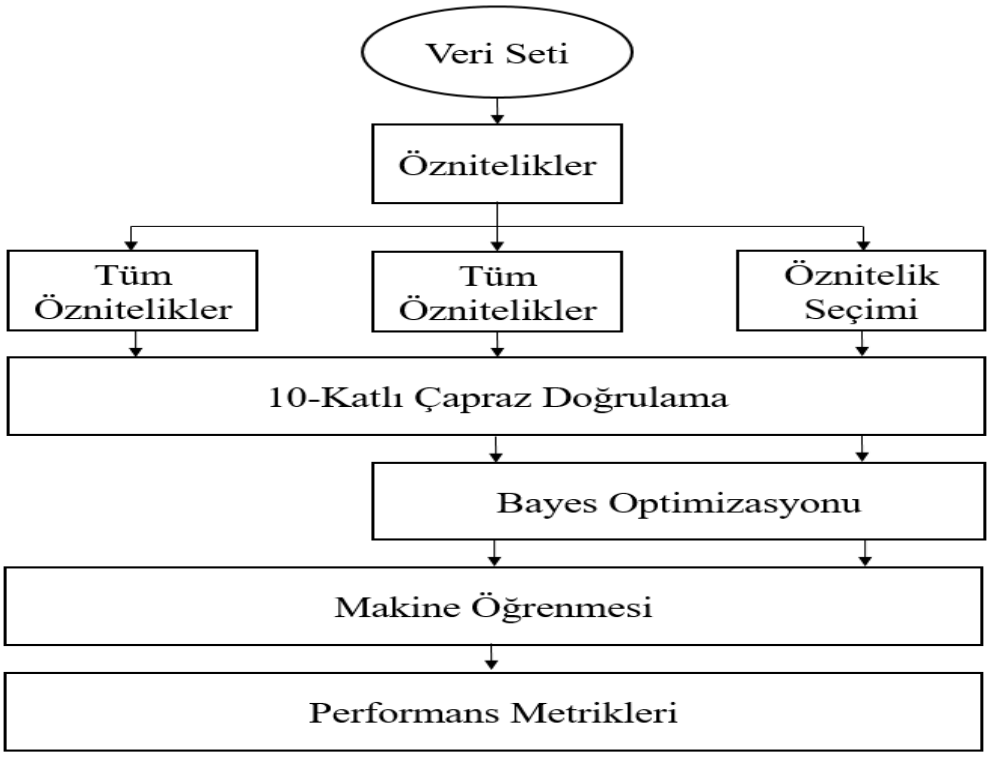
Ghergeout ve diğ. [64] mamografi görüntülerinin sınıflandırılması amacıyla bir yöntem önermiştir. Görüntülerden öznitelik çıkarmak için GSEOM ve GSKUM yöntemlerini kullanılmıştır. Öznitelik seçim yöntemi olarak Relief ve Minimum Artıklık Maksimum İlgililik tekniklerini, sınıflandırma işlemi için Geri Yayılım Algoritmasını kullanılmıştır. Önerilen yöntem MIAS veri seti üzerinde test edilmiştir. Çalışma sonucunda %95,2 iyi huylu tümörler için %92 kötü huylu tümörler için sınıflandırma performansı elde edilmiştir.

Sapate ve diğ. [65] mamografi görüntülerinin iyi-kötü huylu sınıflandırılması amacıyla bir yöntem önermiştir. Çalışmada 460 mamografi görüntüsü kullanılmıştır. Mamografi görüntülerinden 48 adet geometrik ve doku öznitelikleri çıkarılmıştır. Sınıflandırma algoritması olarak K-NN ve DVM tercih edilmiştir. Çalışma sonucunda DVM algoritması %85,56 doğruluk oranına ulaşmıştır.

Loizidou ve diğ. [66] mamografi görüntülerinin iyi-kötü huylu sınıflandırılması amacıyla bir yöntem önermiştir. 320 adet mamografi görüntüsü kullanılmıştır. Görüntülerden çeşitli şekil, histogram ve doku öznitelikleri çıkarılmıştır. Öznitelik seçim algoritması olarak ANOVA ve t-test kullanılmıştır. Sınıflandırma için 6 farklı algoritma kullanılmış ve DVM algoritması %99,55 doğruluk oranına ulaşmıştır.

5. MATERYAL VE YÖNTEM

Bu çalışmada meme kanserinin etkin ve doğru bir şekilde tespiti için önerilen hiper-parametre optimizasyon-öznitelik seçim tabanlı geliştirilmiş makine öğrenme yöntemi tasarlanmıştır. Önerilen sınıflandırma sisteminin çalışma mekanizması Şekil 5.1'de gösterilmiştir.



Şekil 5.1: Sistem akış şeması.

Sistemin çalışma mekanizmasındaki adımlar şöyledir:

1. Veri setinin yüklenmesi
2. Özniteliklerin belirlenmesi
3. Özniteliklerin ölçeklendirilmesi
4. 10-katlı çapraz doğrulama yöntemi ile veri setinin eğitim ve test verisi olarak bölünmesi
5. Sınıflandırma
    a. Öznitelik seçim yöntemleri ve hiperparametre optimizasyonu yöntemi uygulamadan sınıflandırma
    b. BO tabanlı makine öğrenme algoritmalarının kullanılması

c. Öznitelik yöntemleri-BO optimizasyon tabanlı makine öğrenme algoritmalarının kullanılması

6. Deney sonuçlarının performans metrikleri ile karşılaştırılması

5.1 Veri Setleri

Bu çalışmada, meme kanseri teşhisi için iki farklı kanser veri seti kullanılmıştır. Veri setlerinden birincisi literatürde konu ile ilgili çalışmalarda sık kullanılan Wisconsin Meme Kanseri Veri Seti (Wisconsin Diagnostic Breast Cancer-WBCD) veri seti, ikincisi hastane ortamında oluşturulmuş Mamografi Meme Kanseri Veri seti (Mamographic Breast Cancer Dataset-MBCD) veri setidir.

5.1.1 Wisconsin meme kanseri veri seti

Bu veri seti, WBCD Wisconsin Üniversitesi Genel Cerrahi bölümünden Dr.William Wolberg tarafından iğne ucu genişliğindeki bir meme kitlesinin biyopsi ile alınarak görüntülenmesi ve bu görüntülerin Wisconsin Üniversitesi Bilgisayar bölümü araştırmacılarından William Nick Street tarafından sayısallaştırılması ile oluşturulmuştur. Veri seti Kaliforniya Üniversitesi-Irvine'de bulunan Makine Öğrenme Deposunda kamuya açık bir şekilde paylaşıma sunulmuştur. Veri setinde 212 tanesi kötü, 357 tanesi iyi huylu olmak üzere toplam 569 örnek bulunmaktadır. Her bir veri örneği için 30 tanımlayıcı özellik, bir adet teşhis sınıfı, bir adet hasta kimliği olmak üzere toplam 32 adet öznitelik veri setinde bulunmaktadır. 30 adet özniteliğin 10 tanesi tümör hücresinin çekirdeği üzerinden direkt olarak ölçümlenmiş, 20 tanesi ise bunlara bağlı olarak hesaplanmış sayısal değerleri ifade etmektedir [67-69]. Veri seti için tanımlanan öznitelikler Çizelge 5.1'de gösterilmiştir.

5.1.2 Mamografi meme kanseri veri seti

Çalışmada kullanılan ikinci veri seti, retrospektif olarak 2015-2020 yılları arasında Ankara Eğitim ve Araştırma Hastanesi Radyoloji bölümünde tetkik görmüş olguların mamografi görüntülerinden oluşmaktadır. Retrospektif olarak toplanan bu veri seti için Ankara Eğitim ve Araştırma Hastanesi Etik Kurulu'ndan (319/E-20 sayılı karar ile-EK) onay alınmıştır. Retrospektif çalışmanın doğası gereği hasta bilgileri göz ardı edilmiştir.

Çizelge 5.1: WBCD veri seti öznitelikleri.

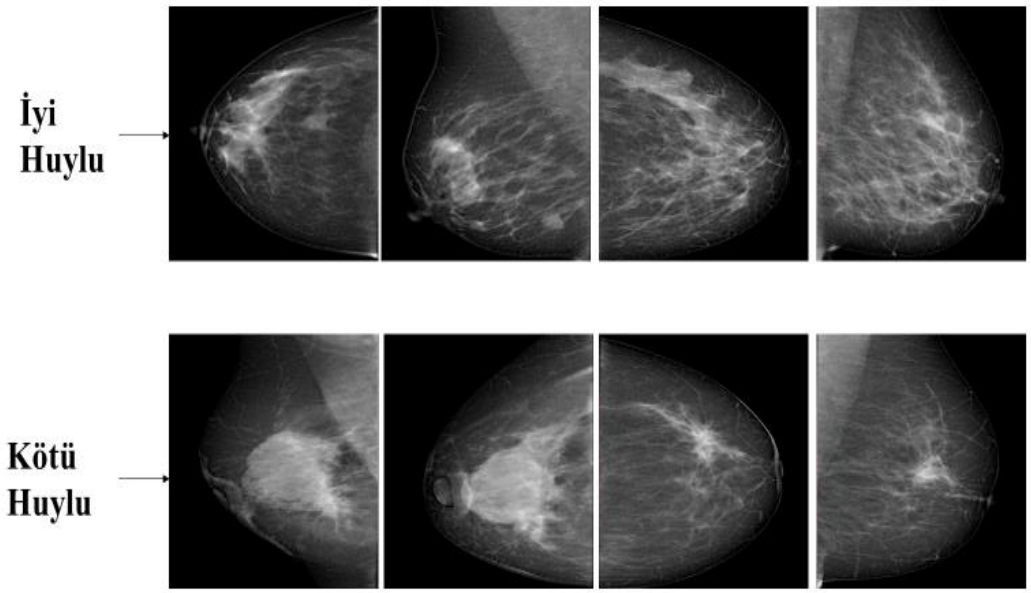
No	Öznitelik	No	Öznitelik	No	Öznitelik
1	Ortalama yarıçap	11	Yarıçap şiddeti	21	En kötü yarıçap
2	Ortalama doku	12	Doku şiddeti	22	En kötü doku
3	Ortalama çevre	13	Çevre şiddeti	23	En kötü çevre
4	Ortalama alan	14	Alan şiddeti	24	En kötü alan
5	Ortalama pürüzsüzlük	15	Pürüzsüzlük şiddeti	25	En kötü pürüzsüzlük
6	Ortalama yoğunluk	16	Yoğunluk şiddeti	26	En kötü yoğunluk
7	Ortalama içbükeylik	17	İçbükeylik şiddeti	27	En kötü içbükeylik
8	Ortalama içbükeylik noktası	18	İçbükeylik noktası Şiddeti	28	En kötü içbükeylik noktası
9	Ortalama simetri	19	Simetri şiddeti	29	En kötü simetri
10	Ortalama fraktal boyut	20	Fraktal boyut şiddeti	30	En kötü fraktal boyut

Mamografi görüntülerinde kitlesi olan olgular, patoloji incelemesi sonucunda meme kitlelerinin iyi veya kötü huylu olarak belirlenmiş olan olgular, patolojik incelemesi yoksa en azından 2 yıllık takip sonrası iyi huylu meme kitlesine sahip olan olgular çalışmaya dahil edilirken, mamografi görüntülenmesinden önce herhangi bir tedavi geçmişi olan olgular ve düşük kaliteye sahip olan mamografi görüntüleri olan olgular çalışma dışında bırakılmıştır. Sonuç olarak 101 olguya ait mamografi görüntüleri çalışmaya dahil edilmiş olup bu hastaların 40 tanesi iyi huylu meme kitlesine sahip iken, 61 tanesi de kötü huylu meme kitlesine sahiptir. Çalışma kapsamında kullanılan bütün görüntüler IMS Giotto (Bologna-İtalya) dijital mamografi cihazı ile kaydedilmiştir. Bütün görüntüler Ankara Eğitim ve Araştırma Hastanesi PACS (Picture Archiving Communication Systems-Görüntü Arşivleme ve İletişim Sistemleri) sisteminden alınmış ve DICOM (Tıpta Dijital Görüntüleme ve İletişim-Digital Imaging and Communications in Medicine) formatında kaydedilmiştir. 101

olguya ait meme kitlesinin göründüğü MLO ve/veya CC projeksiyonunda toplam 195 adet mamografisi kullanılmıştır. Çalışma kapsamında kullanılan MBCD veri setine ait olgu sayısı, yaş bilgisi, görüntü sayısı, kötü ve iyi huylu olgu sayısı bilgileri Çizelge 5.2'de gösterilmiştir. Mamografi veri setindeki iyi ve kötü huylu bazı örnek görüntüler Şekil 5.2'de gösterilmiştir.

Çizelge 5.2: Mamografi meme kanseri veri seti bilgileri.

OLGU POPULASYONU	
Olgu sayısı	101
YAŞ	
Ortalama	57.5
Standart sapma	12.4
En küçük	34
En büyük	89
İYİ-KÖTÜ HUYLU	
İyi Huylu	40
Kötü Huylu	61
GÖRÜNTÜ SAYISI	
İyi Huylu	79
Kötü Huylu	116



Şekil 5.2: Mamografi veri setindeki iyi ve kötü huylu bazı örnek görüntüler.

5.1.2.1 Mamografi görüntülerindeki şüpheli bölgelerin belirlenmesi ve bölütlenmesi

Mamografi görüntülerindeki meme lezyonlarının tespitini yapabilmek için görüntülerden ilgili bölgelerin çıkarılması gerekmektedir. Mamografi görüntülerinden şüpheli meme lezyonlarının bulunduğu ilgili bölgelerin (ROI) çıkarılması işlemine bölütleme işlemi adı verilmektedir. Bölütleme sonucunda sadece lezyonun bulunduğu bölgelerin tutulması ve gereksiz kısımların görüntüden çıkarılması amaçlanmaktadır. Mamografi görüntülerindeki meme lezyonlarının bulunduğu ilgili bölgelerin görüntüden çıkarılması için sunulan algoritma Şekil 5.3'de gösterilmektedir.

Şekil 5.3: Mamografi görüntülerinden şüpheli lezyonların çıkarılması.

Görüntülerin İşaretlenmesi: 20 yıllık ve 5 yıllık iki radyolog tarafından ortak bir fikir birliğine varılması sonucu, RadiANT DICOM Viewer programı yardımıyla meme lezyonlarının sınırları yeşil renkle belirlenmiştir.

Gri Seviye Yoğunluk Eşikleme: Yeşil renkle belirlenen sınırlar üzerinde gri seviye yoğunluk eşikleme algoritması uygulanmıştır. Eşikleme algoritması temel olarak görüntüyü oluşturan piksel matris değerlerinin belli bir değerden küçük veya büyük olanların başka bir değere eşitlenmesidir. Bu şekilde istenmeyen yoğunluk değerine sahip olan piksel değerleri görüntü matrisinden çıkarılmış olmaktadır. Eşik değerinin alt sınırı "minimum yoğunluk eşiği", üst sınırına ise "maksimum yoğunluk eşiği" olarak tanımlanmaktadır. Alt sınır ve üst sınır arasındaki piksel değerleri lezyon olabilecek ilgi alanlarını (ROI) oluşturmaktadır. Bu iki yoğunluk değeri arasında kalan piksel değerleri nodül olabilecek ilgi alanlarını oluşturmaktadır. Minimum yoğunluk

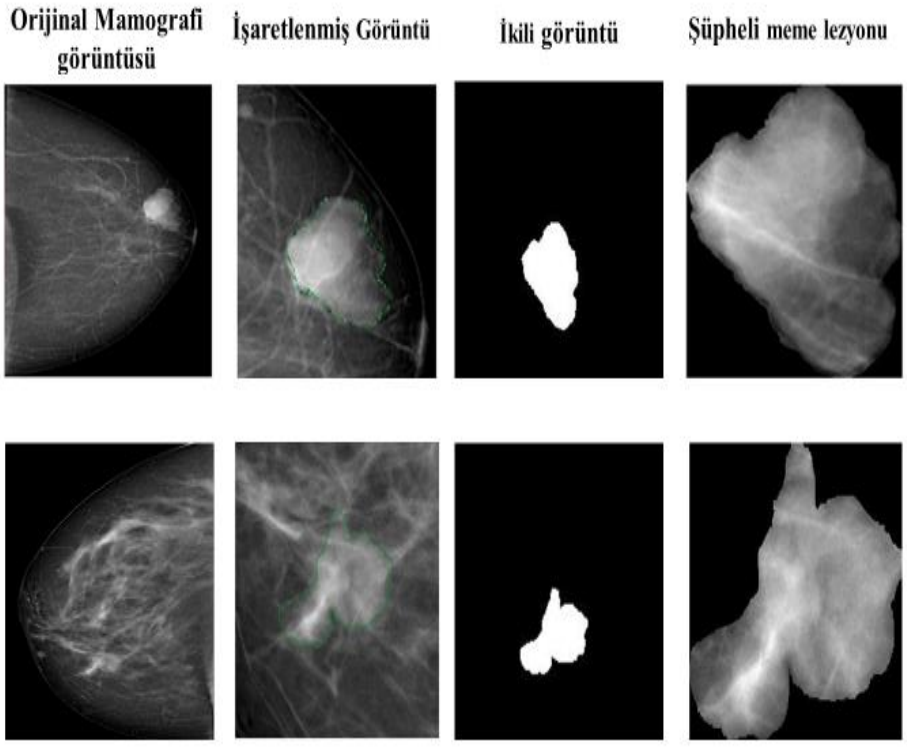
eşiğinden küçük ve maksimum yoğunluk eşiğinden büyük piksel değerleri 0 değerini almakta ve bu alanlar siyah olmaktadır [70]. Bu çalışmada kullanılan mamografi görüntülerinde radyologlar tarafından yeşil renkle işaretlenen kitlelerin gri düzey aralıkları belirlenmiştir. Uzun denemelerden sonra minimum yoğunluk eşiği olarak 100, maksimum yoğunluk eşiği olarak da 145 piksel değerleri belirlenmiş ve bu iki değer için eşikleme yapılmıştır.

İkili görüntüye çevirme: Gri seviye yoğunluk eşikleme yönteminden sonra görüntüler ikili görüntüye çevrilmektedir. Sonuç olarak sadece "1" ve "0" değerlerinden oluşan bir matris oluşmaktadır. İkili görüntü üzerinde komşuluk incelemesi yapılarak bir yapıyı oluşturan pikseller tespit edilebilmektedir.

Morfolojik İşlemler: Eşikleme ve ikili görüntüye çevirme işlemlerinden sonra görüntülerde saçak, girinti veya boşluk gibi bozukluklar meydana gelebilmektedir. Morfolojik işlemler sayesinde görüntü üzerinde oluşan bozuklukları giderilerek yapıyı oluşturan pikseller daha da belirgin hale getirilebilmektedir. Aşındırma ve genleştirme genel olarak iki temel morfolojik işlemdir. Aşındırma işlemi ikili görüntüdeki yapıyı küçültmeye veya inceltmeye çalışmaktadır. Genleştirme işlemi iki görüntüdeki yapıyı büyütmeye veya kalınlaştırmaya çalışmaktadır. Aşındırma işlemi görüntü üzerinde birbirine ince bir gürültü ile bağlanmış olan iki veya daha fazla sayıda yapının birbirinden ayrılması için, genişletme işlemi ise aynı yapının bir gürültü ile ince bir biçimde bölünmesi ile iki ayrı yapı olarak görünmesini engellemek amacıyla kullanılmaktadır. Bu iki işlem aslında birbirinin tersi olarak tanımlanmaktadır. Görüntü üzerindeki bölgelerde aşındırma ve genişletme işlemlerinden birisi uygulandığında komşu bölgeler zıt olan işleme tabi tutulmaktadır. Başka bir deyişle, aşındırma işlemi gerçekleştirilirken komşu alanda genişletme işlemi uygulanmış olmaktadır [71-72].

Lezyonun Çıkarılması: Morfolojik işlemler uygulandıktan sonraki ikili görüntü ile orijinal mamografi görüntüsü çarpıldıktan sonra şüpheli lezyonun görüntüden çıkarılması sağlanır.

Şekil 5.4'de mamografi görüntülerindeki şüpheli meme lezyonların belirlenmesi ve bölütlenmesi ile ilgili örnek görüntüler gösterilmektedir.



Şekil 5.4: Mamografi görüntülerindeki şüpheli meme lezyonların belirlenmesi ve bölütlenmesi ile ilgili örnek görüntüler.

5.1.2.2 Öznitelik çıkarım yöntemleri

Şüpheli meme lezyonlar için öznitelik çıkarımı bilgisayarlı destek sistemlerinin meme kanserini etkin bir biçimde teşhis edilmesi açısından son derece önemlidir. Literatürde pek çok farklı öznitelik çıkarım yöntemi önerilmiştir. İyi ve kötü huylu meme lezyonlarının birbirinden ayrılması için kullanılan öznitelik çıkarım yöntemleri doku ve morfolojik özellikler olmak üzere iki ana grupta incelenebilmektedir [73]. Bu kapsamda, şüpheli meme lezyonları için hesaplanan doku ve morfolojik özellikler Çizelge 5.3'de gösterilmiştir.

Morfolojik özellikler; BI-RADS kriterlerine göre, mamografi görüntülerinde tanımlanan şüpheli lezyonların karakterize edilmesini sağlamaktadır. Meme lezyonlarının iyi huylu veya kötü huylu tümör kategorisine girme olasılığını belirlemede morfolojik özellikler kritik bir öneme sahiptir. Morfolojik özellikler genel olarak şüpheli meme lezyonlarının şekil ve fiziksel özelliklerini yansıtmaktadır.

Çizelge 5.3: MBCD veri seti öznitelikleri.

No	Öznitelik	No	Öznitelik	No	Öznitelik	No	Öznitelik
1	Alan	15	İncelik oranı	29	Aralık	43	Korelasyon bilgisi ölçümü 2
2	Çevre	16	Şekil indeksi	30	Kök kare ortalama	44	Kısa koşu vurgusu
3	Mak. Yarıçap	17	Ortalama	31	Medyan	45	Uzun koşu vurgusu
4	Min yarıçap	18	Standart Sapma	32	Zıtlık	46	Gri seviye düzensizliği
5	Euler sayısı	19	Varyans	33	Korelasyon	47	Koşu uzunluk düzensizliği
6	Dış merkezlilik	20	Yumuşaklık	34	Enerji	48	Koşu yüzdesi
7	Katılık	21	Çarpıklık	35	Homojenlik	49	Düşük gri seviye koşu vurgusu
8	Entropi	22	Basıklık	36	Kareler toplamı	50	Yüksek gri seviye koşu vurgusu
9	Eşlenik çap	23	Ortalama mutlak sapma	37	Toplam ortalama	51	Kısa koşu düşük gri-seviye vurgusu
10	Uzatılmışlık	24	Minimum	38	Toplam varyans	52	Kısa koşu yüksek gri-seviye vurgusu
11	Dairesellik 1	25	Maksimum	39	Toplam entropi	53	Uzun koşu düşük gri-seviye vurgusu
12	Dairesellik 2	26	10.dereceden yüzdelik dilim	40	Varyans farkları	54	Uzun koşu yüksek gri-seviye vurgusu
13	Kompaktlık	27	90.dereceden yüzdelik dilim	41	Entropi farkları		
14	Dağılım	28	Çeyrekler arası aralık	42	Korelasyon bilgisi ölçümü 1		

İyi huylu tümörler oval, yuvarlak ve keskin sınırlı morfolojik özellikler ile kötü huylu tümörler ise düzensiz, mikrolobüle, belirsiz ve spiküler özellikler ile tarif edilmektedir. İyi huylu tümörler, kötü huylu tümörlere göre mamografi görüntülerinde daha küçük bir alan kaplamaktadır [74]. Bu kapsamda mamografi görüntülerinde saptanan şüpheli meme lezyonlarının iyi veya kötü huylu tümör kategorisine girme olasılığını belirlemek için 16 tane morfolojik özellik hesaplanmıştır [74-75]. Bu özellikler, açıklamalar ve formüller Eşitlik 5.1- 5.11 arasında gösterilmiştir.

Alan: Şüpheli meme lezyondaki toplam piksel sayısını ifade etmektedir.

Çevre: Şüpheli meme lezyonunun sınırındaki toplam piksel sayısını ifade etmektedir.

Maximum yarıçap: Lezyonun merkezinin en uzak köşesine olan mesafesidir.

Minimum yarıçap: Lezyonun merkezinin en yakın köşesine olan mesafesidir.

Euler sayısı: Lezyondaki ayrık bölgelerin sayısı ile delik sayısı arasındaki farkı ifade etmektedir.

Dış merkezlilik : Lezyonun eliptik özellikleri ifade etmektedir.

$$Dış\ merkezlilik = \sqrt{1 - \left(\frac{Min\ Yarıçap}{Max\ Yarıçap}\right)^2} \tag{5.1}$$

Katılık: Lezyonun toplam piksel sayısının konveks alan içindeki piksel alanına oranı olarak ifade edilmektedir.

$$Katılık = \left(\frac{Alan}{Konveks\ Alan}\right) \tag{5.2}$$

Entropi : Lezyonun şekil yapısındaki bozukluk miktarını ölçmektedir.

$$Entropi = \sum p\left(log_2(p)\right) \tag{5.3}$$

Eşlenik çap: Lezyonla aynı alana ait dairenin çapıdır. Yuvarlak ve oval düzenli görünümlü lezyonları düzensiz yapıdaki lezyonlardan ayırmak için kullanılmaktadır.

$$Eşlenik\ çap = \sqrt{\frac{4x\ Alan}{\Pi}} \tag{5.4}$$

Uzatılmışlık: Yuvarlak ve oval düzenli görünümlü lezyonları düzensiz lezyonlardan ayırmak için kullanılmaktadır.

$$Uzatılmışlık = \frac{Alan}{(2\ Max.Yarıçap)^2} \tag{5.5}$$

Dairesellik 1: Oval özelliğe sahip iyi huylu lezyonların düzensiz yapıdaki kötü huylu lezyonlardan ayırmak için kullanılmaktadır.

$$Dairesellik\ 1 = \sqrt{\frac{Alan}{\Pi\ Max.Radius^2}} \tag{5.6}$$

Dairesellik 2: Oval özelliğe sahip iyi huylu lezyonların düzensiz yapıdaki kötü huylu lezyonlardan ayırmak için kullanılmaktadır.

$$Dairesellik\ 2 = \sqrt{\frac{Min.Yarıçap}{Max.Yarıçap}} \tag{5.7}$$

Yoğunluk: Lezyonun bulunduğu bölgenin pürüzsüzlük derecesini ölçmek için kullanılmaktadır.

$$Yoğunluk = \frac{2\sqrt{Alan\ \Pi}}{Çevre} \tag{5.8}$$

Saçılma: Kötü huylu lezyonların karakteristik özelliklerini belirlemek için kullanılmaktadır.

$$Saçılma = \frac{Max.Yarıçap}{Alan} \tag{5.9}$$

İncelik oranı: Çizgisel yapıya sahip olan bölgeleri diğer yapıda olan bölgelerden ayırmak için kullanılmaktadır.

$$İncelik\ oranı = \frac{4\Pi Alan}{Çevre^2} \tag{5.10}$$

Şekil indeksi: Lezyonun kenar boşlukları ile ilgili bilgi vermek için kullanılmaktadır.

$$Şekil\ indeksi = \frac{Çevre}{2\ Max.Yarıçap} \tag{5.11}$$

Doku öznitelikleri; görüntünün bir bölgesinde bulunan piksellerin yoğunluklarının istatiksel özniteliklerinin bir kümesi olarak tanımlanabilmektedir [76]. Bu çalışma kapsamında histogram, GSEOM ve GSKUM öznitelik yöntemleri kullanılmıştır.

Histogram öznitelikleri; görüntülerin histogram yoğunluk özelliklerinden farklı istatiksel özelliklerin çıkarılması prensibine dayanmaktadır. Bu öznitelikler birinci dereceden istatiksel özellik olarak tanımlanmakta ve görüntülerdeki piksellerin gri-seviye dağılımlarından hesaplanmaktadır [77]. Bu çalışmada, şüpheli lezyon bölgelerinden ortalama, standart sapma, varyans, yumuşaklık, çarpıklık, basıklık, ortalama mutlak sapma, minimum, maksimum, 10.dereceden yüzdelik dilim, 90.dereceden yüzdelik dilim, çeyrekler arası aralık, aralık, kök kare ortalama ve medyan olmak üzere 15 adet farklı istatiksel özellik çıkarılmıştır. Bu özellikler, özelliklere ait açıklamalar ve Eşitlik 5.12-5.19 arasında gösterilmiştir.

Ortalama (μ): Gri seviye değerlerinin ortalama değerini ifade etmektedir.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{5.12}$$

Standart sapma (σ): Gri seviye değerlerinin standart sapmasıdır.

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu)^2} \tag{5.13}$$

Varyans: Gri seviye değerlerinin varyansıdır.

$$var = \sigma^2 \tag{5.14}$$

Yumuşaklık: Bir bölgedeki yoğunluğun göreceli yumuşaklığının ölçüsüdür.

$$Yumuşaklık = 1 - \frac{1}{1+\sigma^2} \tag{5.15}$$

Çarpıklık: Piksel değerlerinin ortalama etrafındaki simetrikliğinin ölçüsüdür.

$$Çarpıklık = \frac{1}{N}\frac{\sum_{i=1}^{N}(X_i - \mu)^3}{\sigma^3} \tag{5.16}$$

Basıklık: Piksel değerlerinin dağılımındaki sivrilik derecesinin ölçümüdür.

$$Basıklık = \frac{1}{N}\frac{\sum_{i=1}^{N}(X_i - \mu)^4}{\sigma^4} \tag{5.17}$$

Ortalama mutlak sapma: Piksellerin ortalama arasındaki ortalama uzaklıktır. Pikseller arasındaki değişkenliği göstermektedir.

$$Ortalama\ mutlak\ sapma = \frac{1}{N}\sum_{i=1}^{N}|X_i - \mu| \tag{5.18}$$

Minimum: Lezyon içindeki en küçük piksel yoğunluğudur.

Maksimum: Lezyon içindeki en yüksek piksel yoğunluğudur.

10.dereceden yüzdelik dilim: Piksellerin 10.dereceden yüzdelik dilim oranını göstermektedir.

90.dereceden yüzdelik dilim: Piksellerin 90.dereceden yüzdelik dilim oranını göstermektedir.

Çeyrekler arası aralık: Piksel yoğunluk dağılımının orta aralığını ifade etmektedir. Dağılımın bir üst çeyreği ve bir alt çeyreğinin arasındaki fark olarak hesaplanmaktadır.

Aralık: Lezyon içindeki gri seviye yoğunluk aralıkları ifade etmektedir. Maksimum gri seviye değeri ile minimum gri seviye değeri arasındaki fark olarak hesaplanmaktadır.

Kök kare ortalama: Lezyon içindeki gri seviye yoğunluklarının karelerinin aritmetik ortalamasının karekökü olarak tanımlanmaktadır.

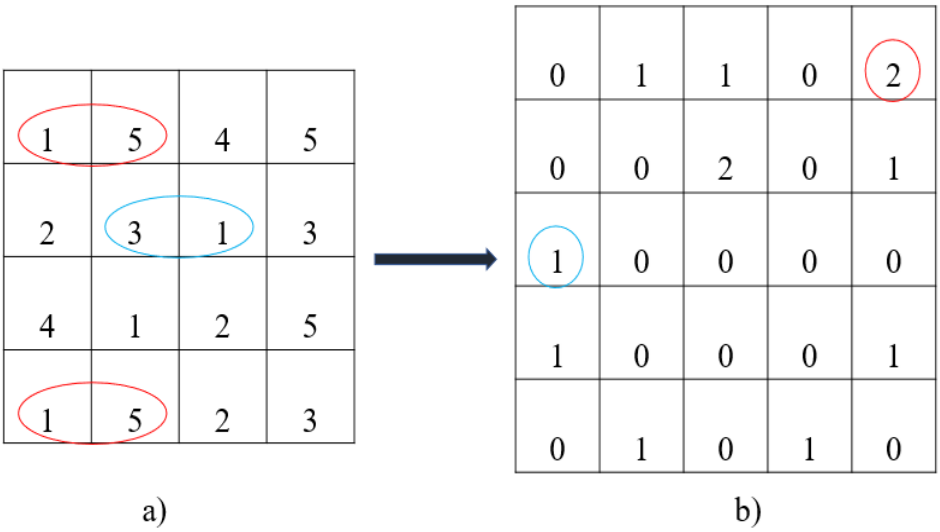
$$Karekök\ ortalama = \sqrt{\frac{1}{N}\sum_{i=1}^{N} X_i} \tag{5.19}$$

Medyan: Şüpheli lezyonun ortanca gri seviye yoğunluğunu ifade etmektedir.

Gri seviye eş oluşum matrisi (GSEOM); Haralick tarafından önerilen görüntünün ikinci derece olasılık fonksiyonlarından yararlanılarak oluşturulan doku analizlerini istatiksel olarak hesaplayan bir yöntemdir. GSEOM piksellerin komşu pikseller ile ilişkisi hakkında bilgi sağlamaktadır. Belli bir yönde ve aralarında belli bir uzaklık bulunan gri seviyeli çift piksellerinin birbirlerine göre oluşum sıklıklarını tanımlamaktadır. GSEOM hesaplanırken görüntü matrislerinin yön ve komşuluk değerlerinden yararlanılmaktadır. *NxN* boyutlu bir görüntü matrisine ait P eş oluşum matrisi Eşitlik 5.20'de gösterilmektedir. Eşitlik 5.20'deki Δ_x x yönündeki piksel değerini ve Δ_y y yönündeki piksel değerini, i merkez pikseli, j ise komşu pikseli belirtmektedir [75, 76].

$$P(i,j) = \sum_{x=1}^{N}\sum_{y=1}^{N} \begin{cases} 1, I(x,y) ve\ I(x+\Delta_x), I(y+\Delta_y) = j\ ise \\ 0, \qquad\qquad diğer\ durumlar \end{cases} \tag{5.20}$$

Şekil 5.5'de (a) matrisi 4x4'lük bir görüntüye ait bir matris gösterilmiştir. (b) matrisi ise (a) matrisinin GSEOM matrisini göstermektedir. (a) matrisindeki en büyük piksel değeri 5 olduğu için (b) matrisi 5x5 boyutundadır. (a) matrisinin ilk satırında 5 değeri iki kez belirtilmiş olup bu nedenle (b) matrisinde ilk satır ve beşinci sütun değeri 2 olmuştur. (b) de belirtilen GSEOM matrisinde (a) matrisindeki piksellerde bir tekrar yoksa ilgili piksel değeri 0 olarak belirlenmiştir. (a) matrisinde 1'den 1'e herhangi bir geçiş olmaması nedeniyle (b) matrisindeki birinci satır ve sütun piksel değerleri 0 olarak belirtilmiştir.



Şekil 5.5: Gri Seviye eş oluşum matrisinin elde edilmesi.

GSEOM matrisinden üretilen öznitelikler Eşitlik 5.21-5.35 arasında gösterilmektedir.

Zıtlık: Görüntüdeki lokal değişim miktarını ifade etmektedir.

$$Zıtlık = \sum_{i,j}^{N} P_{i,j}\,(i-j)^2 \tag{5.21}$$

Korelasyon: Piksel çiftlerinin ortak olasılık oluşumunun ölçüsünü ifade etmektedir.

$$Korelasyon = \sum_{i,j}^{N} \frac{(i-\sigma_x)(j-\sigma_y)P(i,j)}{\sigma_x\sigma_y} \tag{5.22}$$

Enerji: GSEOM matrisi içindeki karesel elemanların toplamını ifade etmektedir.

$$Enerji = \sum_{i,j}^{N} P(i,j)^2 \tag{5.23}$$

Homojenlik: GSEOM matrisi içindeki piksel dağılımlarının diyagonale olan yakınlığının ölçüsünü ifade etmektedir.

$$Homojenlik = \sum_{i,j}^{N} \frac{P(i,j)}{1+(i-j)^2} \tag{5.24}$$

Kareler toplamı: GSEOM matrisi içindeki piksel dağılımlarını ölçmek için kullanılmaktadır.

$$Kareler\ toplamı = \sum_{i,j}^{N} (i - \mu)^2\, P(i,j) \tag{5.25}$$

Toplam ortalama: GSEOM matrisi içindeki gri seviye dağılımının ortalamasını ölçmektedir.

$$Toplam\ ortalama = \sum_{i=2}^{2N} iP_{x+y}\,(i) \tag{5.26}$$

Toplam varyans: GSEOM matrisi içindeki gri seviye dağılımının varyansını ölçmektedir.

$$Toplam\ varyans = \sum_{i=2}^{2N} (i - Toplam\ Ortalama)^2 P_{x+y}(i) \tag{5.27}$$

Toplam entropi: GSEOM matrisi içindeki gri seviye dağılımındaki bozuklukları ölçmektedir.

$$Toplam\ entropi = -\sum_{i=2}^{2N} P_{x+y}(i) \log\left(P_{x+y}(i)\right) \tag{5.28}$$

Varyans farkları: GSEOM matrisi içindeki gri seviye farklarının varyansını ölçmektedir.

$$Varyans\ farkları = \sum_{i=0}^{N-1} (i)^2 P_{x-y}(i) \tag{5.29}$$

Entropi farkları: GSEOM matrisi içindeki gri seviye farklarının entropisini ölçmektedir.

$$Entropi\ farkları = -\sum_{i}^{N-1} P_{x-y}\,(i) \log\Big(\,P_{x-y}(i)\Big) \tag{5.30}$$

Korelasyon bilgisi ölçümü 1: GSEOM matrisi içindeki iki farklı pikselin korelasyon ölçümünü ifade etmektedir.

$$Korelasyon\ bilgisi\ \text{ölçümü}\ 1 = \frac{H_{xy} - H_{xy}1}{\max\left(H_x H_y\right)} \tag{5.31}$$

$$H_{xy} = -\sum_{i,j}^{N} P(i,j)\log\big(P(i,j)\big) \tag{5.32}$$

$$H_{xy}1 = -\sum_{i,j}^{N} P(i,j)\log\Big(P_x(i)P_y(j)\Big) \tag{5.33}$$

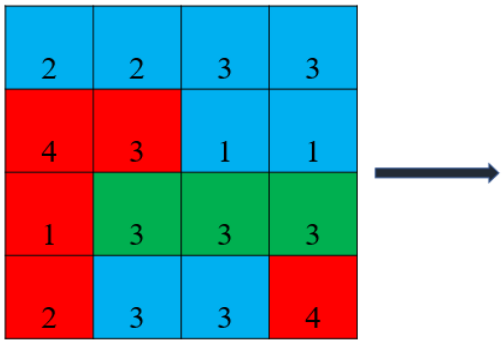
$$H_{xy}2 = -\sum_{i,j}^{N} P_x(i)P_y(j)\log\Big(P_x(i)P_y(j)\Big) \tag{5.34}$$

Korelasyon bilgisi ölçümü 2: GSEOM matrisi içindeki ortak entropiler arasındaki farkı ölçmektedir.

$$Korelasyon\ bilgisi\ \text{ölçümü}\ 2 = \sqrt{1 - e^{-2HXY2-HXY}} \tag{5.35}$$

Gri seviye koşu uzunluğu matrisi (GSKUM); ikinci dereceden istatiksel doku analiz yöntemlerinden birisidir [78]. Bu yöntem görüntüdeki farklı uzunlukların gri seviye sayılarının hesaplanmasına dayanmaktadır. Gri seviye uzunluğu, aynı gri seviye değerlerine sahip doğrusal olan komşu görüntü noktalar dizisi olarak tanımlanmaktadır. Gri seviye koşusu aynı yönde aynı gri seviyeye sahip olan sıralı piksel kümesinden oluşmaktadır. Koşu uzunluğu koşudaki piksel sayısını, koşu uzunluğu değeri ise görüntüdeki koşuların meydana geliş sayısını ifade etmektedir. Piksel yoğunluğu yüksek olan dokular küçük koşma eğilimi gösterirken, piksel yoğunluğu düşük olanlar ise yüksek koşma eğilimi göstermektedir. Bu yöntem ile uzun ve kısa uzunluklarda yüksek ve düşük gri tonlama seviyeleri belirlenebilmektedir. GSKUM yapısında orijinal matris boyutu koşu uzunluğunun maksimum sınır değerini belirtmektedir. Gri seviye belirlenme işlemi matrisdeki her değer için yapılmaktadır [76, 78].

Şekil 5.6'da bir görüntünün GSKUM yapısının oluşturulması gösterilmiştir. (a) matrisi görüntüyü, (b) matrisi ise (a) matrisine ait GSKUM' yu ifade etmektedir. (a) matrisindeki en büyük sayı 4 olduğundan gri seviyesi 1-4 arasında incelenmiştir. 4x4 matrisler için maksimum koşu sayısı 4 değerini alabilmektedir. (a) matrisinde art arda durumu ikinci satır ikinci sütun, üçüncü satır üçüncü sütun ve üçüncü satır dördüncü sütunda piksel 3 değerini göstermekte ve bu nedenle (b) matrisinde 3 gri seviye değeri için koşu uzunluğu 3 değerini almaktadır. (a) matrisindeki tüm piksel değerleri benzer şekilde incelenerek (b) matrisi oluşturulmaktadır.



a)

Gri Seviye	Koşu Uzunluğu			
1	1	1	0	0
2	1	1	0	0
3	1	2	1	0
4	2	0	0	0

b)

Şekil 5.6: Gri seviye koşu uzunluğu matrisinin elde edilmesi.

GSKUM matrisi oluşturulurken bazı özelliklere de ulaşılabilmektedir. Bu özellikler ve özelliklere ait formüller Eşitlik 5.36- 5.46 arasında gösterilmiştir. Eşitliklerdeki p(i,j) koşu uzunluğundaki i ve j.'inci noktadaki değerini, N_g gri seviye sayısını, N_r maksimum koşu uzunluğunu ifade etmektedir.

Kısa koşu vurgusu: Matristeki kısa koşuların dağılımını ölçmektedir. Her koşu uzunluk değerinin uzunluğunun karesine bölünmesi ile hesaplanmaktadır.

$$Kısa\ koşu\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j)}{j^2} \tag{5.36}$$

Uzun koşu vurgusu: Matristeki uzun koşuların dağılımını ölçmektedir. Her koşu uzunluk değerinin uzunluğunun karesi ile çarpılması ile hesaplanmaktadır.

$$Uzun\ koşu\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j)\, j^2 \tag{5.37}$$

Gri seviye düzensizliği: Matris boyunca gri seviyelerin benzerliğini ölçmektedir.

$$Gri\ seviye\ düzensizliği = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} p(i,j))^2 \tag{5.38}$$

Koşu uzunluğu düzensizliği: Matris boyunca koşuların benzerliğini ölçmektedir.

$$Koşu\ uzunluğu\ düzensizliği = \frac{1}{N_{koşu}} \sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} p(i,j))^2 \quad (5.39)$$

Koşu yüzdesi: Matristeki toplam koşu sayısının olası koşu sayısına oranıdır.

$$Koşu\ yüzdesi = \frac{1}{N_{koşu}} \frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} p(i,j))}{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} p(i,j))} \quad (5.40)$$

Düşük gri seviye koşu vurgusu: Düşük gri seviye koşularının dağılımını ölçmektedir.

$$Düşük\ gri\ seviye\ koşu\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j)}{i^2} \quad (5.41)$$

Yüksek gri seviye koşu vurgusu: Yüksek gri seviye koşularının dağılımını ölçmektedir.

$$Yüksek\ gri\ seviye\ koşu\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j)\, j^2 \quad (5.42)$$

Kısa koşu düşük gri seviye vurgusu: Matristeki kısa koşu ve düşük gri seviye değerlerinin ortak dağılımını ölçmektedir.

$$Kısa\ koşu\ düşük\ gri\ seviye\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j)}{i^2 j^2} \quad (5.43)$$

Kısa koşu yüksek gri seviye vurgusu: Matristeki kısa koşu ve yüksek gri seviye değerlerinin ortak dağılımını ölçmektedir.

$$Kısa\ koşu\ yüksek\ gri\ seviye\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j) i^2}{j^2} \quad (5.44)$$

Uzun koşu düşük gri seviye vurgusu: Matristeki uzun koşu ve düşük gri seviye değerlerinin ortak dağılımını ölçmektedir.

$$Uzun\ koşu\ düşük\ gri\ seviye\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j) j^2}{i^2} \quad (5.45)$$

Uzun koşu yüksek gri seviye vurgusu: Matristeki uzun koşu ve yüksek gri seviye değerlerinin ortak dağılımını ölçmektedir.

$$Uzun\ koşu\ yüksek\ gri\ seviye\ vurgusu = \frac{1}{N_{koşu}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j)\, i^2 j^2 \qquad (5.46)$$

5.2 Veri Ölçeklendirilmesi

Veri setlerinde bulunan özniteliklerin farklı ölçeklere sahip olması veri setinin modellenmesini olumsuz olarak etkilemektedir. Bu nedenle özniteliklerin ölçeklendirilmesi algoritmaların veri setleri ile modellenmesi açısından önem arz etmektedir. Öznitelik ölçeklendirme verilerde bulunan bağımsız özellikleri sabit bir aralıkta standartlaştırmak için kullanılan bir tekniktir. Ölçeklendirmedeki temel amaç; farklı özniteliklerdeki sayısal değerleri ortak bir değer aralığına getirmek ve büyük farkları belirli bir düzen içerisinde göstermektir. Z-skor yöntemi verilerin ölçeklendirilmesi amacıyla sıklıkla kullanılan bir metottur. Bu yöntemde her öznitelik için ortalama değerinden uzağa ve nitelik değerindeki standart sapmaya göre yeni değer hesaplanmaktadır. Z-skor yönteminin formülü Eşitlik 5.47'de gösterilmiştir [79]. Eşitlik 5.47'deki v' yeni değeri, v öznitelik değerini, μ ortalamayı ve σ standart sapmayı göstermektedir.

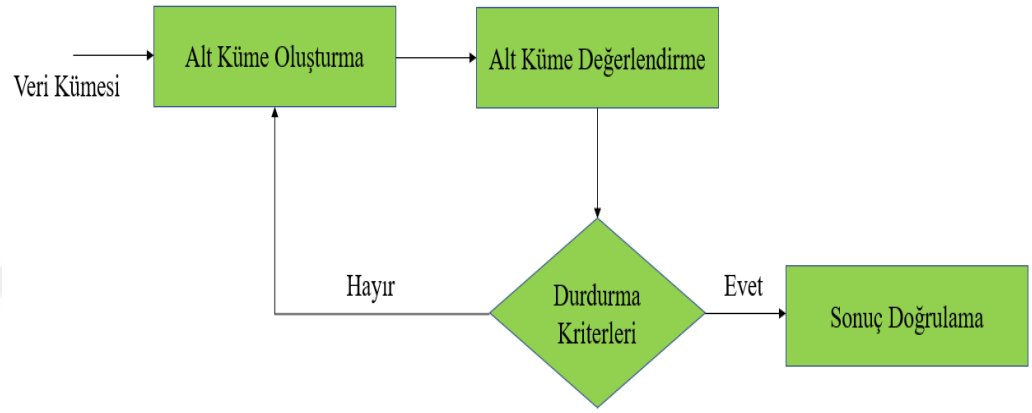
$$v' = \frac{v - \mu}{\sigma} \qquad (5.47)$$

5.3 Öznitelik Seçim Yöntemleri

Yüksek boyutlu veri setlerinde çok boyutluluk sıkça rastlanan problemlerden birisidir. Veri setindeki çok boyutluluk veri setinin karmaşıklığının yanı sıra bir hedef ile ilişkilendirilen özniteliklerin ayırt edici olmayanlarının çok sayıda olması sınıflandırma algoritmalarının öğrenme sürecinde zorluklara sebebiyet vermektedir. Öznitelik seçim yönteminin amacı bir veri seti içerisindeki en faydalı öznitelikleri bularak verideki öznitelik sayısını azaltmaktır. Öznitelik seçim yöntemleri;

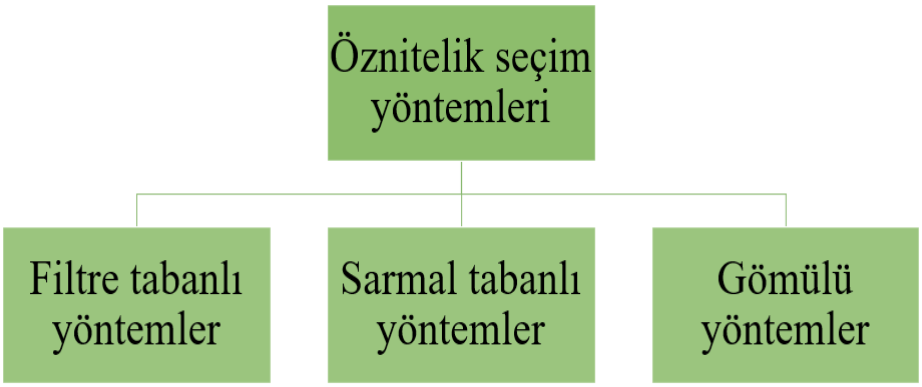
1. Öznitelik kümesinin boyutunu düşürmektedir.
2. İlgili olmayan ve gürültülü özniteliği ortadan kaldırmaktadır.
3. Veri setini sadeleştirmektedir.
4. Veri setinin kalitesini artırmaktadır.
5. Kaynak tasarrufu sağlamaktadır.
6. Sınıflandırma algoritmasının hızını ve başarı oranını artırmaktadır.

Öznitelik seçim işlem süreci Şekil 5.7'deki adımlardan oluşmaktadır. İlk aşamada veri setinden bir öznitelik alt kümesi oluşturulmaktadır. Sonraki aşamada ilgili öznitelik farklı formüller ile seçim durumuna karar verilmektedir. Sonuçta; seçilen öznitelik alt kümeye eklenmektedir. Bu işlem süreci durdurma kriterini sağlanana kadar devam etmektedir [80].



Şekil 5.7: Öznitelik seçim işlem süreçleri.

Genel olarak öznitelik seçim algoritmaları Şekil 5.8'de gösterildiği gibi filtre tabanlı yöntemler, sarmal tabanlı yöntemler ve gömülü yöntemler olmak üzere 3 grupta incelenmektedir [81].



Şekil 5.8: Öznitelik seçim yöntemleri.

5.3.1 Filtre tabanlı yöntemler

Filtre-tabanlı yöntemler özniteliklerin önem derecesini hesaplamak amacıyla öznitelik ile hedef değişken arasındaki ilişkiyi esas olarak almaktadır. Bu yöntemler uzaklık,

tutarlılık ve bilgi gibi istatiksel verileri dikkate alan fonksiyonlar aracılığıyla hesaplama yapmaktadır. Şekil 5.9'da filtre tabanlı öznitelik seçim yönteminin çalışma adımları gösterilmektedir. Bu şekilde çalışan tekniklerde veride bulunan her bir öznitelik için bir değerlendirme fonksiyonu kullanılarak farklı skorlar hesaplanmaktadır. Hesaplanan bu skorlar içerisinde en yüksek değere ulaşan öznitelikler alt kümeye dahil edilmektedir [82-83].

Şekil 5.9: Filtre tabanlı öznitelik yöntemlerinin çalışma adımları.

5.3.1.1 Relief

Relief yöntemi Kira ve Rendel tarafından 1992 yılında geliştirilmiştir. Bu yöntem veri setinden rastgele seçilen bir örneğin bulunduğu sınıfa yakınlığına ve farklı sınıflara olan uzaklığına bağlı bir model oluşturarak öznitelik seçim işlemini gerçekleştirmektedir. Bu yöntemde öncelikle bir özniteliğe bir ağırlık değeri atanmakta daha sonra da belirlenen bir eşik değer üzerinden yararlı bulunan öznitelikler seçilmektedir [84].

Algoritmanın temel çalışma aşamaları şöyledir.

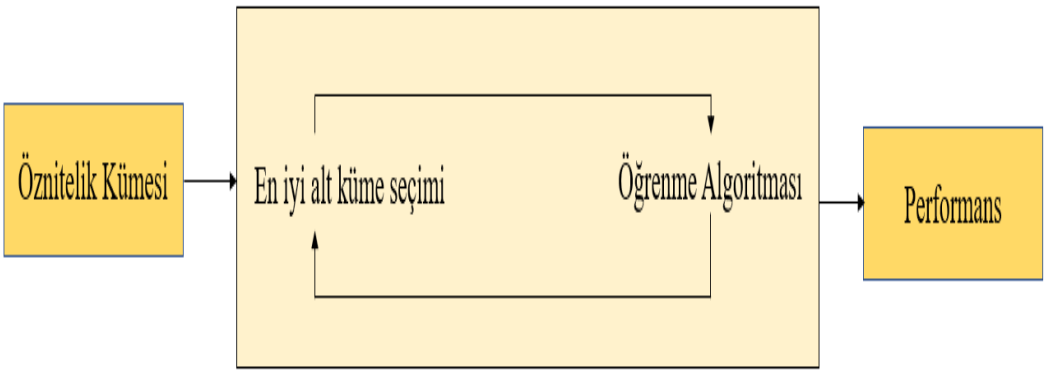
1. Veri setinden sınıflara ait en yakın özniteliklerin seçilmesi
2. Özniteliklerin ağırlıklarının belirlenmesi
3. Hesaplanan ağırlıkların sıralanması
4. En iyi "n" kadar özniteliğin seçilmesi

Algoritmanın ağırlık hesaplama yöntemi Eşitlik 5.48'de gösterilmiştir. Eşitlikte R relief özniteliğinin önem derecesini, As aynı sınıftaki en yakın öznitelik değerini, F_s ise farklı sınıftaki en yakın öznitelik değerini göstermektedir [85].

$$R = R_{i-1} - (x_i - As_i)^2 + (x_i - Fs_i)^2 \tag{5.48}$$

5.3.2 Sarmal tabanlı yöntemler

Şekil 5.10'da sarmal tabanlı yöntemlerin çalışma adımları gösterilmektedir. Öznitelik seçim süreci bu yöntemlerde bir doğru sınıflama oranına bağlıdır. Bu süreç, farklı öğrenme algoritmaları kullanılarak en iyi tahmin oranına sahip olanların seçilmesi ile sağlanmaktadır. En iyi öznitelik alt kümesinin belirlenmesinde sarmal yöntemler, filtre tabanlı yöntemler ile karşılaştırıldığında daha iyi sonuçlar vermektedir. Ancak sarmal tabanlı yöntemler daha yüksek hesaplama işlemi gerektirmektedir [86-88].



Şekil 5.10: Sarmal tabanlı yöntemlerin çalışma adımları.

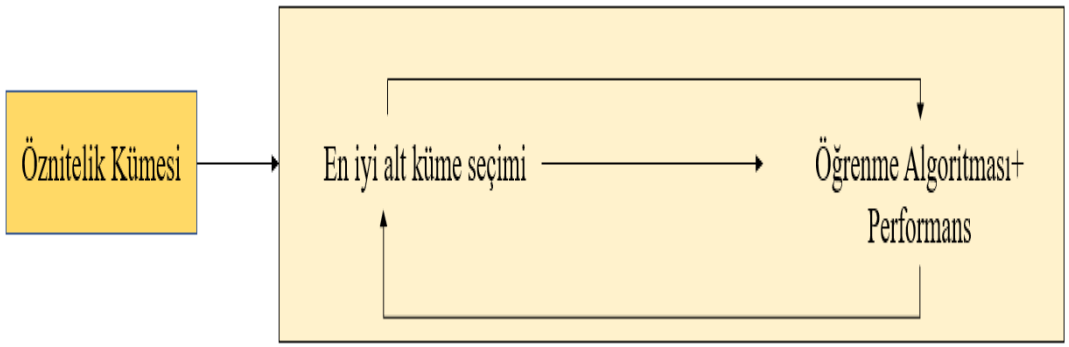
5.3.2.1 Ardışık ileri yönde seçim (AİYS)

Whitney tarafından önerilen sade ve etkili bir sarmal tabanlı öznitelik seçim yöntemidir. Bu yöntem boş bir öznitelik kümesi ile başlamaktadır. Her tekrarlayan süreçte daha önce öznitelik kümeye dahil edilmemiş bir özniteliğin eklenmesi ile öznitelik seçim süreci tamamlanmış olmaktadır. Özniteliklerin kümeye eklenip eklenmemesinde, özniteliğin sınıflandırma başarısına etkisi dikkate alınmaktadır. Her tekrarda sadece bir tane öznitelik eklenerek sınıflandırma yönteminde artış olmayana kadar bu süreç devam etmektedir. AİYS yönteminin aşamaları aşağıdaki gibidir [83].

1. Boş bir öznitelik kümesi belirlenir ($X=\{Ø\}$).
2. Sıradaki en iyi öznitelik seçilir ($x^+=\text{argmax}_x+ [K (X_i+ x^+)]$).
3. Eğer $K (X_i+ x^+)>K(X_i)$ ise, $X_{i+1}= X_i+ x^+$, $i=i+1$ olarak güncellenir.
4. Adım 2
5. Dur.

5.3.3 Gömülü yöntemler

Gömülü yöntemler bir veri setinde en yararlı öznitelikleri seçmek amacıyla öğrenme algoritmalarının kalite ölçütlerinden faydalanmaktadır. Bu yöntemler bünyelerinde hem sınıflandırma algoritması hem de öznitelik seçim algoritmasını birlikte barındırmaktadır. Ayrıca sınıflandırma ve öznitelik seçme işlemlerinin öğrenme süreçleri birbirleri ile eş zamanlı olarak gerçekleştirilmektedir. Bu yöntemin amacı; bir öznitelik kümesinde yer alan alt kümeler arasında bir arama algoritmasından yararlanarak en iyi çözümü bulmaktır [83, 87]. Gömülü yöntemlerin çalışma mantığı Şekil 5.11'de gösterilmiştir.



Şekil 5.11: Gömülü yöntemlerin çalışma adımları.

5.3.3.1 En küçük mutlak büzülme ve seçim operatörü (LASSO)

1996 yılında Tibshirani tarafından önerilmiş bir gömülü öznitelik seçim algoritmasıdır [89]. Yöntem bir regresyon modeli oluşturularak bağımsız değişkenlerin değerlerinden faydalanılarak bağımlı değişkenlerin değerlerini tahmin etmeyi amaçlamaktadır. Oluşturulan regresyon modelinde bağımsız değişkenlerin katsayılarının tahmini ve değişken seçimi eş zamanlı olarak yapılmaktadır. Regresyon modeli oluşturulurken en küçük kareler yöntemi kullanılmaktadır. Bağımsız değişkenler veri kümelerindeki özniteliklere karşılık gelmektedir. Regresyon modeli kurulurken çapraz doğrulama ve ceza parametresi kullanılabilmektedir. Ceza parametresi veri seti için daha az önemli olan özniteliklerin sıfır değerini almasını istemektedir. Bu nedenle ceza parametresi özniteliklerin katsayılarına büzülme uygulamaktadır. Ceza parametesinin değeri özniteliklere uygulanılacak olan büzülme miktarını belirlemektedir. Yöntemde özniteliklerin katsayılarını belirlemek amacıyla L1 ceza fonksiyonu kullanılmaktadır. L1 ceza fonksiyonu özniteliklerin katsayılarının büyüklüklerinin mutlak değerine eşit

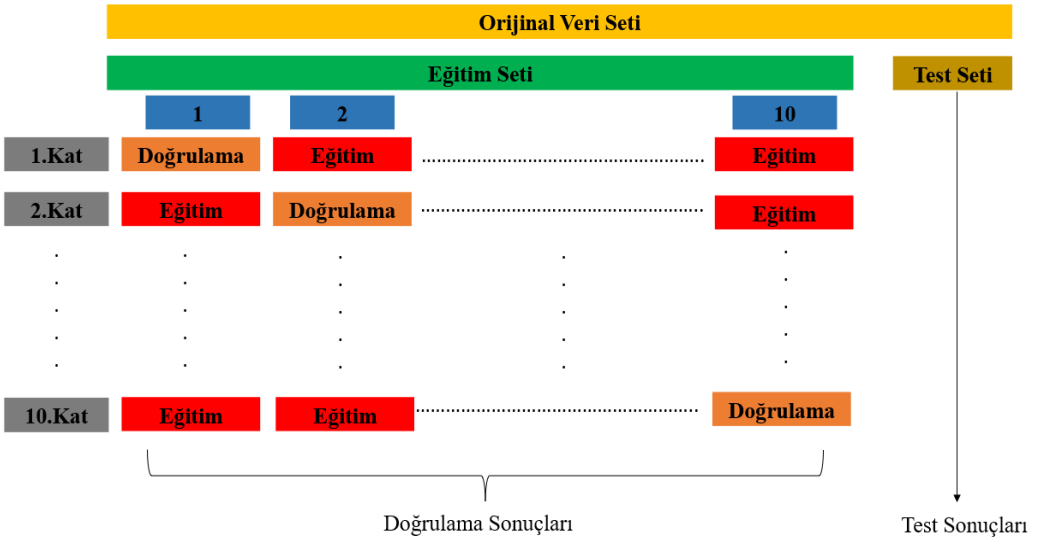
bir fonksiyon eklenmesidir. Lasso yöntemi ile özniteliklerin katsayıların hesaplanması için Eşitlik 5.49'dan yararlanılmaktadır.

$$\beta_{LASSO} = \arg_{\beta} \min \left(\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{P} X_{i,j}\,\beta_j)^2 + \lambda \sum_{j=1}^{P} |\beta_j|\right) \quad (5.49)$$

Eşitlikte 5.49'daki N sayısı gözlem sayısını, y_i çıkış değerini, $x_{i,j}$ öznitelik kümesini, λ ceza parametresini ve β ise lasso katsayılarını ifade etmektedir. Sonuç olarak sıfır katsayısına sahip öznitelikler öznitelik kümesinden elenmesi ile öznitelik seçim işlemi tamamlanmış olmaktadır [89, 90].

5.4 Çapraz Doğrulama

Veri seti genellikle sınıflandırma sürecinden önce eğitim ve test verilerine ayrılmaktadır. Çapraz doğrulama veri setinin eğitim ve test verilerine ayrılma yöntemlerinden birisidir. Eğitim verisi, oluşturulan modelin öğrenme işleminin gerçekleştiği (eğitim ve doğrulama verileri ile), test verisi ise değerlendirme aşamasında modelin performanısını ölçmek üzere kullanılan veri setidir. Bu yöntem eğitim verisi ile oluşturulan modeli eğitirken, geriye kalan veriler (doğrulama) ile de modelin performansını ölçmektedir. Çapraz doğrulama yöntemi eğitim aşamasında ezberlemeyi önleyerek ağın daha verimli bir şekilde öğrenmesini sağlamaktadır. k-katmanlı yöntem en sık kullanılan çapraz doğrulama yöntemlerinden birisidir. Bu yöntem uygulanırken veri seti (k-katlı olarak) k adet alt kümeye bölünmektedir. Bu kümelerden bir tanesi doğrulama verisi olarak k-1 tanesi de eğitim verisi olarak değerlendirilmektedir. Böylelikle veri setinin her noktası en az bir kere doğrulama verisi olarak kullanılmış ve modelin performansı daha doğru bir şekilde değerlendirilmiştir [91]. Şekil 5.12'de 10-katlı-çapraz doğrulama yöntemi ile bir veri setinin eğitim adımları gösterilmiştir. 9 parça eğitim verisi kalan parça da doğrulama verisini tanımlamaktadır. Sonuç olarak, eğitim işlemi 10 doğrulama ölçümünün ortalaması alınarak hesaplanmaktadır.



Şekil 5.12: 10-katlı çapraz doğrulama.

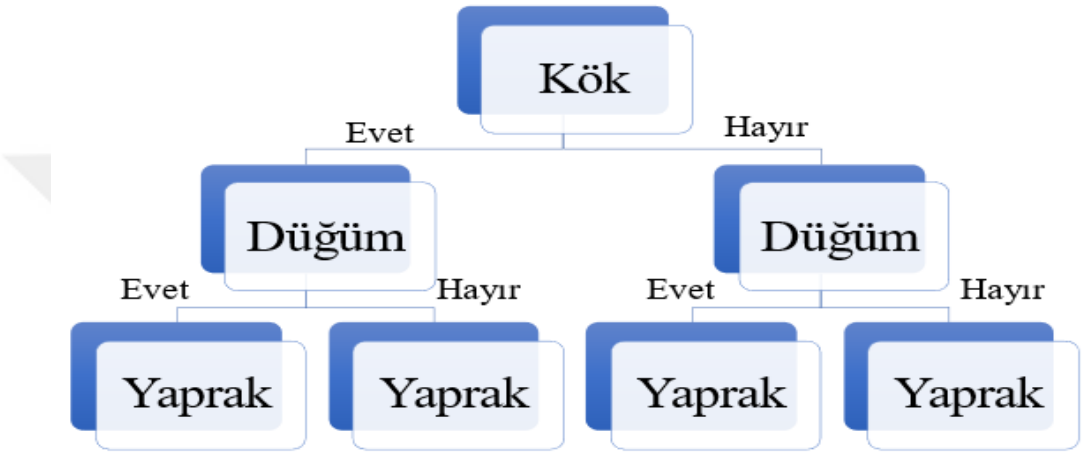
5.5 Sınıflandırma

Sınıflandırma, benzer özelliklere sahip bir veri setinin önceden belirlenmiş veri gruplarından hangisine ait olduğunun tahmin edilmesi işlemidir. Sınıflandırma algoritmaları yeni bir öğenin daha önce tanımlanmış olan kategorilerden hangisine girme olasılığını hesaplamak amacıyla kullanılmaktadır [92]. Literatür çalışmalarında genellikle sınıflandırma işlemi için makine öğrenme algoritma tercih edilmektedir. Bu çalışma kapsamında, sınıflandırma algoritmaları olarak KA, NB, DVM, K-NN ve TÖ yöntemleri uygulanmıştır.

5.4.1 Karar ağacı

Denetimli bir makine algoritması olup genellikle regresyon ve sınıflandırma amacıyla kullanılmaktadır. Bu algoritmanın amacı veri setini belirli bir karar kuralları çerçevesinde daha anlaşılır küçük alt gruplara bölmektir. KA ağacın kurulduğu ve sınıflandırılmanın gerçekleştirildiği iki aşamadan oluşmaktadır. Bir karar ağacı kök, düğümler, dallar ve yaprak (terminal düğüm) olmak üzere dört aşamadan oluşmaktadır. Kök tüm gözlemlerin veya populasyonun bulunduğu bölümdür. Sınıflandırma süreci bu bölümden başlamaktadır. Homojen yapıdaki gözlemler doğal olarak aynı sınıfta yer alacak ve kök dallanma yapmadan sınıflandırma süreci sona erecektir. Hetorojen yapıdaki gözlemlerde kök, gözlemleri sınıflara bölen en iyi niteliğe göre iki veya daha fazla sayıda dala ayırarak yeni düğümleri oluşturmaktadır.

Gözlemleri sınıflara bölen en iyi nitelik, modelde bulunan bağımsız değişkenleri betimlemektedir. Bölme işlemi sonrasında meydana gelen yeni düğümler kendi içlerinde homojen bir yapıya sahip, birbirleri arasında ise hetorojen gözlem gruplarından oluşmaktadır. Yeni düğümler gözlemlerin niteliklerine göre tekrardan dallara ayrılmaktadır. Bu işlem süreci daha iyi bölme kalmayana kadar sürmektedir. Karar ağacının dallanma yapmayan son düğümü terminal düğüm olup gözlemlerin atandıkları sınıfları göstermektedir [93]. Şekil 5.13'de bir karar ağacı modeli gösterilmektedir.



Şekil 5.13: Karar ağacı yapısı.

Ağaçlarda meydana gelen düğümlerin dallanma yapıp yapmayacağının belirlenmesinde en sık kullanılan kriterler entropi ve bilgi kazancıdır. Entropi verilerle ilgili düzensizliğin bir ölçüsüdür. Veri homojense entropi 0 değerini, verideki değerler eşit olarak bölünmüşse 1 değerini almaktadır. Ağaç yapısının entropi değerini en aza indirgeyen bölünmeler yapması beklenmektedir. En iyi bölünmeyi belirlemek amacıyla da bilgi kazancı kullanılmaktadır. Bilgi kazancı entropideki belirsizliği ölçmektedir. Entropi ve bilgi kazancının formülleri Eşitlik 5.50 ve Eşitlik 5.51'de gösterilmektedir.

$$Entropi = \sum_{i=1}^{c} -plog_2\, p_i \tag{5.50}$$

$$Bilgi\ kazancı(S, A) = Entropi(S) - \sum \frac{|S_v|}{|S|} Entropi(S_v) \tag{5.51}$$

Eşitlik 5.50'deki p belirli bir sınıfa ait grubun yüzdesini, Eşitlik 5.51'deki A kümenin bölünmüş bir parçasını, S orijinal veri setini ve S_v ise özniteliği A değeri olan S alt

kümesini göstermektedir. En yüksek bilgi kazancına sahip olan öznitelik, karar ağacında dallanma için tercih edilmektedir [94].

5.4.2 Naive Bayes

Koşullu olasılık kuralını temel esas olarak alarak bir öğenin belirli bir kategoriye girme olasılığını hesaplamaktadır. Denetimli bir makine öğrenme algoritmasıdır. Algoritma bir sınıfta bulunan bir özniteliğin değerinin etkisinin diğer özniteliklerden bağımsız olduğunu varsaymaktadır. Bu varsayım işlemi koşullu olasılık kuralı olarak tanımlanmaktadır. Koşullu olasılık kuralı Bayes Teoreminden faydalanılarak hesaplanmaktadır. Bayes Teoremi daha önce oluşan bir olayın olasılığı dikkate alındığında, bir olayın oluşma olasılığını bulmaktadır. Bayes Teoremi her sınıf için belirli bir öğenin o sınıfa ait olma olasılığını tahmin etmeyi amaçlamaktadır. Daha basit bir deyişle, bir A olasılığı daha önceki bir olayı ve bağımlı olasılığı temsil ediyorsa bir B olayının gerçekleşme olasılığı Bayes Teoremi ile hesaplanabilmektedir. Bayes teoremi Eşitlik 5.52 ile ifade edilebilmektedir [95].

$$P(\mathrm{B/A}) = \frac{\mathrm{P(A/B).\,P\ (B)}}{\mathrm{P(A)}} \tag{5.52}$$

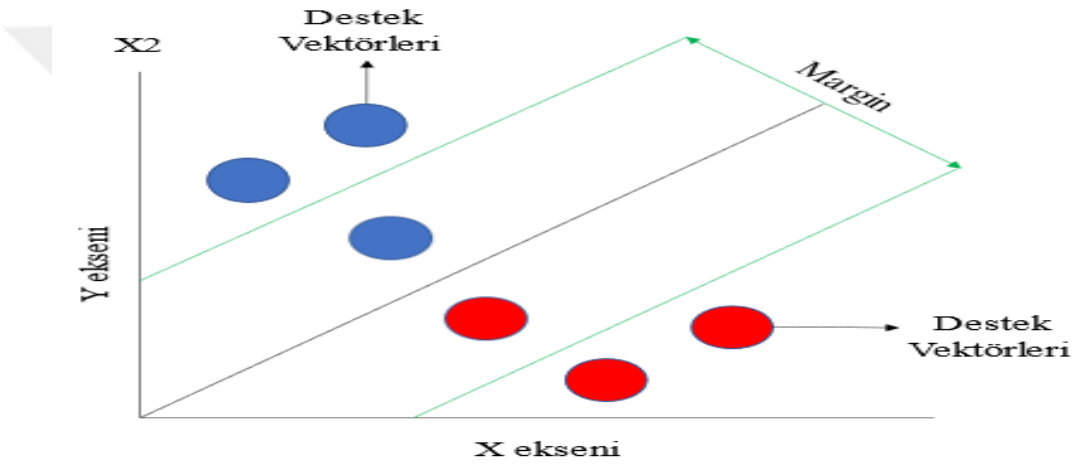
P(B/A) A olayı meydana geldiğinde B olayının olma olasılığını, P(B) B olayının olma olasılığını, P(A/B) A olayı meydana geldiğinde B olayının olma olasılığını ve P(A) olayının olma olasılığını ifade etmektedir. A olayı bilinen bir B olayının olasılığını hesaplamak için Bayes Teoremi A ile B olaylarının birlikte gerçekleştiği vakaları saymakta ve A'nın tek başına gerçekleştiği vaka sayısına bölmektedir. Her sınıfın başka bir olay olasılığı (C_i) ise NB algoritması tarafından Bayes Teoremi ile hesaplanabilmektedir. Sınıflandırma algoritması kendisine bir sınıf verildiğinde A özniteliklerinin bağımsız olduğunu varsaymaktadır. Bu sebeple olasılık sınıfı belirtilen her öğenin bireysel koşullu olasılıklarının çarpımı ile elde edilmektedir [96]. Bu işlem süreci Eşitlik 5.53'de gösterilmektedir.

$$\mathrm{P\ (C_i/A_1 \ldots A_m) = P(C_i).\ P\ (A_i\ /C_i)\ \ldots\ldots P(A_m/C_i).\ P\ (A)} \tag{5.53}$$

Bu algoritma sınıflandırma için az miktarda eğitim verisi kullanması nedeniyle hem ikili hem de çoklu sınıflandırma problemlerinde kullanılabilmektedir.

5.4.3 Destek vektör makineleri

Denetimli makine öğrenme yöntemlerinden birisidir. Bu algoritma bir verinin en iyi sınıflara ayıran bir hiper düzlem eğrisinin bulunmasını amaçlamaktadır. Hiper düzlem eğrisi iki sınıfa da en uzak noktadadır. Eğitim verileri kullanılarak hiper düzlem bulunmaktadır. Test verileri hangi hiper düzlemin hangi tarafında bulunuyorsa o sınıfa atanmaktadır. Hiper düzlem eğrisi doğrusal olmayan örnekleri doğrusal bir şekilde bölmektedir. Hiper düzlem ayrıca değişik örnekler arasındaki maksimum bölme işleminin yapılmasını sağlamaktadır [97]. Şekil 5.14'de destek vektör makineleri algoritmasının çalışma mantığı gösterilmiştir.



Şekil 5.14: Destek vektör makinesi.

Şekil 5.14'de kırmızı ve maviler olmak üzere iki farklı sınıf belirtilmektedir. Sınıflandırmadaki temel amaç yeni gelecek örneğin hangi sınıfa atanacağına karar verilmesi işlemidir. Sınıflandırmanın yapabilmesi için iki sınıfı birbirinden ayıran bir doğru çizilmektedir. İki yeşil doğru arasında kalan bölgeye margin adı verilmektedir. Margin iki farklı sınıf arasında kalan boşluk olarak tanımlanmaktadır. Marginin boyutu ne kadar geniş olursa sınıflar o kadar birbirinden iyi ayrıştırılabilmektedir. Bu yönteminde sınıflara atanma işlemi Eşitlik 5.54 ile bulunmaktadır.

$$y' = \begin{cases} 0 & eğer\ w^T\ x + b < 0 \\ 1 & eğer\ w^T\ x + b \geq 0 \end{cases} \tag{5.54}$$

w ağırlık vektörünü, y' çıkış vektörünü, x giriş vektörünü ve b sapmayı ifade etmektedir. Yeni çıkan değer için sonuç 0'a eşit ya da büyük olursa mavi noktalara yakın, 0'dan küçük olursa kırmızı noktalara daha yakın olacaktır [98]. Verilerin bir

hiper düzlem ile ayrılamaması durumunda çekirdek fonksiyonu kullanılabilmektedir. Çekirdek fonksiyonu örneklerin çok boyutlu ve doğrusal yöntemler ile ayrılabilen bir alana taşıyarak sınıflandırma işlemi yapmaktadır. Sınıflandırma işlemi bir optimizasyon fonksiyonu ile yapılmaktadır. Optimizasyon fonksiyonun formülü Eşitlik 5.55'deki gibidir. $g(x)$ optimizasyon fonskiyonunu, α reel sayılarda herhangi bir değeri ve K ise çekirdek fonksiyonunu ifade etmektedir.

$$g(x) = \sum_{i=1}^{N} \alpha_i \, y_i K(x_i y_i) + b \tag{5.55}$$

Hiper düzlem olmadığı için b sapma değeri ihmal edilebilmektedir. Sapma değeri çekirdek fonksiyonu işlevlerinde kullanılmaktadır. Sınıflandırma işlemlerinde doğrusal çekirdek, polinom çekirdek, sigmoid ve radyal tabanlı çekirdek fonksiyonları kullanılabilmektedir [99].

5.4.4 K-en yakın komşu

Denetimli makine öğrenme algoritmalarından birisidir. Algoritmadaki temel mantık veri setine yeni gelen bir öğe ile k adet komşu arasındaki mesafenin hesaplanmasıdır. Algoritmanın çalışma aşamaları şöyledir.

- k sayısı belirlenir. Bu değer komşu sınıflar arasındaki uzaklık ve sınıflandırma performansına direkt etki etmektedir.
- Veri setlerine verilecek yeni öğenin mevcut olan verilere göre uzaklıkları hesaplanır. Uzaklık hesabı için Minkowski, Öklid, Manhattan ve Chebyschev fonksiyonları kullanılabilmektedir.
- Bulunan uzaklıklardan en yakın k. komşu dikkate alınmakta ve öznitelik değerlerine göre yeni öğenin k. komşu sınıfına atanarak tahmin edilmesi istenen sınıf değeri bulunur [30].

5.4.5 Topluluk öğrenme

İki ya da daha fazla sınıflandırma algoritmasının tahmin sonuçlarının birleştirilmesi ile oluşturulmaktadır. Tahmin sonuçlarının birleştirilmesi oylama ile ya da sonuçların ortalaması alınması şeklinde yapılabilmektedir [100]. Torbalama (bagging) ve yükseltme (boosting) yöntemleri en sık kullanılan topluluk öğrenme yöntemlerinden ikisidir [101]. Bu yöntemler genellikle eğitim sürecinde KA algoritmasından yararlanmaktadır [102].

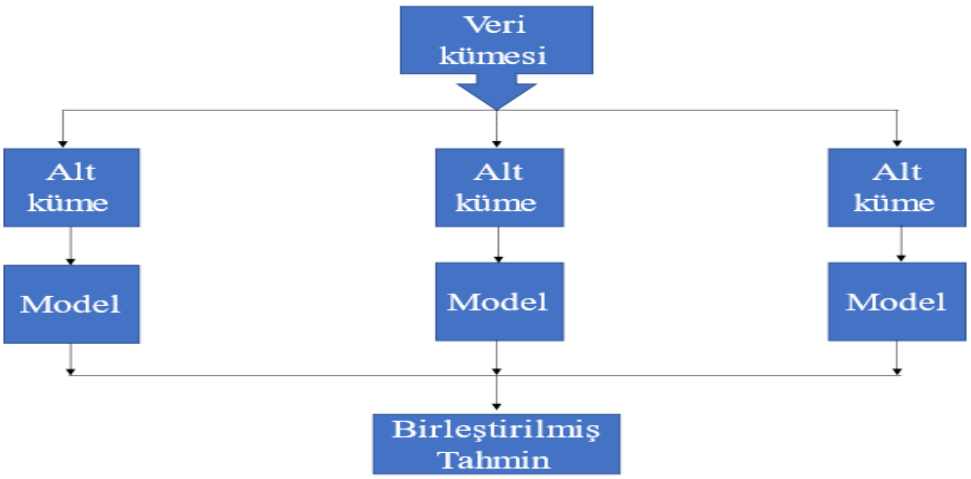
Torbalama yöntemi ilk kez 1996 yılında Breiman tarafından önerilmiştir. Bu yöntemde aynı büyüklükte farklı eğitim veri kümeleri oluşturulmaktadır. Bütün eğitim verileri birbirlerine paralel bir şekilde eş zamanlı olarak çalıştırılmaktadır. Tüm algoritmalar önce aynı veri setine verilerek test edilmektedir. Daha sonra genel sonuçlar oylanarak sınıflandırma işlemi tamamlanmaktadır [103]. Şekil 5.15'de torbalama öğrenme algoritmasının çalışma aşamaları gösterilmiştir. Buna göre;

- İşlem süreci veri setinden çok sayıda alt kümeler oluşturularak başlamaktadır.
- Alt kümeler birbirlerinden bağımsız ve birbirlerine paralel bir biçimde ilerlemektedir. Bu işlem süreci bir model oluşturana kadar devam etmektedir.
- Son tahminler tüm modellerde oluşan tahminlerin birleştirilmesi ile oluşturulmaktadır.

Yükseltme algoritmaları ise torbalama algoritmalarının aksine sıralı bir şekilde çalıştırılmaktadır. Bu algoritmalarda temel amaç veri setlerine farklı ağırlık verilerek elde edilen karar ağaçlarından tahminler yapılmasıdır [101].

Yükseltme algoritmasının çalışma aşamaları şöyledir:

- Veri setinden bir alt küme seçilir.
- Veri setindeki tüm noktalara eşit ağırlık verilir.
- Alt küme için bir model oluşturulur.
- Oluşturulan bu model veri setinin tahminleri için kullanılır.
- Yöntemdeki hatalar gerçek ve tahmin edilen değerlerin birlikte kullanılması ile bulunur.
- Yanlış tahmin edilen örneklere daha fazla ağırlık değeri verilir.
- Yeni bir model oluşturulur ve tekrardan tahmin işlemi gerçekleştirilir.
- Daha önceki modellerin hatalarını düzelten çoklu model mekanizma yapıları oluşturulur.
- Bu tekrarlamalı süreç istenen model sayısında ve doğruluk değerinde bir sınıra ulaşana kadar devam eder.
- Oluşturulan son model tüm modellerin ortalama değerlerini temsil etmektedir.



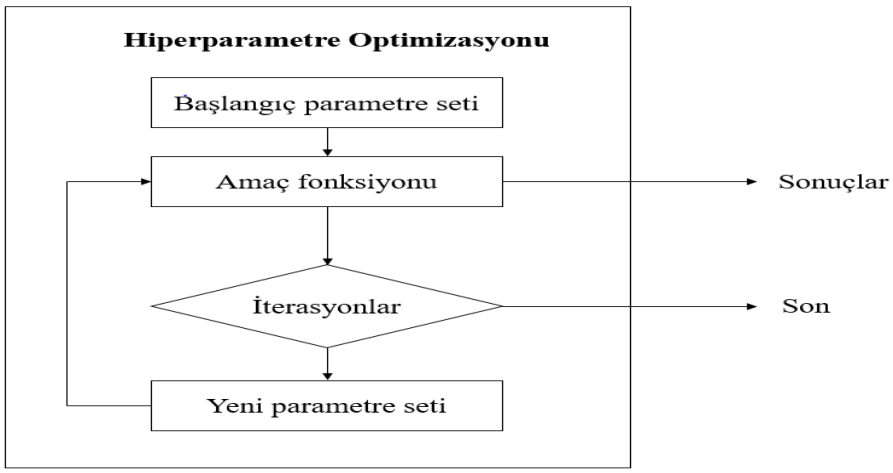
Şekil 5.15: Torbalama yöntemi.

5.6 Hiperparametre Optimizasyonu

Makine öğrenme algoritmaları tasarlanırken kullanılan parametreler eğitim sürecinde veri setinden doğrudan elde edilebilen veya programcı tarafından önceden tanımlanmış olanlar olmak üzere ikiye ayrılmaktadır. Bu parametreler model parametreleri ve hiperparametrelerdir. Model parametreleri veri setlerinden direkt olarak öğrenilebilmektedir. Program tasarımcısının modeli kurarken bu parametreler için ayarlama yapması beklenmemektedir. Model parametreleri öğrenme sürecinin bir parçası olarak kabul edilmektedir. DVM destek vektörleri model parametresine örnektir. Hiperparametrelerin model parametrelerin aksine veri setinden tahmin edilmesi mümkün değildir. Bu nedenle tasarımcı tarafından ayarlanması gerekmektedir. DVM çekirdek sayısı, K-NN k sayısı ve TÖ öğrenme sayısı makine öğrenme algoritmasındaki hiperparametrelere örnek olarak verilebilmektedir. Makine öğrenme algoritmaları oluşturulurken daha yüksek sınıflandırma sonuçları elde edilmesi için hiperparametrelerin optimize edilmesi gerekmektedir [104].

Hiperparametre optimizasyonu bir modelin performansını en üst düzeye çıkaran en doğru hiperparametre kombinasyonun bulunma süreci olarak tanımlanabilmektedir. Makine öğrenme algoritmalarında hiperparametreler manuel ve otomatik olarak ayarlanmaktadır. Manuel olarak hiperparametelerin belirlenmesi işleminde tasarımcı oluşturacağı model için farklı hiperparametre kombinasyonlarını denemekte ve en iyi performans gösteren hiperparametreleri seçmektedir. Hiperparametrelerin manuel olarak belirlenmesi işlemi süreç ile ilgili tasarımcıya daha fazla kontrol imkanı verse de zaman alıcıdır. Dikkate alınması gereken çok sayıda hiperparametre olduğunda

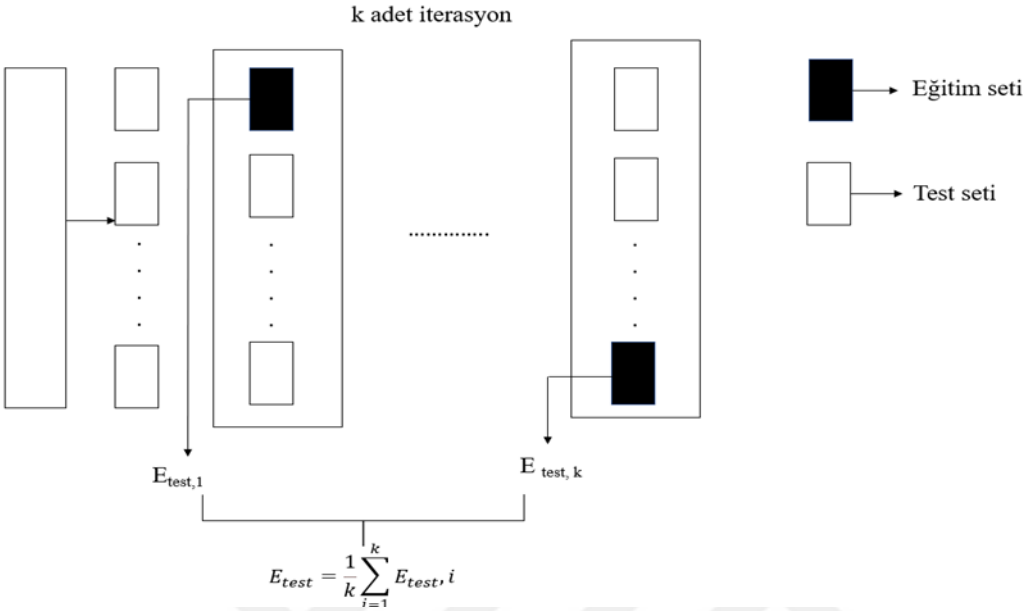
hiperparametrelerin manuel olarak belirlenmesi çok pratik bir yöntem değildir. Hiperparametrelerin otomatik olarak ayarlanması işlemi hali hazırda var olan algoritmaların kullanılması ile gerçekleşmektedir [104]. Şekil 5.16'da genel olarak bir hiperparametre optimizasyon sürecinin akış şeması gösterilmiştir. Genel olarak bütün hiperparametre optimizasyon algoritmaları benzer akış şemasını kullanmakta ve sadece yeni parametre setinin belirlenmesinde farklılıklar göstermektedir. Başlangıç ve yeni oluşturulan parametreler bir amaç fonksiyonu ile değerlendirilmektedir. Parametre setinde bulunan her öğe için hesaplanan performans değeri son olarak bir veri tabanında saklanmaktadır [105].



Şekil 5.16: Genel bir hiperparametre optimizasyon sürecinin şematik akışı.

Hiperparametre optimizasyonu algoritmanın doğruluğunu hesaplamak için çapraz doğrulama işlemini kullanan bir amaç fonksiyonundan yararlanmaktadır. Çapraz doğrulama verinin k alt kümeye bölünmesi ve daha sonra da bu k alt kümenin her birinin test verisi olarak kullanılması anlamına gelmektedir. Böylece oluşturulan alt kümelerin k. ıncı alt küme test verisi olarak, kalan k-1 alt kümeler ise eğitim ve validasyon için kullanılmaktadır. Şekil 5.17 oluşturulan k-katlı çapraz doğrulamayı göstermektedir. Buna göre, ilk olarak veri seti k adet parçaya bölünmektedir. Daha sonra, her bir alt parçanın test verisi olarak kullanıldığı k adet iterasyon gerçekleştirilmektedir. Kalan k-1 parça eğitim ve doğrulama için kullanılmaktadır. Her bir iterasyon için, amaç fonksiyonu olarak karekök ortalama hatası (Root Mean Squared Error-RMSE) hesaplanmaktadır. Eşitlik 5.56 ortalama hata karesinin hesaplanmasını göstermektedir. Eşitlik 5.56'daki $\varphi_{tahmin} = [\varphi_{1,tahmin}, \ldots\ldots \varphi_{n,tahmin}]$ modelin tahminini, $\varphi'_j = [\varphi'_1 \ldots\ldots \varphi'_j]$ test veri setinin referansını ifade etmektedir.

$$E_{test,i}(\varphi_{tahmin}, \varphi') = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\varphi_{j,tahmin} - \varphi' j)^2} \quad (5.56)$$



Şekil 5.17: k-katlı çapraz doğrulama.

Hiperparametre optimizasyonu optimizasyon problemini çözmeyi amaçlamaktadır. Eşitlik 5.57'de hiperparametre optimizasyonun formülü gösterilmektedir. Hiperparametre optimizasyonu, modelin referans verilerini olabildiğince doğru olarak tahmin etmeye çalışmaktadır. Eşitlik 5.57'deki $Etest$ validasyon verisinde değerlendirilen ortalama hata karesine gibi en aza indirilecek olan amaç fonksiyonunu, l^* en düşük değeri veren hiperparametre setini ve l ise L alanındaki herhangi bir değeri ifade etmektedir.

$$l^* = \min_{l} Etest \quad (5.57)$$

Bütün hiperparametre optimizasyon yöntemlerinde amaç fonksiyonun hesaplanması işlemi benzerdir. Amaç fonksiyonundaki her bir iterasyon hiperparametre konfigurasyonun değerlendirilmesine karşılık gelmektedir. Hiperparametre optimizasyon yönteminin yapısına göre yeni hiperparametrelerin seçim kriterleri değişkenlik göstermektedir. Genel olarak literatürde sıklıkla kullanılan hiperparametre optimizasyon yöntemleri ızgara arama, rastgele arama ve BO'dur. Izgara arama yönteminde her bir konfigürasyon parametresi için bir değerler kümesi seçilmekte ve bu değerlerin tüm kombinasyonları değerlendirilmektedir. Bu değerlendirme sürecinin

sonunda kombinasyonların en iyisi döndürülmektedir. Rastgele arama yöntemi rastgele parametreler oluşturmakta ve bu parametrelerin sonuçlarını elde edilen en iyi sonuçla karşılaştırmaktadır. En iyi sonuçtan daha iyi bir sonuç elde edildiği taktirde en iyi parametreler en iyi sonucu veren parameteler ile değiştirilmektedir. Bu işlem süreci durma kriteri sağlanana kadar devam etmektedir. BO en iyi hiperparametre kombinasyonunu bulmak için makine öğrenme algoritmalarından faydalanmaktadır. Bu yöntemde daha önceki sonuçların değerlendirilmesi ile hiperparametreler için yeni sonuçlar tahmin edilmektedir. Bayes optimizasyonu bu işlem sürecini Gauss sürecini kullanarak gerçekleştirmektedir. Durma kriteri sağlanana kadar bahsedilen bu işlem süreci devam etmektedir. En iyi sonuca ulaşılana kadar elde edilen bir önceki sonuçlar kaydedilmektedir. BO olasılıksal bir hesaplama metodu kullandığından ızgara arama ve rastgele arama yöntemlerine göre daha az değerlendirme yaparak sonuca daha hızlı bir şekilde ulaşabilmektedir. Bahsedilen bu avantajlarından dolayı çalışma kapsamında makine öğrenme algoritmalarının hiperparametre opitimizasyonu için BO yöntemi kullanılmıştır [104, 105].

5.6.1 Bayes optimizasyonu

Optimizasyon en genel haliyle; amaç fonksiyonu adı verilen gerçek değerli bir fonksiyonu en aza indiren noktaları bulma işlemidir. Hiperparametre, bir sınıflandırıcının (örneğin bir destek vektör makinesinin kutu kısıtlaması veya güçlü bir topluluk topluluğunun öğrenme hızı gibi) dahili bir parametresidir. Bu parametreler, bir sınıflandırıcının performansını güçlü bir şekilde etkileyebilmektedir. Buna rağmen hiperparametreleri optimize etmek zor veya zaman alıcı bir işlemdir. hiperparametreleri optimize edilmesi bir sınıflandırıcının veya regresyonun çapraz doğrulama kaybını en aza indirmeye çalışmak anlamına gelmektedir.

Bayes optimizasyonunda ızgara ve rastgele arama yöntemlerinden farklı olarak geçmişteki hareketler kaydedilmekte, sonraki hareketlerin belirlenmesinde geçmişteki hareketler referans olarak alınmaktadır. Izgara ve rastgele arama yöntemlerinde her iterasyonda amaç fonksiyonun çağrılması durumu mevcuttur. Bu durum oldukça maliyetli bir işlemdir. Bu problemi düzeltmek için BO yönteminde bir sonraki hiperparametre kombinasyonunu belirlemek için amaç fonksiyonun çağrılması yerine bir vekil fonksiyon tanımlanmaktadır. Vekil fonksiyon oluşturmanın faydası; işlev amaç fonksiyonuna yapılan önceki çağrılara dayanarak, değerlendirmek için yalnızca

en umut verici hiperparametre setini seçerek amaç fonksiyonuna yapılan çağrı sayısı azaltılmasıdır. BO yöntemi genel olarak giriş olarak verilen bir amaç fonksiyonunun maksimum veya minimum yapan noktaların bulunmasını amaçlamaktadır. BO yönteminde amaç fonksiyonun bir olasılıksal modelini oluşturulmaktır. Oluşturulan bu olasılık modeli asıl amaç fonksiyonunda kullanılmak üzere en iyi sonuç veren hiperparametreleri bulmak için kullanılmaktadır. Algoritma, olasılıksal tahmin modelinin oluşturulması için Bayes Teoreminden yararlanmaktadır. Hiperparametre optimizasyonu için kullanılan genel formül Eşitlik 5.58'de gösterilmiştir.

$$x^{+} = argmax\, f(x) \quad x \in \mathrm{H} \tag{5.58}$$

Burada f(x) amaç fonksiyonunu, H hiperparametreler kümesini, x H uzayındaki herhangi bir alt kümeyi ve x^{+} yeni hiperparametre alt kümesini ifade etmektedir. Bu eşitlikteki amaç validasyon setinde en iyi sonuçları verebilecek olan hiperparametreleri seçmektir [106, 107].

Bayes optimizasyonunda olasılıksal vekil modeli ve bir edinim fonksiyonu olmak üzere iki temel bileşen bulunmaktadır. Vekil model bütün gözlemlenen noktaları amaç fonksiyonuna sığdırmayı amaçlamaktadır. Olasılıksal vekil modelin tahmine dayalı dağılımı elde edildikten sonra edinim fonksiyonu keşif ve sömürü arasındaki dengeyi bulmaktadır. Böylece edinim fonksiyonu kullanılarak farklı noktalar kullanılabilmektedir. Keşif, fonksiyonun yüksek belirsizlik alanlarından örneklenmesi olarak tanımlanmaktadır. Sömürü, fonksiyonun yüksek değerlere sahip olan noktalardan örneklenmesi işlemidir. Keşif ve sömürü arasında kurulan denge örnekleme sayısının azalmasını sağlamaktadır. Bu durumda, fonksiyon birden çok yerel maksimum noktasına sahip olsa da modelin performansı artmaktadır.

Bayes optimizasyonunda vekil model olarak Gauss süreci sıklıkla kullanılmaktadır. Bayes optimizasyonu veriyi fit etmek ve sonsal dağılımı güncellemek amacıyla Gauss sürecini kullanmaktadır. Gauss süreci parametrik olmayan bir model olup parametre sayısı sadece girdi değerlerine bağlıdır. Gauss süreci, bir Gauss dağılımının sonsuz boyutlu bir rastlantısal sürece bir uzantısı olarak tanımlanmaktadır. Gauss süreci temel olarak ön olasılık dağılımını varsayarak son olasılık dağılımını tahmin etmekte ve eğitim verilerine dayanarak ön olasılık dağılımını güncellemeyi amaçlamaktadır. Gauss sürecinin temel amacı verilen x değerlerinin üzerinde tahmine bir dayalı bir dağılımı öğrenerek tahminlere göre yeni değerlendirmelerin yapılabilme olanağının

sağlanabilmesidir. Gauss süreci Gauss dağılımına benzer olarak kovaryans ve ortalama fonksiyonları ile tanımlanmaktadır. Gauss sürecinde bilinmeyen f fonksiyonunun ortalama fonksiyonu (m(x) ve kovaryans fonksiyonu (k (x, x') olan bir süreci takip ettiği varsayılmaktadır. Bir Gauss süreci Eşitlik 5.59'daki gibi tanımlanabilmektedir.

$$f(x) \sim GS\big(m(x), k(x, x')\big) \tag{5.59}$$

Gauss sürecinde ortalama fonksiyonu sıfır olarak kabul edilmektedir (m(x)=0). Kovaryans fonksiyonu ise Eşitlik 5.60'daki gibi ifade edilmektedir.

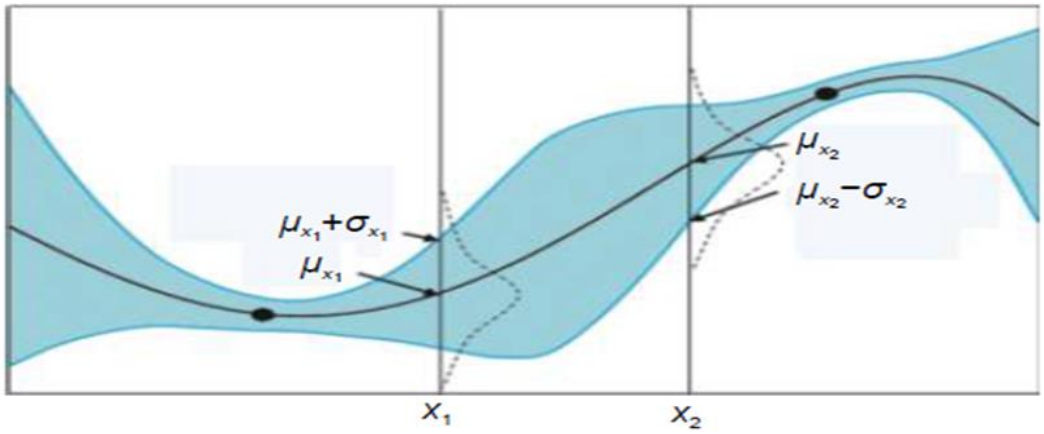
$$k\big(x_i, x_j\big) = \exp\left(-\frac{1}{2}||x_i - x_j||^2\right) \tag{5.60}$$

Eşitlikteki x_i ve x_j i. ve j. dereceden örnek noktalarını, $k\big(x_i, x_j\big)$ ise kovaryans fonksiyonunu ifade etmektedir. Eğer x_i ve x_j değerleri birbirine yakınsa kovaryans matris değeri 1 yakınsarken aksi durumda ise 0 değerine yakınsamaktadır.

f(x) amaç fonksiyonun sonsal dağılımını elde etmek amacıyla izlenen süreç şöyledir:

1. Öncelikle f fonksiyonun değerleri gözlemlenen eğitim veri setine (D) atanmaktadır. $(D = \{x_i, f_i(x)\}_{i=1}^{t})$. Daha sonra f fonksiyonun değerleri çok değişkenli normal dağılıma göre çizdirilmektedir. (f~ N (0, K)). Bu durum $k\big(x_i, x_j\big)$=1 olduğunda geçerlidir.
2. f fonksiyonuna göre yeni örnek noktasındaki $f_{t+1}=f(x_{t+1})$ değerleri hesaplanmaktadır. Gauss sürecinin varsayımına göre, $f_{1:t}$ eğitim setindeki ek noktayı f_{t+1} ise boyutlu normal dağılımı temsil etmektedir [107].

Şekil 5.18'de iki gözlem için tek boyutlu bir Gauss sürecini göstermektedir. Şekildeki siyah noktalar veri noktalarının gözlem değerlerini, siyah eğri amaç fonksiyonun tahmin edilen ortalama değerlerini ifade etmektedir. Mavi alanlar tahmin edilen vekil modelin standart sapma derecesini göstermektedir. İki gözlem noktasındaki aralık darsa standart sapma değerleri küçülmektedir. Bu durum, henüz araştırılmamış bir yerde daha büyük fonksiyon değerlerinin olabileceği anlamına gelmektedir [106, 107].



Şekil 5.18: Tek boyutlu gauss süreci [106].

Bu durumda vekil fonksiyonu ve kümülatif dağılım fonksiyonu kullanılarak "İyileştirme Olasılığı" (Probability of Improvement-PI) hesaplanabilmektedir. PI fonksiyonu Eşitlik 5.61'deki gibi hesaplanabilmektedir. PI fonksiyonu mevcut optimal değerden daha iyi bir değere sahip noktanın bulunmasını amaçlamaktadır. Yeni örnek model değeri ile mevcut optimal değer arasındaki fark ε kadar ise yeni örnek model mevcut optimal değerinin yerini almaktadır.

$$PI(x) = P(f(x) \geq f(x^+) + \varepsilon) = \varphi\left(\frac{(\mu(x) - f(x^+) - \varepsilon)}{\sigma(x)}\right) \tag{5.61}$$

Edinim fonksiyonu olarak "Beklenen İyileştirme" (Expected Improvement-EI) terimi kullanılmaktadır. İyileştirmenin derecesi (I) yeni örnek nokta değeri ile mevcut optimal değer arasındaki fark olarak tanımlanmaktadır. İyileştirmenin derecesi Eşitlik 5.62 ile ifade edilmektedir.

$$I = \max\left(0, f_{t+1}(x) - f(x^+)\right) \tag{5.62}$$

Burada $f_{t+1}(x)$ yeni örnek nokta değerini, $f(x^+)$ ise mevcut optimal değeri ifade etmektedir. Yeni örnek nokta değeri mevcut optimal değerinden daha az ise iyileştirme fonksiyonu sıfır değerini almaktadır. EI fonksiyonun optimizasyon stratejisine göre, EI değerinin mevcut optimal değerine göre maksimize edilmesi amaçlanmaktadır. Bu durum Eşitlik 5.63 ile ifade edilmektedir.

$$x = argmax E\left(max f_{t+1}(x) - f(x^+)\right) \tag{5.63}$$

Eşitlik 5.63'deki E terimi beklenen değeri ifade etmekte ve olasılık yoğunluk fonksiyonu ile hesaplanmaktadır. Sonuç olarak, eldeki var olan gözlemlere dayanarak

bir sonraki adımda seçilecek hiperparametre kümesinin belirlenmesi için her iterasyonda belirlenen sayıda aday çözümler bulunmaktadır [106]. Bayes optimizasyon algoritması, sınırlı bir etki alanında x için bir skaler amaç fonksiyonunu f(x) en aza indirmeye çalışmaktadır. Bu amaç fonksiyonu deterministik ya da stokastik olabilmektedir. Amaç fonksiyonu aynı x noktasında değerlendirildiğinde farklı sonuçlara ulaşabilmektedir. BO yöntemindeki minimizasyon işleminin temel aşamaları şöyledir:

- f(x) amaç fonksiyonun Gauss süreç modelinin belirlenmesi,
- Bayes Teoremi kullanılarak güncelleme prodesürünün oluşturulması,
- Bu güncelleme prosedürünün her yeni f(x) değerlendirilmesinde Gauss süreç modelini değiştirmesi için kullanılması,
- Yeni x noktasının bulunması için maksimize edilmiş bir edinim fonksiyonun oluşturulmasıdır.

Bayes optimizasyon algoritması değişken sınırlar içerisinde rastgele alınan ilk değerlendirme noktalarının (x_i noktaları) $y_i=f(x_i)$ fonksiyonun değerlendirilme sürecinden oluşmaktadır. Bu süreçte değerlendirme hataları mevcutsa, süreç başarılı ilk değerlendirme noktalarına ulaşana kadar rastgele nokta alınmaktadır. Bayes optimizasyon algoritmasındaki daha sonraki adımlar şöyledir:

- Fonksiyonlar üzerinde sonsal bir dağılım elde etmek için f (x)' in Gauss süreç modeli güncellenmektedir.
- Edinim işlevini maksimize eden yeni x noktalarının bulunması amaçlanmaktadır.

Bayes optimizasyonu maksimum iterasyon sayısı, sabit bir zaman aralığına ve durdurma kriterine ulaşma gibi kriterleri sağlandığında algoritma sonlanmış olmaktadır. Bu çalışma kapsamında makine öğrenme algoritmalarının BO yöntemi ile hiperparametelerinin optimize edilmesi için maksimum iterasyon sayısı 30 olarak seçilmiştir [106]. Çalışmada makine öğrenmesi için kullanılan parametreler Çizelge 5.4'de gösterilmiştir.

Çizelge 5.4: Makine öğrenmesinde kullanılan hiperparametreler.

Algoritma	**Hiperparametreler**	**Arama Aralığı**
KA	Mak. Düğüm Sayısı	[1-568]
	Düğüm Kriteri	Gini indeksi, Maksimum sapma azaltma
NB	Dağılım Adı	Gauss, Çekirdek
	Çekirdek Tipi	Gauss, Kutu, Epanechnikov, Üçgen
DVM	Çekirdek Fonksiyonu	Gauss, Doğrusal, Quadratik, Kübik
	Çekirdek Ölçeği	[0,001-1000]
	Kutu Kısıtlama Seviyesi	[0,001-1000]
	Standartlaştırılmış Veri	Doğru/Yanlış
K-NN	Komşu Sayısı	[1-285]
	Uzaklık Metriği	City Block, Correlation, Euclidian, Hamming, Jaccard, Mahalanobis, Minkowski, Spearman
	Standartlaştırılmış Veri	Doğru/Yanlış
TÖ	Topluluk Metodu	Bag, GentleBoost, LogitBoost, AdaBoost, RUSBoost
	Öğrenme Sayısı	[10-500]
	Öğrenme Oranı	[0,001-1]
	Mak. Düğüm Sayısı	[1-568]

5.7 Sınıflandırma Algoritmalarının Performans Değerlendirme Kriterleri

Sınıflandırma algoritmalarının etkinliği, karmaşıklık matrisinden elde edilen doğru pozitif (DP), yanlış pozitif (YP), doğru negatif (DN) ve yanlış negatif (YN) değerlendirme kriterlerini dikkate alarak hesaplanmaktadır. Karmaşıklık matrisi DP, YP, DN ve YN ifadelerini bir tablo şekinde sunmaktadır [108]. Çizelge 5.5'de karmaşıklık matrisi gösterilmiştir.

Çizelge 5.5: Kamaşıklık Matrisi

Tahmin	**Gerçek**	
	Pozitif	Negatif
Pozitif	DP	YP
Negatif	DN	YN

DP: Bir öğenin pozitif olduğu doğru tahminlerin sayısını ifade etmektedir. Gerçekte hasta olan ve algoritmaya göre hasta olarak sınıflandırılan örnek sayısıdır.

YP: Bir öğenin negatif olduğu yanlış tahminlerin sayısını ifade etmektedir. Gerçekte hasta olmayan ancak algoritmaya göre hasta olarak sınıflandırılan örnek sayısıdır.

DN: Bir öğenin pozitif olduğu yanlış tahminlerin sayısını ifade etmektedir. Gerçekte hasta olan ancak algoritmaya göre hasta değil olarak sınıflandırılan örnek sayısıdır.

YN: Bir öğenin negatif olduğu yanlış tahminlerin sayısını ifade etmektedir. Gerçekte de algoritmaya göre hasta değil olarak sınıflandırılan örnek sayısıdır.

Doğruluk: Bir modeldeki doğru tahmin edilen örneklerin, toplam veri seti sayısına oranı olarak ifade edilmektedir. Sınıflandırma algoritmasının vakaları ne kadar iyi tahmin edebildiğini ölçmektedir. Eşitlik 5.64-Eşitlik 5.67'de sırasıyla doğruluk, kesinlik, duyarlılık ve F1-skoru hesaplamaları gösterilmiştir.

$$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN} \tag{5.64}$$

Kesinlik: Doğru pozitif olarak bulunan örnek sayısının tahmin edilen tüm pozitif örnek sayısına oranıdır.

$$Kesinlik = \frac{DP}{DP + YP} \tag{5.65}$$

Duyarlılık: Doğru pozitif olarak tahmin edilen örneklerin gerçek doğrulara oranıdır.

$$Duyarlılık = \frac{DP}{DP + YN} \tag{5.66}$$

F1-skoru: Duyarlılık ile kesinlik parametrelerinin harmonik ortalamasıdır.

$$F1 - skoru = \frac{2.Kesinlik.Duyarlılık}{Kesinlik + Duyarlılık} \tag{5.67}$$

6. DENEYSEL ÇALIŞMALAR VE SONUÇLAR

Çalışma kapsamında KA, NB, DVM, K-NN ve TÖ öğrenme algoritmaları olmak üzere 5 farklı makine öğrenme yöntemi meme kanserinin teşhisi amacıyla kullanılmıştır. Çalışma kapsamında kullanılan ilk veri meme kanseri bulgularını içeren Wisconsin Meme Kanseri Veri Seti (WBCD)' dir. İkinci veri seti ise Ankara Eğitim ve Araştırma Hastanesi Radyoloji bölümünden 101 hastaya ait mamografi görüntülerinden oluşmaktadır (MBCD-Mamografik Meme Kanseri Veri Seti). İkinci verideki mamografi görüntüleri üzerinde radyologlar tarafından işaretlenen şüpheli meme lezyonlarının saptanması (ROI'lerin belirlenmesi) için gri seviye eşikleme yöntemi ve morfolojik operatörler kullanılmıştır. Her bir ROI için 54 tane morfolojik ve doku öznitelik hesaplanmıştır. İki farklı meme kanseri veri seti üzerinde meme kanserinin tespiti için en yüksek sınıflandırma oranlarını elde etmek için farklı deneyler yapılmıştır.

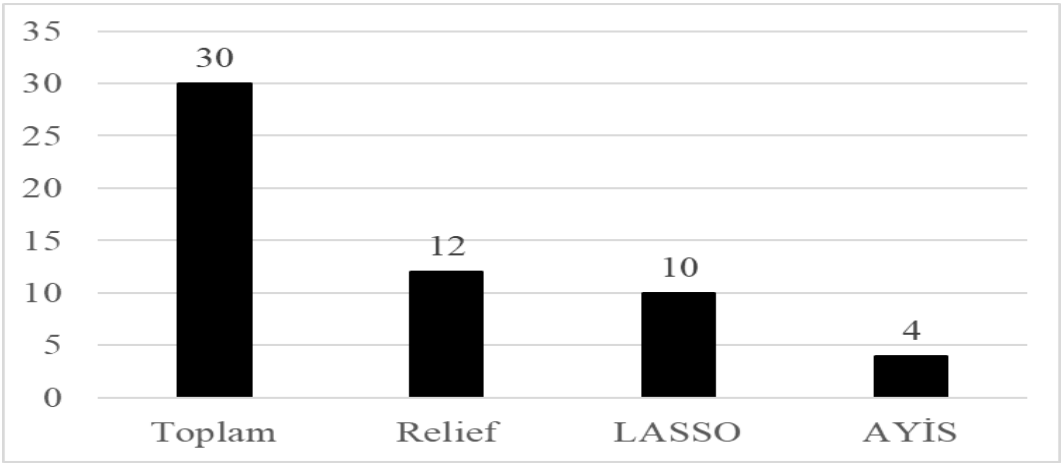
- İlk önce her iki veri setindeki bütün öznitelikler (BÖ) makine öğrenme algoritmalarına giriş verisi olarak verilmiş ve MATLAB programı tarafından geliştirilen fonksiyonlarından yararlanılarak sınıflandırma işlemi yapılmıştır.
- Daha sonra tekrardan bütün öznitelikler (BÖ) makine öğrenme algoritmalarına giriş olarak verilmiştir. Makine öğrenme algoritmalarının hiperparametreleri BO yöntemi kullanılarak optimize edilmiştir. BO yöntemi için MATLAB Statistics and Machine Learning Toolbox arayüzü kullanılmıştır [109].
- Son olarak ise sırasıyla veri setindeki ayırt edici öznitelikler sırasıyla Relief, LASSO ve AYİS öznitelik seçim yöntemleri kullanarak belirlenmiştir. BO yöntemi kullanarak makine öğrenme yöntemlerinin hiperparametreleri optimize edilmiştir.

Öznitelik seçim yöntemleri için MATLAB tarafından geliştirilen "relieff", "lasso" ve "sequentialfs" fonksiyonlarından yararlanılmıştır. "relief" fonksiyonu en yakın 10 komşuyu kullanarak özniteliklere ağırlık vererek önem sırasına göre sıralamaktadır. Yüksek ağırlık değerine sahip olan öznitelikler sınıflandırma sonucuna daha fazla etki ederken, düşük ağırlık değerine sahip olanlar daha az etki etmektedir. Tüm ağırlıklar hesaplandıktan sonra ayırt edici özellikleri seçmek için bir eşik değeri

uygulanmaktadır [Url-3]. Bu çalışmada eşik değeri olarak 0 seçilmiştir. Özniteliklerin ağırlıkları 0'dan büyük olan öznitelikler seçilmiştir. “lasso” fonksiyonu 10 katlı çapraz doğrulama yöntemini kullanarak doğrusal bir regresyonun katsayılarının belirlenmesine olanak sağlamaktadır. Lasso tekniği veri setlerinde bulunan öznitelik katsayılarını belirlerken cezalandırılmış regresyon yöntemi olarak da adlandırılan bir Büzülme yöntemi kullanmaktadır. Bu yöntemdeki temel amaç hata kareler toplamını minimize eden katsayıları, katsayılara ceza uygulayarak bulmaktır. İlgisiz olan özniteliklerin katsayıları 0’ a eşitlenmektedir. 0 değerinden büyük katsayılara sahip olan öznitelikler seçilmiştir [Url-4]. MATLAB’da bulunan “sequentialfs” fonksiyonu AİYS modelini oluşturmak için kullanmaktadır. Bu fonksiyon boş bir öznitelik kümesinden başlayarak, henüz seçilmemiş özniteliklerin herbirini sırayla ekleyerek aday öznitelik alt kümeleri oluşturmaktadır. Her aday öznitelik kümesi için eğitim ve test verilerinin giriş ve çıkış alt kümelerini art arda çağırarak 10 katlı çapraz doğrulama işlemini gerçekleştirmektedir. Eğitim verilerinin giriş ve çıkış alt kümeleri X ve Y satırlarının aynı alt kümesini içerirken test verilerinin giriş ve çıkış alt kümeleri tamamlayıcı satır alt kümesini içermektedir. Bir modeli eğitmek için eğitim verileri (giriş ve çıkış) kullanılmaktadır. Oluşturulan bu model kullanılarak giriş test verileri tahmin edilmekte ve son olarak da çıkış test verilerinden tahmin edilen değerlerin kayıp ölçüsü döndürülmektedir. Genel olarak kayıp ölçümleri regresyon modelleri için hata karelerinin toplamını ve sınıflandırma modelleri için yanlış sınıflandırılmış gözlemlerin sayısını içermektedir. Her bir aday özellik alt kümesi için ortalama kriter değerlerini hesapladıktan sonra, “sequentials” fonksiyonu ortalama kriter değerini en aza indiren aday özellik alt kümesini seçmektedir. Bu işlem daha fazla öznitelik eklemek kriterini düşürmeyene kadar devam etmektedir [Url-5].

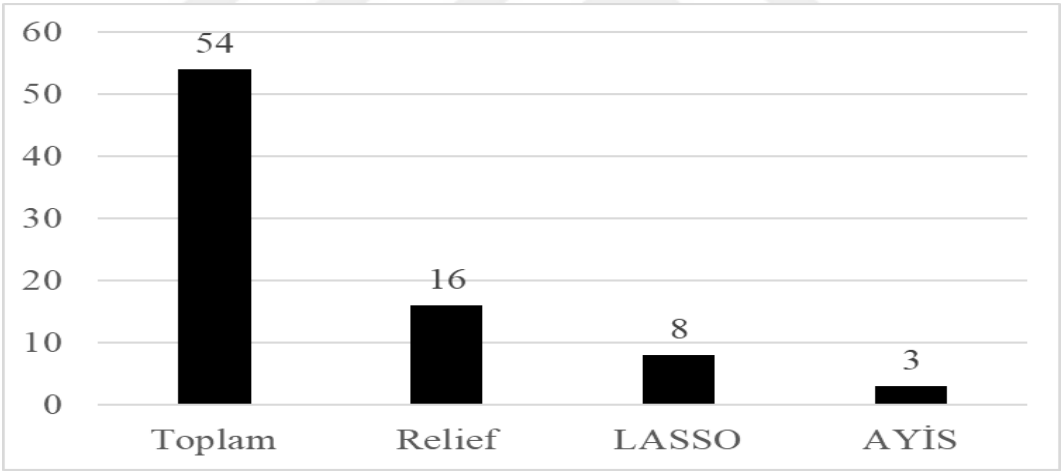
6.1 Öznitelik Seçim Yöntemi Sonrasında Elde Edilen Ayırt Edici Öznitelikler

Şekil 6.1’de WBCD veri setine öznitelik yöntemleri uygulandıktan sonra veri seti için seçilen ayırt edici öznitelik sayıları gösterilmiştir. Veri setine ait 30 adet öznitelikten sırasıyla Relief, LASSO ve AİYS yöntemleri uygulandıktan sonra sırasıyla 12, 10 ve 4 öznitelik ayırt edici öznitelik olarak bulunmuştur.



Şekil 6.1: WBCD veri seti için öznitelik yöntemleri uygulandıktan sonra ayırt edici öznitelikler.

MBCD veri seti için ise öznitelik yöntemleri uygulandıktan sonra seçilen ayırt edici öznitelik sayıları Şekil 6.2'de gösterilmiştir. 54 öznitelikten sırasıyla Relief, LASSO ve AİYS yöntemleri uygulandıktan sonra 16, 8 ve 3 adet ayırt edici öznitelik bulunmuştur.



Şekil 6.2: MBCD veri seti için öznitelik yöntemleri uygulandıktan sonra ayırt edici öznitelikler.

Çizelge 6.1'de WBCD ve MBCD veri setleri için Relief yöntemi uygulandıktan sonra seçilen öznitelikler belirtilmiştir.

Çizelge 6.1: WBCD ve MBCD veri setleri için Relief yönteminden sonra seçilen ayırt edici öznitelikler.

Seçim Yöntemi	WBCD	MBCD
Relief	Ortalama yarıçap	Çevre
	Ortalama doku	Min yarıçap
	Ortalama içbükeylik	Dış merkezlilik
	Yarıçap şiddeti	Katılık
	Çevre şiddeti	Uzatılmışlık
	Alan şiddeti	Dairesellik 2
	İçbükeylik şiddeti	Dağılım
	Simetri şiddeti	Varyans
	En kötü yarıçap	Ortalama mutlak sapma
	En kötü doku	90.dereceden yüzdelik dilim
	En kötü çevre	Korelasyon
	En kötü simetri	Varyans farkları
		Korelasyon bilgi ölçümü 2
		Gri seviye düzensizliği
		Koşu yüzdesi
		Düşük gri seviye koşu vurgusu
		Kısa koşu düşük gri-seviye vurgusu

Çizelge 6.2'de WBCD ve MBCD verileri için LASSO yöntemi uygulandıktan sonra seçilen öznitelikler belirtilmiştir.

Çizelge 6.2: WBCD ve MBCD veri setleri için LASSO yönteminden sonra seçilen ayırt edici öznitelikler.

Seçim Yöntemi	WBCD	MBCD
LASSO	Ortalama yarıçap	Çevre
	Ortalama doku	Katılık
	Ortalama içbükeylik noktası	Şekil indeksi
	Yarıçap şiddeti	Ortalama
	Pürüzsüzlük şiddeti	Varyans
	İçbükeylik noktası şiddeti	Medyan
	En kötü yarıçap	Homojenlik
	En kötü doku	Koşu uzunluk düzensizliği
	En kötü pürüzsüzlük	
	En fötü fraktal boyut	

Çizelge 6.3'de WBCD ve MBCD veri setleri için AİYS yöntemi uygulandıktan sonra seçilen öznitelikler belirtilmiştir.

Çizelge 6.3:WBCD ve MBCD veri seti için AİYS yönteminden sonra seçilen ayırt edici öznitelikler.

Seçim Yöntemi	WBCD	MBCD
AYİS	Ortalama yarıçap	Katılık
	En kötü yarıçap	Kontrast
	En kötü doku	Koşu yüzdesi
	En kötü çevre	

6.2 Karar Ağacı Algoritmasının Sonuçları

Çizelge 6.4'de WBCD veri seti için KA sonuçları gösterilmiştir. KA; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %92,10 doğruluk, %90,10 kesinlik, %88,4 duyarlılık ve %89,46 F1-skor oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %92,80 doğruluk, %91,04 kesinlik, %89,77 duyarlılık ve %90,4 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca sırasıyla %94,03 doğruluk, %91,51 kesinlik, %92,38 duyarlılık ve %91,94 F1-skoru elde edilmiştir LASSO öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca sırasıyla %95,43 doğruluk, %94,33 kesinlik, %93,46 duyarlılık ve %93,90 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca sırasıyla %94,55 doğruluk, %92,45 kesinlik, %92,89 duyarlılık ve %92,67 F1-skoru elde edilmiştir. Çizelge 6.4'deki tüm sonuçlar karşılaştırıldığında, KA yöntemi için **LASSO-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.4:WBCD veri seti için KA yöntemi sınıflandırma sonuçları.

Deney	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
BÖ	92,10	90,10	88,84	89,46
BÖ-BO	92,80	91,04	89,77	90,40
Relief-BO	94,03	91,51	92,38	91,94
LASSO-BO	95,43	94,33	93,46	93,90
AYİS-BO	94,55	92,45	92,89	92,67

Çizelge 6.5'de MBCD veri seti için KA sonuçları gösterilmiştir. KA; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %91,28 doğruluk, %92,24 kesinlik, %93,04

duyarlılık ve %92,64 F1-skor oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %92,31 doğruluk, %93,10 kesinlik, %93,91 duyarlılık ve %93,51 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca sırasıyla %95,38 doğruluk, %96,55 kesinlik, %95,73 duyarlılık ve %96,14 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca kullanılınca sırasıyla %94,36 doğruluk, %95,69 kesinlik, %94,87 duyarlılık ve %95,28 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca sırasıyla %93,85 doğruluk, %95,69 kesinlik, %94,07 duyarlılık ve %94,87 F1-skoru elde edilmiştir. Çizelge 6.5'deki tüm sonuçlar karşılaştırıldığında, KA yöntemi için **Relief-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.5: MBCD veri seti için KA yöntemi sınıflandırma sonuçları.

Deney	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
BÖ	91,28	92,24	93,04	92,64
BÖ-BO	92,31	93,10	93,91	93,51
Relief-BO	95,38	96,55	95,73	96,14
LASSO-BO	94,36	95,69	94,87	95,28
AYİS-BO	93,85	95,69	94,07	94,87

6.3 Naive Bayes Algoritmasının Sonuçları

Çizelge 6.6'da WBCD veri seti için NB sonuçları gösterilmiştir. NB; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %93,50 doğruluk, %90,10 kesinlik, %92,27 duyarlılık ve %91,17 F1-skor oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %94,38 doğruluk, %92,45 kesinlik, %92,45 duyarlılık ve %92,45 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %96,66 doğruluk, %97,17 kesinlik, %94,06 duyarlılık ve %95,59 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %95,43 doğruluk, %93,87 kesinlik, %93,87 duyarlılık ve %93,87 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak sırasıyla kullanılınca %95,08 doğruluk, %92,92 kesinlik, %93,81 duyarlılık ve %93,36 F1-skoru elde edilmiştir. Çizelge 6.6'daki tüm sonuçlar karşılaştırıldığında, NB yöntemi için **Relief-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.6: WBCD veri seti için NB yöntemi sınıflandırma sonuçları.

Deney	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
BÖ	93,50	90,10	92,27	91,17
BÖ-BO	94,38	92,45	92,45	92,45
Relief-BO	96,66	97,17	94,06	95,59
LASSO-BO	95,43	93,87	93,87	93,87
AYİS-BO	95,08	92,92	93,81	93,36

Çizelge 6.7'de MBCD veri seti için NB sonuçları gösterilmiştir. NB; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %90,26 doğruluk, %90,51 kesinlik, %92,92 duyarlılık ve %91,70 F1-skor oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %91,28 doğruluk, %92,24 kesinlik, %93,04 duyarlılık ve %92,64 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %93,85 doğruluk, %95,69 kesinlik, %94,06 duyarlılık ve %94,87 F1-skoru elde edilmiştir. LASSO öznitelik BO yöntemi tekniği birlikte hibrit olarak kullanılınca %94,36 doğruluk, %94,83 kesinlik, %95,66 duyarlılık ve %95,24 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO tekniği birlikte hibrit olarak kullanılınca %95,9 doğruluk, %95,69 kesinlik, %97,37 duyarlılık ve %96,52 F1-skoru elde edilmiştir. Çizelge 6.7'deki tüm sonuçlar karşılaştırıldığında, NB yöntemi için **AYİS-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.7: MBCD veri seti için NB yöntemi sınıflandırma sonuçları.

Deney	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
BÖ	90,26	90,51	92,92	91,70
BÖ-BO	91,28	92,24	93,04	92,64
Relief-BO	93,85	95,69	94,06	94,87
LASSO-BO	94,36	94,83	95,66	95,24
AYİS-BO	95,9	95,69	97,37	96,52

6.4 Destek Vektör Makine Algoritmasının Sonuçları

Çizelge 6.8'de WBCD veri seti için DVM sonuçları gösterilmiştir. DVM; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %95,96 doğruluk, %93,4 kesinlik, %95,66

duyarlılık ve 94.52% F1-skoru oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %96,84 doğruluk, %93,4 kesinlik, %98,02 duyarlılık ve %95,66 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %98,77 doğruluk, %100 kesinlik, %96,81 duyarlılık ve %98,38 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %98,95 doğruluk, %97,17 kesinlik, %100 duyarlılık ve %98,57 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve tekniği birlikte hibrit olarak kullanılınca sırasıyla sırasıyla %97,19 doğruluk, %94,37 kesinlik, %98,05 duyarlılık ve %96,18 F1-skoru elde edilmiştir. Çizelge 6.8'deki tüm sonuçlar karşılaştırıldığında, DVM yöntemi için **LASSO-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.8: WBCD veri seti için DVM yöntemi sınıflandırma sonuçları.

Deney	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
BÖ	95,96	93,4	95,66	94,52
BÖ-BO	96,84	93,4	98,02	95,66
Relief-BO	98,77	100	96,81	98,38
LASSO-BO	98,95	97,17	100	98,57
AYİS-BO	97,19	94,37	98,05	96,18

Çizelge 6.9'da MBCD veri seti için DVM sonuçları gösterilmiştir. DVM; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %90,77 doğruluk, %91,38 kesinlik, %92,99 duyarlılık ve %92,18 F1-skoru oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %92,82 doğruluk, %93,97 kesinlik, %93,97 duyarlılık ve %93,97 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca %95,9 doğruluk, %96,56 kesinlik, %96,56 duyarlılık ve %96,56 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak sırasıyla %97,95 doğruluk, 98,28 kesinlik, %98,28 duyarlılık ve %98,28 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %96,41 doğruluk, % 97,42 kesinlik, %96,59 duyarlılık ve %97 F1-skoru elde edilmiştir. Çizelge 6.9'daki tüm sonuçlar karşılaştırıldığında, DVM yöntemi için **LASSO-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.9: MBCD veri seti için DVM yöntemi sınıflandırma sonuçları.

Deney	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
BÖ	90,77	91,38	92,99	92,18
BÖ-BO	92,82	93,97	93,97	93,97
Relief-BO	95,9	96,56	96,56	96,56
LASSO-BO	97,95	98,28	98,28	98,28
AYİS-BO	96,41	97,42	96,59	97

6.5 K-En Yakın Komşu Algoritmasının Sonuçları

Çizelge 6.10'da WBCD veri seti için K-NN sonuçları gösterilmiştir. K-NN; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %95,43 doğruluk, %92,46 kesinlik, %95,15 duyarlılık ve %93,78 F1-skoru oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %95,78 doğruluk, %92,93 kesinlik, %95,64 duyarlılık ve %94,26 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla sırasıyla %96,48 doğruluk, %93,87 kesinlik, %96,61 duyarlılık ve %95,22 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi ve tekniği birlikte hibrit olarak kullanılınca sırasıyla %97,19 doğruluk, %93,4 kesinlik, %99 duyarlılık ve %96,12 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %98,06 doğruluk, %95,29 kesinlik, %99,51 duyarlılık ve %97,35 F1-skoru elde edilmiştir. Çizelge 6.10'daki tüm sonuçlar karşılaştırıldığında, K-NN yöntemi için **AYİS-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.10: WBCD veri seti için K-NN yöntemi sınıflandırma sonuçları.

Deney	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru
BÖ	95,43	92,46	95,15	93,78
BÖ-BO	95,78	92,93	95,64	94,26
Relief-BO	96,48	93,87	96,61	95,22
LASSO-BO	97,19	93,4	99	96,12
AYİS-BO	98,06	95,29	99,51	97,35

Çizelge 6.11'de MBCD veri seti için K-NN sonuçları gösterilmiştir. K-NN; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %85,13 doğruluk, %88,8 kesinlik, %86,56 duyarlılık ve %87,66 F1-skoru oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %90,26 doğruluk, %92,25 kesinlik, %91,46 duyarlılık ve %91,85 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak kullanılınca sırasıyla %94,36 doğruluk, %95,69 kesinlik, %94,88 duyarlılık ve %95,28 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi BO yöntemi tekniği birlikte hibrit olarak sırasıyla %95,38 doğruluk, %95,69 kesinlik, %96,53 duyarlılık ve %96,11 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak sırasıyla %96,92 doğruluk, %98,28 kesinlik, %96,62 duyarlılık ve %97,44 F1-skoru elde edilmiştir. Çizelge 6.11'deki tüm sonuçlar karşılaştırıldığında, K-NN yöntemi için **AYİS-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.11: MBCD veri seti için K-NN yöntemi sınıflandırma sonuçları.

Deney	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
BÖ	85,13	88,8	86,56	87,66
BÖ-BO	90,26	92,25	91,46	91,85
Relief-BO	94,36	95,69	94,88	95,28
LASSO-BO	95,38	95,69	96,53	96,11
AYİS-BO	96,92	98,28	96,62	97,44

6.6 Topluluk Öğrenme Algoritmasının Sonuçları

Çizelge 6.12'de WBCD veri seti için TÖ sonuçları gösterilmiştir. TÖ; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %95,08 doğruluk, %92,45 kesinlik, %94,23 duyarlılık ve %93,33 F1-skoru oranını göstermiştir. BO yöntemi kullanılıp bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %95,61 doğruluk, %93,4 kesinlik, %94,74 duyarlılık ve %94,06 F1-skoru oranını göstermiştir. Relief öznitelik yöntemi ve BO tekniği birlikte hibrit bir yöntem olarak kullanılınca sırasıyla %96,30 doğruluk, %93,87 kesinlik, %96,14 duyarlılık ve %94,99 F1-skoru oranı göstermiştir. LASSO öznitelik yöntemi ve BO tekniği birlikte hibrit bir yöntem olarak sırasıyla %98,24 doğruluk, %98,11 kesinlik, %97,19 duyarlılık ve %97,65 F1-skoru oranı göstermiştir.

AYİS öznitelik yöntemi ve BO yöntemi birlikte hibrit bir yöntem olarak kullanılınca sırasıyla %97,01 doğruluk, %95,28 kesinlik, %96,65 duyarlılık ve %95,96 F1-skoru oranı göstermiştir. Çizelge 6.12'deki tüm sonuçlar karşılaştırıldığında, TÖ yöntemi için **LASSO-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.12: WBCD veri seti için TÖ yöntemi sınıflandırma sonuçları.

Deney	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
BÖ	95,08	92,45	94,23	93,33
BÖ-BO	95,61	93,40	94,74	94,06
Relief-BO	96,30	93,87	96,14	94,99
LASSO-BO	98,24	98,11	97,19	97,65
AYİS-BO	97,01	95,28	96,65	95,96

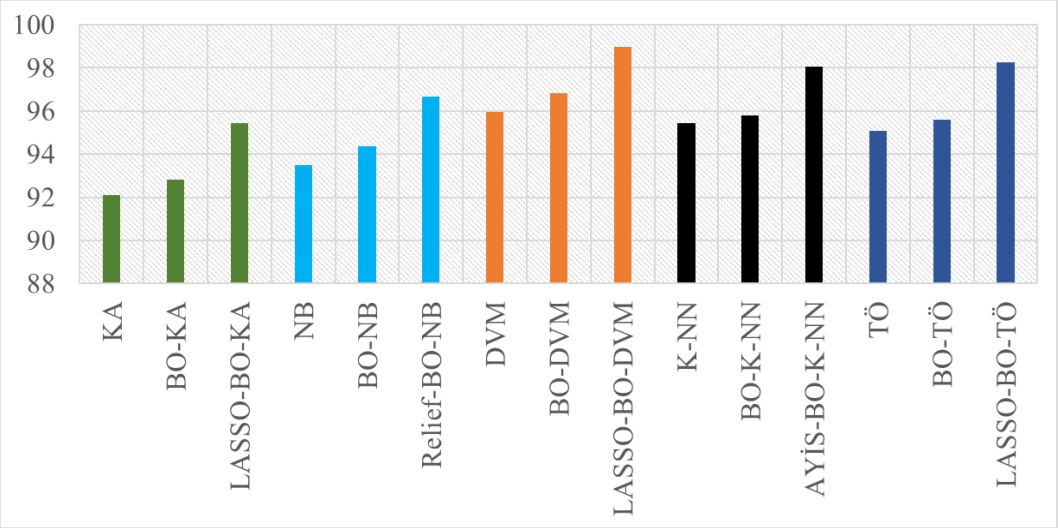
Çizelge 6.13'de MBCD veri seti için TÖ sonuçları gösterilmiştir. TÖ; herhangi bir öznitelik seçim yöntemi ve optimizasyon tekniği kullanmadan bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %89,23 doğruluk, %91,38 kesinlik, %90,6 duyarlılık ve %90,99 F1-skoru oranı göstermiştir. BO yöntemi kullanılıp, bütün öznitelikler giriş verisi olarak kullanılınca sırasıyla %91,79 doğruluk, %93,1 kesinlik, %93,1 duyarlılık ve %93,1 F1-skoru elde edilmiştir. Relief öznitelik yöntemi ve BO yöntemi tekniği birlikte BO yöntemi tekniği birlikte hibrit olarak sırasıyla %95,38 doğruluk, %95,70 kesinlik, %96,55 duyarlılık ve %96,10 F1-skoru elde edilmiştir. LASSO öznitelik yöntemi ve BO tekniği tekniği birlikte hibrit olarak %97,43 doğruluk, %98,28 kesinlik, %98,24 duyarlılık ve %97,85 F1-skoru elde edilmiştir. AYİS öznitelik yöntemi ve BO yöntemi tekniği birlikte hibrit olarak sırasıyla %95,9 doğruluk, %97,41 kesinlik, %95,76 duyarlılık ve %96,58 F1-skoru elde edilmiştir. Çizelge 6.7'deki tüm sonuçlar karşılaştırıldığında, TÖ yöntemi için **LASSO-BO** yöntemi en yüksek sınıflandırma oranına ulaşmıştır.

Çizelge 6.13: MBCD veri seti için TÖ yöntemi sınıflandırma sonuçları.

Deney	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
BÖ	89,23	91,38	90,60	90,99
BÖ-BO	91,79	93,10	93,10	93,10
Relief-BO	95,38	95,70	96,55	96,10
LASSO-BO	97,43	98,28	98,24	97,85
AYİS-BO	95,9	97,41	95,76	96,58

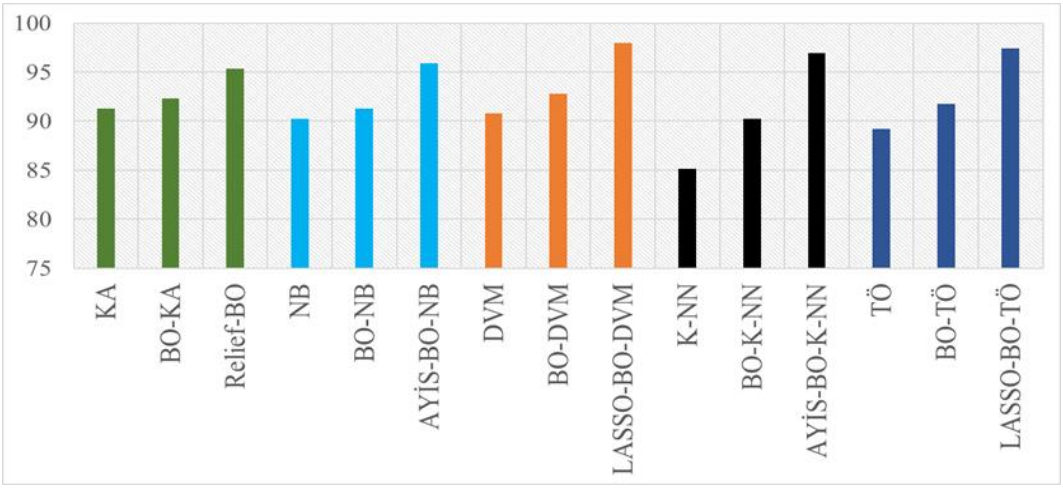
6.7 Öznitelik Seçim Yöntemlerinin ve Bayes Optimizasyonun Sınıflandırma Algoritmalarına Etkisi

Şekil 6.3'te WBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi doğruluk oranları açısından gösterilmiştir. BO metodu sırasıyla KA, NB, DVM, K-NN ve TÖ algoritmalarının doğruluklarını %0,7, %0,88, %0,88, %0,35 ve %0,53 oranında artırmıştır. LASSO-BO yöntemi KA'nın doğruluk oranını %2,63, Relief-BO yöntemi NB'nın doğruluk oranını %2,28, LASSO-BO yöntemi DVM'nın doğruluk oranını %2,11, AYİS-BO yöntemi K-NN'nın doğruluk oranını %2,28 ve LASSO-BO yöntemi TÖ'nın doğruluk oranını %2,63 artırmıştır.



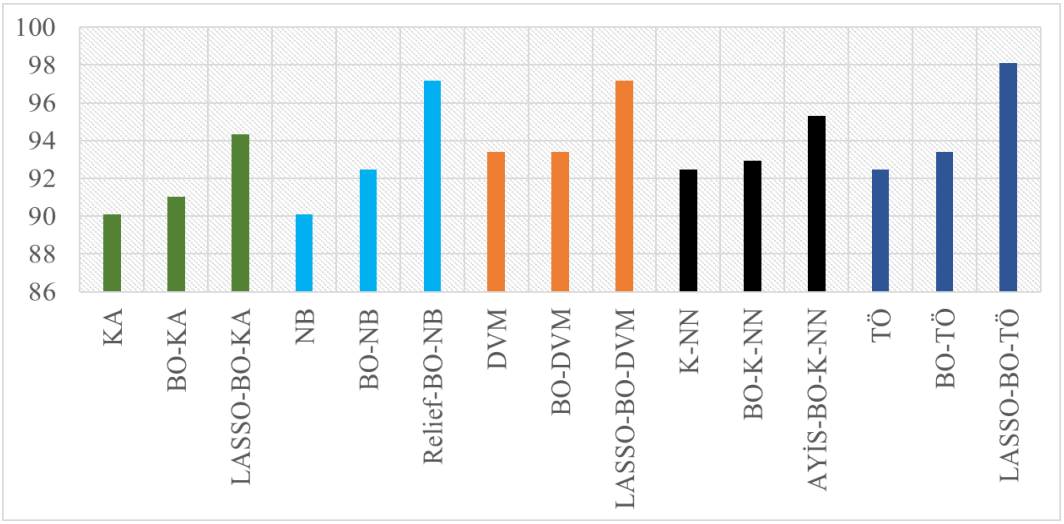
Şekil 6.3:WBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına doğruluk açısından etkisi.

Şekil 6.4'te MBCD veri seti öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi doğruluk oranları açısından gösterilmiştir. BO metodu sırasıyla KA, NB, DVM, K-NN ve TÖ algoritmalarının doğruluklarını %1,03, %1,02, %2,05, %5,13 ve %2,56 oranında artırmıştır. Relief-BO yöntemi KA'nın doğruluk performansını %3,07, AYİS-BO yöntemi NB'nın doğruluk performansını %4,62, LASSO-BO DVM'nın doğruluk performansını %5,13, AYİS yöntemi K-NN'nın doğruluk performansını %6,66 ve LASSO-BO yöntemi TÖ'nın doğruluk performansını %5,64 oranında artırmıştır.



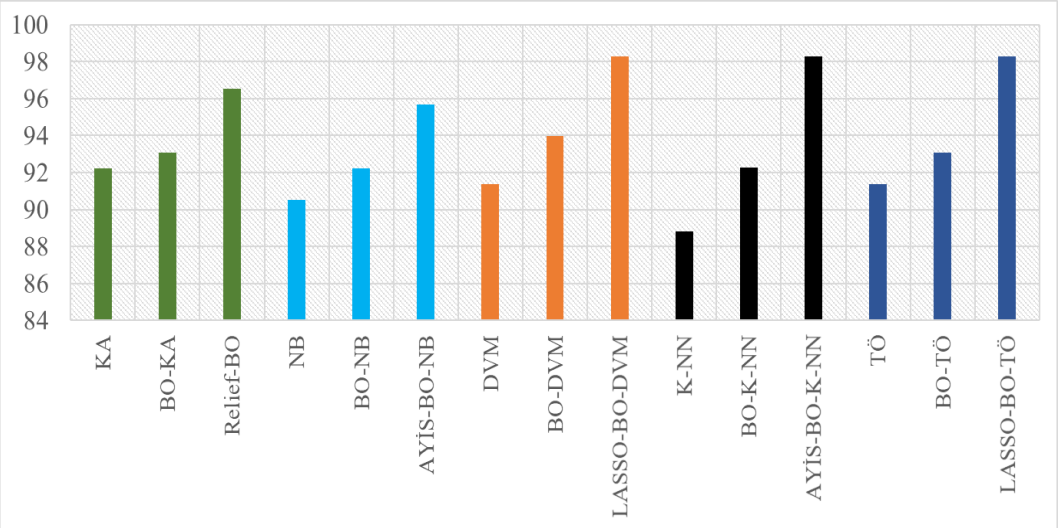
Şekil 6.4: MBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına doğruluk açısından etkisi

Şekil 6.5'te WBCD veri seti öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi kesinlik oranları açısından gösterilmiştir. BO sırasıyla KA'nın kesinlik performansını %0,94, NB'nin kesinlik performansını %2,35, DVM'nin kesinlik performansını %0,0, K-NN'nin kesinlik performansını %0,47 ve TÖ'nin kesinlik performansını %0,95 oranlarında artırmıştır. LASSO-BO yöntemi KA'nın kesinlik performansını %3,29, Relief-BO yöntemi NB'nın kesinlik performansını %4,72, LASSO-BO DVM'nın kesinlik performansını %3,77, AYİS-BO yöntemi K-NN'nın kesinlik performansını %2,36 ve LASSO-BO yöntemi TÖ'nın kesinlik performansını %4,71 oranında artırmıştır.



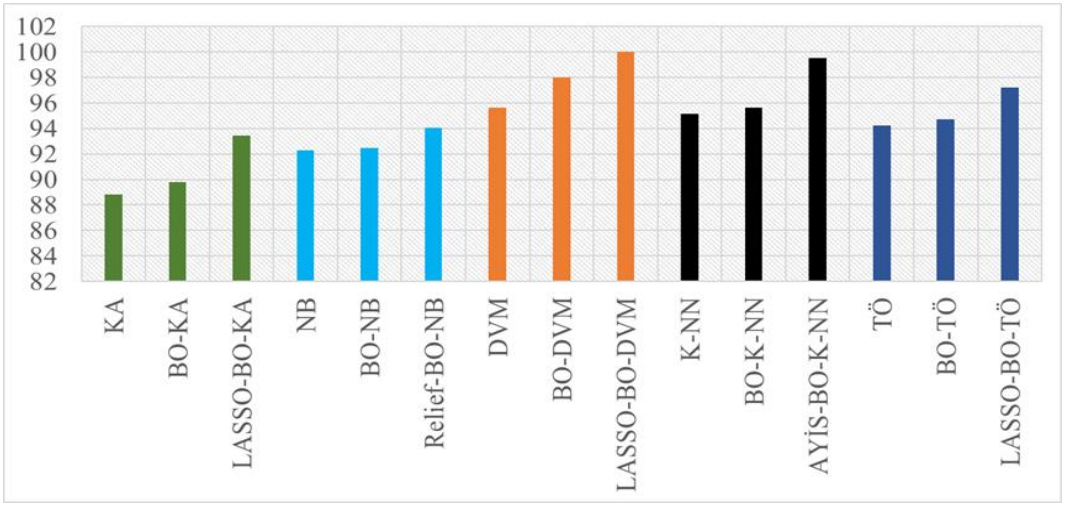
Şekil 6.5: WBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına kesinlik açısından etkisi.

Şekil 6.6'da MBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi kesinlik oranları açısından gösterilmiştir. BO sırasıyla KA'nın kesinlik performansını %0,86, NB'nin kesinlik performansını %1,73, DVM'nin kesinlik performansını %2,59, K-NN'nin kesinlik performansını %3,45 ve TÖ'nin kesinlik performansını %1,72 oranlarında artırmıştır. Relief-BO yöntemi KA'nın kesinlik performansını %3,45, AYİS-BO yöntemi NB'nın kesinlik performansını %3,45, LASSO-BO yöntemi DVM'nın kesinlik performansını %4,31, AYİS-BO yöntemi K-NN'nın kesinlik performansını %6,03 ve LASSO-BO yöntemi TÖ'nın kesinlik performansını %5,18 oranında artırmıştır.



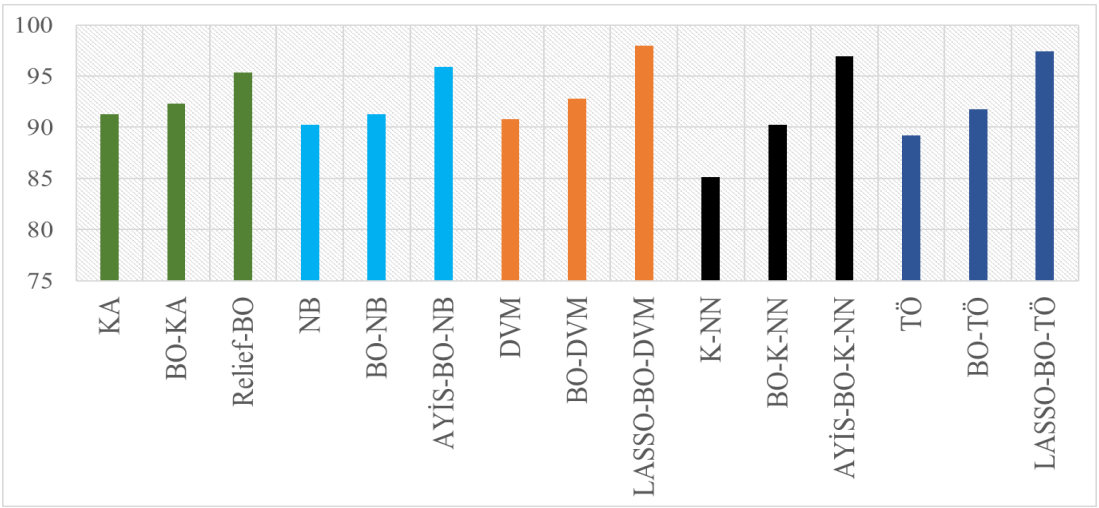
Şekil 6.6: MBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına kesinlik açısından etkisi.

Şekil 6.7'de WBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi duyarlılık oranları açısından gösterilmiştir. BO sırasıyla KA'nın duyarlılık performansını %0,93, NB'nin duyarlılık performansını %0,18, DVM'nin duyarlılık performansını %2,36, K-NN'nin duyarlılık performansını %0,49 ve TÖ'nin duyarlılık performansını %0,51 oranlarında artırmıştır. LASSO-BO yöntemi KA'nın duyarlılık performansını %3,69, Relief-BO yöntemi NB'nın duyarlılık performansını %1,61, LASSO-BO DVM'nın duyarlılık performansını %1,98, AYİS-BO yöntemi K-NN'nın duyarlılık performansını %3,87 ve LASSO-BO yöntemi TÖ'nın duyarlılık performansını %2,45 oranında artırmıştır.



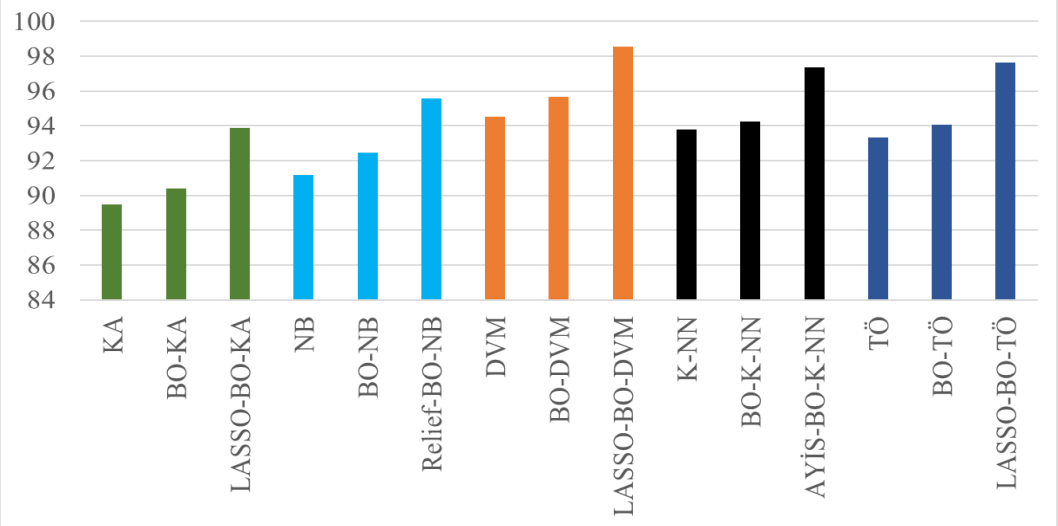
Şekil 6.7: WBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına duyarlılık açısından etkisi

Şekil 6.8'de MBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi duyarlılık oranları açısından gösterilmiştir. BO sırasıyla KA'nın duyarlılık performansını %0,87, NB'nin duyarlılık performansını %0,12, DVM'nin duyarlılık performansını %0,98, K-NN'nin duyarlılık performansını %4,9 ve TÖ'nin duyarlılık performansını %2,5 oranlarında artırmıştır. Relief-BO yöntemi KA'nın duyarlılık performansını %1,82, AYİS-BO yöntemi NB'nın duyarlılık performansını %4,33, LASSO-BO DVM'nın duyarlılık performansını %4,31, AYİS-BO yöntemi K-NN'nın duyarlılık performansını %5,16 ve LASSO-BO yöntemi TÖ'nın duyarlılık performansını %5,14 oranında artırmıştır.



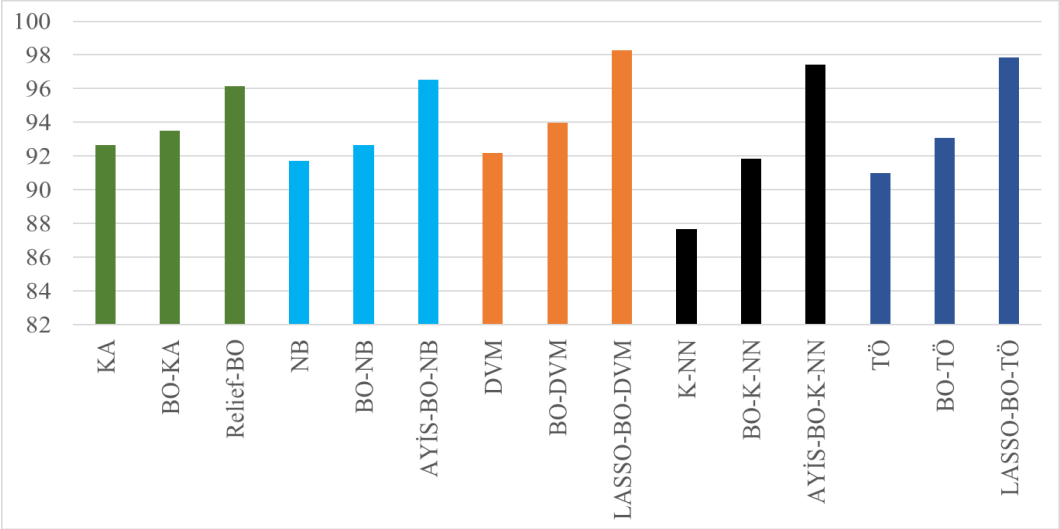
Şekil 6.8: MBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına duyarlılık açısından etkisi

Şekil 6.9'da WBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi F1-skoru oranları açısından gösterilmiştir. BO sırasıyla KA'nın F1-skoru performansını %0,94, NB'nin F1-skoru performansını %1,28, DVM'nin F1-skoru performansını %1,14, K-NN'nin F1-skoru performansını %0,48 ve TÖ'nin F1-skoru performansını %0,73 oranlarında artırmıştır. LASSO-BO yöntemi KA'nın F1-skoru performansını %3,5, Relief-BO yöntemi NB'nın duyarlılık performansını %3,14, LASSO-BO DVM'nın F1-skoru performansını %2,91, AYİS-BO yöntemi K-NN'nın F1-skoru performansını %3,09 ve LASSO-BO yöntemi TÖ'nın F1-skoru performansını %3,59 oranında artırmıştır.



Şekil 6.9: WBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına F1-skoru açısından etkisi

Şekil 6.10'da MBCD veri seti için öznitelik seçim yöntemlerinin ve BO tekniğinin sınıflandırma algoritmalarına etkisi F1-skoru oranları açısından gösterilmiştir. BO sırasıyla KA'nın F1-skor performansını %0,87, NB'nin F1-skor performansını %0,94, DVM'nin F1-skor performansını %1,79, K-NN'nin F1-skor performansını %4,19 ve TÖ'nin F1-skor performansını %2,11 oranlarında artırmıştır. Relief-BO yöntemi KA'nın F1-skoru performansını %2,63, AYİS-BO yöntemi NB'nın F1-skor performansını %3,88, LASSO-BO DVM'nın F1-skor performansını %4,31, AYİS-BO yöntemi K-NN'nın F1-skor performansını %5,59 ve LASSO-BO yöntemi TÖ'nın F1-skor performansını %4,75 oranında artırmıştır.



Şekil 6.10: MBCD veri seti için hibrit yöntemlerin sınıflandırma algoritmalarına F1-skoru açısından etkisi

7. TARTIŞMA

Bu çalışmada meme kanserinin etkin bir şekilde sınıflandırılması amacıyla hibrit bir sınıflandırma sistemi önerilmiştir. Önerilen hibrit sınıflandırma sistemi geliştirilmiş makine öğrenme algoritmalarını kullanmaktadır. Öznitelik yöntemleri ve BO yöntemi birleştirilerek geliştirilmiş makine öğrenme algoritmaları oluşturulmuştur. Öznitelik yöntemleri olarak sırasıyla Relief, LASSO ve AİYS kullanılırken makine öğrenme algoritmaları olarak ise KA, NB, DVM, K-NN ve TÖ yöntemleri tercih edilmiştir. Makine öğrenme algoritmalarının hiperparametrelerini belirlemek için BO yöntemi kullanılmıştır. Çalışmanın ana amacı; öznitelik seçimi ve hiperparametre optimizasyon yöntemlerinin makine öğrenmesi algoritmalarının performansları üzerindeki etkisini araştırarak meme kanseri teşhisi için en yüksek performansa sahip hibrit modelin belirlenmesidir. Çalışma kapsamında önerilen hibrit modeller iki farklı meme kanseri setinde çeşitli deneyler yapılarak test edilmiştir. Meme kanserinin teşhisi üç farklı sınıflandırma süreci uygulanmıştır. İlk aşamada, hiçbir şekilde optimizasyon ve öznitelik yöntemi kullanmadan bütün öznitelikler sınıflandırıcılara giriş olarak verilmiş ve makine öğrenme yöntemlerinin performansları performans kriterleri ile değerlendirilmiştir. İkinci aşamada, BO yöntemi kullanarak yine bütün özellikler sınıflandırıcılara giriş olarak verilmiş ve algoritmaların performansları tekrardan değerlendirilmiştir. En son aşamada, sırasıyla Relief, LASSO ve AİYS yöntemleri kullanılarak veri setleri için ayırt edici öznitelikler belirlenmiş ve seçilen öznitelikler algoritmalara giriş olarak verilmiş ve makine öğrenme yöntemlerinin sınıflandırma performansları değerlendirilmiştir. Son aşamada da ikinci aşamada olduğu gibi makine öğrenme algoritmalarının hiperparametreleri BO yöntemi kullanarak ayarlanmıştır. WBCD veri setinde tanımlanmış olan 30 adet öznitelikten Relief, LASSO ve AİYS yöntemleri uygulandıktan sonra sırasıyla 16, 8 ve 3 öznitelik seçilmiştir. MBCD veri seti için ise tanımlanmış olan 54 adet öznitelikten Relief, LASSO ve AİYS yöntemleri uygulandıktan sonra da sırasıyla 12, 10 ve 4 öznitelik seçilmiştir. KA algoritmasının WBCD veri seti için LASSO-BO ve MBCD veri seti için Relief-BO kombinasyonları, NB WBCD veri seti için Relief-BO ve MBCD veri seti için AYİS-BO kombinasyonları, DVM algoritmasının WBCD ve MBCD veri setleri için LASSO-BO kombinasyonları, K-NN algoritmasının WBCD ve MBCD veri setleri için AYİS-BO

kombinasyonları, TÖ algoritmasının WBCD ve MBCD veri setleri için LASSO-BO kombinasyonları en yüksek sınıflandırma oranlarına ulaşmıştır. Şekil 6.3-6.10 arasında BO ve her sınıflandırıcı için en yüksek başarı oranına sahip öznitelik-BO hibrit yöntemlerinin makine öğrenme algoritmalarının sınıflandırma performanslarına etkisi incelenmiştir. Yapılan karşılaştırmalar neticesinde BO optimizasyonu makine öğrenme algoritmalarının performanslarını doğruluk, kesinlik, duyarlılık ve F1-skor değerlendirme ölçütleri açısından artırdığı görülmüştür. Öznitelik seçim yöntemleri ve BO yöntemi birlikte kullanılarak oluşturulan hibrit modellerin ise makine öğrenme algoritmalarının sınıflandırma performanslarını önemli oranda artırdığı görülmüştür. Makine öğrenme algoritmalarının performanslarındaki bahsedilen bu artışlar doğruluk, kesinlik, duyarlılık ve F1-skor performans değerlendirme ölçütleri cinsinden Şekil 6.3-6.10 arasında gösterilmiştir. Çizelge 7.1-7.2 ve Şekil 7.1-7.2'de WBCD ve MBCD veri setleri için en yüksek sınıflandırma oranlarına sahip hibrit modelleri göstermektedir. Her iki veri seti içinde LASSO-BO-SVM hibrit modeli en yüksek doğruluk, kesinlik, duyarlılık ve F1-skor oranına sahiptir (WBCD %98,95, %97,17, %100, %98,57; MBCD %97,95, %98,28, %98,28, %98,28).

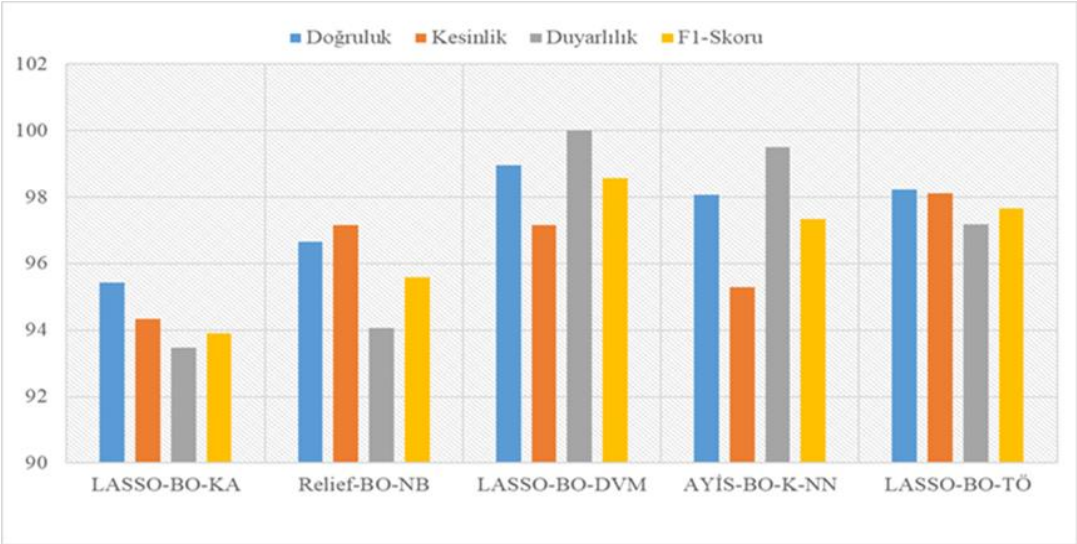
Çizelge 7.1: WBCD veri seti için en başarılı hibrit yöntemler

Yöntem	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
LASSO-BO-KA	95,43	94,33	93,46	93,9
Relief-BO-NB	96,66	97,17	94,06	95,59
LASSO-BO-DVM	98,95	97,17	100	98,57
AYİS-BO-K-NN	98,06	95,29	99,51	97,35
LASSO-BO-TÖ	98,24	98,11	97,19	97,65

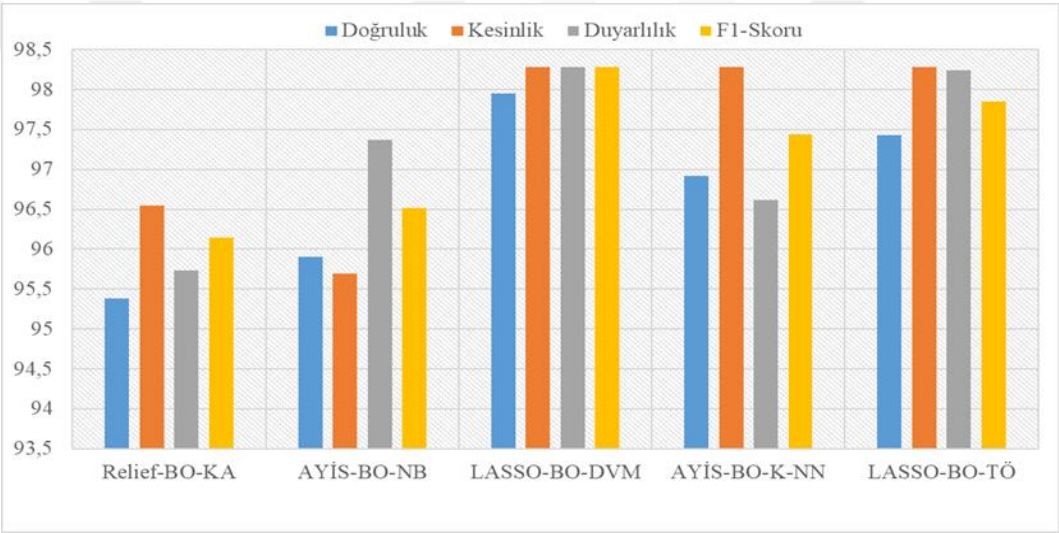
Çizelge 7.2: MBCD veri seti için en başarılı hibrit yöntemler

Yöntem	**Doğruluk**	**Kesinlik**	**Duyarlılık**	**F1-Skoru**
Relief-BO-KA	95,38	96,55	95,73	96,14
AYİS-BO-NB	95,9	95,69	97,37	96,52
LASSO-BO-DVM	97,95	98,28	98,28	98,28
AYİS-K-NN	96,92	98,28	96,62	97,44
LASSO-BO-TÖ	97,43	98,28	98,24	97,85

Literatürde yapılan benzer çalışmalar ile LASSO-BO-SVM hibrit yönteminin karşılaştırılması Çizelge 7.3'de sunulmuştur.



Şekil 7.1: WBCD veri seti için en başarılı hibrit yöntemler



Şekil 7.2: MBCD veri seti için en başarılı hibrit yöntemler.

Çizelge 7.3: LASSO-BO-SVM yönteminin literatürdeki benzer çalışmalar ile karşılaştırılması

Referans	**Metot**	**Veri Seti**	**Doğruluk**
Mate ve diğ, [50]	BO-Aşırı Ağaç Sınıflandırıcı	WBCD	%96,2
Kumar ve diğ, [48]	BO-RO	WBCD	%97,9
Asri ve diğ, [34]	DVM	WBCD	%97,13
Bensaoucha [49]	BO-DVM	WBCD	%96,52
Khandezemin ve diğ, [40]	LR-VİGY	WBCD	%97,9
Önerilen Yöntem	**LASSO-BO-SVM**	**WBCD** **MBCD**	**%98,95** **%97,95**

Mate ve diğ, [50] WBCD veri seti üzerinde BO-aşırı ağaç sınıflandırıcı yöntemini kullanmış ve %96,2 doğruluk oranı elde etmişlerdir. Kumar ve diğ. [48] WBCD veri setinde BO-RO yöntemini uygulamış ve %97,9 doğruluk oranına ulaşmıştır. Asri ve diğ. WBCD veri setinde DVM yöntemini uygulamış ve %97,13 doğruluk oranı elde etmiştir. Bensoucha [49] WBCD veri setinde meme kanserinin sınıflandırılması amacıyla BO-DVM yöntemi uygulamış ve %96,52 doğruluk oranı elde etmiştir. Khandezemin ve diğ. [40] LR-VİGY yöntemini WBCD veri setinde test etmiş ve %97,9 doğruluk oranı elde etmiştir. Literatürdeki makine öğrenme yöntemleri kullanarak meme kanseri tespiti için yapılan çalışmalar incelendiğinde, bu çalışma kapsamında önerilen LASSO-BO-DVM hibrit yönteminin yüksek bir sınıflandırma oranına sahip olduğu görülmüştür.

8. SONUÇ VE ÖNERİLER

Bu bölümde çalışma kapsamında denenmiş olan bir çok deney sonucunda varılan değerlendirmeler sunulmuştur. Bu değerlendirmeler şöyledir:

- Makine öğrenme algoritmaları pek çok hiperparametre içermektedir. Tüm hiperparametre kombinasyonlarını tek tek denemek ve en uygun kombinasyonun seçilmesi bulmak tasarımcılar için oldukça zaman alıcı ve zor bir işlemdir. Bu nedenle, hiperparametre optimizasyon yöntemlerinin kullanılması en uygun hiperparametre kombinasyonun bulunması için son derece yararlı bir işlemdir. Çalışma kapsamında makine öğrenme algoritmalarının hiperparametrelerinin ayarlanması için BO yöntemi kullanılmış ve bu yöntem makine öğrenme algoritmalarının performanslarını artırmıştır. Bu yüzden BO yöntemi makine öğrenme algoritmalarının en uygun hiperparametrelerinin seçilmesi açısından etkili bir teknik olarak değerlendirilebilir.
- Veri setlerindeki belirlenen özniteliklerinin sayısı makine öğrenme algoritmalarının sınıflandırma oranlarını önemli oranda etkilemektedir. Hem veri setlerindeki öznitelik boyutunu azaltmak hem de makine öğrenme algoritmalarının sınıflandırma performanslarını artırmak için farklı öznitelik seçim yöntemlerinin kullanılması faydalıdır. Çalışma kapsamında kullanılan öznitelik seçim yöntemleri hem veri setlerindeki öznitelik sayısını azaltmaya hemde sınıflandırma oranlarına olumlu etkisi bulunmaktadır.
- Öznitelik seçim yöntemleri ve hiperparametre optimizasyon yöntemin uygulanması makine öğrenme algoritmalarının performanslarını olumlu etkilemektedir. Bu çalışmada öznitelik seçim yöntemleri ve hiperparametre optimizasyon yöntemi birleştirilerek farklı hibrit modeler oluşturulmuştur. Oluşturulan bu hibrit modeller çalışmada kullanılan herbir makine öğrenme algoritmalarının performanslarını artırmıştır.
- Her iki meme kanseri veri setinde de yapılan bir çok uygulama sonucunda LASSO öznitelik seçim yöntemi ve DVM algoritması en yüksek sonuca ulaşmıştır.

- Sonuç olarak, çalışma kapsamında önerilen hibrit modellerden LASSO-BO-DVM yöntemi literatürde meme kanserinin teşhisi için yapılan çalışmalar ile karşılaştırıldığında iyi bir sınıflandırma oranına sahiptir. Bu nedenle önerilen yöntem başka verilerin sınıflandırılması içinde kullanılabilir.
- Bu çalışma kapsamında meme kanserinin sınıflandırılması için önerilen hibrit yöntem umut verici sonuçlara ulaşsmış olsa da, bu çalışmanın bazı sınırlamaları bulunmaktadır. Çalışmada kullanılan mamografi görüntülerinin tek bir merkezden alınması ve örnek sayısının sınırlı olması bir dezavantaj olarak kabul edilebilir. Mamografi görüntülerinden ROI'lerin çıkarılması için literatürde bazı çalışmalarda otomatik segmentasyon yöntemleri de tercih edilmiştir [110-111]. Bununla birlikte yarı-otomatik segmentasyon yöntemi kullanılarak %90 doğruluk oranlarına ulaşan çalışmalar da literatürde bulunmaktadır [59, 61, 112]. Yüksek sayıda örnek sayısına ulaşıldığında, derin öğrenme algoritmaları sınıflandırma işlemleri için yüksek performans gösterebilmektedir. Bu nedenle makine öğrenme algoritmalarına alternatif olarak derin öğrenme yöntemleri de gelecek çalışmalarda kullanılabilir. Örnek sayısının az olduğu çalışmalarda, bu tezde önerilen hibrit yöntemlerle geliştirilmiş makine öğrenme algoritmalarının kullanılabileceğini düşünmekteyim.

KAYNAKLAR

[1] **Sung H., Ferlay J., Siegel R. L., Laversanne M., Soerjomataram I., Jemal A., Bray F.,** (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA:Cancer J. Clin.*, 71(3), 209-249.

[2] **Houssein E. H., Emam M. M., Ali A. A., Suganthan P. N**., (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Syst. Appl.*, 167, 114161.

[3] **Wang L.,** (2017). Early diagnosis of breast cancer, *Sensors*, 17(7), 1572.

[4] **Zebari, D. A., Ibrahim, D. A., Zeebaree, D. Q., Haron, H., Salih, M. S., Damaševičius, R., Mohammed, M. A**. (2021). Systematic review of computing approaches for breast cancer detection- based computer aided diagnosis using mammogram images. *Appl.Artif.Intell.*, 35(15), 2157-2203.

[5] **Fatima N., Liu L., Hong S., Ahmed H.,** (2020). Prediction of breast cancer, comparative review of machine learning techniques and their analysis, *IEEE Access*, 8, 150360-150376.

[6] **Pulat, M., Kocakoç, İ. D.** (2021). Türkiye'de Makine Öğrenmesi ve Karar Ağaçları Alanında Yayınlanmış Tezlerin Bibliyometrik Analizi. *Yönetim ve Ekonomi Dergisi*, 28(2), 287-308.

[7] **Miao J., Niu. L**., (2016). A survey on feature selection, *Procedia Comput. Sci.*, 91, 919-926.

[8] **Claesen M., De Moor B.,** (2015). Hyperparameter search in machine learning, *arXiv preprint* arXiv:1502.02127.

[9] **Radzi S. F. M., Karim M. K. A., Saripan M. I., Rahman M. A. A., Isa I. N. C., Ibahim. M. J.,** (2021). Hyperparameter tuning and pipeline optimization via grid search method and tree-based autoML in breast cancer prediction, *J Pers Med.*, 11(10), 978.

[10] **Toz G., Erdoğmuş P.,** (2020). *Meme kanserinin teşhisinde kullanılan görüntü işleme teknikleriyle ilgili bir derleme çalışması*, Dr. Öğr. Üy. E.Avuçlu, Dr.Öğr. Üy. D. Ekmekçi (Ed.), Geleceğin Dünyasında Bilimsel ve Mesleki Çalışmalar, (Sf.70), Bursa, Ekin Basın Yayım Dağıtım, (Mart, 2020).

[11] **Biçer R. M. B.,** (2014). Meme kanseri görüntülemesinde mikrodalganın yeri, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, 30(4). 257-263.

[12] **Bulut İ., Oğuzöncül A. F., Kara K. T.,** (2021). Kanser erken teşhis tarama ve eğitim merkezi'ne ait meme ve serviks kanserlerini tarama programı sonuçları, *ESTÜDAM Halk Sağlığı Dergisi*, *6*(2), 182-190.

[13] **Meisner A. L., Fekrazad, M. H., Royce M. E.,** (2008). Breast disease: benign and malignant, *Med Clin North Am*., *92*(5), 1115-1141.

[14] **Sharma G. N., Dave R. Sanadya J., Sharma P., Sharma K**., (2010). Various types and management of breast cancer: an overview, *J. Adv Pharm Technol Res*., 1(2), 109.

[15] **Karabulut Gül. Ş., Oruç A. F., Mayadağlı A.,** (2013). Duktal karsinoma In

Situ, *Journal of Kartal Training & Research Hospital/Kartal Egitim ve Arastirma Hastanesi Tip Dergisi*, *24*(2).

[16] **Koçak S., Çelik L., Özbaş S., Sak S. D., Tükün A., Yalçın B.,** (2011). Meme kanserinde risk faktörleri, riskin değerlendirilmesi ve prevansiyon: İstanbul 2010 konsensus raporu, *Meme Sagligi Dergisi/Journal of Breast Health*, 7(2).

[17] **Açıkgöz A., Yıldız. E. A.,** (2017). Meme kanseri etiyolojisi ve risk faktörleri, *Ergoterapi ve Rehabilitasyon Dergisi*, 5(1), 45-56.

[18] **Parlar S., Kaydul N., Ovayolu N.**, (2005). Meme kanseri ve kendi kendine meme muayenesinin önemi, *Anadolu Hemşirelik ve Sağlık Bilimleri Dergisi*, 8(1), 72-83.

[19] **Sohbet R., Karasu F.,** (2017). Kadınların meme kanserine yönelik bilgi. davranış ve uygulamalarının incelenmesi, *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*, 6(4), 113-121.

[20] **Joshi P., Singh N., Raj G., Singh R., Malhotra, K. P., Awasthi N. P.,** (2022). Performance evaluation of digital mammography. digital breast tomosynthesis and ultrasound in the detection of breast cancer using pathology as gold standard: an institutional experience, *Egypt J Radiol Nucl Med.,* 53(1),1-11.

[21] **Shen R., Yan K., Tian K., Jiang C., Zhou K.,** (2019). Breast mass detection from the digitized X-ray mammograms based on the combination of deep active learning and self-paced learning, *Future Gener. Comput. Syst* , 101, 668-679.

[22] **Köşüş N., Köşüş A., Duran M., Simavlı S., Turhan N.,** (2010). Comparison of standard mammography with digital mammography and digital infrared thermal imaging for breast cancer screening, *J Turk Ger Gynecol Assoc.,* 11(3), 152.

[23] **Nam K. J., Han B. K., Ko E. S., Choi J. S., Ko E. Y., Jeong D. W., Choo K. S.,** (2015). Comparison of full-field digital mammography and digital breast tomosynthesis in ultrasonography-detected breast cancers, *The Breast*, 24(5), 649-655.

[24] **Türe H. (2021)**. Mamografi görüntülerindeki kitlelerin ve pektoral kas bölgesinin süperpozisyon etkisini dikkate alarak belirlenmesi, (doktora tezi), Adres: http://acikerisim.ktu.edu.tr/jspui/handle/123456789/4012

[25] **Turgut A. T., Hasırcıoğlu F., Koşar U.**, (2000). Meme hastalıklarının tanısında mamografi, STED, 9, 12.

[26] **Pektaş F., (2016).** Mamografide tek projeksiyonda görülen fokal asimetrik opasitelerin meme MRG ile değerlendirilmesi, (uzmanlık tezi), https://acikerisim.uludag.edu.tr/bitstream/11452/10229/1/429011.pdf

[27] **Spak D. A., Plaxco J. S., Santiago L., Dryden M. J Dogan B. E.,** (2017). BI-RADS® fifth edition: A summary of changes. *Diagn. Interv. Imaging*, 98(3), 179-190.

[28] **Kiarashi N., Samei E.,** (2013). Digital breast tomosynthesis: a concise overview, *Imaging in Medicine*, 5(5), 467.

[29] **Rikabi A., Hussain S.,** (2013). Diagnostic usefulness of tru-cut biopsy in the diagnosis of breast lesions, *Oman Med J.*, 28(2), 125.

[30] **Metlek S., Kayaalp K.,** *Makine öğrenmesinde. teoriden örnek matlab uygulamalarına kadar destek vektör makineleri*, İksad Yayınevi, Türkiye (2020).

[31] **Öztürk K., Şahin M. E.,** (2018). Yapay sinir ağları ve yapay zekâ'ya genel bir bakış, *Takvim-i Vekayi*, 6(2), 25-36.

[32] **Metlek S. Çetiner H,** *MATLAB ortamında derin öğrenme uygulamaları*. İksad Yayınevi, Türkiye, (2021).

[33] **Doğan F., Türkoğlu. İ.**, (2019). Derin öğrenme modelleri ve uygulama alanlarına ilişkin bir derleme, *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 10(2), 409-445.

[34] **Asri H., Mousannif H., Al Moatassime H., Noel T.,** (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis, *Procedia Comput. Sci.*, 83:1064-1069.

[35] **Naji M. A., El Filali S., Aarika K., Benlahmar E. H., Abdelouhahid R. A., Debauche O.,** (2021). Machine learning algorithms for breast cancer prediction and diagnosis, *Procedia Comput. Sci.*, 191: 487-492.

[36] **Amrane M., Oukid S., Gagaoua I., Ensari T.,** (2018). Breast cancer classification using machine learning, *In 2018 Electric Electronics. Computer Science, Biomedical Engineerings' Meeting (EBBT), (*pp. 1-4), IEEE.

[37] **Ak M. F.,** (2020) A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications, *In: Healthcare. MDPI,* 8(2):111.

[38] **Khan M. M., Islam S., Sarkar S., Ayaz F. I., Ananda M. K., Tazin T., Albraikan A. A, Almalki F. A.,** (2022). Machine learning based comparative analysis for breast cancer prediction, *J. Healthc. Eng.*, 2022:4365855.

[39] **Omondiagbe D. A., Veeramani. S., Sidhu. A. S.,** (2019. April). Machine learning classification techniques for breast cancer diagnosis, *In IOP Conference Series: Materials Science and Engineering*, 495(1):012033.

[40] **Khandezamin Z. Naderan M., Rashti M. J.,** (2020). Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier, *J. Biomed. Inform.*, 111, 103591.

[41] **Haq A. U., Li. J. P., Saboor A., Khan J., Wali. S., Ahmad S., Ali A., Khan G.H., Zhou W.,** (2021). Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques, *IEEE Access*, 9, 22090-22105.

[42] **Bacha S., Taouali O.,** (2022). A novel machine learning approach for breast cancer diagnosis, *Measurement*, 187, 110233.

[43] **Vadivel A., Surendiran B.,** (2013). A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories, *Comput. Biol. Med.*, 43(4), 259-267.

[44] **Jadoon M. M., Zhang Q., Haq I. U., Butt S., Jadoon A**., (2017). Three-class mammogram classification based on descriptive CNN features, *BioMed Res. Int.*, 3640901:11.

[45] **Punitha S., Amuthan A., Joseph K.S.,** (2018). Benign and malignant breast cancer segmentation using optimized growing technique, *Future Computing and Informatics Journal*, 3(2), 348-358

[46] **Bajcsi A., Andreica A., Chira C.,** (2021). Towards feature selection for digital mammogram classification, *Procedia Comput. Sci.*, 192: pp:632-641.

[47] **Wang H., Yoon S.W.,** (2015). Breast cancer prediction using data mining method, *In Proceedings of the IIE Annual Conference, Institute of Industrial and Systems Engineers* (IISE), Nashville, TN.,USA, 31

May–3 June, p. 818.

[48] **Kumar P., Nair G. G.,** (2021). An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization, *Biomed. Signal Process. Control., 68:102682*

[49] **Bensaoucha S.,** (2021). Breast cancer diagnosis using optimized machine learning algorithms, *In 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*, Tebessa, Algeria, pp. 1-6.

[50] **Mate Y., Somai. N.,** (2021). Hybrid feature selection and Bayesian optimization with machine learning for breast cancer prediction, In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp:612-619.

[51] **Dhanya R., Paul I. R., Akula S. S., Sivakumar M., Nair. J. J.,** (2019). A comparative study for breast cancer prediction using machine learning and feature selection, *In 2019 International Conference on Intelligent Computing and Control Systems* (ICCS), (pp. 1049-1055), IEEE.

[52] **Kumari L. K., Jagadesh B. N. ,** (2022). A robust feature extraction technique for breast cancer detection using digital mammograms based on advanced GLCM approach, *EAI Endorsed Trans. on Pervasive Health Technol.*, 8(30), e3-e3.

[53] **Vijayarajeswari R., Parthasarathy P., Vivekanandan S., Basha A. A.,** (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform, *Measurement*, 146, 800-805.

[54] **Ancy C. A., Nair L. S.,** (2017. April). An efficient CAD for detection of tumour in mammograms using SVM, In *2017 International Conference on Communication and Signal Processing (ICCSP)*, (pp. 1431-1435), IEEE.

[55] **Alshammari M. M., Almuhanna A., Alhiyafi J., (**2021). Mammography image-based diagnosis of breast cancer using machine learning: A pilot study, *Sensors*, 22(1), 203.

[56] **Farid A. A., Selim G., Khater H.,** (2020). A composite hybrid feature selection learning-based optimization of genetic algorithm for breast cancer detection, *Preprints*, 25, 1-21.

[57] **Ergin S., Esener İ. I., Yüksel T.,** (2016). A genuine GLCM-based feature extraction for breast tissue classification on mammograms, *International Journal of Intelligent Systems and Applications in Engineering*, 4(Special Issue-1), 124-129.

[58] **Farhan A. H., Kamil M. Y.,** (2020. November). Texture analysis of breast cancer via LBP, HOG and GLCM techniques, IOP Conf. Ser.: Mater. Sci. Eng., 928(7):072098.

[59] **Wang G., Shi. D., Guo Q., Zhang H., Wang S., Ren K.,** (2022). Radiomics based on digital mammography helps to identify mammographic masses suspicious for cancer, *Front Oncol.*, 12:8438436.

[60] **Li M., Zhu L., Zhou G., He J., Jiang Y., Chen Y.,** (2021). Predicting the pathological status of mammographic microcalcifications through a radiomics approach, *Intelligent Medicine*, 1(03). 95-103.

[61] **Stelzer, P. D., Steding, O., Raudner, M. W., Euller, G., Clauser, P., Baltzer, P. A. T**., (2020). Combined texture analysis and machine

learning in suspicious calcifications detected by mammography: Potential to avoid unnecessary stereotactical biopsies, *Eur J Radiol.*, 132:109309.

[62] **Nugroho, H. A., Faisal, N., Soesanti, I., Choridah, L.,** (2014, October). Identification of malignant masses on digital mammogram images based on texture feature and correlation-based feature selection, In *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, Indonesia, *pp.*1-6.

[63] **Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., Basha, A. A.** (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform, *Measurement*, 146, 800-805.

[64] **Gherghout, Y., Tlili, Y., Souici, L.,** (2021). Classification of breast mass in mammography using anisotropic diffusion filter by selecting and aggregating morphological and textural features, *Evolving Systems*, 12(2), 273-302.

[65] **Sapate, S. G., Mahajan, A., Talbar, S. N., Sable, N., Desai, S., Thakur, M.,** (2018). Radiomics based detection and characterization of suspicious lesions on full field digital mammograms, *Comput. Methods Programs Biomed.*, 163, 1-20.

[66] **Loizidou, K., Skouroumouni, G., Nikolaou, C., Pitris, C**. (2020). An automated breast micro-calcification detection and classification technique using temporal subtraction of mammograms, *IEEE Access*, 8, 52785-52795.

[67] **Wolberg W.H., Street W.N., Mangasarian O.L., (**1995). Wisconsin Breast Cancer Database, University of Wisconsin Hospitals, Madison, Wisconsin, USA.

[68] **M. Lichman,** (2013) UCI Machine Learning Repository, University of California. School of Information and Computer Science. Irvine, California. USA**.**

[69] **Wolberg W.H., Street W.N., Heisey D.M., Mangasarian O.L.,** (1995). Computerized breast cancer diagnosis and prognosis from fine needle aspirates. *Arch. Surg.* vol. 130(5): 511- 516.

[70] **Demir Ö., Çamurcu A.Y.,** (2015) Computer aided detection of lung nodules using outer surface features, *Biomed. Mater. Eng.* 26.s1: S1213-S1222.

[71] **Doğan B., Demir Ö.,Kazdal Çalık S.,** (2016) Computer-aided detection of brain tumors using morphological reconstruction, *Uludağ University Journal of The Faculty of Engineering*, 21(2): 257-268.

[72] **Parker J.R., (2010).** Algorithms for Image Processing and Computer vision, *Wiley Publishing, New York.*

[73] **Chaieb R., Bacha A., Kalti K., Lamine F. B.,** (2014. August). Image features extraction for masses classification in mammograms, *In 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 203-208, Tunis, Tunisia.

[74] **Surendiran B., Vadivel. A.,** (2012). Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer, *Int. J. Med. Eng. Inform.*, 4(1), 36-54.

[75] **Çetinel G., Mutlu F., Gül S.,** (2019. April). Breast lesion classification based on shape features, In *2019 27th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

[76] **Chaieb R., Kalti K., (**2019). Feature subset selection for classification of malignant and benign breast masses in digital mammography, *Pattern Analysis and Applications*, 22(3), 803-829.

[77] **Haralick, R. M., Shanmugam K., Dinstein I. H.**, (1973). Textural features for image classification. *IEEE Trans. Syst. Man. Cybern. Syst*, (6), 610-621.

[78] **Galloway M. M.,** (1975). Texture analysis using gray level run lengths, *Computer graphics and image processing*, 4(2), 172-179.

[79] **Kotsiantis S. B., Kanellopoulos D., Pintelas P. E.,** (2006). Data preprocessing for supervised learning, *International Journal of Computer Science*, 1(2), 111-117.

[80] **Dash M., Liu H.,** (1997). Feature selection for classification, *Intelligent data analysis*, 1(1-4), 131-156.

[81] **Remeseiro B., Bolon-Canedo V.,** (2019). A review of feature selection methods in medical applications, *Comput. Biol. Med.*, 112, 103375.

[82] **Cherrington M., Thabtah F., Lu J., Xu Q. I.,** (2019). Feature selection: filter methods performance challenges, In *2019 International Conference on Computer and Information Sciences (ICCIS)*, Sakaka, Saudi Arabia, p. 1-4.

[83] **Budak H.,** (2018). Özellik seçim yöntemleri ve yeni bir yaklaşım, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 21-31.

[84] **Urbanowicz R. J., Meeker M., La Cava W., Olson R. S., Moore J. H.,** (2018). Relief-based feature selection: Introduction and review, *J. Biomed. Inform.*, 85, 189-203.

[85] **Yılmaz A., Sümer E.,** (2021). Relief özellik seçim yöntem tabanlı önerilen hibrit model ile kalp hastalığı teşhisi, *Avrupa Bilim ve Teknoloji Dergisi*, (31), 609-615.

[86] **Trivedi. S. K.,** (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, 63, 101413.

[87] **Guyon I., Elisseeff A.,** (2003). An introduction to variable and feature selection, *J. Mach. Learn. Res.*, *3*(Mar), 1157-1182.

[88] **Whitney A. W.,** (1971). A direct method of nonparametric measurement selection. *IEEE Trans. Comput.*, 100(9). 1100-1103.

[89] **Tibshirani R., (1996)** Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

[90] **Fonti V., Belitser E.,** (2017). Feature selection using lasso, *VU Amsterdam research paper in business analytics*, 30, 1-25.

[91] **Refaeilzadeh P., Tang L., Liu H,** (2009). Cross-validation, *Encyclopedia of database systems*, Springer, Boston, MA, 5, 532-538.

[92] **Fawcett T.,** (2006). An introduction to ROC analysis, *Pattern recognition letters*, 27(8), 861-874.

[93] **Rokach L., Maimon O.**, (2005). Decision trees, In *Data mining and knowledge discovery handbook* , Springer, Boston, MA., 165-192.

[94] **Çalış A., Kayapınar S., Çetinyokuş T.,** (2014). Veri madenciliğinde karar ağacı algoritmaları ile bilgisayar ve internet güvenliği üzerine bir uygulama, *Endüstri Mühendisliği*, 25(3), 2-19.

[95] **Webb G. I., Keogh E., Miikkulainen R.,** (2010). Naive Bayes, *Encyclopedia of machine learning*, 15, 713-714.

[96] **Swinburne R.,** (2004). Bayes' Theorem, *Revue Philosophique de la France Et de* l. 194(2).

[97] **Suthaharan S.,** (2016). Support vector machine, In *Machine learning models and algorithms for big data classification,* (pp. 207-235), Springer, Boston, MA.

[98] **Yang. J., Awan A. J., Vall-Llosera G.,** (2019). Support vector machines on noisy intermediate scale quantum computers, *arXiv preprint arXiv:1909.11988*.

[99] **Dağdeviren B. M. E., Orman Z., (2013).** *El yazısı rakam tanıma için destek vektör makinelerinin ve yapay sinir ağlarının karşılaştırması,* (yüksek lisans tezi), http://nek.istanbul.edu.tr/ :4444/ekos/TEZ/50092.pdf

[100] **Polikar R.,** (2012). Ensemble learning, In *Ensemble machine learning*, Springer: Boston, MA, pp. 1-34.

[101] **Buhlmann P.,** (2012). Bagging, boosting and ensemble methods, In *Handbook of computational statistics*, Springer, Berlin, Heidelberg, 985-1022.

[102] **Plaia A., Buscemi S., Fürnkranz J., Mencia E. L.,** (2022). Comparing boosting and bagging for decision trees of rankings, *J Classif.*, 39(1), 78-99.

[103] **Tuv E.,** (2006). Ensemble learning, In: Guyon. I., Nikravesh. M., Gunn. S. Zadeh. L.A. (eds*) Feature Extraction, Studies in Fuzziness and Soft Computing*, vol 207, Springer, Berlin, Heidelberg.

[104] **Tanyıldızı E., Demirtaş. F.,** (2019. November). Hiper parametre optimizasyonu hyper parameter optimization, In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, Ankara, Turkey, pp. 1-5).

[105] **Blume, S., Benedens, T., Schramm, D.** (2021). Hyperparameter optimization techniques for designing software sensors based on artificial neural networks, *Sensors*, 21(24), 8435.

[106] **Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., Deng, S. H.** (2019). Hyperparameter optimization for machine learning models based on Bayesian Optimization, *J. Electron. Sci. Technol.*, 17(1), 26-40.

[107] **Kim H.C., Kang M.J**., (2020). Comparison of hyperparameter optimization methods for deep neural networks, *Journal of IKEEE*, 24(4), 969-974.

[108] **Susmaga R.,** (2004). Confusion matrix visualization, In *Intelligent information processing and web mining* (pp. 107-116). Springer, Berlin, Heidelberg.

[109] **MATLAB and Statistics Toolbox Release 2020a**, The MathWorks Inc., Natick, Massachusetts, United States.

[110] **Patil, R. S., Biradar, N., Pawar, R.,** (2022). A new automated segmentation and classification of mammogram images, *Multimed. Tools and Appl.*, 81(6), 7783-7816.

[111] **Praveen Kumar, C., Rajendra Prasad, K.,** (2021). Multi-ROI segmentation for effective texture features of mammogram images, *J. Discrete Math. Sci. Cryptogr.*, 24(8), 2461-2469.

[112] **Saleck M. M., El Moutaouakkil, A., Rmili, M.,** (2018). Semi-automatic segmentation of breast masses in mammogram images, In *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence (PRAI)* (pp. 59-62).

Url-1< https://www.celalsaglam.com/meme-kanseri/ alındığı tarih:23.08.2022.

Url-2< https://ivekakademi.org/blog/meme-kanseri-turlerinden-invaziv-

duktal-karsinom-idc-ve-invaziv-lobuler-karsinom-ilclarin-birbirinden-ayirt-edilmesinde-biyoinformatik-analizlerin-yeri-ve-onemi, alındığı tarih:23.08.2022.

Url-3< https://www.mathworks.com/help/stats/relieff.html/ alındığı tarih:21.11.2022.

Url-4< https://www.mathworks.com/help/stats/lasso.html/ alındığı tarih:21.11.2022.

Url-5< https://www.mathworks.com/help/stats/sequentialfs.html/ alındığı tarih:21.11.2022.

EK: ETİK KURUL ONAYI

T.C.
ANKARA VALİLİĞİ
İL SAĞLIK MÜDÜRLÜĞÜ
Sağlık Bakanlığı Ankara Eğitim ve Araştırma Hastanesi
Klinik Araştırmalar Etik Kurul Başkanlığı

Sayı : E.Kurul –E-20

319-no'lu çalışma

SBÜ Ankara Eğitim ve Araştırma Hastanesi Radyoloji Kliniği'nden "Mamografide Saptanan Kitle Lezyonlarının Bening-Maling Ayrımında Yapay Zeka Algoritamaları Kullanılmasının Tanıya Katkısı" konulu çalışma incelenmiş olup, Etik açıdan oy birliğiyle uygun görülmüştür.

12/07/2020