

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA İLE MEME KANSERİ TEŞHİSİ

YÜKSEK LİSANS TEZİ

İlker ÇAKAR

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Elektronik Mühendisliği Bilim Dalı

ŞUBAT 2025

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA İLE MEME KANSERİ TEŞHİSİ

YÜKSEK LİSANS TEZİ

İlker ÇAKAR

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Elektronik Mühendisliği Bilim Dalı

Tez Danışmanı: Doç. Dr. Muhammed Kürşad UÇAR

ŞUBAT 2025

İlker ÇAKAR tarafından hazırlanan “Yapay Zeka ile Meme Kanseri Teşhisi” adlı tez çalışması 07.02.2025 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim **Dalı Elektronik Mühendisliği Bilim Dalı**’nda **Yüksek Lisans tezi** olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Doç. Dr. M. Kürşad UÇAR (Danışman)
	Sakarya Üniversitesi
Jüri Üyesi :	Doç. Dr. Selçuk EMİROĞLU
	Sakarya Üniversitesi
Jüri Üyesi :	Doç. Dr. Kutlucan GÖRÜR
	Bandırma Onyedli Eylül Üniversitesi



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “YAPAY ZEKA İLE MEME KANSERİ TEŞHİSİ” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(...../...../2025).

(imza)

İlker ÇAKAR





Güzel eşime ve canım aileme ...



TEŞEKKÜR

Tez konusunun belirlenmesinde, gerekli çalışma şartlarının sağlanmasında ve çalışmalarımın her adımında bana desteğini esirgemeyen danışman hocam Sayın Doç. Dr. Muhammed Kürşad UÇAR'a ve tüm hayatım boyunca gerek maddi gerek manevi olarak yanımda olan canım eşime, anneme, babama ve ablama teşekkürlerimi sunarım.



İlker ÇAKAR



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxv
1. GİRİŞ.....	1
1.1. Tezin Amacı ve Kapsamı	1
1.2. Literatür Araştırması	4
2. MEME KANSERİ.....	7
2.1. İstatiksel Veriler	7
2.2. Türleri	12
2.3. Belitiler	13
2.4. Nedenler	14
2.5. Risk Faktörleri	14
2.6. Teşhis	15
2.7. Tedavi	16
3. MATERYAL VE YÖNTEM.....	17
3.1. Veri Setleri	17
3.2. Veri Ön işleme	18
3.2.1. Verilerin gruplandırılması	18
3.2.1.1. 1. veri seti (UMMC) için verilerin gruplandırılması	18
3.2.1.2. 2. veri seti (Breast Cancer) için verilerin gruplandırılması	21
3.2.1.3. 3. veri seti (Coimbra) için verilerin gruplandırılması	23
3.2.1.4. 4. veri seti (WPBC) için verilerin gruplandırılması	23
3.2.1.5. 5. veri seti (WBC) için verilerin gruplandırılması	25
3.2.1.6. 6. veri seti (WDBC) için verilerin gruplandırılması	26
3.2.2. Veri dengeleme	28
3.2.2.1. 1. veri seti (UMMC) için veri dengeleme	28
3.2.2.2. 2. veri seti (Breast Cancer) için veri dengeleme	28
3.2.2.3. 3. veri seti (Coimbra) için veri dengeleme	29
3.2.2.4. 4. veri seti (WPBC) için veri dengeleme	29
3.2.2.5. 5. veri seti (WBC) için veri dengeleme	30
3.2.2.6. 6. veri seti (WDBC) için veri dengeleme	31
3.3. Özellik Çıkarımı	31
3.3.1. 1. veri seti (UMMC) için özellik çıkarımı	34
3.3.2. 2. veri seti (Breast Cancer) için özellik çıkarımı	34
3.3.3. 3. veri seti (Coimbra) için özellik çıkarımı	35

3.3.4. 4. veri seti (WPBC) için özellik çıkarımı	35
3.3.5. 5. veri seti (WBC) için özellik çıkarımı	37
3.3.6. 6. veri seti (WDBC) için özellik çıkarımı	37
3.4. Özellik Seçme Algoritması.....	38
3.4.1. Seçim kriteri 1 – e-score.V1.....	39
3.4.2. Seçim kriteri 2 – e-score.V2.....	39
3.4.3. 1. veri seti (UMMC) için özellik seçme algoritmasının uygulanılması ...	40
3.4.4. 2. veri seti (Breast Cancer) için özellik seçme algoritmasının uygulanılması	41
3.4.5. 3. veri seti (Coimbra) için özellik seçme algoritmasının uygulanılması ..	42
3.4.6. 4. veri seti (WPBC) için özellik seçme algoritmasının uygulanılması	43
3.4.7. 5. veri seti (WBC) için özellik seçme algoritmasının uygulanılması.....	46
3.4.8. 6. veri seti (WDBC) için özellik seçme algoritmasının uygulanılması....	47
3.5. Makine Öğrenmesi Algoritmaları.....	50
3.5.1. Ensemble karar ağaçları	50
3.5.2. K-en yakın komşu algoritması	51
3.5.3. Destek vektör makineleri	51
3.5.4. Hibrit yapay zeka	52
3.6. Performans Değerlendirme Kriterleri.....	54
4. SONUÇ VE ÖNERİLER.....	61
4.1. 1. Veri Seti (UMMC) için Sonuçlar	61
4.2. 2. Veri Seti (Breast Cancer) için Sonuçlar	80
4.3. 3. Veri Seti (Coimbra) için Sonuçlar	87
4.4. 4. Veri Seti (WPBC) için Sonuçlar.....	90
4.5. 5. Veri Seti (WBC) için Sonuçlar.....	106
4.6. 6. Veri Seti (WDBC) için Sonuçlar	111
5. TARTIŞMA VE KARAR	117
KAYNAKLAR.....	141
ÖZGEÇMİŞ.....	147

KISALTMALAR

ANN	: Yapay Sinir Ağları (Artificial Neural Network)
AUC	: Alıcı İşletim Karakteristikleri Alanı (Area Under the Curve)
C4.5	: Karar Ağaçları Sınıflandırması (Decision Tree Classifier)
CART	: Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
CNN	: Evrişimsel Sinir Ağı (Convolutional Neural Network)
Ctree	: Koşullu Karar Ağacı (Conditional Decision Tree)
DCIS	: Bir Meme Kanseri Türü (Duktal Karsinoma In Situ)
FN Oranı	: Yanlış Negatif Oranı (False Negative Oranı)
FP Oranı	: Yanlış Pozitif Oranı (False Positive Oranı)
IBk	: Öznitelik Tabanlı Öğrenme (Simple Instance-Based Learner)
kNN	: k-En Yakın Komşular (k-Nearest Neighbors)
L	: Liste (List)
LR	: Doğrusal Regresyon (Linear Regression)
MCC	: Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient)
ML	: Makine Öğrenmesi (Machine Learning)
MLP	: Çok Katmanlı Algılayıcı (Multilayer Perceptron)
NB	: Naive Bayes
NE	: İç İç Topluluk Tekniği (Nested Community Technique)
RF	: Rastgele Ağaç Tekniği (Random Forest)
SAE	: Seyrek Otoenkoder (The Sparse Autoencoder)
SSAE	: Yığılan Seyrek Otoenkoder (Stacked Sparse Autoencoder)
SVMs	: Destek Vektör Makineleri (Support Vector Machines)
TP Oranı	: Doğru Pozitif Oranı (True Positive Oranı)
UMMC	: Malaya Üniversitesi Tıp Merkezi Meme Kanseri Kayıtları (Breast Cancer Registry of University Malaya Medical Centre)
WBC	: Wisconsin Meme Kanseri Kayıtları (Wisconsin Breast Cancer)
WDBC	: Wisconsin Tanısal Meme Kanseri Kayıtları (Wisconsin Diagnostic Breast Cancer)
WEKA	: Waikato Bilgi Analizi Ortamı (Waikato Environment for Knowledge Analysis)

WPBC : Wisconsin Öngörüşel Meme Kanseri Kayıtları (Wisconsin Pragnostic Breast Cancer)



SİMGELER

A	: Aktivite – Hjort Parametreleri (Activity Hjort Parameters)
C	: Karmaşıklık - Hjort Parametreleri (Complexity – Hjort Parameters)
CL	: Ortalama Eğri Uzunluğu (Average Curve Length)
CM	: Merkezi Momentler (Central Moments)
DK	: Varyans Katsayısı (Variance Coefficient)
E	: Ortalama Enerji (Energy Mean)
Eta	: Eta Değeri (Eta Value)
Eta_{mean}	: Ortalama Eta Değeri (Eta Value Mean)
G	: Geometrik Ortalama (Geometric Mean)
H	: Harmonik Ortalama (Harmonic Mean)
M	: Mobilite – Hjort Parametreleri (Mobility Hjort Parameters)
MAD	: Ortalama Mutlak Sapma (Mean Absolute Deviation)
IQR	: Çeyrek Arası Açıklık (Interquartile Range)
p	: Eleman Sayısı Denklemi (Number of Elements Equation)
r_p^b	: Eta Korolasyon Katsayısı (Eta Correlation Coefficient)
S	: Standart Sapma (Standart Deviation)
S_{x̄}	: Standart Hata (Standart Error)
S_y	: Her İki Sınıf için Standart Sapma (Standart Deviation)
SF	: Şekil Faktörü (Shape Factor)
SVD	: Tekil Değer Ayırıştırması (Singular Value Decomposition)
TE	: Ortalama Teager Enerjisi (Avarage Teager Energy)
T50	: %50 Kırılmış Ortalama (%50 Tripped Mean)
T25	: %25 Kırılmış Ortalama (%25 Tripped Mean)
x	: Medyan veya Ortalama (Median or Mean)
$\tilde{x}$	: Ortanca (Median)
$\bar{x}$	: Ortalama (Mean)
x_{kur}	: Basıklık (Kurtosis)
x_{max}	: Maksimum (Maximum)
x_{min}	: Minimum
x_{ske}	: Çarpıklık (Skewness)

X_{rms} : Ortalama Karekök Deęeri (Root Mean Square)



TABLO LİSTESİ

Sayfa

Tablo 3.1. 1. Veri Setindeki Nominal Özelliklerin Tanımları.....	19
Tablo 3.2. 1. Veri Setindeki Sayısal Özelliklerin Tanımları.....	21
Tablo 3.3. 2. Veri Setindeki Özelliklerin Tanımı.....	22
Tablo 3.4. 3. Veri Setindeki Özelliklerin Tanımı.....	23
Tablo 3.5. 4. Veri Setindeki Özelliklerin Tanımı.....	24
Tablo 3.6. 5. Veri Setindeki Özelliklerin Tanımı.....	25
Tablo 3.7. 6. Veri Setindeki Özelliklerin Tanımı.....	26
Tablo 3.8. 1. Veri Setindeki Özelliklerin Dengelenmesi.....	28
Tablo 3.9. 2. Veri Setindeki Özelliklerin Dengelenmesi.....	29
Tablo 3.10. 3. Veri Setindeki Özelliklerin Dengelenmesi.....	29
Tablo 3.11. 4. Veri Setindeki Özelliklerin Dengelenmesi.....	30
Tablo 3.12. 5. Veri Setindeki Özelliklerin Dengelenmesi.....	30
Tablo 3.13. 6. Veri Setindeki Özelliklerin Dengelenmesi.....	31
Tablo 3.14. Özellikler ait denklemler.....	32
Tablo 3.15. 1. Veri Setindeki Özelliklerin Çıkarımı.....	34
Tablo 3.16. 2. Veri Setindeki Özelliklerin Çıkarımı.....	34
Tablo 3.17. 3. Veri Setindeki Özelliklerin Çıkarımı.....	35
Tablo 3.18. 4. Veri Setindeki Özelliklerin Çıkarımı.....	35
Tablo 3.19. 5. Veri Setindeki Özelliklerin Çıkarımı.....	37
Tablo 3.20. 6. Veri Setindeki Özelliklerin Çıkarımı.....	37
Tablo 3.21. 1. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	40
Tablo 3.22. 1. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	41
Tablo 3.23. 2. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	42
Tablo 3.24. 2. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	42
Tablo 3.25. 3. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	42
Tablo 3.26. 3. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	43
Tablo 3.27. 4. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	43
Tablo 3.28. 4. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	46
Tablo 3.29. 5. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	47
Tablo 3.30. 5. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	47
Tablo 3.31. 6. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.....	48
Tablo 3.32. 6. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.	49
Tablo 3.33. Hibrit Yapay Zeka Tablo Örneği.....	54
Tablo 3.34. Doğruluk, özgüllük ve duyarlılık karşılaştırma tablosu.....	55
Tablo 3.35. Kappa katsayıları sınır değerleri.....	56

Tablo 3.36. 1. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .57	57
Tablo 3.37. 2. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .57	57
Tablo 3.38. 3. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .58	58
Tablo 3.39. 4. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .58	58
Tablo 3.40. 5. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .59	59
Tablo 3.41. 6. Veri Seti için Eğitim-Test Dağılımı (Örnekleme Dağılımı Tablosu). .59	59
Tablo 4.1. A Veri Seti Özet Tablo - En İyi Yöntemler.	64
Tablo 4.2. B Veri Seti Özet Tablo - En İyi Yöntemler.....	67
Tablo 4.3. C Veri Seti Özet Tablo - En İyi Yöntemler.....	69
Tablo 4.4. D Veri Seti Özet Tablo - En İyi Yöntemler.	71
Tablo 4.5. E Veri Seti Özet Tablo - En İyi Yöntemler.....	73
Tablo 4.6. F Veri Seti Özet Tablo - En İyi Yöntemler.	75
Tablo 4.7. A Veri Seti Özet Tablo - En İyi Yöntemler.	82
Tablo 4.8. B Veri Seti Özet Tablo - En İyi Yöntemler.....	83
Tablo 4.9. C Veri Seti Özet Tablo - En İyi Yöntemler.....	84
Tablo 4.10. Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.	89
Tablo 4.11. A Veri Seti Özet Tablo - En İyi Yöntemler.	92
Tablo 4.12. B Veri Seti Özet Tablo - En İyi Yöntemler.....	96
Tablo 4.13. C Veri Seti Özet Tablo - En İyi Yöntemler.....	99
Tablo 4.14. A Veri Seti Özet Tablo - En İyi Yöntemler.	107
Tablo 4.15. B Veri Seti Özet Tablo - En İyi Yöntemler.....	108
Tablo 4.16. Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.	112
Tablo 5.1. Çalışmamızda Yer Alan Tüm Veri Setleri ve Değerlerinin Yüzdesel Performans Sonuçları.	137
Tablo 5.2. Literatürde Araştırması Yapılan Tüm Veri Setleri ve Değerlerinin Yüzdesel Performans Sonuçları.	138

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1. Tez Akış Diyagramı.....	3
Şekil 2.1. Meme kanserinin dünya genelinde vaka sayısı ve ölüm oranı (Ferlay ve ark., 2021).	8
Şekil 2.2. Meme kanserinin diğer kanserlere göre kıyasla vaka sayısı (Ferlay ve ark., 2021).	8
Şekil 2.3. Meme kanserinin diğer kanserlere göre kıyasla ölüm oranı (Ferlay ve ark., 2021).	9
Şekil 2.4. Meme kanserinin kıtalara göre vaka sayısı (Ferlay ve ark., 2021).	10
Şekil 2.5. Meme kanserinin kıtalara göre ölüm sayısı (Ferlay ve ark., 2021).	10
Şekil 2.6. Meme kanserinin kıtalara göre vaka yaygınlığı (Ferlay ve ark., 2021).	11
Şekil 2.7. Meme kanserinin dünya genelinde bölgeler arası yaşa göre standardize edilmiş vaka ve ölüm oranları (Ferlay ve ark., 2021).	11
Şekil 2.8. Meme kanserinin dünya genelinde bölgelere arası cinsiyete göre standardize edilmiş vaka oranı (Ferlay ve ark., 2021).	12
Şekil 2.9. Meme kanseri belirtileri (Meme Kanseri Belirtileri ve İşaretleri, 2024)...	13
Şekil 3.1. E-Score Akış Diyagramı.....	39
Şekil 3.2. Ensemble Karar Ağaçları (Ulgen, 2017).	50
Şekil 3.3. k-En Yakın Komşu Algoritması (k-en Yakın Komşu Algoritması, 2024).51	
Şekil 3.4. Destek Vektör Makineleri (Uçar, 2017).	52
Şekil 3.5. Hibrit Yapay Zeka (Wang ve ark., 2021).	53
Şekil 4.1. A Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	77
Şekil 4.2. B Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	77
Şekil 4.3. C Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	78
Şekil 4.4. D Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	78
Şekil 4.5. E Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	79
Şekil 4.6. F Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	79
Şekil 4.7. A Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	86
Şekil 4.8. B Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	86
Şekil 4.9. C Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	87
Şekil 4.10. Kullanılan Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	90
Şekil 4.11. A Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	103
Şekil 4.12. B Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	104
Şekil 4.13. C Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	105
Şekil 4.14. A Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	110
Şekil 4.15. B Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	110
Şekil 4.16. Kullanılan Veri Seti Tüm Sınıflandırma Modellerinin Performansı.	116



YAPAY ZEKA İLE MEME KANSERİ TEŞHİSİ

ÖZET

Meme kanseri, memede bulunan hücrelerin kontrolden çıkması olarak tanımlanan bir hastalıktır. Meme kanseri dünya genelinde kadınlar arasında rastlanılan en popüler hastalıklar arasındadır. Çoğu meme kitlesinin iyi huylu olduğunu veya kanser olmadığını anlamak erken teşhis açısından oldukça önemlidir. Meme kanserinin erken teşhisi hastalığın tedavisi açısından ise ayrı bir öneme sahiptir. Kontrol edilmezse kötü huylu tümörler vücuda yayılıp diğer organlara sıçrayabilir ve ölümcül boyutlara ulaşabilir. Tedavinin erken aşamada tespiti hastalıktan doğacak ölümlerin önüne geçmektedir. Meme kanserinin herkes için uygun olacak tek bir tedavi yaklaşımı yoktur. Tedavi için meme kanserinin tipi, evresi ve kişinin yaşamı gibi faktörler göz önünde alınmaktadır. Meme kanseri için genel olarak beş tedavi seçeneği vardır ve çoğu tedavi planı aşağıdakilerin bir kombinasyonunu içerir: ameliyat, radyasyon, hormon tedavisi, kemoterapi ve hedefe yönelik tedavilerdir. Bazıları lokaldir ve yalnızca tümörün etrafındaki alanı hedef alır. Diğerleri sistemiktir ve kanserle mücadele eden ajanlarla tüm vücudunuzu hedef alır. Tüm bu tedavi yöntemlerine rağmen kanser vücudun diğer bölgelerine yayılmışsa genellikle tedavi edilemez, ancak normalde uzun süre etkili bir şekilde kontrol edilebilir.

Son zamanlarda meme kanserinin erken evrede teşhisini ve doğruluğunu sağlamak amacıyla bilgisayar destekli sistemler geliştirilmektedir. Makine öğrenimi yaklaşımlarıyla geliştirilen bilgisayar destekli sistemler meme kanserinin teşhisindeki sürece büyük katkı sağlamaktadır. Makine öğrenimi yaklaşımlarıyla geliştirilen bilgisayar destekli akıllı ve otomatik tanı sistemleri, analizde önemli araçlardır ve tıp uzmanlarını meme kanseri konusunda destekleyebilmekte ve tıbbi karar verme sürecinde yer almaktadır. Çok sayıda yöntem gösterilmiş olmasına rağmen, yöntemlerin çoğu doğru ve tutarlı bir sonuç sağlayamamaktadır. Ayrıca mevcut sistemler daha yüksek doğruluk oranı ve daha az hesaplama süresi gerektirmektedir. Ancak tüm bu mevcut çalışmalar gene de tutarlı bir doğruluk oranı sağlayamamaktadır. Makine öğreniminin meme kanseri üzerinde temel amacı, bir sistemin insan müdahalesi olmadan öğrenmesini sağlamaktır, bu da karar verme için otomatik bir sistem tasarlamaya yardımcı olmaktadır. Ancak makine öğrenimi modelinin tahmin doğruluğunun iyileştirilmesi büyük bir zorluk ve araştırma boşluğu olduğu görülmüştür. Tüm bunlara rağmen meme kanserinin sınıflandırma zorluğu için çok sayıda makine öğrenimi tekniği sunulmuştur.

Bu çalışmadaki veri setleri herkese açık olarak sunulan internet sitelerinden elde edilmiştir. Çalışmada kullanılan 1. veri seti setinde 4024 farklı kişiden 14 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 4024 farklı kişinin yaşama ve ölme durumları sınıflandırılmıştır. Çalışmada kullanılan 2. veri seti setinde 286 farklı kişiden 9 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 286 farklı kişinin meme kanserinin tekrarlama ve tekrarlama durumu sınıflandırılmıştır. Çalışmada kullanılan 3. veri seti setinde 116 farklı kişiden 9 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 116 farklı kişinin sağlıklı ya da hasta olma durumları sınıflandırılmıştır.

Çalışmada kullanılan 4. veri seti setinde 198 farklı kişiden 33 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 198 farklı kişinin meme kanserinin tekrarlama ve tekrarlamama durumları sınıflandırılmıştır. Çalışmada kullanılan 5. veri seti setinde 699 farklı kişiden 9 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 699 farklı kişinin meme kanserinin iyi huylu veya kötü huylu olma durumları sınıflandırılmıştır. Çalışmada kullanılan 6. veri seti setinde 569 farklı kişiden 30 özellik bilgisi alınmıştır. Bu bilgiler doğrultusunda 569 farklı kişinin meme kanserinin iyi huylu veya kötü huylu olma durumları sınıflandırılmıştır.

Bu çalışmada öncelikle veri ön işleme ile tüm veri setlerindeki özellik değerleri tanım ve ifadelerine göre incelenip, sayısal ve nominal değerler olarak ayrılmıştır. Sayısal olmayan nominal değerler veri işleme açısından sayısal değerlere atanılmıştır. Sayısal değerler ise olduğu gibi bırakılmış ya da kendi içlerinde gruplandırma yapılmıştır. Daha sonra tüm sisteme Eta özellik seçme algoritması uygulanmıştır. Daha sonra 6 veri seti üzerinde de sistemin daha verimli çalışabilmesi için veri dengeleme işlemi yapılmıştır. Veri dengeleme işleminden sonra ise veriler eğitim-test olacak şekilde 1. veri setinde, 2. veri setinde ve 4. veri setinde %50-%50, 3. veri setinde %85-%15, 5. veri setinde ve 6. veri setinde ise %80-%20 şeklinde bölünme oranlarına ayrılmıştır. Daha sonra tüm 6 veri seti üzerinde makine öğrenmesi doğrultusunda Ensemble, kNN (k En Yakın Komşu), SVMs (Destek Vektör Makineleri) ve Hibrit Yapay Zeka gibi makine öğrenmesi algoritmaları uygulanmıştır. Elde edilen sonuçlar ile meme kanseri algoritması oluşturulmuştur ve görüntüler elde edilmiştir.

6 farklı veri setinde de makine öğrenmesi algoritmaları ile doğruluk, özgüllük, duyarlılık, kappa sayısı, F-Score gibi performans değerlendirme kriterleri elde edilmiştir. Tüm veri setlerinde bulunan değerlendirme kriterleri öncelikle hangi veri seti üzerinde çalışma yapıldıysa o veri seti üzerindeki makine öğrenmesi teknikleriyle daha sonra ise tüm veri setleri üzerinde bulunan değerlendirme kriterleri ile karşılaştırma yapılmıştır.

Bu çalışmada elde edilen tüm veri setlerinde bulunan sonuçlar karşılaştırıldığında ise en yüksek doğruluk oranı, F-Score, Kappa ve AUC değerleri çalışmada 6. veri seti olarak nitelendirilen Wisconsin Diagnostic Breast Cancer veri setinde Ensemble yöntemi ile %100, en yüksek duyarlılık oranı çalışmada bulunan 6 veri setinden 1. veri seti olarak nitelendirilen Breast Cancer Registry of University Malaya Medical Centre veri setinde Hibrit yöntem ile %100, 2. veri seti olarak nitelendirilen veri setinde SVMs, Ensemble, kNN ve Hibrit yöntemleri ile %100, 3. veri seti olarak nitelendirilen Coimbra veri setinde Ensemble ve Hibrit yöntemleri ile %100, 6. veri seti olarak nitelendirilen Wisconsin Diagnostic Breast Cancer veri setinde Ensemble yöntemi ile %100 ve en yüksek özgüllük oranı çalışmada bulunan 6 veri setinden 3. veri seti olarak nitelendirilen Coimbra veri setinde SVMs, Ensemble ve Hibrit yöntemleri ile %100, 5. veri seti olarak nitelendirilen Wisconsin Breast Cancer veri setinde SVMs, kNN, Ensemble ve Hibrit yöntemleri ile %100, 6. veri seti olarak nitelendirilen Wisconsin Diagnostic Breast Cancer veri setinde Ensemble yöntemi ile %100 olarak bulunmuştur.

Bu sonuçlar doğrultusunda Matlab ortamında 6 farklı veri seti üzerinde uygulaması yapılan bu çalışmanın makine öğrenmesi algoritmalarını etkili şekilde kullanıldığı görülmüştür.

Sonuç olarak bu çalışmanın farklı makine öğrenmesi teknikleri ile daha yüksek yüzdelerde doğruluk, özgüllük ve duyarlılık oranları bulunabileceğini ispatlanmıştır. Bu durum da çalışmamızın meme kanseri teşhisinde hastalıklı ve sağlıklı kişilerin

bulunmasında güvenilir bir çalışma olduğunu ortaya koyarak sađlık alanında daha uygulanabilir ve elverişli bir çalışma olduğunu göstermiştir.





BREAST CANCER DIAGNOSIS WITH ARTIFICIAL INTELLIGENCE

SUMMARY

Breast cancer is a disease characterized by the uncontrolled growth of cells in the breast. It is among the most commonly diagnosed illnesses in women worldwide. Recognizing that most breast lumps are benign or non-cancerous is vital for early diagnosis. Early detection of breast cancer holds particular significance in terms of effective treatment. If left unchecked, malignant tumors may spread to other parts of the body, potentially reaching life-threatening stages. Detecting and addressing the disease at an early stage plays a crucial role in preventing cancer-related deaths.

There is no universal treatment approach suitable for every individual diagnosed with breast cancer. Factors such as the type and stage of cancer, as well as the individual's overall lifestyle and health, are taken into consideration when determining the best course of treatment. Recently, computer-aided systems have been developed to improve the early detection and diagnostic accuracy of breast cancer. These systems, powered by machine learning approaches, have made substantial contributions to the process of diagnosing breast cancer.

Machine learning-based computer-aided intelligent and automated diagnostic systems are essential tools in analyzing breast cancer data. They not only assist healthcare professionals in understanding and diagnosing breast cancer but also play a pivotal role in the medical decision-making process. Despite the development of numerous methods, many fail to provide accurate and consistent results. Current systems also demand higher accuracy rates and faster computational times. Even with these improvements, achieving consistent accuracy across varying datasets and conditions remains a significant challenge.

The primary aim of machine learning in breast cancer research is to enable systems to learn and adapt without direct human intervention, facilitating the development of automated decision-making tools. However, improving the predictive accuracy of machine learning models continues to present a significant challenge and identifies a notable research gap in the field. Despite these difficulties, many machine learning techniques have been proposed to address the challenges associated with breast cancer classification.

Expanding on this, the importance of machine learning in breast cancer cannot be overstated. It offers an opportunity to revolutionize how we approach diagnosis and treatment. Machine learning algorithms can process vast datasets, identify patterns that may not be immediately apparent to human experts, and deliver predictions with remarkable speed. For example, deep learning algorithms have been applied to mammographic images, significantly improving the ability to detect anomalies that could indicate early-stage cancer.

Moreover, these systems can incorporate patient data such as genetic information, lifestyle factors, and treatment histories to personalize diagnostics and predict treatment outcomes. This individualized approach ensures that each patient receives

care tailored to their unique circumstances, potentially improving prognosis and quality of life.

However, the integration of these technologies into routine clinical practice is not without obstacles. Ethical considerations, data privacy issues, and the need for extensive validation across diverse populations remain critical areas of concern. Furthermore, the computational demands of advanced machine learning models require robust hardware and software infrastructures, which may not be readily available in all healthcare settings.

Despite these challenges, the field continues to advance, driven by the promise of better outcomes for patients. Future research should focus on developing models that not only achieve high accuracy but also exhibit generalizability across different datasets and populations. Additionally, collaborative efforts between machine learning researchers and medical practitioners will be essential in translating theoretical advancements into practical solutions that benefit patients worldwide.

In conclusion, machine learning holds the potential to transform breast cancer diagnosis and treatment. By overcoming existing limitations and addressing current gaps, these technologies can provide a powerful ally in the fight against breast cancer, ultimately leading to earlier diagnoses, better treatment options, and improved survival rates.

The datasets used in this study were obtained from publicly available online sources. In the first dataset used in the study, data on 14 attributes were collected from 4024 individuals. Based on this information, the survival and mortality status of these 4024 individuals were classified. In the second dataset, data on 9 attributes were collected from 286 individuals. Based on this information, the recurrence and non-recurrence status of breast cancer among these 286 individuals were classified. In the third dataset, data on 9 attributes were collected from 116 individuals. Based on this information, the health status of these 116 individuals (healthy or diseased) was classified. In the fourth dataset, data on 33 attributes were collected from 198 individuals. Based on this information, the recurrence and non-recurrence status of breast cancer among these 198 individuals were classified. In the fifth dataset, data on 9 attributes were collected from 699 individuals. Based on this information, the benign or malignant status of breast cancer among these 699 individuals was classified. In the sixth dataset, data on 30 attributes were collected from 569 individuals. Based on this information, the benign or malignant status of breast cancer among these 569 individuals was classified.

In this study, all feature values in the datasets were initially examined through data preprocessing according to their definitions and expressions, and were categorized as numerical or nominal values. Non-numerical nominal values were assigned numerical equivalents for processing purposes, while numerical values were either left as they were or grouped internally. Subsequently, the Eta feature selection algorithm was applied to the entire system. Data balancing was performed on all six datasets to enhance system efficiency. After balancing, the data were divided into training and testing sets with specific split ratios: 50%-50% for the 1st, 2nd, and 4th datasets; 85%-15% for the 3rd dataset; and 80%-20% for the 5th and 6th datasets. Following these preprocessing steps, machine learning algorithms such as Ensemble, k-Nearest Neighbors (kNN), Support Vector Machines (SVMs), and Hybrid Artificial Intelligence were applied to all six datasets. The results obtained were used to develop a breast cancer algorithm, and visual representations were generated based on the findings.

Performance evaluation metrics such as accuracy, specificity, sensitivity, kappa statistic, and F-Score were obtained for all six datasets using machine learning algorithms. These evaluation metrics were initially analyzed for each dataset individually, focusing on the machine learning techniques applied to that specific dataset. Subsequently, the evaluation metrics across all datasets were compared to assess the overall performance of the applied techniques.

When the results obtained from all datasets in this study were compared, the highest accuracy, F-Score, Kappa, and AUC values were achieved with the Ensemble method on the 6th dataset, referred to as the Wisconsin Diagnostic Breast Cancer dataset, reaching 100%. The highest sensitivity rate was observed on the 1st dataset, known as the Breast Cancer Registry of University Malaya Medical Centre dataset, with the Hybrid method achieving 100%. Similarly, the 2nd dataset achieved a sensitivity rate of 100% with SVMs, Ensemble, kNN, and Hybrid methods. The 3rd dataset, referred to as the Coimbra dataset, also reached 100% sensitivity using Ensemble and Hybrid methods. The 6th dataset, the Wisconsin Diagnostic Breast Cancer dataset, achieved 100% sensitivity with the Ensemble method. The highest specificity rate was found in the 3rd dataset, the Coimbra dataset, where SVMs, Ensemble, and Hybrid methods achieved 100% specificity. Likewise, the 5th dataset, referred to as the Wisconsin Breast Cancer dataset, achieved 100% specificity with SVMs, kNN, Ensemble, and Hybrid methods. The 6th dataset, the Wisconsin Diagnostic Breast Cancer dataset, also achieved 100% specificity with the Ensemble method.

Based on the results obtained in this study, it is evident that the machine learning algorithms applied across six distinct datasets using the MATLAB environment were utilized with exceptional efficiency. This research highlights the significant potential of integrating advanced computational methods into medical applications, particularly in the field of breast cancer diagnosis. By leveraging diverse datasets and sophisticated algorithms, the study underscores the ability of machine learning techniques to deliver accurate and reliable outcomes. These outcomes are not only scientifically significant but also have practical implications for real-world healthcare scenarios.

In conclusion, this study has convincingly demonstrated that higher accuracy, specificity, and sensitivity rates can be achieved through the implementation of various machine learning techniques. The results obtained serve as a testament to the capability of machine learning systems to distinguish between diseased and healthy individuals with precision. The high performance of these algorithms reinforces the reliability of this study as an innovative and effective approach in breast cancer diagnosis. Furthermore, it underlines the importance of adopting computational tools in medical research, showcasing their applicability and practicality in addressing critical healthcare challenges.

One of the most commendable aspects of this study is its use of six different datasets, each presenting unique characteristics and challenges. By systematically analyzing these datasets and applying advanced algorithms, the research demonstrates its robustness and adaptability. The inclusion of datasets with varying attributes ensures that the findings are not limited to a single context, enhancing their generalizability and relevance across diverse scenarios. This comprehensive approach reflects the study's commitment to rigor and excellence in research.

Moreover, the integration of techniques such as Ensemble methods, k-Nearest Neighbors (kNN), Support Vector Machines (SVMs), and Hybrid Artificial Intelligence showcases the breadth of machine learning methodologies explored in this

study. These algorithms, known for their strong predictive capabilities, have been effectively applied to achieve high-performance metrics, including accuracy, specificity, and sensitivity. This demonstrates not only the technical expertise of the research but also its focus on delivering results that are both meaningful and actionable in clinical settings.

The study also addresses a critical gap in current breast cancer diagnostic methods by emphasizing the importance of consistent and reliable outcomes. While traditional diagnostic approaches may suffer from variability and limited scalability, the machine learning-driven framework proposed in this study offers a more standardized and scalable solution. By automating the diagnostic process and reducing reliance on subjective human interpretations, the research paves the way for more objective and reproducible results. This advancement has the potential to significantly improve early detection rates, leading to better patient outcomes and more effective treatment planning.

Additionally, this study's contribution extends beyond its technical achievements. It serves as a powerful example of how interdisciplinary collaboration between computational science and medicine can yield transformative solutions to pressing healthcare problems. By aligning the strengths of machine learning with the needs of medical practitioners, the research provides a pathway for future innovations in diagnostic and therapeutic strategies.

In the context of breast cancer, where early detection and accurate classification are crucial for successful treatment, the significance of this study cannot be overstated. Breast cancer remains one of the leading causes of morbidity and mortality worldwide, making the development of reliable diagnostic tools a priority. This study not only meets this need but also sets a benchmark for future research in the field. Its findings contribute to the growing body of evidence supporting the use of machine learning in oncology, reinforcing its role as a critical tool in the fight against cancer.

In summary, this research exemplifies the power and promise of machine learning in healthcare. Through meticulous design, comprehensive analysis, and innovative application of advanced algorithms, the study has achieved outcomes that are both scientifically valuable and clinically relevant. It stands as a testament to the potential of computational methods to revolutionize breast cancer diagnosis, offering hope for earlier detection, more accurate classifications, and ultimately, improved survival rates. This study is not only a significant academic achievement but also a meaningful contribution to the ongoing efforts to enhance healthcare delivery and patient outcomes worldwide.

1. GİRİŞ

1.1. Tezin Amacı ve Kapsamı

Meme kanseri, memedeki hücrelerin kontrolden çıktığı bir hastalıktır (Centers for Disease Control and Prevention, 2020). Meme kanseri tüm dünyada kadınlarda görülen kanserlerin başında gelmektedir (Islam ve ark., 2020). Meme kanseri ABD'li kadınlar arasında en sık teşhis edilen kanserdir (Breastcancer, 2023). Her yıl kadınlarda yeni teşhis edilen kanserlerin yaklaşık %30'u meme kanseridir (Breastcancer, 2023). Çoğu meme kitlesinin iyi huylu olduğunu ve kanser (kötü huylu) olmadığını anlamak önemlidir (American Cancer Society, 2021). İyi belirlenmiş bir prognostik faktör olan tümör histolojik derecesi, klinik uygulamada uygun tedaviyi yönlendirmek için çok önemlidir (Fan ve ark., 2020). Ayrıca hastalığın erken aşamalarda tespit edilmesi artan ölümlerin önlenmesine yardımcı olabilir (Bray ve ark., 2018). Kontrol edilmezse kötü huylu tümörler vücuda yayılabilir ve ölümcül olabilir (Krag, 2020).

Meme kanseri için herkese uyacak tek bir tedavi yaklaşımı yoktur (Content on Early Breast Cancer, 2024). Tedavi seçeneği için meme kanseri tipi, evresi ve kişinin yaşam tarzı gibi faktörler dikkate alınır (Content on Early Breast Cancer, 2024). Son yıllarda tanı ve tedavinin doğruluğunu ve etkinliğini artırmak amacıyla bilgisayar destekli tekniklerin meme kanseri alanında entegrasyonuna yönelik artan bir eğilim vardır (Cheng ve ark., 2023). Makine öğrenimi teknikleri ve tıbbi görüntüleme bu süreçte yardımcı olmaktadır (Houssein ve ark., 2021). Makine öğrenimi yaklaşımlarıyla geliştirilen bilgisayar destekli akıllı ve otomatik tanı sistemleri, analizde önemli araçlardır ve tıp uzmanlarını meme kanseri konusunda destekleyebilir ve tıbbi karar verme sürecinde yer alır (Kadam ve ark., 2019). Son zamanlarda, makine öğrenmesi tekniklerinin, (Ayon ve ark., 2022; Haque ve ark., 2018; Muhammad ve ark., 2020), yanında derin öğrenme gibi çeşitli teknikler tıbbi alanda kullanılmıştır (Islam ve ark., 2019; Zabirul ve ark., 2020). Ayrıca veri madenciliği tekniği de verileri anlama ve tahmin etmede basit bir yol olarak görülmüştür (Ak, 2020). Mikrodalga görüntüleme de erken evre tarama ve meme kanserinin izlenmesi için yaygın görüntüleme

tekniklerindendir (Conceição ve ark., 2020). Çok sayıda yöntem gösterilmiş olmasına rağmen, yöntemlerin çoğu doğru ve tutarlı bir sonuç sağlayamamaktadır (Islam ve ark., 2020). Ayrıca mevcut sistemler daha yüksek doğruluk oranı ve daha az hesaplama süresi gerektirmektedir (Muduli ve ark., 2020). Ancak tüm bu mevcut çalışmalar gene de tutarlı bir doğruluk oranı sağlayamamıştır.

Meme kanseri teşhisinde kullanılan makine öğrenmesi, kolayca tanımlanamayan gizli bilgileri keşfetmek için verileri kullanma süreci olarak tanımlanır (Huang ve Chen, 2022). Makine öğreniminin temel amacı, bir sistemin insan müdahalesi olmadan öğrenmesini sağlamaktır, bu da karar verme için otomatik bir sistem tasarlamaya yardımcı olur (Kumar Jakhar ve ark., 2022). Literatürde, önceki çalışmalarda da Makine Öğrenmesi (ML) kullanımını önerilmiştir (Agarap, 2018). Ancak makine öğrenimi modelinin tahmin doğruluğunun iyileştirilmesi büyük bir zorluk ve araştırma boşluğu olduğu görülmüştür (Haq ve ark., 2021). Tüm bunlara rağmen meme kanserinin sınıflandırma zorluğu için araştırmacılar önceki çalışmalarda çok sayıda makine öğrenimi tekniği sunmuştur (Naseem ve ark., 2022).

Bu tez amacına uygun olarak işleme şekli Şekil 1.1’de özetlenmiştir. Bu işleyişe göre tez içeriği yazılmıştır.

Bu çalışmada makine öğrenmesi doğrultusunda Ensemble, kNN, SVMs ve Hibrit Yapay Zeka gibi sınıflandırma işlemleri uygulanmıştır. Elde edilen sonuçlar ile Meme Kanseri teşhis algoritması oluşturulmuştur.

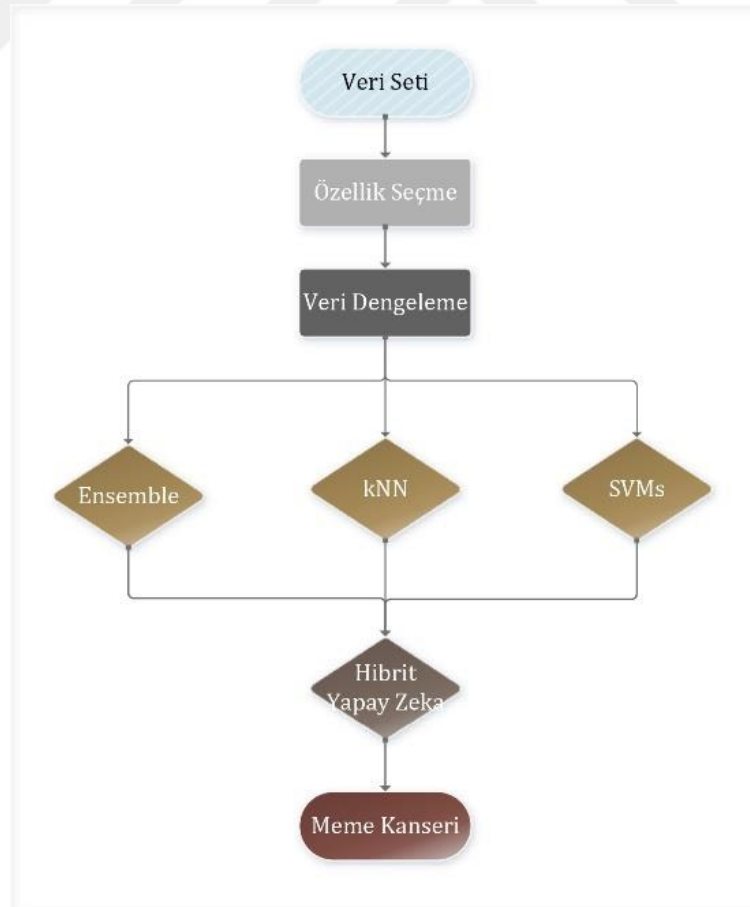
Çalışmamızda da literatürdeki gibi doğruluk oranı, duyarlılık, özgüllük, kappa sayısı, F-Score, AUC gibi performans değerlendirme kriterleri uygulanmıştır (Uçar ve ark., 2017, 2018).

Çalışmamızda meme kanseri ile ilgili 6 farklı veri seti kullanılmıştır. Çalışmamızda elde edilen sonuçlarda tüm veri setleri üzerindeki değerler kıyaslandığında 1. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, en yüksek özgüllük oranı Ensemble yöntemi ile %98,7 ve en yüksek duyarlılık oranı birçok yöntem ile %100 olarak bulunmuştur. 2. veri setine göre en yüksek doğruluk oranı SVMs yöntemi ile 77,63, en yüksek özgüllük oranı Hibrit ve Ensemble yöntemi ile %88,10 ve en yüksek duyarlılık oranı tüm yöntemler ile %100 olarak bulunmuştur. 3. veri setine göre en yüksek doğruluk oranı SVMs, Ensemble ve Hibrit yöntemi ile %99,33, en yüksek özgüllük oranı SVMs, Ensemble ve Hibrit yöntemi ile %100 ve en yüksek duyarlılık

oranı Ensemble ve Hibrit yöntemi ile %100 olarak bulunmuştur. 4. veri setine göre en yüksek doğruluk oranı SVMs yöntemi ile %85,42, en yüksek özgüllük oranı SVMs ile %91,3 ve en yüksek duyarlılık oranı SVMs ve Ensemble yöntemi ile %88 olarak bulunmuştur. 5. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %98,94, en yüksek özgüllük oranı tüm yöntemler ile %100 ve en yüksek duyarlılık oranı yöntemi ile Ensemble %97,83 olarak bulunmuştur. 6. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %100, en yüksek özgüllük oranı Ensemble yöntemi ile %100 ve en yüksek duyarlılık oranı Ensemble yöntemi ile %100 olarak bulunmuştur.

Çalışmamızda 1. veri seti, 2. veri seti ve 4. veri setinde %50 eğitim - %50 test veri seti, 3. veri setinde %85 eğitim - %15 test veri seti, 5. veri seti ve 6. veri setinde %80 eğitim - %20 test veri seti uygulaması yapılmıştır.

Sonuçlarda 1. veri seti, 2. veri seti, 3. veri seti, 5. veri seti ve 6. veri setinde yapmış olduğumuz makine öğrenmesi uygulaması meme kanseri teşhisinde çok üstün sonuçlara ulaştığı görülmüştür.



Şekil 1.1. Tez Akış Diyagramı.

1.2. Literatür Araştırması

Literatürde makine öğrenmesi teknikleri olarak Yapay Sinir Ağları (ANN), Destek Vektör Makineleri (SVMs), Naive Bayes (NB), Sınıflandırma ve Regresyon Ağaçları (CART), k-En Yakın Komşu (kNN), Doğrusal Regresyon (LR) ve Çok Katmanlı Algılayıcı (MLP), Rastgele Orman (Random Forest), Aşırı Güçlendirme (Extreme Boost), Karar Ağaçları (C4.5), Lojistik Regresyon, Doğrusal Ayrımcı Analiz (Linear Discriminant Analysis), Boosting ve AdaBoost, Bagging Algoritması, IBk (Bazı Parametrelerle Öznitelik Tabanlı Öğrenme Bazı), Rastgele Komite Algoritması kullanılmıştır.

Literatürde makine öğrenmesi modelinin performans değerlendirmesi için F-Score, AUC, ROC eğrisi altındaki alan, doğruluk (accuracy), duyarlılık değeri (recall), duyarlılık (sensitivity), özgüllük (specificity), kappa istatistiği, TP Oranı (Doğru Pozitif Oran), FP Oranı (Yanlış Pozitif Oran), MCC, Alıcı İşletim Karakteristikleri Alanı (AUC), zaman karmaşıklığı (Time Complexity), Lift Eğrisi (Lift Curve) ve Kalibrasyon Grafiği (Calibration Plot), Rekürsif Özellik İyileştirme Tekniği (Recursive Feature Elimination) gibi farklı performans metrikleri kullanılmıştır.

Çalışmamızda da gene literatürdeki gibi doğruluk oranı, duyarlılık, özgüllük, kappa sayısı, F-Score, AUC gibi performans değerlendirme kriterleri uygulanmıştır (Uçar ve ark., 2017, 2018).

Literatürdeki çalışmalarda R programlama, Weka, Spark ve Python gibi farklı platformlarda makine öğrenmesi algoritmaları geliştirilmiştir (Bharati ve ark., 2018; Keleş, 2019; Sharma ve ark., 2018).

Literatürde yapılan çalışmaların büyük çoğunluğu çalışmamızda kullanılan 6. Veri Seti ile aynı olan Wisconsin Diagnostic Breast Cancer (WDBC) veri seti üzerinde gerçekleşmiştir (Abdar ve ark., 2020; Bharat ve ark., 2018; Fatima ve ark., 2020).

Diğer çalışmalar ise WDBC'den farklı veri setleri üzerinde gerçekleşmiştir (LG ve AT, 2013; Maliha ve ark., 2019).

Çalışmamızda olduğu gibi literatürdeki bazı çalışmalarda özellik azaltılarak alt veri kümeleri oluşturma ve veri temizleme, özellik seçimi, özellik çıkarımı gibi ortak aşamalar uygulanmış ancak yüksek doğruluk oranlarına ulaşamamıştır (Israni, 2019; Memon ve ark., 2019; Said ve ark., 2018).

Literatürde bazı çalışmalarda veri seti %80 eğitim, %20 test olarak bölünmüştür. Literatürdeki bu çalışmalarda olduğu gibi çoğu çalışmada özellik azaltılarak alt veri kümeleri oluşturulmuş ve özellik seçimi yapılmıştır (Chaurasia ve Pal, 2020). Literatürde bir diğer çalışmada eğitim ve test veri setleri %66'ya %33 olarak bölünmüştür. Böylece deneysel sonuçlarda SVMs %96,9957 ile en iyi doğruluk performansını göstermiştir (Bayrak ve ark., 2019).

Literatürde en yüksek doğruluk oranı WDBC veri setinde SVMs yöntemi ile %99,68, WPBC veri setinde %99,68 bulunmuştur (Abdel ve Eldeib, 2016; Fatima ve ark., 2020). Literatürde en yüksek duyarlılık değeri WPBC veri setinde %99,68 bulunmuştur (Abdel ve Eldeib, 2016). Literatürde en yüksek özgüllük değeri WPBC veri setinde %99,47 bulunmuştur (Abdel ve Eldeib, 2016). Literatürde en yüksek F-Score değeri WDBC ve WBC veri setinde %98,1 bulunmuştur (Abdar ve ark., 2020; Israni, 2019). Literatürde en yüksek Kappa değeri Coimbra veri setinde %70,39 bulunmuştur (Sen, 2021). Literatürde en yüksek AUC değeri WBC veri setinde %99,5 bulunmuştur (Israni, 2019).



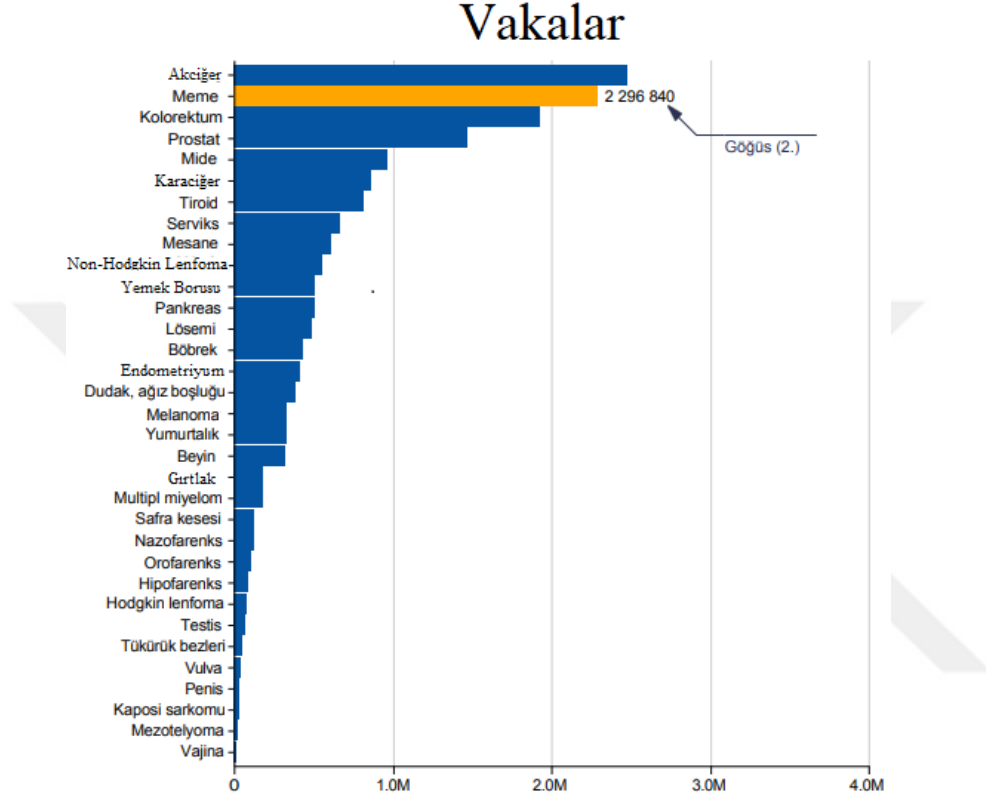
2. MEME KANSERİ

2.1. İstatiksel Veriler

Meme kanseri ile ilgili 2022 yılında Global Cancer Observatory verilerine göre Şekil 2.1, Şekil 2.2 ve Şekil 2.3’de görüldüğü üzere meme kanseri dünya genelinde kanser türleri sıralamasında vakalar arasında 2296840 vaka ile 2. sırada, ölüm oranları göz önünde bulundurulduğunda ise 666103 ölüm sayısı ile 4. sırada yer almaktadır (Ferlay ve ark., 2021). Meme kanserinin kıtalar arası vaka sayıları ve oranları göz önüne alındığında Şekil 2.4’de görülebileceği gibi Asya kıtası 985817 vaka sayısı ve %42.9 oranıyla kıtalar arasında en çok vaka görülen kıta olmuş ve arkasından Avrupa, Kuzey Amerika, Latin Amerika ve Karayipler, Afrika ve Okyanusya kıtaları gelmiştir (Ferlay ve ark., 2021). Şekil 2.5’de meme kanseri ölüm sayıları ve oranları göz önüne alındığında ise vaka sayıları ve oranlarında olduğu gibi Asya kıtası 315309 ölüm sayısı ve %47.3 oranıyla kıtalar arasında en çok ölüm görülen kıta olmuş ve arkasından Avrupa, Afrika, Latin Amerika ve Karayipler, Kuzey Amerika ve Okyanusya gelmiştir (Ferlay ve ark., 2021). Şekil 2.6’de meme kanseri vaka yaygınlık sayısı ve oranları göz önüne alındığında ise vaka sayıları ve oranları ile ölüm sayıları ve oranlarında olduğu gibi Asya kıtası 3197043 vaka yaygınlık sayısı ve %39.1 oranıyla kıtalar arasında en çok vaka yaygınlığı görülen kıta olmuş ve arkasından Avrupa, Kuzey Amerika, Latin Amerika ve Karayipler, Afrika ve Okyanusya gelmiştir (Ferlay ve ark., 2021). Meme kanserinin dünya genelinde bölgeler arası yaşa göre standardize edilmiş vaka ve ölüm oranları incelendiğinde ise Şekil 2.7’de görülebileceği gibi en çok vaka görülen bölge Avusturalya-Yeni Zelanda olmuş ve en çok ölüm görülen bölge ise Melanezya olmuştur (Ferlay ve ark., 2021). Meme kanserinin dünya genelinde bölgelere arası cinsiyete göre standardize edilmiş vaka oranı Şekil 2.8’de görülebileceği gibi kadınlar arasında en çok vaka görülen bölge Avusturya-Yeni Zelanda olmuştur (Ferlay ve ark., 2021).

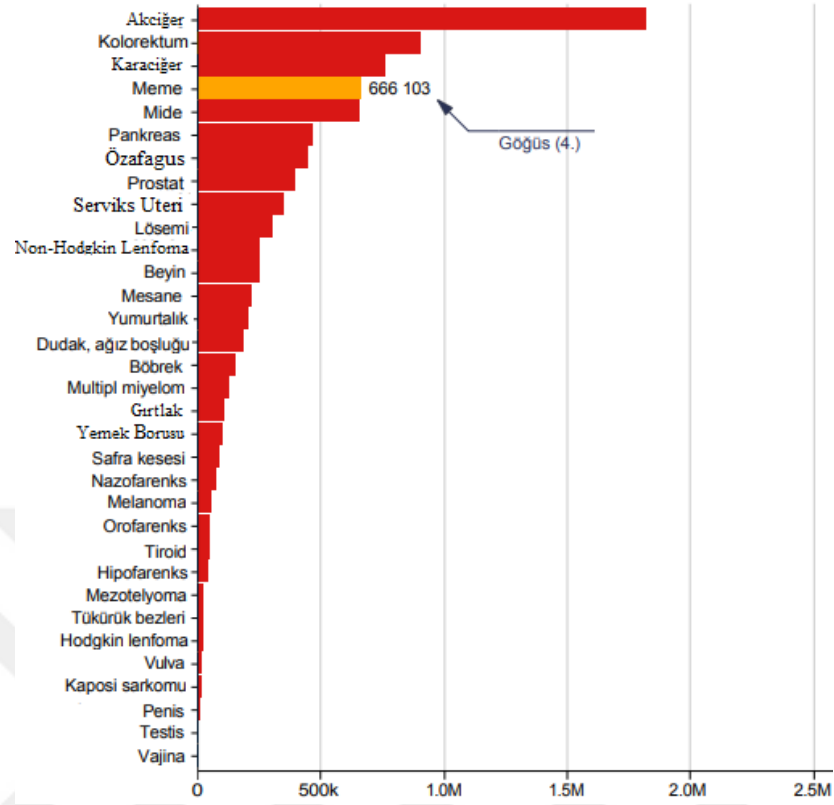
Vakalar			Ölüm Oranı		
Rütbe	Vakalar	ASR (Dünya)	Rütbe	Ölümler	ASR (Dünya)
2	2 296 840	46.8	4	666 103	12.7

Şekil 2.1. Meme kanserinin dünya genelinde vaka sayısı ve ölüm oranı (Ferlay ve ark., 2021).



Şekil 2.2. Meme kanserinin diğer kanserlere göre kıyasla vaka sayısı (Ferlay ve ark., 2021).

Ölüm Oranı



Şekil 2.3. Meme kanserinin diğer kanserlere göre kıyasla ölüm oranı (Ferlay ve ark., 2021).

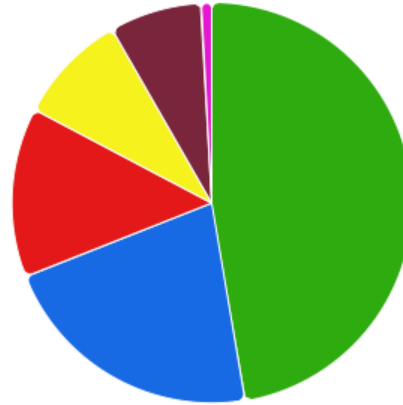
Vakalar



Kıta	Vakalar	Yüzde
Asya	985 817	42.9%
Avrupa	557 532	24.3%
Kuzey Amerika	306 307	13.3%
Latin Amerika ve Karayipler	220 124	9.6%
Afrika	198 553	8.6%
Okyanusya	28 507	1.2%

Şekil 2.4. Meme kanserinin kıtalara göre vaka sayısı (Ferlay ve ark., 2021).

Ölüm oranı



Kıta	Ölümler	Yüzde
Asya	315 309	47.3%
Avrupa	144 439	21.7%
Afrika	91 252	13.7%
Latin Amerika ve Karayipler	59 876	9.0%
Kuzey Amerika	49 744	7.5%
Okyanusya	5 483	0.82%

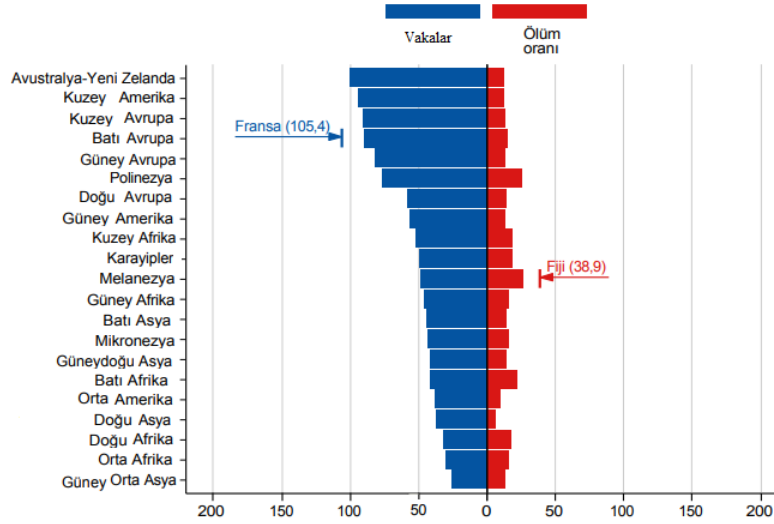
Şekil 2.5. Meme kanserinin kıtalara göre ölüm sayısı (Ferlay ve ark., 2021).

Yaygınlık

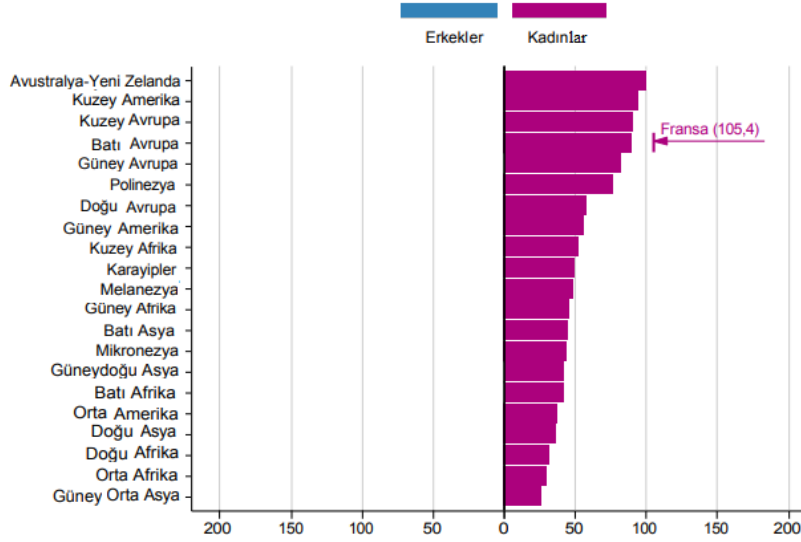


Kıta	Vaka Yaygınlığı	Yüzde
Asya	3 197 043	39.1%
Avrupa	2 296 495	28.1%
Kuzey Amerika	1 332 343	16.3%
Latin Amerika ve Karayipler	725 017	8.9%
Afrika	507 659	6.2%
Okyanusya	119 836	1.5%

Şekil 2.6. Meme kanserinin kıtalara göre vaka yaygınlığı (Ferlay ve ark., 2021).



Şekil 2.7. Meme kanserinin dünya genelinde bölgeler arası yaşa göre standardize edilmiş vaka ve ölüm oranları (Ferlay ve ark., 2021).



Şekil 2.8. Meme kanserinin dünya genelinde bölgelere arası cinsiyete göre standardize edilmiş vaka oranı (Ferlay ve ark., 2021).

2.2. Türleri

Meme kanserinin, memenin farklı bölgelerinde gelişebilen birkaç farklı türü bulunmaktadır (Breast cancer (female), 2024). Meme kanseri genellikle invaziv olmayan ve invaziv türler olarak ikiye ayrılmaktadır (Breast cancer (female), 2024). Bu türler aşağıdaki gibidir (Breast cancer (female), 2024).

2.2.1. İnvaziv olmayan meme kanseri

İnvaziv olmayan meme kanseri, kanser veya karsinoma in situ olarak da bilinir. Bu kanser, memenin kanallarında bulunur ve memenin dışına yayılma yeteneği geliştirmemiştir. Bu kanser türü, memede hissedilebilen bir kitle olarak nadiren görülür ve genellikle mamogramda bulunur. İnvaziv olmayan kanserin en yaygın türü duktal karsinoma in situ'dur (DCIS).

2.2.2. İnvaziv meme kanseri

İnvaziv kanserin memenin dışına yayılma yeteneği vardır, ancak bu mutlaka yayıldığı anlamına gelmez. Meme kanserinin en yaygın biçimi, meme kanallarını kaplayan hücrelerde gelişen invaziv duktal meme kanseridir. İnvaziv duktal meme kanseri, tüm meme kanseri vakalarının yaklaşık %80'ini oluşturur ve bazen “özel bir tür yok” olarak adlandırılır.

2.3. Belitiler

Meme kanserinin birçok belirtisi vardır. Ancak bunlar arasında en çok rastlanılan belirtiler aşağıdaki gibidir (Symptoms of Breast Cancer, 2024).

- Memede veya koltuk altında yeni kitle
- Memenin bir kısmının kalınlaşması veya şişmesi
- Meme derisinde tahriş veya çukurlaşma
- Meme ucunda veya göğüste kızarıklık veya pullanma
- Meme ucunun içeri çekilmesi veya meme ucunda ağrı
- Anne sütü dışındaki meme başından gelen akıntı (kan dahil)
- Memenin büyüklüğünde veya şeklinde herhangi bir değişiklik
- Memenin herhangi bir bölgesinde ağrı



Şekil 2.9. Meme kanseri belirtileri (Meme Kanseri Belirtileri ve İşaretleri, 2024).

Şekil 2.9’da belirtilen ve yukarıda anlatılmış bu semptomların kanser olmayan diğer rahatsızlıklarda da ortaya çıkabilme ihtimali de bulunmaktadır (Symptoms of Breast

Cancer, 2024). Bu semptomların çoğu iyi huylu (kanseri olmayan) meme rahatsızlıklarından da kaynaklanabilir (Breast Cancer Signs and Symptoms, 2024). Yine de, yeni meme kitlesi, yumru veya diğer değişikliklerin deneyimli bir sağlık uzmanı tarafından kontrol edilmesi önemlidir, böylece neden bulunabilir ve gerekirse tedavi edilebilir (Breast Cancer Signs and Symptoms, 2024).

2.4. Nedenler

Çoğu meme kanserinin kesin nedeni bilinmemektedir. Araştırmacılar ise bu konuda meme kanseri riskini artıran şeyler arasında hormonlar, yaşam tarzı seçimleri ve çevredeki şeyler yer aldığını belirtmektedir (Meme kanseri, 2024). Ayrıca meme kanserinin genetik yapınız ve etrafınızdaki dünyanın karmaşık bir etkileşimi yoluyla meydana gelmesi muhtemel olduğu söylenmektedir (Meme kanseri, 2024).

2.5. Risk Faktörleri

Meme kanseri riskini artırabilecek faktörler aşağıdaki gibidir (Breast Cancer Risk Factors You Can't Change, 2024).

- Ailede meme kanseri öyküsü: Ailede biri meme kanserine yakalanmışsa, meme kanseri riski artmaktadır. Ailede genç yaşta meme kanseri olma öyküsü varsa risk daha yüksektir. Ailede meme kanseri olan birden fazla kişi varsa risk daha yüksek olmaktadır. Yine de meme kanseri teşhisi konulan çoğu kişinin ailesinde hastalık öyküsü bulunmamaktadır.
- Kişisel meme kanseri geçmişi: Eğer bir göğüslerden birinde kanser olduysa, diğer göğüsde de kanser olma riski artmaktadır.
- Kişisel meme rahatsızlıkları geçmişi: Bazı meme rahatsızlıkları, meme kanseri riskinin daha yüksek olduğunun belirteçlerindendir.
- Adet döneminin daha genç yaşta başlaması: Adet dönemini 12 yaşından önce başladı ise meme kanseri riskinizi artmaktadır.
- İleri yaşta menopoza girmek: 55 yaşından sonra menopoza girmek meme kanseri riskini arttırmaktadır.
- Kadın olmak: Kadınların meme kanserine yakalanma olasılığı erkeklerden çok daha fazladır. Herkes bir miktar meme dokusuyla doğar, bu nedenle herkes meme kanserine yakalanabilir.

- Yoğun meme dokusu: Meme dokusu yağ dokusu ve yoğun dokudan oluşmaktadır. Yoğun doku süt bezlerinden, süt kanallarından ve lifli dokudan oluşmaktadır. Yoğun memelerde yağ dokusundan daha fazla yoğun doku bulunmaktadır. Yoğun memelere sahip olmak, mamogramda meme kanserini tespit etmeyi zorlaştırabilir. Mamogramda yoğun memelerin olduğu gösterilirse, meme kanseri riski artmaktadır.
- Alkol içmek: Alkol içmek meme kanseri riskini arttırmaktadır.
- İlk çocuğu ilerliyen yaşta doğurmak: İlk çocuğu 30 yaşından sonra doğurmak meme kanseri riskini artırabilir.
- Hiç hamile kalmamış olmak: Bir veya daha fazla kez hamile kalmış olmak meme kanseri riskini azaltmaktadır. Hiç hamile kalmamış olmak riski arttırmaktadır.
- Yaşın ilerlemesi: Meme kanseri riski yaş ilerledikçe artmaktadır.
- Kanser riskini artıran kalıtsal DNA değişiklikleri: Meme kanseri riskini artıran belirli DNA değişiklikleri ebeveynlerden çocuklara geçebilir. En bilinen değişiklikler BRCA1 ve BRCA2 olarak adlandırılır. Bu değişiklikler meme kanseri ve diğer kanserlere yakalanma riskinizi büyük ölçüde artırabilir, ancak bu DNA değişikliklerine sahip olan herkes kanser olmayabilir.
- Menopoz hormon tedavisi: Menopoz semptomlarını kontrol etmek için belirli hormon tedavisi ilaçları almak meme kanseri riskini artırabilir. Risk, östrojen ve progesteronu birleştiren hormon tedavisi ilaçlarıyla bağlantılıdır. Bu ilaçlar bırakıldığında risk azalmaktadır.
- Obezite: Obezitesi olan kişilerde meme kanseri riski daha yüksek olmaktadır.
- Radyasyon maruziyeti: Çocukken veya genç yetişkinde göğsünüze radyasyon tedavisi aldıysanız, meme kanseri riskiniz daha yüksek olmaktadır.

2.6. Teşhis

Meme kanserinin teşhisi aşağıdaki gibidir (Breast Cancer, 2024).

1. Kendi kendine meme muayenesi: Meme derisindeki anormallikleri veya memelerde ve koltuk altlarında oluşan kitleleri tespit etmeye yardımcı olabilir.

2. Mamografi: Tarama veya tanı testi olarak küçük, elle muayenede fark edilemeyen kitlelerin veya kanserin tespit edilmesine yardımcı olabilir.
3. Meme ultrasonu: Memede katı kitleler veya sıvı dolu kistlerin olup olmadığını tespit edebilir.
4. Biyopsi: Meme dokularınızın bir örneğini kanser hücreleri açısından analiz etmeye yardımcı olabilir. Kanserlerin türünü ve derecesini ve kanser hücrelerinin hormon reseptörlerinin durumunu belirleyerek uygun bir tedavi planlayabilir.
5. Meme MR'ı: Tümörün veya kanserin boyutunu, yerini ve yaygınlığını belirlemeye yardımcı olabilir.

2.7. Tedavi

Meme kanseri için genel olarak beş tedavi seçeneği vardır ve çoğu tedavi planı aşağıdakilerin bir kombinasyonunu içerir: ameliyat, radyasyon, hormon tedavisi, kemoterapi ve hedefe yönelik tedavilerdir (Breast Cancer Treatment Options, 2020). Bazıları lokaldir ve yalnızca tümörün etrafındaki alanı hedef alır (Breast Cancer Treatment Options, 2020). Diğerleri sistemiktir ve kanserle mücadele eden ajanlarla tüm vücudunuzu hedef alır (Breast Cancer Treatment Options, 2020). Tüm bu tedavi yöntemlerine rağmen kanser vücudun diğer bölgelerine yayılmışsa genellikle tedavi edilemez, ancak normalde uzun süre etkili bir şekilde kontrol edilebilir (Chaurasia ve ark., 2018).

3. MATERYAL VE YÖNTEM

3.1. Veri Setleri

Bu çalışmadaki veri setleri herkese açık olarak sunulan “www.archive.ics.uci.edu” ve www.kaggle.com internet sitelerinden elde edilmiştir (UCI Machine Learning Repository, 2024; Your Machine Learning and Data Science Community, 2024).

Çalışmada 6 farklı veri seti kullanılmıştır.

Çalışmada kullanılan 1. veri setinde 4024 farklı kişiden yaş, ırk, evlilik durumu, T evresi, N evresi, 6. evre, A evresi, tümör boyutu, östrojen durumu, progesteron durumu, incelenen bölgesel lenf düğümü, pozitif bölgesel lenf düğümü ve yaşama süresi (aylar) bilgileri alınmıştır. Bu bilgiler sonucunda 4024 farklı kişinin yaşama ve ölme durumları sınıflandırılmıştır.

Çalışmada kullanılan 2. veri setinde 286 farklı kişiden yaş, menopoz, tümör boyutu, invaziv-düğümeleri, düğüm-kapakları, kötü huylu derecesi, göğüs, meme-dörtlü ve ıslanma bilgileri alınmıştır. Bu bilgiler sonucunda 286 farklı kişinin meme kanserinin tekrarlayıp tekrarlamama durumları sınıflandırılmıştır.

Çalışmada kullanılan 3. veri setinde 116 farklı kişiden yaş, vücut kitle indeksi, glikoz, insülin, HOMA, leptin, adiponektin, direnç ve MCP.1 bilgileri alınmıştır. Bu bilgiler sonucunda 104 farklı kişinin meme kanseri konusunda hasta ya da sağlıklı olma durumları sınıflandırılmıştır.

Çalışmada kullanılan 4. veri setinde 198 farklı kişiden zaman, yarıçap1, doku1, çevre1, alan1, pürüzsüzlük1, sıklık1, içbükeylik1, içbükey noktalar1, simetri1, fraktal boyut1, yarıçap2, doku2, çevre2, alan2, pürüzsüzlük2, sıklık2, içbükeylik2, içbükey noktalar2, simetri2, fraktal boyut2, yarıçap3, doku3, çevre3, alan3, pürüzsüzlük3, sıklık3, içbükeylik3, içbükey noktalar3, simetri3, fraktal boyut3, tümör boyutu ve lenf bezi durumu bilgileri alınmıştır. Bu bilgiler sonucunda 197 farklı kişinin meme kanserinin tekrarlayıp tekrarlamama durumları sınıflandırılmıştır.

Çalışmada kullanılan 5. veri setinde 699 farklı kişiden küme kalınlığı, hücre boyutunun benzerliği, hücre şeklinin benzerliği, marjinal yapışkanlık, tek epitelyal

hücre boyutu, boş çekirdekler, yumuşak kromatin, normal nükleoller ve mitozlar bilgileri alınmıştır. Bu bilgiler sonucunda 699 farklı kişinin meme kanserinin iyi ya da kötü huylu olma durumları sınıflandırılmıştır.

Çalışmada kullanılan 6. veri setinde 569 farklı kişiden yarıçap1, doku1, çevre1, alan1, pürüzsüzlük1, sıklık1, içbükeylik1, içbükey noktalar1, simetri1, fraktal boyut1, yarıçap2, doku2, çevre2, alan2, pürüzsüzlük2, sıklık2, içbükeylik2, içbükey noktalar2, simetri2, fraktal boyut2, yarıçap3, doku3, çevre3, alan3, pürüzsüzlük3, sıklık3, içbükeylik3, içbükey noktalar3, simetri3 ve fraktal boyut3 bilgileri alınmıştır. Bu bilgiler sonucunda 424 farklı kişinin meme kanserinin tekrarlayıp tekrarlamama durumları sınıflandırılmıştır.

3.2. Veri Önleme

Verilerin analizler için hazır hale getirilmiştir. Literatürde veri ön işleme (data preprocessing) olarak bilinen veri hazırlama süreci Han ve Kamber tarafından oluşturulmuş başlıklar altında ayrıntılı bir biçimde ele alınmıştır (Kartal, 2015). Çalışmada kullanılan veri önleme adımları sırayla aşağıda belirtilmiştir.

3.2.1. Verilerin gruplandırılması

3.2.1.1. 1. veri seti (UMMC) için verilerin gruplandırılması

Ham veri setindeki 4024 kişinin sahip olduğu 14 özellik belirli matematiksel değerlerde gösterilmiştir. 14 özellik yaş, ırk, medeni durum, T evresi, N evresi, 6. evre, derece, A evresi, tümör boyutu, östrojen durumu, progesteron durumu, incelenen bölgesel lenf düğümü, pozitif bölgesel lenf düğümü ve yaşama süresi (aylar) şeklindedir. Bu özelliklerden yaş, tümör boyutu, incelenen bölgesel lenf düğümü, pozitif bölgesel lenf düğümü ve yaşama süresi (aylar) sayısal değerlere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Ancak bu özelliklerin dışındakiler sayısal veri olmadığı için her biri sayısal bir değere atanılmıştır. Bu atamalar Tablo 3.1 ve Tablo 3.2’de gösterilmiştir.

Tablo 3.1. 1. Veri Setindeki Nominal Özelliklerin Tanımları.

Özellik	Tanım	İfade
İrk	1=Beyaz (%85), 2=Siyah (%7), 3=Diğer (Amerikan Yerlisi/AK Yerlisi, Asyalı/Pasifik Adalı) (%8)	Bu, kişinin ırkını belirtir.
Medeni Durum	1=Bekar (hiç evlenmemiş) (%15), 2=Boşanmış (%12), 3=Ayrı (%1), 4=Evli (kanuni birliktelikler dahil) (%66), 5=Dul (%6)	Kişinin medeni durumunu belirtir.
T Evresi	1=T1 (%40), 2=T2 (%44), 3=T3 (%13), 4=T4 (%3)	TNM sisteminde "T" harfi ve ardından gelen harf veya sayı (0-4 arası), tümörün boyutunu ve yerini tanımlar. Tümör boyutu santimetre (cm) cinsinden ölçülür.
N Evresi	1=N1 (%68), 2=N2 (%20), 3=N3 (%12)	"N" harfi, lenf düğümlerini temsil eder. Kanserin başladığı bölgeye yakın lenf düğümleri "bölgesel" olarak adlandırılır.
6. Evre	1=IIA (%32), 2=IIB (%28), 3=IIIA (%26), 4=IIIB (%2), 5=IIIC (%12)	Cerrahiden sonra kanserin evresi genellikle 5-7 gün içinde belirlenir. Neoadjuvan tedavi uygulandığında evre, klinik olarak belirlenir. Evre I ve IIA "erken evre", IIB ile III ise "yerel ileri evre" olarak adlandırılabilir.

Tablo 3.1. (Devamı) 1. Veri Setindeki Nominal Özelliklerin Tanımları.

Özellik	Tanım	İfade
Derece	1=İyi diferansiye; Derece I (%13), 2=Orta derecede diferansiye; Derece II (%58), 3=Kötü diferansiye; Derece III (%28), 4=Farklılaşmamış; Anaplastik; Derece IV (%1)	Kanser hücrelerinin mikroskop altında nasıl görüldüğünü ve normal hücrelere ne kadar benzer olduğunu belirtir.
A Evresi	1=Bölgesel (%98), 2=Uzak (%2)	SEER veritabanı, kanserin ne kadar yayıldığına göre 5 yıllık sağkalım oranlarını takip eder. Veritabanı, kanserleri AJCC TNM evrelerine göre gruplandırmaz, bunun yerine "bölgesel" ve "uzak" evreler olarak sınıflandırır.
Östrojen Durumu	1=Pozitif (%93), 2=Negatif (%7)	Kanserde östrojen hormonunun varlığı veya yokluğu belirtilir.
Progesteron Durumu	1=Pozitif (%83), 2=Negatif (%17)	Kanserde progesteron hormonunun varlığı veya yokluğu belirtilir.
Sınıf	1=Yaşayan (%85), 2=Ölü (%15)	Kişinin hayatta olup olmadığını belirtir.

Tablo 3.2. 1. Veri Setindeki Sayısal Özelliklerin Tanımları.

Özellik	Tanım						İfade
	Min	25%	50%	75%	Max	Ortalama	
Yaş	39	47	54	61	69	54	Yaş
Tümör Boyutu	1	16	25	38	140	30,5	Tümör Boyutu
İncelenen Bölgesel Lenf Dügümü	1	9	14	19	61	14,4	İncelenen Bölgesel Lenf Dügümü
Pozitif Bölgesel Lenf Dügümü	1	1	2	5	46	4,16	Pozitif Bölgesel Lenf Dügümü
Yaşama Süresi (Aylar)	1	56	73	90	107	71,3	Yaşama Süresi (Aylar)

3.2.1.2. 2. veri seti (Breast Cancer) için verilerin gruplandırılması

Ham veri setindeki 286 kişinin sahip olduğu 9 özellik belirli matematiksel değerlerde gösterilmiştir. 9 özellik yaş, menopoz, tümör boyutu, invaziv-dügümleri, düğüm-kapakları, kötü huylu derecesi, göğüs, meme-dörtlü ve ışınlama şeklindedir. Bu özelliklerden yaş, menopoz, tümör boyutu, invaziv-dügümleri, meme-dörtlü ve sınıf kategorik, düğüm-kapakları, göğüs ve ışınlama ikili, kötü huylu derecesi sayısal değerlere sahiptir. Kategorik değerler sabit bir sayısal veri olmadığı için her biri sayısal bir değere atanılmıştır. İkili değerler sayısal veri olmadığı için her biri sayısal bir değere atanılmıştır. Ancak bu özelliklerin dışındakiler sayısal değerlere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Bu atamalar Tablo 3.3’de gösterilmiştir.

Tablo 3.3. 2. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanımı	İfade
Yaş	1=10-19, 2=20-29, 3=30-39, 4=40-49, 5=50-59, 6=60-69, 7=70-79, 8=80-89, 9=90-99 (20-49 (%))	Yaş (yıl olarak)
Menopoz	1=lt40, 2=Ge40, 3=Premeno	Tanı anında hastanın menopoz durumu
Tümör Boyutu	1=0-4, 2=5-9, 3=10-14, 4=15-19, 5=20-24, 6=25-29, 7=30-34, 8=35-39, 9=40-44, 10=45-49, 11=50-54, 12=55-59	Tümör boyutu (mm)
İnvaziv-Düğümleri	1=0-2, 2=3-5, 3=6-8, 4=9-11, 5=12-14, 6=15-17, 7=18-20, 8=21-23, 9=24-26, 10=27-29, 11=30-32, 12=33-35, 13=36-39	İnvaziv nodlar (histolojik inceleme sırasında BC gösteren lenf nodlarının sayısı olarak raporlanmıştır)
Düğüm-Kapakları	1=Evet, 2=Hayır	Tümörün lenf düğümü kapsülüne penetrasyonu
Kötü Huylu Derecesi	1, 2, 3	Tümörün histolojik derecesi (burada; Derece 1: normal meme hücrelerine en çok benzeyen ve genellikle yavaş büyüyen, Derece 2: normal hücrelere daha az benzeyen ve daha hızlı büyüyen ve Derece 3: normal meme hücrelerinden farklı görünen ve genellikle hızlı büyüyen)
Göğüs Meme-Dörtlü Işınlama	1=Sol, 2=Sağ 1=Sol-Yukarı, 2=Sol-Aşağı, 3=Sağ-Yukarı, 4=Sağ-Aşağı, 5=Merkez 1=Evet, 2=Hayır	Meme kanserinden memenin konumu Meme kanserinden etkilenen memenin spesifik konumu Işınlama (hastanın radyasyon tedavisi geçmişi)
Sınıf	1=Tekrarlanmayan Olaylar, 2=Tekrarlanan Olaylar	Hastalığın nüks olup olmadığının sınıflandırılması (bağımlı değişken sınıftır)

3.2.1.3. 3. veri seti (Coimbra) için verilerin gruplandırılması

Ham veri setindeki 116 kişinin sahip olduğu 9 özellik belirli matematiksel değerlerle gösterilmiştir. 9 özellik yaş, vücut kitle indeksi, glikoz, insülin, HOMA, leptin, adiponektin, direnç ve MCP.1 şeklindedir. Bu özelliklerden yaş, vücut kitle indeksi, insülin, HOMA, leptin, adiponektin, direnç ve MCP.1 rasyonel, glikoz tam sayı ve sınıf kategorik değerlere sahiptir. Kategorik değer sabit bir sayısal veri olmadığı için her biri sayısal bir değerlere atanılmıştır. Ancak bu özelliklerin dışındakiler sayısal değerlere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Bu atamalar Tablo 3.4'de gösterilmiştir.

Tablo 3.4. 3. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanım	İfade
Yaş	24 - 89	Yaş (yıl olarak)
Vücut indeksi	kitle 18,37 - 38,58	Vücut kitle endeks değeri (kg/m ²)
Glikoz	60 - 201	Glikoz değeri (mg/dL)
İnsülin	2,43 - 58,46	İnsülin değeri (μU/mL)
HOMA	0,47 - 25,05	İnsülin direnci testi değeri
Leptin	4,31 - 90,28	Tokluk hormon değeri (ng/mL)
Adiponektin	1,66 - 38,04	İnsülin direncine karşı çalışan, insülin hassasiyetini artıran; yağ hücreleri tarafından salgılanan hormon değeri (μU/mL)
Direnç	3,21 - 82,10	İnsülin direnci (ng/mL)
MCP.1	45,84 - 1698,44	Serum monosit kemoatraktan protein-1 değeri (pg/dL)
Sınıflandırma	1=Sağlıklılar, 2=Hastalar	Kişilerin sağlıklı ya da hasta olup olmadığının sınıflandırması

3.2.1.4. 4. veri seti (WPBC) için verilerin gruplandırılması

Ham veri setindeki 198 kişinin sahip olduğu 33 özellik belirli matematiksel değerlerle gösterilmiştir. 33 özellik zaman, yarıçap1, doku1, çevre1, alan1, pürüzsüzlük1, sıklık1, içbükeylik1, içbükey noktalar1, simetri1, fraktal boyut1, yarıçap2, doku2, çevre2, alan2, pürüzsüzlük2, sıklık2, içbükeylik2, içbükey noktalar2, simetri2, fraktal boyut2, yarıçap3, doku3, çevre3, alan3, pürüzsüzlük3, sıklık3, içbükeylik3, içbükey noktalar3, simetri3, fraktal boyut3, tümör boyutu ve lenf bezi durumu şeklindedir. Bu özelliklerden zaman tam sayı, sınıf kategorik ve diğer özellikler rasyonel sayı değerlerine sahiptir. Kategorik değer sabit bir sayısal veri olmadığı için her biri sayısal

bir değere atanılmıştır. Ancak bu özelliklerin dışındakiler sayısal değere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Bu atamalar Tablo 3.5'de gösterilmiştir.

Tablo 3.5. 4. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanım	İfade
Zaman	1 - 125	Zaman (Alan 2 = R İse Tekrarlama Zamanı, Alan 2 = N İse Hastalıksız Zaman)
Yarıçap1	10,95 - 27,22	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)
Doku1	10,38 - 39,28	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre1	71,9 - 182,1	Çevre
Alan1	361,6 - 2250	Alan
Pürüzsüzlük1	0,07497 0,1447	- Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık1	0,04605 0,3114	- Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik1	0,02398 0,4268	- İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)
İçbükey Noktalar1	0,02031 0,2012	- İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri1	0,1308 - 0,304	Simetri
Fraktal Boyut1	0,05025 0,09744	- Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Yarıçap2	0,1938 - 1,819	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)
Doku2	0,3621 - 3,503	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre2	1,153 - 13,28	Çevre
Alan2	13,99 - 316	Alan
Pürüzsüzlük2	0,002667 0,03113	- Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık2	0,007347 0,1354	- Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik2	0,01094 0,1438	- İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)
İçbükey Noktalar2	0,005174 0,03927	- İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri2	0,007882 0,06041	- Simetri
Fraktal Boyut2	0,001087 0,01256	- Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Yarıçap3	12,84 - 35,13	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)

Tablo 3.5. (Devamı) 4. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanım	İfade
Doku3	16,67 - 49,54	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre3	85,1 - 232,2	Çevre
Alan3	508,1 - 3903	Alan
Pürüzsüzlük3	0,08191 - 0,2226	Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık3	0,05131 - 1,058	Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik3	0,02398 - 1,17	İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)
İçbükey Noktalar3	0,02899 - 0,2903	İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri3	0,1565 - 0,6638	Simetri
Fraktal Boyut3	0,05504 - 0,2075	Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Tümör Boyutu	0,4 - 10	Tümör Boyutu - Çıkarılan Tümörün Santimetre Cinsinden Çapı
Lenf Bezi Durumu	0 - 27	Lenf Nodu Durumu - Ameliyat Sırasında Gözlenen Pozitif Aksiller Lenf Nodu Sayısı
Sınıf	1=Tekrarlamayan, 2=Tekrarlayan	Kişilerin hastalığının tekrarlayıp tekrarlamama durumu

3.2.1.5. 5. veri seti (WBC) için verilerin gruplandırılması

Ham veri setindeki 699 kişinin sahip olduğu 9 özellik belirli matematiksel değerlerle gösterilmiştir. 9 özellik küme kalınlığı, hücre boyutunun benzerliği, hücre şeklinin benzerliği, marjinal yapışkanlık, tek epitelyal hücre boyutu, boş çekirdekler, yumuşak kromatin, normal nükleoller ve mitozlar şeklindedir. Bu özelliklerden sadece sınıf kategorik olup diğer özellikler tam sayı değerlere sahiptir. Kategorik değer sabit bir sayısal veri olmadığı için her biri sayısal bir değere atanmıştır. Ancak bu özelliklerin dışındakiler sayısal değere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Bu atamalar Tablo 3.6'da gösterilmiştir.

Tablo 3.6. 5. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanımı	İfade
Küme Kalınlığı	1 - 10	Küme Kalınlığı Değeri
Hücre Boyutunun Benzerliği	1 - 10	Hücre Boyutunun Benzerlik Değeri
Hücre Şeklinin Benzerliği	1 - 10	Hücre Şeklinin Benzerlik değeri
Marjinal Yapışkanlık	1 - 10	Marjinal Yapışma Değeri
Tek Epitelyal Hücre Boyutu	1 - 10	Tek Epitel Hücre Boyutu Değeri
Boş Çekirdekler	1 - 10	Boş Çekirdeklerin Değeri

Tablo 3.6. (Devamı) 5. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanımı	İfade
Yumuşak Kromatin	1 - 10	Yumuşak Kromatin Değeri
Normal Nükleoller	1 - 10	Normal Nükleollerin Değeri
Mitozlar	1 - 10	Her biri aynı sayı ve türde kromozomlara sahip iki kardeş hücre oluşması ile sonuçlanan hücre bölünmesi değeri
Sınıf	1 = İyi Huylu, 2 = Kötü Huylu	Meme kanserinin iyi huylu ya da kötü huylu olma durumu

3.2.1.6. 6. veri seti (WDBC) için verilerin gruplandırılması

Ham veri setindeki 569 kişinin sahip olduğu 30 özellik belirli matematiksel değerlerle gösterilmiştir. 30 özellik yarıçap1, doku1, çevre1, alan1, pürüzsüzlük1, sıklık1, içbükeylik1, içbükey noktalar1, simetri1, fraktal boyut1, yarıçap2, doku2, çevre2, alan2, pürüzsüzlük2, sıklık2, içbükeylik2, içbükey noktalar2, simetri2, fraktal boyut2, yarıçap3, doku3, çevre3, alan3, pürüzsüzlük3, sıklık3, içbükeylik3, içbükey noktalar3, simetri3 ve fraktal boyut3 şeklindedir. Bu özelliklerden sınıf kategorik olup diğer özellikler rasyonel sayı değerlerine sahiptir. Kategorik değer sabit bir sayısal veri olmadığı için her biri sayısal bir değerlere atanılmıştır. Ancak bu özelliklerin dışındakiler sayısal değerlere sahiptir ve bu değerler veri setinde herhangi bir değere atanılmamıştır. Bu atamalar Tablo 3.7'de gösterilmiştir.

Tablo 3.7. 6. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanım	İfade
Yarıçap1	10,95 - 27,22	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)
Doku1	10,38 - 39,28	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre1	71,9 - 182,1	Çevre
Alan1	361,6 - 2250	Alan
Pürüzsüzlük1	0,07497 0,1447	- Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık1	0,04605 0,3114	- Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik1	0,02398 0,4268	- İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)

Tablo 3.7. (Devamı) 6. Veri Setindeki Özelliklerin Tanımı.

Özellik	Tanım	İfade
İçbükey Noktalar1	0,02031 0,2012	- İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri1	0,1308 - 0,304	Simetri
Fraktal Boyut1	0,05025 0,09744	- Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Yarıçap2	0,1938 - 1,819	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)
Doku2	0,3621 - 3,503	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre2	1,153 - 13,28	Çevre
Alan2	13,99 - 316	Alan
Pürüzsüzlük2	0,002667 0,03113	- Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık2	0,007347 0,1354	- Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik2	0,01094 0,1438	- İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)
İçbükey Noktalar2	0,005174 0,03927	- İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri2	0,007882 0,06041	- Simetri
Fraktal Boyut2	0,001087 0,01256	- Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Yarıçap3	12,84 - 35,13	Yarıçap (Merkezden Çevredeki Noktalara Olan Mesafelerin Ortalaması)
Doku3	16,67 - 49,54	Doku (Gri Tonlama Değerlerinin Standart Sapması)
Çevre3	85,1 - 232,2	Çevre
Alan3	508,1 - 3903	Alan
Pürüzsüzlük3	0,08191 0,2226	- Pürüzsüzlük (Yarıçap Uzunluklarındaki Yerel Varyasyon)
Sıklık3	0,05131 1,058	- Sıklık ($\text{Çevre}^2 / \text{Alan} - 1,0$)
İçbükeylik3	0,02398 - 1,17	İçbükeylik (Konturun İçbükey Kısımlarının Şiddeti)
İçbükey Noktalar3	0,02899 0,2903	- İçbükey Noktalar (Konturun İçbükey Kısımlarının Sayısı)
Simetri3	0,1565 0,6638	- Simetri
Fraktal Boyut3	0,05504 0,2075	- Fraktal Boyut ("Sahil Şeridi Yaklaşımı" - 1)
Sınıf	1 = İyi Huylu, 2 = Kötü Huylu	Meme kanserinin iyi huylu ya da kötü huylu olma durumu

3.2.2. Veri dengeleme

3.2.2.1. 1. veri seti (UMMC) için veri dengeleme

Veri setindeki 4024 kişinin sahip olduğu 14 özelliğe göre sonuçlanan sınıflamalar "Sınıf" adı altında "Yaşayan" ve "Ölü" şeklinde tanımlanmıştır. Ancak Yaşayan kişi sayısı ile Ölü kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmalarından sonuçları daha kesin almak amacıyla veri seti veri dengeleme işleminden geçirilmiştir. Sonuç olarak Yaşayan kişi sayısı veri setinde sıralı olarak 6 parçaya bölünmüştür. Her bir parçaya Ölü kişi sayıları eklenmiştir. Böylelikle Yaşayan 568 kişiden oluşan 6 grup, Ölü 616 kişiden oluşan 1 grup şeklinde her Yaşayan gruba Ölü grubu eklenerek toplamda 1184 kişilik 6 farklı alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.8'de gösterilmiştir.

Tablo 3.8. 1. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Total
	Yaşayan	Ölü	
A Veri Seti	568	616	1184
B Veri Seti	568	616	1184
C Veri Seti	568	616	1184
D Veri Seti	568	616	1184
E Veri Seti	568	616	1184
F Veri Seti	568	616	1184

3.2.2.2. 2. veri seti (Breast Cancer) için veri dengeleme

Veri setindeki 286 kişinin sahip olduğu 9 özelliğe göre sonuçlanan sınıflamalar "Sınıf" adı altında "Tekrarlanmayan Olaylar" ve "Tekrarlayan Olaylar" şeklinde tanımlanmıştır. Ancak Tekrarlanmayan Olaylar kişi sayısı ile Tekrarlayan Olaylar kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmaların sonuçları daha kesin almak amacıyla veri seti veri dengeleme işleminden geçirilmiştir. Sonuç olarak Tekrarlayan Olaylar kişi sayısı veri setinde sıralı olarak 3 parçaya bölünmüştür. Her bir parçaya Tekrarlanmayan Olaylar kişi sayıları eklenmiştir. Böylelikle Tekrarlayan Olaylar 67 kişiden oluşan 3 grup, Tekrarlanmayan Olaylar 85 kişiden oluşan 1 grup şeklinde her Tekrarlayan Olaylar grubuna Tekrarlanmayan Olaylar grubu eklenerek toplamda 152 kişilik 3 farklı alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.9'da gösterilmiştir.

Tablo 3.9. 2. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Toplam
	Tekrarlayan Olaylar	Tekrarlanmayan Olaylar	
A Veri Seti	67	85	152
B Veri Seti	67	85	152
C Veri Seti	67	85	152

3.2.2.3. 3. veri seti (Coimbra) için veri dengeleme

Veri setindeki 116 kişinin sahip olduğu 9 özelliğe göre sonuçlanan sınıflamalar “Sınıf” adı altında “Sağlıklılar” ve “Hastalar” şeklinde tanımlanmıştır. Ancak Sağlıklılar kişi sayısı ile Hastalar kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmaların sonuçları daha kesin almak amacıyla veri seti veri dengeleme işleminden geçirilmiştir. Sonuç olarak Hastalar kişi sayısı veri setindeki 64 kişi arasından 52 kişi seçilmiştir. Oluşturulan 52 kişilik bu yeni parçaya Sağlıklılar kişi sayıları eklenmiştir. Böylelikle Hastalar 52 kişiden oluşan 1 grup, Sağlıklılar 52 kişiden oluşan 1 grup şeklinde Hastalar grubuna Sağlıklılar grubu eklenerek toplamda 104 kişilik bir alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.10’da gösterilmiştir.

Tablo 3.10. 3. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Toplam
	Sağlıklılar	Hastalar	
Veri Seti	52	52	104

3.2.2.4. 4. veri seti (WPBC) için veri dengeleme

Veri setindeki 198 kişinin sahip olduğu 33 özelliğe göre sonuçlanan sınıflamalar “Sınıf” adı altında “Tekrarlanmayan Olaylar” ve “Tekrarlayan Olaylar” şeklinde tanımlanmıştır. Ancak Tekrarlanmayan Olaylar kişi sayısı ile Tekrarlayan Olaylar kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmaların sonuçları daha kesin almak amacıyla veri seti veri dengeleme işleminden geçirilmiştir. Sonuç olarak Tekrarlanmayan Olaylar kişi sayısı veri setinde sıralı olarak 3 parçaya bölünmüştür. Her bir parçaya Tekrarlayan Olaylar kişi sayıları eklenmiştir. Böylelikle

Tekrarlamayan Olaylar 50 kişiden oluşan 3 grup, Tekrarlayan Olaylar 47 kişiden oluşan 1 grup şeklinde her Tekrarlamayan Olaylar grubuna Tekrarlayan Olaylar grubu eklenerek toplamda 97 kişilik 3 farklı alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.11’de gösterilmiştir.

Tablo 3.11. 4. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Toplam
	Tekrarlanmayan Olaylar	Tekrarlayan Olaylar	
A Veri Seti	50	47	97
B Veri Seti	50	47	97
C Veri Seti	50	47	97

3.2.2.5. 5. veri seti (WBC) için veri dengeleme

Veri setindeki 699 kişinin sahip olduğu 9 özelliğe göre sonuçlanan sınıflamalar "Sınıf" adı altında "İyi Huylu" ve "Kötü Huylu" şeklinde tanımlanmıştır. Ancak İyi Huylu kişi sayısı ile Kötü Huylu kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmalarından sonuçları daha kesin almak amacıyla veri seti veri dengeleme işlenminden geçirilmiştir. Sonuç olarak İyi Huylu kişi sayısı veri setinde sıralı olarak 2 parçaya bölünmüştür. Her bir parçaya Kötü Huylu kişi sayıları eklenmiştir. Böylelikle İyi Huylu 229 kişiden oluşan 2 grup, Kötü Huylu 241 kişiden oluşan 1 grup şeklinde her İyi Huylu gruba Kötü Huylu grubu eklenerek toplamda 470 kişilik 2 farklı alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.12’de gösterilmiştir.

Tablo 3.12. 5. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Toplam
	İyi Huylu	Kötü Huylu	
A Veri Seti	229	241	470
B Veri Seti	229	241	470

3.2.2.6. 6. veri seti (WDBC) için veri dengeleme

Veri setindeki 569 kişinin sahip olduğu 30 özelliğe göre sonuçlanan sınıflamalar “Sınıf” adı altında “İyi Huylu” ve “Kötü Huylu” şeklinde tanımlanmıştır. “Ancak İyi Huylu kişi sayısı ile Kötü Huylu kişi sayısı arasında çok fark olduğundan yapay zeka tabanlı algoritmaların sonuçları daha kesin almak amacıyla veri seti veri dengeleme işleminden geçirilmiştir. Sonuç olarak İyi Huylu kişi sayısı veri setindeki 357 kişi arasından 212 kişi seçilmiştir. Oluşturulan 212 kişilik bu yeni parçaya Kötü Huylu kişi sayıları eklenmiştir. Böylelikle İyi Huylu 212 kişiden oluşan 1 grup, Kötü Huylu 212 kişiden oluşan 1 grup şeklinde İyi Huylu grubuna Kötü Huylu grubu eklenerek toplamda 424 kişilik bir alt veri seti oluşturulmuştur. Bu veri setinin dağılımı Tablo 3.13’de gösterilmiştir.

Tablo 3.13. 6. Veri Setindeki Özelliklerin Dengelenmesi.

Veri Seti	Sınıf		Toplam
	İyi Huylu	Kötü Huylu	
Veri Seti	212	212	424

3.3. Özellik Çıkarımı

Elimizdeki ham veri setindeki bilgileri muhafaza ederken işlenebilecek sayısal değerlere çevirme süreci özellik çıkarımı olarak isimlendirilir. Yapay zeka algoritmalarını doğrudan veri seti ile işlem yapmak yerine daha iyi modeller oluşturabilmek için özellik çıkarımından yararlanılır. Sinyal ve görüntülerin herhangi bir özelliğe sahip olmaması elde edilecek verilerin daha karmaşık olacağını ortaya çıkartmaktadır. Bu tarz verilerde kullanan kişinin tercihinine bağlı farklı özellik çıkarımları uygulanmaktadır. Bu uygulamada EKG sinyallerini daha anlaşılır yapmak ve tahmin değerlerini yüksek doğrulağa sahip olmasını sağlamak için 25 farklı parametre özellik olarak kullanılmıştır. Tüm bu parametreler Tablo 3.14’de gösterilmiştir.

Tablo 3.14. Özellikler ait denklemler.

Sayı	Özellik	Denklem
1	Basıklık	$x_{kur} = \frac{\sum_{i=1}^n (x(i) - \bar{x})^4}{(n-1)S^4}$
2	Çarpıklık	$x_{ske} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)S^3}$
3	* Çeyrekler Arası Açıklık	$IQR = iqr(x)$
4	Varyans Katsayısı	$DK = (S/\bar{x})100$
5	Geometik Ortalama	$G = \sqrt[n]{x_1 + \dots + x_n}$
6	Harmonik ortalama	$H = n / (\frac{1}{x_1} + \dots + \frac{1}{x_n})$
7	Aktivite – Hjort Parametresi	$A = S^2$
8	Mobilite - Hjort Parametresi	$M = S_1^2 / S^2$
9	Karmaşıklık – Hjort Parametresi	$C = \sqrt{(S_2^2 / S_1^2)^2 - (S_1^2 / S^2)^2}$
10	* Maksimum	$x_{max} = \max(x_i)$
11	Medyan	$\tilde{x} = \begin{cases} \frac{x_{n+1}}{2} & : x \text{ odd} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & : x \text{ even} \end{cases}$
12	* Ortalama Mutlak Sapma	$MAD = mad(x)$
13	* Minimum	$x_{min} = \min(x_i)$
14	* Merkezi Momentler	$CM = moment(x, 10)$

Tablo 3.14. (Devamı) Özellikler ait denklemler.

Sayı	Özellik	Denklem
15	Ortalama	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$
16	Ortalama Eğri Uzunluğu	$CL = \frac{1}{n} \sum_{i=2}^n x_i - x_{i-1} $
17	Ortalama Enerji	$E = \frac{1}{n} \sum_{i=1}^n x_i^2$
18	Ortalama Karakök Değeri	$X_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i ^2}$
19	Standart Hata	$S_{\bar{x}} = S/\sqrt{n}$
20	Standart Sapma	$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
21	Şekil Faktörü	$SF = X_{rms} / (\frac{1}{n} \sum_{i=1}^n \sqrt{ x_i })$
22	* Tekil değer ayrışımı	$SVD = \text{svd}(x)$
23	* %25 kırılmış ortalama	$T25 = \text{trimmean}(x, 25)$
24	* %50 kırılmış ortalama	$T50 = \text{trimmean}(x, 50)$
25	Ortalama Teager enerjisi	$TE = \frac{1}{n} \sum_{i=3}^n (x_{i-1}^2 - x_i x_{i-2})$

* simgesi Matlab ile hesaplanan ifadeleri belirtir

S^2 : x sinyalinin varyansı

S_1^2 : x sinyalinin 1.türevinin varyansı

S_2^2 : x sinyalinin 2.türevinin varyansı

3.3.1. 1. veri seti (UMMC) için özellik çıkarımı

4024 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.15'den yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.15. 1. Veri Setindeki Özelliklerin Çıkarımı.

Numara	Özellik	Birim	Veri Tipi
1	Yaş	30 – 69	Sayısal
2	İrk	1, 2, 3	Nominal
3	Medeni Durum	1, 2, 3, 4, 5	Nominal
4	T Evresi	1, 2, 3, 4	Nominal
5	N Evresi	1, 2, 3	Nominal
6	6. Evre	1, 2, 3, 4, 5	Nominal
7	Derece	1, 2, 3, 4	Nominal
8	A Evresi	1, 2	Nominal
9	Tümör Boyutu	1 – 140	Sayısal
10	Östrojen Durumu	1, 2	Nominal
11	Progesteron Durumu	1, 2	Nominal
12	İncelenen Bölgesel Lenf Düğümü	1 – 61	Sayısal
13	Pozitif Bölgesel Lenf Düğümü	1 – 46	Sayısal
14	Yaşama Süresi (Aylar)	1 – 107	Sayısal
15	Sınıf	1, 2	Nominal

3.3.2. 2. veri seti (Breast Cancer) için özellik çıkarımı

286 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.16'dan yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.16. 2. Veri Setindeki Özelliklerin Çıkarımı.

Sa y1	Özellik	Birimi	Veri Tipi
1	Yaş	10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99	Kategorik
2	Menopoz	lt40, Ge40, Premeno	Kategorik
3	Tümör Boyutu	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59	Kategorik
4	İnvaziv-Düğümleri	0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39	Kategorik
5	Düğüm-Kapakları	Evet, Hayır	İkili

Tablo 3.16. (Devamı) 2. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
6	Kötü Derecesi	1, 2, 3	Tam sayı
7	Göğüs	Sol, Sağ	İkili
8	Meme-Dörtlü	Sol-Yukarı, Sol-Aşağı, Sağ-Yukarı, Sağ-Aşağı, Merkez	Kategorik
9	Işınlama	Evet, Hayır	İkili
10	Sınıf	1, 2	Kategorik

3.3.3. 3. veri seti (Coimbra) için özellik çıkarımı

116 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.17'den yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.17. 3. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
1	Yaş	24 - 89	Rasyonel
2	Vücut Kitle İndeksi	18,37 - 38,58	Rasyonel
3	Glikoz	60 - 201	Tam sayı
4	İnsülin	2,43 - 58,46	Rasyonel
5	HOMA	0,47 - 25,05	Rasyonel
6	Leptin	4,31 - 90,28	Rasyonel
7	Adiponektin	1,66 - 38,04	Rasyonel
8	Direnç	3,21 - 82,10	Rasyonel
9	MCP.1	45,84 - 1698,44	Rasyonel
10	Sınıf	1, 2	Kategorik

3.3.4. 4. veri seti (WPBC) için özellik çıkarımı

198 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.18'den yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.18. 4. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
1	Zaman	1 - 125	Tam Sayı
2	Yarıçap1	10,95 - 27,22	Rasyonel

Tablo 3.18. (Devamı) 4. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
3	Doku1	10,38 - 39,28	Rasyonel
4	Çevre1	71,9 - 182,1	Rasyonel
5	Alan1	361,6 - 2250	Rasyonel
6	Pürüzsüzlük1	0,07497 - 0,1447	Rasyonel
7	Sıklılık1	0,04605 - 0,3114	Rasyonel
8	İçbükeylik1	0,02398 - 0,4268	Rasyonel
9	İçbükey Noktalar1	0,02031 - 0,2012	Rasyonel
10	Simetri1	0,1308 - 0,304	Rasyonel
11	Fraktal Boyut1	0,05025 - 0,09744	Rasyonel
12	Yarıçap2	0,1938 - 1,819	Rasyonel
13	Doku2	0,3621 - 3,503	Rasyonel
14	Çevre2	1,153 - 13,28	Rasyonel
15	Alan2	13,99 - 316	Rasyonel
16	Pürüzsüzlük2	0,002667 - 0,03113	Rasyonel
17	Sıklılık2	0,007347 - 0,1354	Rasyonel
18	İçbükeylik2	0,01094 - 0,1438	Rasyonel
19	İçbükey Noktalar2	0,005174 - 0,03927	Rasyonel
20	Simetri2	0,007882 - 0,06041	Rasyonel
21	Fraktal Boyut2	0,001087 - 0,01256	Rasyonel
22	Yarıçap3	12,84 - 35,13	Rasyonel
23	Doku3	16,67 - 49,54	Rasyonel
24	Çevre3	85,1 - 232,2	Rasyonel
25	Alan3	508,1 - 3903	Rasyonel
26	Pürüzsüzlük3	0,08191 - 0,2226	Rasyonel
27	Sıklılık3	0,05131 - 1,058	Rasyonel
28	İçbükeylik3	0,02398 - 1,17	Rasyonel
29	İçbükey Noktalar3	0,02899 - 0,2903	Rasyonel
30	Simetri3	0,1565 - 0,6638	Rasyonel
31	Fraktal Boyut3	0,05504 - 0,2075	Rasyonel
32	Tümör Boyutu	0,4 - 10	Rasyonel
33	Lenf Bezi Durumu	0 - 27	Rasyonel
34	Sınıf	1, 2	Kategorik

3.3.5. 5. veri seti (WBC) için özellik çıkarımı

699 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.19'dan yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.19. 5. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
1	Küme Kalınlığı	1 - 10	Tam sayı
2	Hücre Boyutunun Benzerliği	1 - 10	Tam sayı
3	Hücre Şeklinin Benzerliği	1 - 10	Tam sayı
4	Marjinal Yapışkanlık	1 - 10	Tam sayı
5	Tek Epitelyal Hücre Boyutu	1 - 10	Tam sayı
6	Boş Çekirdekler	1 - 10	Tam sayı
7	Yumuşak Kromatin	1 - 10	Tam sayı
8	Normal Nükleoller	1 - 10	Tam sayı
9	Mitozlar	1 - 10	Tam sayı
10	Sınıf	1, 2	Kategorik

3.3.6. 6. veri seti (WDBC) için özellik çıkarımı

569 farklı kişiye ait veri seti özellik formüllerine göre hesaplanarak özellik çıkarımı yapılmıştır. Özellik çıkarımı için formül hesaplamaları bulunan Tablo 3.20'den yararlanılmıştır. Sınıf etiketleri genel olarak sırasız nitel değişkenlerdir.

Tablo 3.20. 6. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
1	Yarıçap1	10,95 - 27,22	Rasyonel
2	Doku1	10,38 - 39,28	Rasyonel
3	Çevre1	71,9 - 182,1	Rasyonel
4	Alan1	361,6 - 2250	Rasyonel
5	Pürüzsüzlük1	0,07497 - 0,1447	Rasyonel
6	Sıklık1	0,04605 - 0,3114	Rasyonel
7	İçbükeylik1	0,02398 - 0,4268	Rasyonel
8	İçbükey Noktalar1	0,02031 - 0,2012	Rasyonel
9	Simetri1	0,1308 - 0,304	Rasyonel
10	Fraktal Boyut1	0,05025 - 0,09744	Rasyonel
11	Yarıçap2	0,1938 - 1,819	Rasyonel
12	Doku2	0,3621 - 3,503	Rasyonel
13	Çevre2	1,153 - 13,28	Rasyonel
14	Alan2	13,99 - 316	Rasyonel
15	Pürüzsüzlük2	0,002667 - 0,03113	Rasyonel
16	Sıklık2	0,007347 - 0,1354	Rasyonel
17	İçbükeylik2	0,01094 - 0,1438	Rasyonel

Tablo 3.20. (Devamı) 6. Veri Setindeki Özelliklerin Çıkarımı.

Sayı	Özellik	Birimi	Veri Tipi
18	İçbükey Noktalar2	0,005174 - 0,03927	Rasyonel
19	Simetri2	0,007882 - 0,06041	Rasyonel
20	Fraktal Boyut2	0,001087 - 0,01256	Rasyonel
21	Yarıçap3	12,84 - 35,13	Rasyonel
22	Doku3	16,67 - 49,54	Rasyonel
23	Çevre3	85,1 - 232,2	Rasyonel
24	Alan3	508,1 - 3903	Rasyonel
25	Pürüzsüzlük3	0,08191 - 0,2226	Rasyonel
26	Sıklık3	0,05131 - 1,058	Rasyonel
27	İçbükeylik3	0,02398 - 1,17	Rasyonel
28	İçbükey Noktalar3	0,02899 - 0,2903	Rasyonel
29	Simetri3	0,1565 - 0,6638	Rasyonel
30	Fraktal Boyut3	0,05504 - 0,2075	Rasyonel
31	Sınıf	1, 2	Kategorik

3.4. Özellik Seçme Algoritması

Litaratürde birçok korelasyon hesaplama yöntemleri mevcuttur (Alpar, 2010). Ancak her veri grubu kendine uygun korelasyon formülüne ihtiyaç duyar (Uçar, 2019).

Makine öğrenimi alanında farklı türde veriler bulunmaktadır. Sınıf etiketleri genellikle sırasız nitel değişkenlerdir. (3.1)'de de görülebileceği üzere nitel ve sürekli sayısal değişkenler arasındaki korelasyon katsayısı hesaplanırken Eta Korelasyon Katsayısı (r_{pb}) kullanılır (Alpar, 2010). Veri türü değiştiğinde kullanılan yöntem de değişiklik göstermektedir (Alpar, 2010).

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{S_y} \sqrt{P_0 P_1} \quad (3.1)$$

(3.2), $\bar{Y}_1$ ve $\bar{Y}_0$ sırasıyla sınıf 0 ve sınıf 1'deki verilerin ortalamasını ifade etmektedir. S_y ise her iki sınıftaki tüm verilerin standart sapmasını temsil etmektedir.

$$s_y = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n}} \quad (3.2)$$

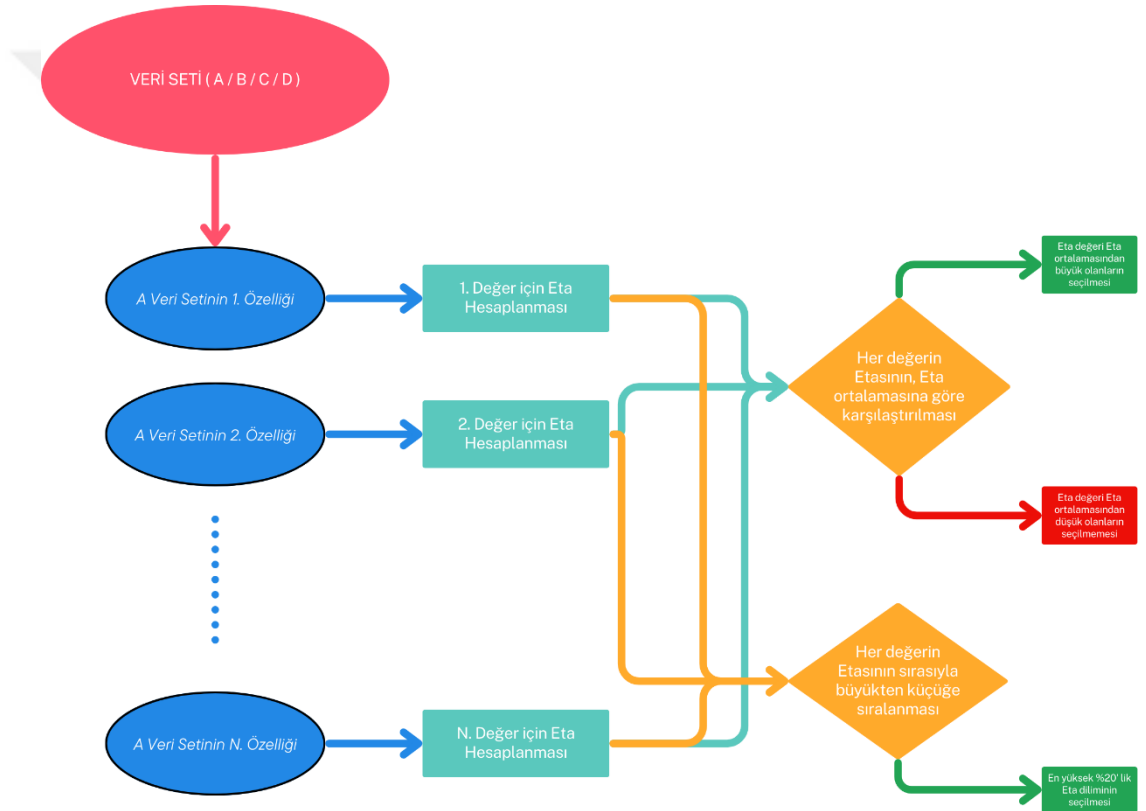
N , N_0 ve N_1 sırasıyla toplam veri setindeki, Sınıf 0'daki ve Sınıf 1'deki eleman sayılarını ifade etmektedir. (3.3)'de p_0 ve p_1 gösterilmektedir.

$$p_0 = \frac{N_0}{N_1}, p_1 = \frac{N_1}{N} \quad (3.3)$$

Bu çalışmada, Eta korelasyon katsayısına dayalı bir özellik seçimi algoritması geliştirilmiştir. Algoritmanın iki versiyonu bulunmaktadır (E-Score.V1 - E-Score.V2).

3.4.1. Seçim kriteri 1 – e-score.V1

Şekil 3.1’de E-Score.V1 süreci gösterilmiştir. Öncelikle her bir özellik için Eta korelasyon katsayısı hesaplanmıştır (Eta ya da r_{pb}). Daha sonra eta eşik değeri bulunmuştur (Eta ya da r_{pb}). Eğer, $Eta > Eta_{mean}$ ise ilgili özellik seçilmiştir.



Şekil 3.1. E-Score Akış Diyagramı.

$$Eta_{mean} = \frac{Eta_1 + Eta_2 + Eta_3 + \dots + Eta_N}{N} \quad (3.4)$$

3.4.2. Seçim kriteri 2 – e-score.V2

Şekil 3.1’de, E-Score.V2 süreç adımlarını göstermektedir. İlk olarak, özelliklerin Eta değeri yüksek değerden düşük değere doğru sıralanır. Daha sonra en yüksek değerli %20 özellik seçilir. Bir kitabın değerinin %80’inin, sayfaların %20’sinde saklı olduğu

ifade edilmektedir (80/20 Principle, 2024). Bu çalışmanın amacı, veri boyutunu %80 oranında azaltmak ve sistem performansını iyileştirmektir. Buna ek olarak, performans değerlendirmeleri, özelliklerin %1 ile %100 arasındaki farklı oranlarda seçilmesiyle yapılmıştır.

Sonuç olarak bu çalışmada Eta özellik seçme algoritması kullanılmıştır. Eta sınıf etiketleri ile sayısal değişkenler arasında kullanılan bir yöntemdir. Bu yöntemde F-Score değeri eşik değeri olarak hesaplanılır. Daha sonra iki farklı yöntem ile özellikler seçilir. 1. yöntem eşik değerine göre üstündekiler alınır. 2. yöntem özellikler büyükten küçüğe doğru sıralanıp, en iyi %20' si alınır. Yani ilk öncelikle bütün özellikler için bir eta değeri bulunur. Daha sonra ortalama bir eta değeri hesaplanılır. Daha sonra bu ortalamaya göre 1. yöntem; eşik değerinin üzerindeki özellikler seçilir, 2. yöntem; ise en iyi yüzde %20'lik dilimdeki eta değerine sahip özellikler alınır.

3.4.3. 1. veri seti (UMMC) için özellik seçme algoritmasının uygulanması

4024 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo 3.21'de gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.22'de gösterilmiştir.

Tablo 3.21. 1. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	7	14
2	14	14, 6
3	21	14, 6, 13
4	29	14, 6, 13, 5
5	36	14, 6, 13, 5, 10
6	43	14, 6, 13, 5, 10, 11
7	50	14, 6, 13, 5, 10, 11, 7
8	57	14, 6, 13, 5, 10, 11, 7, 4

Tablo 3.21. (Devamı) 1. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
9	64	14, 6, 13, 5, 10, 11, 7, 4, 9
10	71	14, 6, 13, 5, 10, 11, 7, 4, 9, 8
11	79	14, 6, 13, 5, 10, 11, 7, 4, 9, 8, 1
12	86	14, 6, 13, 5, 10, 11, 7, 4, 9, 8, 1, 12
13	93	14, 6, 13, 5, 10, 11, 7, 4, 9, 8, 1, 12, 3
14	100	14, 6, 13, 5, 10, 11, 7, 4, 9, 8, 1, 12, 3, 2

Tablo 3.22. 1. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Katsayısı
14	0,4765
6	0,2576
13	0,2566
5	0,2558
10	0,1847
11	0,1771
7	0,1614
4	0,1547
9	0,1342
8	0,0966
1	0,0559
12	0,0348
3	0,0315
2	0,0042

3.4.4. 2. veri seti (Breast Cancer) için özellik seçme algoritmasının uygulanılması

286 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo 3.23'de gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.24'de gösterilmiştir.

Tablo 3.23. 2. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	11	6
2	22	6, 4
3	33	6, 4, 9
4	44	6, 4, 9, 3
5	56	6, 4, 9, 3, 1
6	67	6, 4, 9, 3, 1, 7
7	78	6, 4, 9, 3, 1, 7, 2
8	89	6, 4, 9, 3, 1, 7, 2, 8
9	100	6, 4, 9, 3, 1, 7, 2, 8, 5

Tablo 3.24. 2. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Kat Sayısı
6	0,2994
4	0,2759
9	0,1939
3	0,1751
1	0,0717
7	0,0586
2	0,0490
8	0
5	0

3.4.5. 3. veri seti (Coimbra) için özellik seçme algoritmasının uygulanılması

116 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo 3.25'da gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.26'da gösterilmiştir.

Tablo 3.25. 3. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	11	3
2	22	3, 5
3	33	3, 5, 4

Tablo 3.25. (Devamı) 3. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
4	44	3, 5, 4, 8
5	56	3, 5, 4, 8, 2
6	67	3, 5, 4, 8, 2, 9
7	78	3, 5, 4, 8, 2, 9, 1
8	89	3, 5, 4, 8, 2, 9, 1, 7
9	100	3, 5, 4, 8, 2, 9, 1, 7, 6

Tablo 3.26. 3. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Katsayısı
3	0,3843
5	0,2840
4	0,2768
8	0,2273
2	0,1326
9	0,0914
1	0,0436
7	0,0195
6	0,0011

3.4.6. 4. veri seti (WPBC) için özellik seçme algoritmasının uygulanması

198 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo 3.27'da gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.28'de gösterilmiştir.

Tablo 3.27. 4. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	3	1
2	6	1, 25
3	9	1, 25, 22
4	12	1, 25, 22, 24
5	15	1, 25, 22, 24, 5
6	18	1, 25, 22, 24, 5, 32

Tablo 3.27. (Devamı) 4. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
7	21	1, 25, 22, 24, 5, 32, 4
8	24	1, 25, 22, 24, 5, 32, 4, 2
9	27	1, 25, 22, 24, 5, 32, 4, 2, 15
10	30	1, 25, 22, 24, 5, 32, 4, 2, 15, 14
11	33	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12
12	36	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9
13	39	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11
14	42	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10
15	45	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13
16	48	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30
17	52	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29
18	55	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19
19	58	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3
20	61	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18
21	64	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31
22	67	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8
23	70	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16
24	73	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23
25	76	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20

Tablo 3.27. (Devamı) 4. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
26	79	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21
27	82	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26
28	85	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6
29	88	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6, 27
30	91	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6, 27, 28
31	94	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6, 27, 28, 17
32	97	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6, 27, 28, 17, 7
33	100	1, 25, 22, 24, 5, 32, 4, 2, 15, 14, 12, 9, 11, 10, 13, 30, 29, 19, 3, 18, 31, 8, 16, 23, 20, 21, 26, 6, 27, 28

Tablo 3.28. 4. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Katsayısı
1	0,3513
25	0,2353
22	0,2332
24	0,2320
5	0,1899
32	0,1773
4	0,1765
2	0,1741
15	0,1518
14	0,1416
12	0,1325
9	0,1182
11	0,1124
10	0,0998
13	0,0762
30	0,0747
29	0,0743
19	0,0656
3	0,0643
18	0,0604
31	0,0552
8	0,0549
16	0,0522
23	0,0511
20	0,0443
21	0,0428
26	0,0385
6	0,0208
27	0,0201
28	0,0176
17	0,0095
7	0,0008
33	0,0000

3.4.7. 5. veri seti (WBC) için özellik seçme algoritmasının uygulanılması

699 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo

3.29'da gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.30'da gösterilmiştir.

Tablo 3.29. 5. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	11	3
2	22	3, 2
3	33	3, 2, 7
4	44	3, 2, 7, 1
5	56	3, 2, 7, 1, 8
6	67	3, 2, 7, 1, 8, 4
7	78	3, 2, 7, 1, 8, 4, 5
8	89	3, 2, 7, 1, 8, 4, 5, 9
9	100	3, 2, 7, 1, 8, 4, 5, 9, 6

Tablo 3.30. 5. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Katsayısı
3	0,82
2	0,82
7	0,76
1	0,72
8	0,71
4	0,70
5	0,68
9	0,42
6	0,00

3.4.8. 6. veri seti (WDBC) için özellik seçme algoritmasının uygulanması

569 farklı kişiye ait veri seti Eta özellik seçme algoritmasına göre veri setinde seçilen özellikler ve yüzdesel gösterimi performans değerlendirme sıralamasına göre Tablo 3.31'de gösterilmiştir. Eta'ya göre seçilen özelliklerin hesaplanan korelasyon değerleri ise Tablo 3.32'de gösterilmiştir.

Tablo 3.31. 6. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
1	3	28
2	7	28, 23
3	10	28, 23, 8
4	13	28, 23, 8, 21
5	17	28, 23, 8, 21, 3
6	20	28, 23, 8, 21, 3, 24
7	23	28, 23, 8, 21, 3, 24, 1
8	27	28, 23, 8, 21, 3, 24, 1, 4
9	30	28, 23, 8, 21, 3, 24, 1, 4, 7
10	33	28, 23, 8, 21, 3, 24, 1, 4, 7, 27
11	37	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6
12	40	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26
13	43	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11
14	47	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13
15	50	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14
16	53	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22
17	57	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25
18	60	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29
19	63	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2
20	67	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18
21	70	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5
22	73	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9
23	77	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30
24	80	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16
25	83	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17
26	87	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17, 20

Tablo 3.31. (Devamı) 6. Veri Setinde Eta Kriterine Göre Seçilen Özellik Sayısı.

Özellik Sayısı	Özellik Yüzdesi (%)	Seçilmiş Özellikler
27	90	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17, 20, 15
28	93	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17, 20, 15, 10
29	97	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17, 20, 15, 10, 12
30	100	28, 23, 8, 21, 3, 24, 1, 4, 7, 27, 6, 26, 11, 13, 14, 22, 25, 29, 2, 18, 5, 9, 30, 16, 17, 20, 15, 10, 12, 19

Tablo 3.32. 6. Veri Setinde Eta kriterine göre seçilen özelliklerin korelasyon değerleri.

Özellik Numaraları	Etaya Göre Korolasyon Katsayısı
28	0,79
23	0,78
8	0,78
21	0,78
3	0,74
24	0,73
1	0,73
4	0,71
7	0,70
27	0,66
6	0,60
26	0,59
11	0,57
13	0,56
14	0,55
22	0,46
25	0,42
29	0,42
2	0,42
18	0,41
5	0,36
9	0,33
30	0,32
16	0,29
17	0,25
20	0,08
15	0,07
10	0,01
12	0,01
19	0,01

3.5. Makine Öğrenmesi Algoritmaları

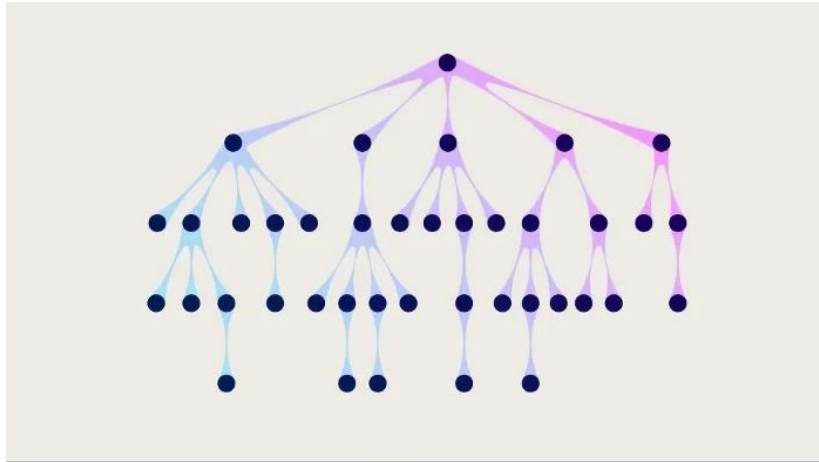
Yapılan bu çalışmada Ensemble, kNN, SVMs ve Hibrit Yapay Zeka makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır.

Tüm veri setleri üzerinde sınıflandırma işlemi sırası ile Tablo 3.8, Tablo 3.9, Tablo 3.10, Tablo 3.11, Tablo 3.12 ve Tablo 3.13’de belirtilen dengelenmiş alt veri setleri üzerinde ayrı şekilde bir kısmı model oluşturma, diğer kısmı ise modeli test etme aşamasında kullanılmıştır. Her bir veri grubu için eğitim veri seti oluşturulmuştur. Geriye kalan veri grubu ise test aşaması için kullanılmıştır. Test verileri üzerinden oluşturulmuş modelin performans değerlendirme kriterleri ile tecrübe edilmiştir.

3.5.1. Ensemble karar ağaçları

Ensemble karar ağaçları, özellikle büyük veri setlerinde veri yorumlama için önemli olan, algoritmalar temelinde çalışan sistemlerdir. Bu ağaçlar, sistematik bir yönlendirme sağlayan algoritmalar üretir. Standart karar ağaçlarının aksine, ensemble karar ağaçları, sistemdeki tüm karar ağaçlarını karşılaştırarak bunları tek bir birleşik karar ağacına entegre eder.

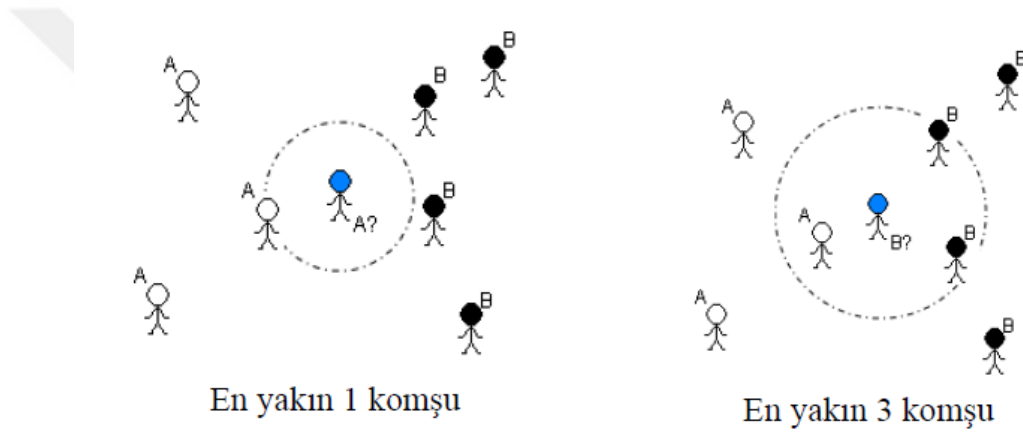
Örneğin: Bir ev kiralamak mı yoksa satın almak mı gerektiğine karar verirken, fiyat, oda sayısı ve metrekare gibi değişkenler göz önünde bulundurulur. Çıktı değişkeni, evet (kirala veya satın al) ya da hayır şeklinde ikili bir sonuç verir. X fiyatı altında ve en az Y metrekarelik bir alana sahip olma gibi kriterleri belirlerken, bir karar ağacı modeli etkin bir şekilde oluşturulmuş olur (Ensemble Karar Ağaçları, 2024). Şekil 3.2’de karar ağacı modeli gösterilmiştir.



Şekil 3.2. Ensemble Karar Ağaçları (Ulgen, 2017).

3.5.2. K-en yakın komşu algoritması

K-En Yakın Komşuluk Algoritması'nda bir veriyi anlamak için verilerin dağılımına göre o verinin etrafındaki en yakın verilere bakılır. "k" ifadesi ise değeri bilinmek istenilen verinin en yakınındaki kaç değere bakılması gerektiğinin sayısal ifadesidir. Genel olarak "k" değeri tek sayılardan oluşmaktadır. Böylelikle sistemdeki karar verme mekanizması daha sağlıklı çalışmaktadır (KNN, 2024). Şekil 3.3'de belirtildiği gibi değeri bilinmek istenilen verinin komşuları belirlenen çap içerisinde en çok hangi değeri kapsıyorsa o sınıfa dahil edilir. "k" değeri genel olarak tek seçildiği için eşitlik durumu oluşmamaktadır. Değerler arası mesafelerin belirlenmesinde Öklid, Manhattan gibi uzaklık fonksiyonları kullanılmaktadır.

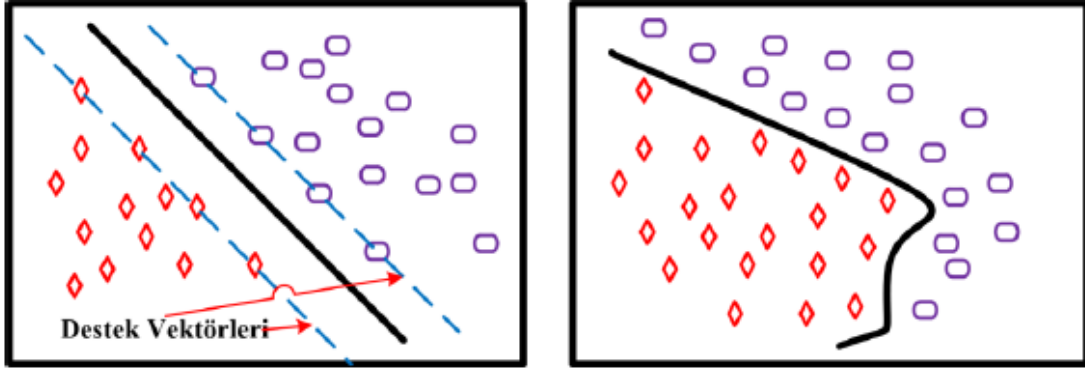


Şekil 3.3. k-En Yakın Komşu Algoritması (k-en Yakın Komşu Algoritması, 2024).

3.5.3. Destek vektör makineleri

kNN'de olduğu gibi tüm veriler x - y ekseninde özellik bazında dizilir. Daha sonra bu iki özelliği birbirinden ayıran bir eğri uydurulur. Uydurulan eğri veri setindeki özellikleri optimize edilerek oluşturulmaktadır. Bu eğri lineer ya da non-lineer olabilir. Daha sonra eğrinin altında ve üstünde kalan alanlardaki veriler sınıflandırılmaktadır. Bu eğri mesela bir çember bile olsa çemberin içinde ya da çemberin dışındaki veriler diye gruplandırılmaya tabii tutulabilmektedir (Alhasadi, 2016).

Şekil 3.4'de iki eksen bazında 2 özellik ile oluşturulmuş lineer ve non-lineer özelliklere sahip Destek Vektör Makineleri Örneği görülmektedir.

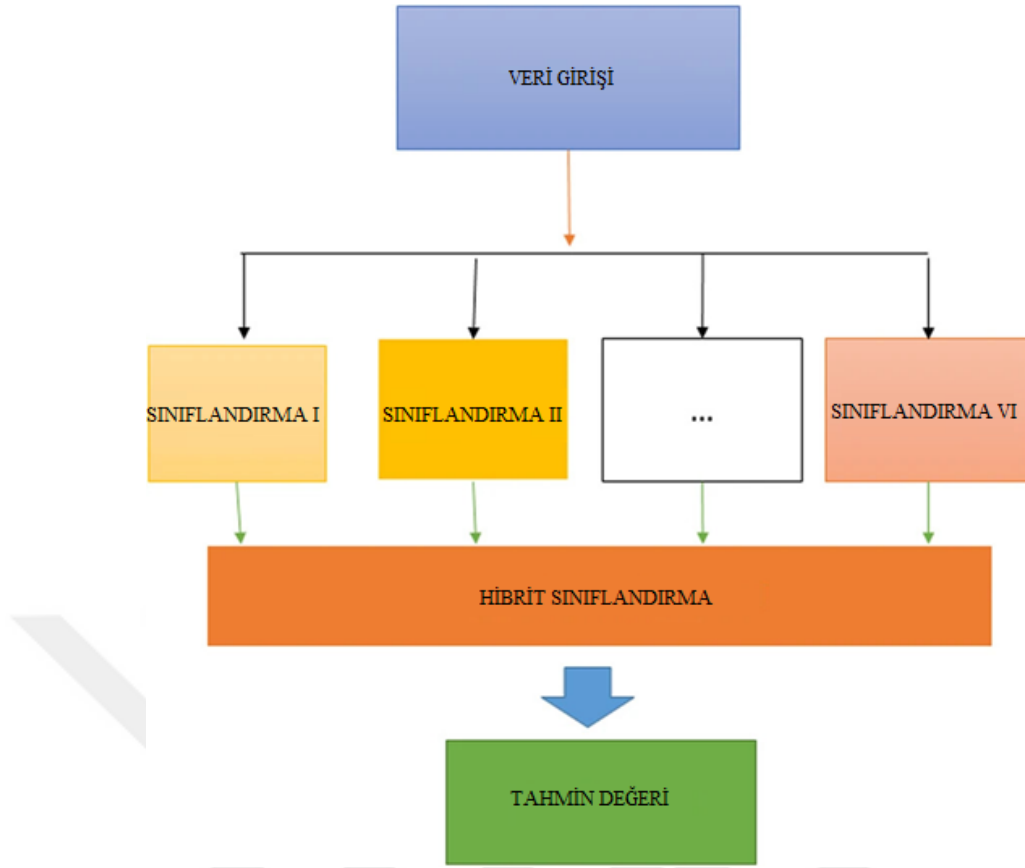


Şekil 3.4. Destek Vektör Makineleri (Uçar, 2017).

3.5.4. Hibrit yapay zeka

Bu yöntem, kNN, Karar Ağaçları ve Destek Vektör Makineleri (SVMs) algoritmalarının verilere uygulanarak elde edilen sınıflandırma kararlarının aritmetik ortalamasına dayanır. Temelde, Hibrit Yapay Zeka'nın çıktısı, bu yöntemler arasında çoğunluk kararına karşılık gelir. Bireysel sınıflandırma algoritmalarının zayıflıkları, Hibrit Yapay Zeka'daki nihai kararı etkilemez. Amaç, zayıf sınıflandırıcıları bir araya getirerek, Hibrit Yapay Zeka adı altında güçlü bir sınıflandırıcı oluşturmaktır. Kullanılan sınıflandırma algoritmalarının sayısı arttıkça performans da artar.

Hibrit yöntem, Ensemble Karar Ağaçları yöntemine oldukça benzer. Hibrit yöntemin diğer bir adı "Ensemble" olabilir. Her iki yöntem de genellikle 99 algoritma sayısına sahip olan birden fazla yapay zeka algoritmasını birleştirmeyi amaçlar. Ensemble yönteminde yalnızca karar ağaçları birleştirilirken, hibrit yöntemin amacı farklı algoritmaları entegre etmektir. Hibrit yöntemde kullanılan algoritmaların sayısı genellikle tek sayıdır (örneğin, 1, 3, 5, 7, ... n). Hibrit Yapay Zeka ile Ensemble Karar Ağaçları arasındaki temel fark, birincisinin karar ağaçlarını içermemesi ve bunun yerine Hibrit Yapay Zeka'nın kararını farklı algoritmaların çıktılarının ortalamasını alarak vermesidir; yani bir karar ağacı türetmeden karar verir.



Şekil 3.5. Hibrit Yapay Zeka (Wang ve ark., 2021).

Şekil 3.5’de görüldüğü üzere bir örnek üzerinden açıklamak gerekirse, n veri noktası (örneğin, veri noktaları 1, 2 ve 3) sisteme giriş olarak verildiğinde, Hibrit yöntemde, veri noktalarının sayısının n ’nin tek sayı olması önemlidir, çünkü çift sayıda veri noktası belirsiz bir sonuca yol açabilir ve bu da Hibrit yöntemin etkinliğini ortadan kaldırır.

Örneğin, veri noktası 1 için, kNN yöntemi bireyi hasta olarak sınıflandırabilir ve etiket 1 verebilir; Koşullu Karar Ağacı (Ctree) yöntemi bireyi sağlıklı olarak sınıflandırabilir ve etiket 2 verebilir; ve SVMs yöntemi bireyi hasta olarak sınıflandırabilir ve etiket 1 verebilir. Hibrit yöntem, bu üç yöntemin ortalamasını hesaplar ve sonuç olarak 1.33333 değeri elde edilir, bu değer Tablo 3.33’de gösterildiği gibi yuvarlanır. Bu yuvarlanmış değer üzerinden, çoğunluk kuralına dayanarak, veri noktası 1 için nihai sınıflandırma sağlıklı olarak belirlenir.

Hibrit yöntemde, ister yöntem etiketlerinin ortalaması alınsın isterse çoğunluk kuralı uygulansın, sonuç tutarlıdır.

Hibrit yöntemin temel prensibi, çoğunluğu yansıtan kararı kabul etmektir.

Tablo 3.33. Hibrit Yapay Zeka Tablo Örneği.

Veri	kNN	Ctree	SVMs	Hibrit	Gerçek
1	1	2	1	1.333333	1
2	2	2	1	1.666667	2
3	2	1	1	1.333333	1

Tablo 3.33'de gösterildiği gibi, çalışmada kullanılan her yöntem, üretilen etiketleri analiz ederken bazı zayıflıklar sergilemektedir. Örneğin, Ctree etiket 1'de bir zayıflık gösterirken, SVMs'ler etiket 2'de, kNN ise etiket 1'de zayıf kalmaktadır.

Hibrit yönteminin temel amacı, zayıf sınıflandırıcıları (Ctree, SVMs ve kNN) bir araya getirerek daha güçlü bir sınıflandırıcı oluşturmaktır. Her bir yöntem veri noktalarına 1, 2 ve 3 ayrı ayrı uygulandığında, her bir yöntem üç veri noktasından ikisini doğru sınıflandırır. Ancak Hibrit yöntem uygulandığında, üç veri noktası da gerçek değerlere göre doğru sınıflandırılmaktadır. Böylece, Hibrit yöntemin amacı, zayıf sınıflandırıcıları birleştirerek güçlü bir sınıflandırıcıya dönüştürmektir.

Bu kavram şu örnekle açıklanabilir: Dört ayağı olan bir masa düşünün. Yeterli gücü olmayan bir kişi masayı kaldırmaya çalıştığında, bunu başaramaz. Ancak, her biri benzer güçte olan dört kişi masayı birlikte kaldırırsa, başarılı olurlar. Masanın tek bir kişi tarafından kaldırılması durumunda dengesiz olabileceği halde, dört kişi birlikte kaldırdığında düşme olasılığı azalır. Hibrit yönteminin prensibi buna benzer şekilde çalışır. Yani, sınıflandırıcı sayısı arttıkça, Hibrit yöntemin performansı da buna paralel olarak artar.

3.6. Performans Değerlendirme Kriterleri

Çalışmada, doğruluk, özgüllük, duyarlılık, AUC, kappa istatistiği ve F ölçütü gibi performans değerlendirme kriterleri (Uçar ve ark., 2017, 2018) kullanılmıştır.

Duyarlılık, bir testin hastalar arasında hastalık durumunu doğru bir şekilde ayırt etme becerisini ifade eder ve 0 ile 1 arasında bir değere sahiptir. Teşhis amaçlı bir testin duyarlılığının 1 olması ideal kabul edilir. Duyarlılığın 1 olduğu durum, testin tüm hastaları eksiksiz şekilde doğru teşhis edebildiğini gösterir. Özgüllük ise, testin sağlıklı bireyler arasında sağlıklı olanları doğru şekilde ayırt etme yeteneğini tanımlar ve bu

da 0 ile 1 arasında bir değer alır. Özgüllük genellikle hastalığın doğrulanmasının önemli olduğu durumlarda kullanılır. Bir testin özgüllük değerinin 1 olması, o testin tüm sağlıklı bireyleri eksiksiz ve doğru bir şekilde tespit edebildiği anlamına gelir. Duyarlılık ve özgüllük parametreleri (3.6) ve (3.7)'deki gibi bulunmuştur. Çalışmadaki doğruluk oranları ise (3.5)'deki gibi bulunmuştur. (3.5), (3.6) ve (3.7)'de kullanılan parametreler; sırasıyla TP, TN, FP ve FN kısaltılmıştır ve Gerçek Pozitif (True Positives, TP), Gerçek Negatif (True Negatives, TN), Yanlış Pozitif (False Positives, FP) ve Yanlış Negatif (False Negatives, FN) anlamlarına gelmektedir. Tablo 3.34'de doğruluk, özgüllük ve duyarlılık için karşılaştırma tablosu verilmiştir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3.5)$$

$$\text{Duyarlılık} = \frac{TP + TN}{TP + FN} \times 100 \quad (3.6)$$

$$\text{Özgüllük} = \frac{TN}{TN + FP} \times 100 \quad (3.7)$$

Tablo 3.34. Doğruluk, özgüllük ve duyarlılık karşılaştırma tablosu.

		Tahmin Edilen	
		Pozitif	Negatif
Gerçek Durum	Pozitif	TP	FN
	Negatif	FP	TN

F-Score, oluşturulan modelin performansını değerlendirmek için kullanılan bir yöntemdir. Bu ölçüm, duyarlılık ve özgüllük değerlerinin ağırlıklı bir ortalaması olarak hesaplanır. F-Score denklemi (3.8)'de yer almaktadır. 0 ile 1 arasında bir değer aralığına sahiptir. Bu ölçümde 1, modelin kusursuz olduğunu, 0 ise modelin performansının son derece yetersiz olduğunu ifade eder.

$$F = 2 \times \frac{\text{Özgüllük} \times \text{Duyarlılık}}{\text{Özgüllük} + \text{Duyarlılık}} \quad (3.8)$$

AUC değeri, hastalığa yönelik tanı testlerinin performansını değerlendirmek için kullanılan bir ölçüttür.

Kappa katsayısı, yalnızca tesadüfe dayalı "şans uyumlarını" düzelterek ölçümlerin güvenilirliği hakkında bilgi sağlayan bir göstergedir. Literatürde, uyum derecesine bağlı olarak Kappa katsayısı için farklı sınır değerleri belirlenmiştir (Polat ve Güneş, 2009). Tablo 3.35’de Kappa katsayılarının sınır değerleri belirtilmiştir.

Tablo 3.35. Kappa katsayıları sınır değerleri.

Kappa Katsayısı	Açıklama
0,81 - 1,00	Çok iyi düzeyde uyum
0,61 - 0,80	İyi düzeyde uyum
0,41 - 0,60	Orta düzeyde uyum
0,21 - 0,40	Düşük düzeyde uyum
0,00 - 0,20	Zayıf uyum
< 0,00	Çok zayıf uyum

Tüm tanımlanmış bu kriterler, Ensemble, kNN, SVMs ve Hibrit Yapay Zeka sınıflandırıcılarına uygulanmıştır.

Bu çalışmada bazı veri setleri sınıflandırmaları için eğitim-test bölünme oranı %50-%50 olarak belirlenmiştir ve bu oran her bir veri seti için Tablo 3.36, Tablo 3.37 ve Tablo 3.39’da gösterilmektedir.

Bu çalışmada bazı veri setleri sınıflandırmaları için eğitim-test bölünme oranı %85 – %15 olarak belirlenmiştir ve bu oran her bir veri seti için Tablo 3.38’de gösterilmektedir.

Bu çalışmada bazı veri setleri sınıflandırmaları için eğitim-test bölünme oranı %80 – %20 olarak belirlenmiştir ve bu oran her bir veri seti için Tablo 3.40 ve Tablo 3.41’de gösterilmektedir.

Tablo 3.36. 1. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	A Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 1)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184
Sınıf	B Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 2)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184
Sınıf	C Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 3)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184
Sınıf	D Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 4)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184
Sınıf	E Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 5)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184
Sınıf	F Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Yaşayan (Grup 6)	284	284	568
Ölü	308	308	616
Toplam	592	592	1184

Tablo 3.37. 2. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	A Veri Seti		
	Eğitim (%50)	Test (%50)	Toplam
Tekrarlayan Olaylar (Grup 1)	33	34	67
Tekrarlanmayan Olaylar	43	42	85
Toplam	76	76	152

Tablo 3.37. (Devamı) 2. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	B Veri Seti		Toplam
	Eğitim (%50)	Test (%50)	
Tekrarlayan Olaylar (Grup 2)	33	34	67
Tekrarlanmayan Olaylar	43	42	85
Toplam	76	76	152

Sınıf	C Veri Seti		Toplam
	Eğitim (%50)	Test (%50)	
Tekrarlayan Olaylar (Grup 3)	33	34	67
Tekrarlanmayan Olaylar	43	42	85
Toplam	76	76	152

Tablo 3.38. 3. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	Veri Seti		Toplam
	Eğitim (%85)	Test (%15)	
Sağlıklılar	44	8	52
Hastalar	45	7	52
Toplam	89	15	104

Tablo 3.39. 4. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	A Veri Seti		Toplam
	Eğitim (%50)	Test (%50)	
Tekrarlamayan Olaylar (Grup 1)	25	25	50
Tekrarlayan Olaylar	24	23	47
Toplam	49	48	97

Sınıf	B Veri Seti		Toplam
	Eğitim (%50)	Test (%50)	
Tekrarlamayan Olaylar (Grup 2)	25	25	50
Tekrarlayan Olaylar	24	23	47
Toplam	49	48	97

Sınıf	C Veri Seti		Toplam
	Eğitim (%50)	Test (%50)	
Tekrarlamayan Olaylar (Grup 3)	25	25	50
Tekrarlayan Olaylar	24	23	47
Toplam	49	48	97

Tablo 3.40. 5. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	A Veri Seti		
	Eğitim (%80)	Test (%20)	Toplam
İyi Huylu (Grup 1)	183	46	229
Kötü Huylu	193	48	241
Toplam	376	94	470

Sınıf	B Veri Seti		
	Eğitim (%80)	Test (%20)	Toplam
İyi Huylu (Grup 2)	183	46	229
Kötü Huylu	193	48	241
Toplam	376	94	470

Tablo 3.41. 6. Veri Seti için Eğitim-Test Dağılımı (Örneklem Dağılımı Tablosu).

Sınıf	Veri Seti		
	Eğitim (%80)	Test (%20)	Toplam
İyi Huylu	169	43	212
Kötü Huylu	171	41	212
Toplam	340	84	424



4. SONUÇ VE ÖNERİLER

4.1. 1. Veri Seti (UMMC) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (14) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanmış, en iyi özellikten başlayarak 14 özellik sıralanmıştır. Bu sıralamaya göre 14 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf”a göre dengesiz bir veri olduğu için oluşturulan alt veri setlerinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 84 başka bir alt veri seti oluşturulmuştur. 84 alt veri seti, ham sınıf veri setinden 6 farklı veri seti elde edilerek 6 farklı veri setinin en iyi özellikten başlayarak 14 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.1’e göre, Tablo 3.36’da yer alan A Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 14 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar azalmış sonrasında genel olarak 14 özelliğe kadar artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 13 özelliğe sahip veri setinde görülmüştür. Şekil 4.1’de ise en iyi 13 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.2’ye göre, Tablo 3.36’da yer alan B Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 14 özelliğe kadar sıralandığında genel olarak en iyi 13 özelliğe kadar azalmış sonrasında 14. özellikte

artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ancak “Ensemble” sınıflandırma yönteminin yanında “kNN” ve “Hibrit” sınıflandırma yöntemleri içerisinde eşit şekilde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 1 özelliğe sahip veri setinde görülmüştür. Şekil 4.2’de ise en iyi 1 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble, kNN ve Hibrit” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 4.3’e göre, Tablo 3.36’da yer alan C Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 14 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar azalmış sonrasında genel olarak 14 özelliğe kadar artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ancak “Ensemble” sınıflandırma yönteminin yanında “kNN ve Hibrit” sınıflandırma yöntemleri içerisinde eşit şekilde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 1 özelliğe sahip veri setinde görülmüştür. Şekil 4.3’de ise en iyi 1 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble, kNN ve Hibrit” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 4.4’e göre, Tablo 3.36’da yer alan D Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 14 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar azalmış sonrasında genel olarak 14 özelliğe kadar artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Hibrit” olduğu görülmüştür. Ancak “Hibrit” sınıflandırma yönteminin yanında “kNN ve Ensemble” sınıflandırma yöntemleri içerisinde eşit şekilde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 1 özelliğe sahip veri setinde görülmüştür. Şekil 4.4’de ise en iyi 1 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble, kNN ve Hibrit” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 4.5'e göre, Tablo 3.36'da yer alan E Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özelliğten 14 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar azalmış sonrasında genel olarak 14 özelliğe kadar artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “SVMs” olduğu görülmüştür. Ancak “Ensemble” sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 1 özelliğe sahip veri setinde görülmüştür. Şekil 4.5'de ise en iyi 1 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 4.6'ya göre, Tablo 3.36'da yer alan F Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özelliğten 14 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar azalmış sonrasında genel olarak 14 özelliğe kadar artmıştır. 14 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 2 özelliğe sahip veri setinde görülmüştür. Şekil 4.6'da ise en iyi 2 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 3.36'ya göre, tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre “Duyarlılık” değeri “Doğruluk” değerinden yüksek olduğu için hastalıkların teşhis oranı daha yüksek olduğu gözlemlenmiştir. Böylece sistemin meme kanseri olan kişilerin bulunması için etkili olduğu görülmüştür. Sistem, meme kanseri olan kişilerin kanser teşhisinin hemen bulunmasında etkilidir. Tablo 4.5 ve Şekil 4.5'e göre; Tablo 3.36'daki E Veri Setindeki en iyi 1 özelliğe sahip “Ensemble” yöntemi, en iyi 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 ve 13 özelliklerine sahip “SVMs” sınıflandırma yönteminin kullanıldığı veri setleri %100 ile en yüksek ve en çok duyarlılık oranına sahiptir. Ancak tüm veri setlerinde uygulanan yöntemlere göre sistemin genel performansında özgüllük değeri duyarlılık değerinden daha düşük olduğu için sistemin sağlıklı kişileri bulması kanserli kişileri bulmasına göre daha

düşüktür. Böylece meme kanseri olan bireyin sistemde sağlıklı gözükme olasılığı düşük olup, kanserli hastaların çoğu teşhis edilmeye çalışılmıştır.

Yapılan tüm bu uygulamada doğruluk oranı, Tablo 3.36'ya göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı genel olarak %90'nın üzerinde olup, sağlık alanında bu sistemin kullanılması oldukça elverişlidir. Özellikle, Tablo 4.6 ve Şekil 4.6'daki F Veri Seti çoğunluğu “Ensemble” sınıflandırma yöntemi içerisinde en yüksek doğruluk oranlarına sahip veri setidir. F Veri Setindeki en iyi 2 özelliğe sahip “Ensemble” sınıflandırma yönteminin kullanıldığı veri setleri %99.3 ile en yüksek doğruluk oranına sahiptir.

Çalışmada Tablo 4.2 ve Şekil 4.2; Tablo 4.3 ve Şekil 4.3; Tablo 4.4 ve Şekil 4.4; ve Tablo 4.5 ve Şekil 4.5'de en iyi 1 özellik kullanılarak, Tablo 4.6 ve Şekil 4.6'da en iyi 2 özellik kullanılarak özellik çıkarımı yapılmıştır. Bu da sistemin az özellik çıkartılarak daha az iş yükü ile verimli sonuca ulaşılabilceği kanıtlanmış ve görselleştirilmiştir.

Tablo 4.1. A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	100,0	100,0	0,0	0,0	0,0	50,0
			kNN	72,5	97,5	49,4	65,5	45,9	73,4
			Ensemble	82,1	79,6	84,4	81,9	64,1	82,0
			Hibrit	82,1	79,6	84,4	81,9	64,1	82,0
2	2	14,28	SVMs	63,7	67,3	60,4	63,6	27,5	63,8
			kNN	76,2	91,9	61,7	73,8	52,9	76,8
			Ensemble	80,4	82,0	78,9	80,4	60,8	80,5
			Hibrit	79,4	87,7	71,8	78,9	59,0	79,7
3	3	21,42	SVMs	59,3	87,7	33,1	48,1	20,3	60,4
			kNN	78,4	85,6	71,8	78,1	57,0	78,7
			Ensemble	79,2	78,9	79,5	79,2	58,4	79,2
			Hibrit	78,4	84,2	73,1	78,2	56,9	78,6

Tablo 4.1. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
4	4	28,56	SVMs	63,3	74,3	53,2	62,0	27,3	63,8
			kNN	78,0	84,9	71,8	77,8	56,3	78,3
			Ensemble	78,4	79,9	76,9	78,4	56,8	78,4
			Hibrit	78,2	82,7	74,0	78,1	56,5	78,4
5	5	35,7	SVMs	61,1	86,3	38,0	52,7	23,8	62,1
			kNN	77,9	85,2	71,1	77,5	56,0	78,2
			Ensemble	79,4	81,0	77,9	79,4	58,8	79,5
			Hibrit	78,7	84,5	73,4	78,5	57,6	78,9
6	6	42,84	SVMs	64,9	81,0	50,0	61,8	30,6	65,5
			kNN	78,2	86,6	70,5	77,7	56,7	78,5
			Ensemble	79,2	79,6	78,9	79,2	58,4	79,2
			Hibrit	79,2	83,8	75,0	79,2	58,6	79,4
7	7	49,98	SVMs	66,9	81,0	53,9	64,7	34,5	67,4
			kNN	77,7	85,6	70,5	77,3	55,6	78,0
			Ensemble	77,5	76,8	78,2	77,5	55,0	77,5
			Hibrit	78,5	81,7	75,6	78,6	57,2	78,7
8	8	57,12	SVMs	66,7	78,2	56,2	65,4	34,0	67,2
			kNN	77,0	83,1	71,4	76,8	54,2	77,3
			Ensemble	78,5	78,5	78,6	78,5	57,1	78,5
			Hibrit	78,9	81,0	76,9	78,9	57,8	79,0
9	9	64,26	SVMs	66,4	79,2	54,5	64,6	33,4	66,9
			kNN	76,9	81,0	73,1	76,8	53,8	77,0
			Ensemble	81,1	82,7	79,5	81,1	62,2	81,1
			Hibrit	80,6	84,2	77,3	80,6	61,2	80,7

Tablo 4.1. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
10	10	71,4	SVMs	66,2	78,9	54,5	64,5	33,1	66,7
			kNN	76,9	81,0	73,1	76,8	53,8	77,0
			Ensemble	81,1	82,4	79,9	81,1	62,2	81,1
			Hibrit	80,9	83,8	78,2	80,9	61,9	81,0
11	11	78,54	SVMs	66,2	77,8	55,5	64,8	33,0	66,7
			kNN	74,8	77,8	72,1	74,8	49,7	74,9
			Ensemble	81,4	82,4	80,5	81,4	62,8	81,5
			Hibrit	79,7	83,1	76,6	79,7	59,5	79,9
12	12	85,68	SVMs	67,4	78,2	57,5	66,2	35,3	67,8
			kNN	71,6	74,6	68,8	71,6	43,3	71,7
			Ensemble	82,3	85,2	79,5	82,3	64,6	82,4
			Hibrit	79,1	83,8	74,7	79,0	58,2	79,2
13	13	92,82	SVMs	68,6	79,2	58,8	67,5	37,6	69,0
			kNN	72,3	76,1	68,8	72,3	44,7	72,4
			Ensemble	83,1	86,3	80,2	83,1	66,3	83,2
			Hibrit	80,7	85,2	76,6	80,7	61,6	80,9
14	14	100	SVMs	68,2	77,8	59,4	67,4	36,9	68,6
			kNN	71,8	75,4	68,5	71,8	43,7	71,9
			Ensemble	83,1	85,2	81,2	83,1	66,2	83,2
			Hibrit	80,1	82,7	77,6	80,1	60,2	80,2

Tablo 4.2. B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	78,2	100,0	58,1	73,5	57,1	79,1
			kNN	93,4	100,0	87,3	93,2	86,9	93,7
			Ensemble	93,4	100,0	87,3	93,2	86,9	93,7
			Hibrit	93,4	100,0	87,3	93,2	86,9	93,7
2	2	14,28	SVMs	78,0	93,7	63,6	75,8	56,6	78,6
			kNN	93,6	100,0	87,7	93,4	87,2	93,8
			Ensemble	93,4	97,2	89,9	93,4	86,8	93,6
			Hibrit	93,9	99,3	89,0	93,8	87,9	94,1
3	3	21,42	SVMs	77,4	92,6	63,3	75,2	55,2	78,0
			kNN	92,4	95,4	89,6	92,4	84,8	92,5
			Ensemble	91,6	92,3	90,9	91,6	83,1	91,6
			Hibrit	92,2	94,7	89,9	92,3	84,5	92,3
4	4	28,56	SVMs	76,5	93,0	61,4	73,9	53,6	77,2
			kNN	92,4	95,4	89,6	92,4	84,8	92,5
			Ensemble	91,7	93,3	90,3	91,8	83,4	91,8
			Hibrit	92,6	95,8	89,6	92,6	85,2	92,7
5	5	35,7	SVMs	76,4	93,0	61,0	73,7	53,3	77,0
			kNN	92,1	95,4	89,0	92,1	84,1	92,2
			Ensemble	90,7	91,5	89,9	90,7	81,4	90,7
			Hibrit	91,6	94,4	89,0	91,6	83,1	91,7
6	6	42,84	SVMs	76,4	92,3	61,7	73,9	53,2	77,0
			kNN	91,4	94,4	88,6	91,4	82,8	91,5
			Ensemble	90,9	91,9	89,9	90,9	81,7	90,9
			Hibrit	91,0	93,7	88,6	91,1	82,1	91,1
7	7	49,98	SVMs	76,2	89,8	63,6	74,5	52,8	76,7
			kNN	91,0	93,3	89,0	91,1	82,1	91,1
			Ensemble	88,9	88,0	89,6	88,8	77,7	88,8
			Hibrit	90,2	91,9	88,6	90,2	80,4	90,3

Tablo 4.2. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	57,12	SVMs	76,2	89,8	63,6	74,5	52,8	76,7
			kNN	91,2	93,7	89,0	91,3	82,4	91,3
			Ensemble	89,4	89,4	89,3	89,4	78,7	89,4
			Hibrit	90,4	92,3	88,6	90,4	80,7	90,4
9	9	64,26	SVMs	76,0	90,5	62,7	74,0	52,5	76,6
			kNN	87,8	90,5	85,4	87,9	75,7	87,9
			Ensemble	91,9	94,4	89,6	91,9	83,8	92,0
			Hibrit	91,2	96,1	86,7	91,2	82,5	91,4
10	10	71,4	SVMs	76,0	90,5	62,7	74,0	52,5	76,6
			kNN	87,8	90,5	85,4	87,9	75,7	87,9
			Ensemble	91,4	93,7	89,3	91,4	82,8	91,5
			Hibrit	90,7	95,8	86,0	90,6	81,5	90,9
11	11	78,54	SVMs	76,4	88,7	64,9	75,0	53,1	76,8
			kNN	86,7	89,8	83,8	86,7	73,3	86,8
			Ensemble	91,2	93,7	89,0	91,3	82,4	91,3
			Hibrit	88,9	93,0	85,1	88,8	77,7	89,0
12	12	85,68	SVMs	75,8	89,1	63,6	74,2	52,1	76,4
			kNN	83,8	87,0	80,8	83,8	67,6	83,9
			Ensemble	91,7	94,4	89,3	91,8	83,5	91,8
			Hibrit	88,7	94,4	83,4	88,6	77,4	88,9
13	13	92,82	SVMs	76,2	89,1	64,3	74,7	52,8	76,7
			kNN	83,8	86,6	81,2	83,8	67,6	83,9
			Ensemble	91,4	94,4	88,6	91,4	82,8	91,5
			Hibrit	88,2	92,6	84,1	88,1	76,4	88,3
14	14	100	SVMs	75,3	88,7	63,0	73,7	51,1	75,9
			kNN	83,4	85,9	81,2	83,5	66,9	83,5
			Ensemble	92,1	94,7	89,6	92,1	84,1	92,2
			Hibrit	88,2	93,0	83,8	88,1	76,4	88,4

Tablo 4.3. C Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	86,7	100,0	74,4	85,3	73,6	87,2
			kNN	96,1	100,0	92,5	96,1	92,2	96,3
			Ensemble	96,1	100,0	92,5	96,1	92,2	96,3
			Hibrit	96,1	100,0	92,5	96,1	92,2	96,3
2	2	14,28	SVMs	86,5	100,0	74,0	85,1	73,2	87,0
			kNN	95,6	99,6	91,9	95,6	91,2	95,8
			Ensemble	95,9	99,3	92,9	96,0	91,9	96,1
			Hibrit	95,6	99,6	91,9	95,6	91,2	95,8
3	3	21,42	SVMs	85,6	98,6	73,7	84,3	71,5	86,1
			kNN	94,8	97,5	92,2	94,8	89,5	94,9
			Ensemble	94,8	96,1	93,5	94,8	89,5	94,8
			Hibrit	94,6	97,2	92,2	94,6	89,2	94,7
4	4	28,56	SVMs	85,3	97,9	73,7	84,1	70,9	85,8
			kNN	94,8	97,5	92,2	94,8	89,5	94,9
			Ensemble	94,8	96,1	93,5	94,8	89,5	94,8
			Hibrit	94,6	97,2	92,2	94,6	89,2	94,7
5	5	35,7	SVMs	85,3	97,9	73,7	84,1	70,9	85,8
			kNN	94,4	96,8	92,2	94,5	88,9	94,5
			Ensemble	94,8	96,5	93,2	94,8	89,5	94,8
			Hibrit	94,6	97,2	92,2	94,6	89,2	94,7
6	6	42,84	SVMs	85,5	97,9	74,0	84,3	71,2	86,0
			kNN	94,4	96,5	92,5	94,5	88,9	94,5
			Ensemble	93,9	94,0	93,8	93,9	87,8	93,9
			Hibrit	94,4	96,5	92,5	94,5	88,9	94,5
7	7	49,98	SVMs	86,8	97,9	76,6	86,0	73,8	87,3
			kNN	93,4	95,1	91,9	93,4	86,8	93,5
			Ensemble	92,4	91,5	93,2	92,4	84,8	92,4
			Hibrit	93,8	95,8	91,9	93,8	87,5	93,8

Tablo 4.3. (Devamı) C Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	57,12	SVMs	86,7	97,5	76,6	85,8	73,5	87,1
			kNN	93,6	94,7	92,5	93,6	87,2	93,6
			Ensemble	93,6	94,0	93,2	93,6	87,1	93,6
			Hibrit	93,9	96,1	91,9	94,0	87,8	94,0
9	9	64,26	SVMs	85,8	96,5	76,0	85,0	71,8	86,2
			kNN	91,9	94,4	89,6	91,9	83,8	92,0
			Ensemble	95,3	97,2	93,5	95,3	90,5	95,3
			Hibrit	94,1	97,5	90,9	94,1	88,2	94,2
10	10	71,4	SVMs	86,0	96,8	76,0	85,1	72,2	86,4
			kNN	91,9	94,4	89,6	91,9	83,8	92,0
			Ensemble	95,1	97,2	93,2	95,1	90,2	95,2
			Hibrit	93,8	97,2	90,6	93,8	87,5	93,9
11	11	78,54	SVMs	86,5	96,1	77,6	85,9	73,1	86,9
			kNN	89,5	93,0	86,4	89,5	79,1	89,7
			Ensemble	95,4	96,8	94,2	95,5	90,9	95,5
			Hibrit	92,9	98,2	88,0	92,8	85,8	93,1
12	12	85,68	SVMs	86,3	96,8	76,6	85,6	72,8	86,7
			kNN	89,7	94,0	85,7	89,7	79,4	89,9
			Ensemble	95,9	98,2	93,8	96,0	91,9	96,0
			Hibrit	92,9	98,9	87,3	92,8	85,9	93,1
13	13	92,82	SVMs	86,1	96,8	76,3	85,3	72,5	86,6
			kNN	89,5	94,0	85,4	89,5	79,1	89,7
			Ensemble	95,9	98,2	93,8	96,0	91,9	96,0
			Hibrit	92,9	98,9	87,3	92,8	85,9	93,1
14	14	100	SVMs	86,0	96,5	76,3	85,2	72,2	86,4
			kNN	89,5	94,0	85,4	89,5	79,1	89,7
			Ensemble	95,8	98,2	93,5	95,8	91,6	95,9
			Hibrit	92,7	98,9	87,0	92,6	85,5	93,0

Tablo 4.4. D Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	91,2	100,0	83,1	90,8	82,5	91,6
			kNN	95,9	100,0	92,2	95,9	91,9	96,1
			Ensemble	95,9	100,0	92,2	95,9	91,9	96,1
			Hibrit	95,9	100,0	92,2	95,9	91,9	96,1
2	2	14,28	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	95,3	100,0	90,9	95,2	90,6	95,5
			Ensemble	95,9	100,0	92,2	95,9	91,9	96,1
			Hibrit	95,3	100,0	90,9	95,2	90,6	95,5
3	3	21,42	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,4	95,1	91,9	93,4	86,8	93,5
			Ensemble	93,1	93,0	93,2	93,1	86,1	93,1
			Hibrit	93,4	95,4	91,6	93,5	86,8	93,5
4	4	28,56	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,6	95,1	92,2	93,6	87,2	93,6
			Ensemble	93,6	94,0	93,2	93,6	87,1	93,6
			Hibrit	93,6	95,4	91,9	93,6	87,2	93,7
5	5	35,7	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,2	94,7	91,9	93,3	86,5	93,3
			Ensemble	93,1	93,3	92,9	93,1	86,1	93,1
			Hibrit	93,2	95,1	91,6	93,3	86,5	93,3
6	6	42,84	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,1	94,4	91,9	93,1	86,1	93,1
			Ensemble	92,4	91,9	92,9	92,4	84,8	92,4
			Hibrit	93,1	94,7	91,6	93,1	86,1	93,1
7	7	49,98	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,1	94,4	91,9	93,1	86,1	93,1
			Ensemble	92,4	91,5	93,2	92,4	84,8	92,4
			Hibrit	93,2	94,7	91,9	93,3	86,5	93,3

Tablo 4.4. (Devamı) D Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	57,12	SVMs	91,6	100,0	83,8	91,2	83,2	91,9
			kNN	93,2	94,4	92,2	93,3	86,5	93,3
			Ensemble	93,2	93,3	93,2	93,2	86,5	93,2
			Hibrit	93,8	95,4	92,2	93,8	87,5	93,8
9	9	64,26	SVMs	91,4	100,0	83,4	91,0	82,9	91,7
			kNN	93,1	96,8	89,6	93,1	86,2	93,2
			Ensemble	93,9	94,7	93,2	93,9	87,8	94,0
			Hibrit	93,8	98,2	89,6	93,7	87,5	93,9
10	10	71,4	SVMs	91,6	100,0	83,8	91,2	83,2	91,9
			kNN	93,1	96,8	89,6	93,1	86,2	93,2
			Ensemble	93,6	94,7	92,5	93,6	87,2	93,6
			Hibrit	94,1	98,6	89,9	94,1	88,2	94,3
11	11	78,54	SVMs	91,7	99,6	84,4	91,4	83,5	92,0
			kNN	90,4	93,0	88,0	90,4	80,8	90,5
			Ensemble	93,6	94,4	92,9	93,6	87,2	93,6
			Hibrit	92,9	97,9	88,3	92,9	85,8	93,1
12	12	85,68	SVMs	91,7	99,6	84,4	91,4	83,5	92,0
			kNN	90,0	92,6	87,7	90,1	80,1	90,1
			Ensemble	94,6	96,1	93,2	94,6	89,2	94,7
			Hibrit	93,1	98,2	88,3	93,0	86,2	93,3
13	13	92,82	SVMs	91,7	99,6	84,4	91,4	83,5	92,0
			kNN	90,0	93,0	87,3	90,1	80,1	90,1
			Ensemble	93,8	94,7	92,9	93,8	87,5	93,8
			Hibrit	92,9	98,2	88,0	92,8	85,8	93,1
14	14	100	SVMs	91,9	99,6	84,7	91,6	83,9	92,2
			kNN	90,0	93,0	87,3	90,1	80,1	90,1
			Ensemble	94,4	96,1	92,9	94,5	88,8	94,5
			Hibrit	92,7	97,9	88,0	92,7	85,5	92,9

Tablo 4.5. E Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	97,3	100,0	94,8	97,3	94,6	97,4
			Ensemble	97,5	100,0	95,1	97,5	94,9	97,6
			Hibrit	97,3	100,0	94,8	97,3	94,6	97,4
2	2	14,28	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	96,8	98,9	94,8	96,8	93,6	96,9
			Ensemble	97,1	98,9	95,5	97,2	94,3	97,2
			Hibrit	96,8	98,9	94,8	96,8	93,6	96,9
3	3	21,42	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	95,6	95,8	95,5	95,6	91,2	95,6
			Ensemble	95,6	95,1	96,1	95,6	91,2	95,6
			Hibrit	95,4	95,8	95,1	95,5	90,9	95,5
4	4	28,56	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	95,4	95,4	95,5	95,4	90,9	95,4
			Ensemble	95,9	95,8	96,1	95,9	91,9	95,9
			Hibrit	95,4	95,8	95,1	95,5	90,9	95,5
5	5	35,7	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	95,3	95,4	95,1	95,3	90,5	95,3
			Ensemble	95,4	95,1	95,8	95,4	90,9	95,4
			Hibrit	95,3	95,8	94,8	95,3	90,5	95,3
6	6	42,84	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	95,6	96,5	94,8	95,6	91,2	95,6
			Ensemble	95,8	95,8	95,8	95,8	91,5	95,8
			Hibrit	95,6	96,8	94,5	95,6	91,2	95,7
7	7	49,98	SVMs	96,3	100,0	92,9	96,3	92,6	96,4
			kNN	95,4	95,8	95,1	95,5	90,9	95,5
			Ensemble	95,1	94,4	95,8	95,1	90,2	95,1
			Hibrit	95,3	96,1	94,5	95,3	90,5	95,3

Tablo 4.5. (Devamı) E Veri Seti Özet Tablo - En İyi Yöntemler.

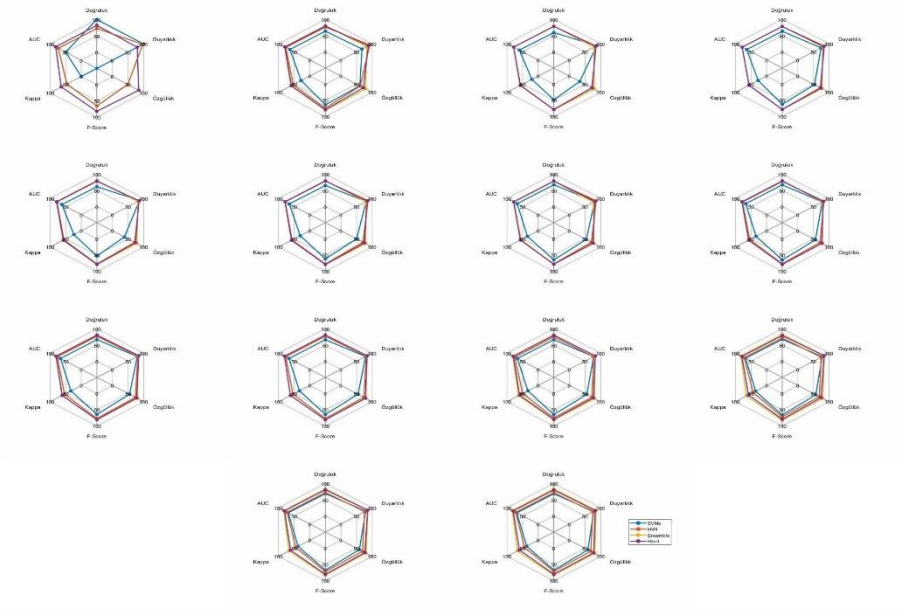
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	57,12	SVMs	96,3	100,0	92,9	96,3	92,6	96,4
			kNN	95,3	95,4	95,1	95,3	90,5	95,3
			Ensemble	95,4	95,1	95,8	95,4	90,9	95,4
			Hibrit	95,3	96,1	94,5	95,3	90,5	95,3
9	9	64,26	SVMs	95,9	100,0	92,2	95,9	91,9	96,1
			kNN	93,2	95,1	91,6	93,3	86,5	93,3
			Ensemble	94,9	95,4	94,5	94,9	89,9	95,0
			Hibrit	94,8	97,5	92,2	94,8	89,5	94,9
10	10	71,4	SVMs	95,9	100,0	92,2	95,9	91,9	96,1
			kNN	93,2	95,1	91,6	93,3	86,5	93,3
			Ensemble	94,8	95,1	94,5	94,8	89,5	94,8
			Hibrit	94,8	97,5	92,2	94,8	89,5	94,9
11	11	78,54	SVMs	95,9	100,0	92,2	95,9	91,9	96,1
			kNN	93,1	96,1	90,3	93,1	86,2	93,2
			Ensemble	95,4	96,1	94,8	95,5	90,9	95,5
			Hibrit	95,6	99,3	92,2	95,6	91,2	95,8
12	12	85,68	SVMs	96,1	100,0	92,5	96,1	92,2	96,3
			kNN	91,0	93,7	88,6	91,1	82,1	91,1
			Ensemble	95,8	97,5	94,2	95,8	91,6	95,8
			Hibrit	95,8	99,3	92,5	95,8	91,6	95,9
13	13	92,82	SVMs	95,9	100,0	92,2	95,9	91,9	96,1
			kNN	91,2	94,0	88,6	91,2	82,4	91,3
			Ensemble	95,9	97,9	94,2	96,0	91,9	96,0
			Hibrit	95,4	98,9	92,2	95,5	90,9	95,6
14	14	100	SVMs	95,8	99,6	92,2	95,8	91,6	95,9
			kNN	91,6	94,7	88,6	91,6	83,1	91,7
			Ensemble	96,5	98,9	94,2	96,5	92,9	96,5
			Hibrit	95,8	99,6	92,2	95,8	91,6	95,9

Tablo 4.6. F Veri Seti Özet Tablo - En İyi Yöntemler.

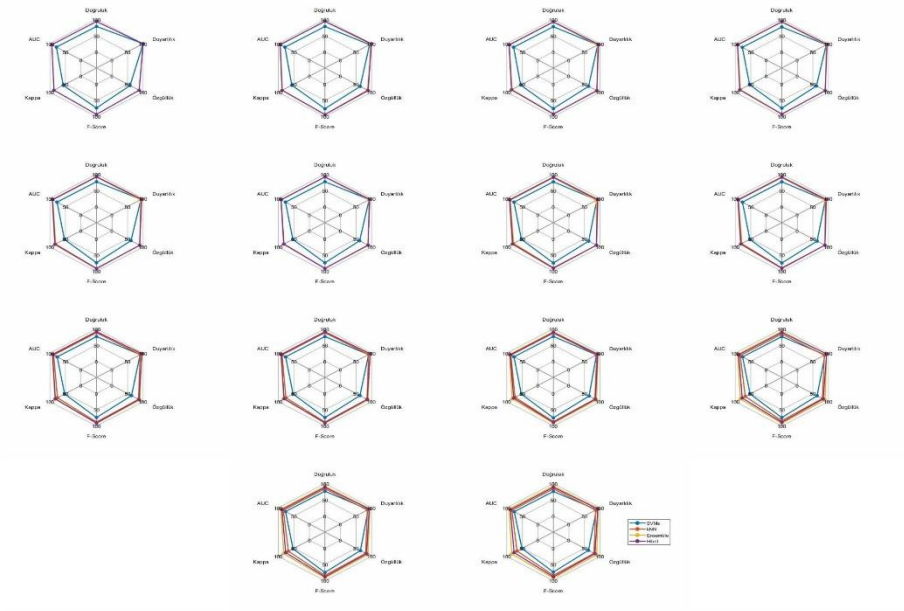
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	7,14	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	99,2	100,0	98,4	99,2	98,3	99,2
			Ensemble	77,7	54,9	98,7	70,6	54,6	76,8
			Hibrit	99,2	100,0	98,4	99,2	98,3	99,2
2	2	14,28	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	99,0	100,0	98,1	99,0	98,0	99,0
			Ensemble	99,3	100,0	98,7	99,3	98,6	99,4
			Hibrit	99,0	100,0	98,1	99,0	98,0	99,0
3	3	21,42	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	98,8	99,6	98,1	98,8	97,6	98,8
			Ensemble	99,2	99,6	98,7	99,2	98,3	99,2
			Hibrit	98,8	99,6	98,1	98,8	97,6	98,8
4	4	28,56	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	98,8	99,6	98,1	98,8	97,6	98,8
			Ensemble	99,2	99,6	98,7	99,2	98,3	99,2
			Hibrit	98,8	99,6	98,1	98,8	97,6	98,8
5	5	35,7	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	98,8	99,6	98,1	98,8	97,6	98,8
			Ensemble	99,0	99,3	98,7	99,0	98,0	99,0
			Hibrit	98,8	99,6	98,1	98,8	97,6	98,8
6	46	42,84	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	98,6	99,6	97,7	98,7	97,3	98,7
			Ensemble	99,0	99,3	98,7	99,0	98,0	99,0
			Hibrit	98,6	99,6	97,7	98,7	97,3	98,7
7	7	49,98	SVMs	98,8	100,0	97,7	98,9	97,6	98,9
			kNN	98,6	99,6	97,7	98,7	97,3	98,7
			Ensemble	98,5	99,3	97,7	98,5	97,0	98,5
			Hibrit	98,6	99,6	97,7	98,7	97,3	98,7

Tablo 4.6. (Devamı) F Veri Seti Özet Tablo - En İyi Yöntemler.

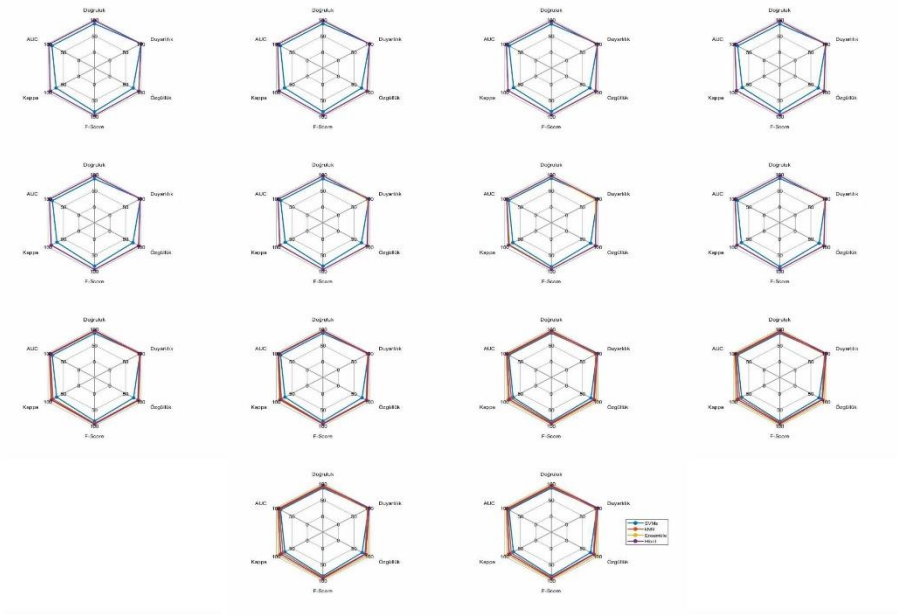
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	57,12	SVMs	98,6	99,6	97,7	98,7	97,3	98,7
			kNN	98,3	98,9	97,7	98,3	96,6	98,3
			Ensemble	97,6	97,5	97,7	97,6	95,3	97,6
			Hibrit	98,5	99,3	97,7	98,5	97,0	98,5
9	9	64,26	SVMs	98,6	99,6	97,7	98,7	97,3	98,7
			kNN	97,5	97,5	97,4	97,5	94,9	97,5
			Ensemble	99,0	99,3	98,7	99,0	98,0	99,0
			Hibrit	98,5	99,3	97,7	98,5	97,0	98,5
10	10	71,4	SVMs	98,6	99,6	97,7	98,7	97,3	98,7
			kNN	97,5	97,5	97,4	97,5	94,9	97,5
			Ensemble	99,0	99,3	98,7	99,0	98,0	99,0
			Hibrit	98,5	99,3	97,7	98,5	97,0	98,5
11	11	78,54	SVMs	98,6	99,6	97,7	98,7	97,3	98,7
			kNN	97,3	98,9	95,8	97,3	94,6	97,4
			Ensemble	99,2	99,6	98,7	99,2	98,3	99,2
			Hibrit	98,8	100,0	97,7	98,9	97,6	98,9
12	12	85,68	SVMs	98,5	99,3	97,7	98,5	97,0	98,5
			kNN	96,3	97,2	95,5	96,3	92,6	96,3
			Ensemble	99,2	99,6	98,7	99,2	98,3	99,2
			Hibrit	98,8	100,0	97,7	98,9	97,6	98,9
13	13	92,82	SVMs	98,5	99,3	97,7	98,5	97,0	98,5
			kNN	96,6	97,9	95,5	96,7	93,2	96,7
			Ensemble	98,0	97,2	98,7	97,9	95,9	97,9
			Hibrit	98,8	100,0	97,7	98,9	97,6	98,9
14	14	100	SVMs	98,6	99,6	97,7	98,7	97,3	98,7
			kNN	96,8	98,2	95,5	96,8	93,6	96,8
			Ensemble	98,3	97,9	98,7	98,3	96,6	98,3
			Hibrit	98,8	100,0	97,7	98,9	97,6	98,9



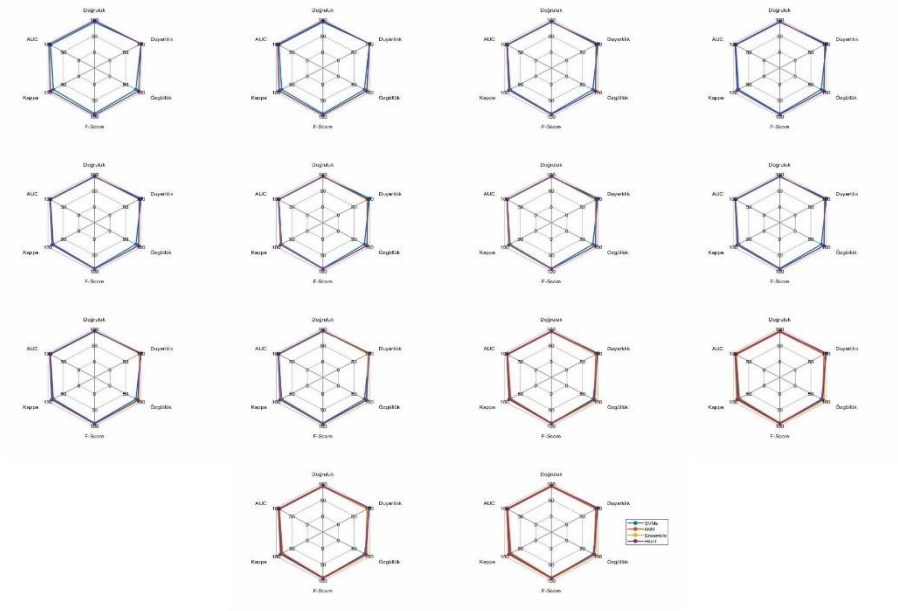
Şekil 4.1. A Veri Seti Sınıflandırma Modellerinin Performansı.



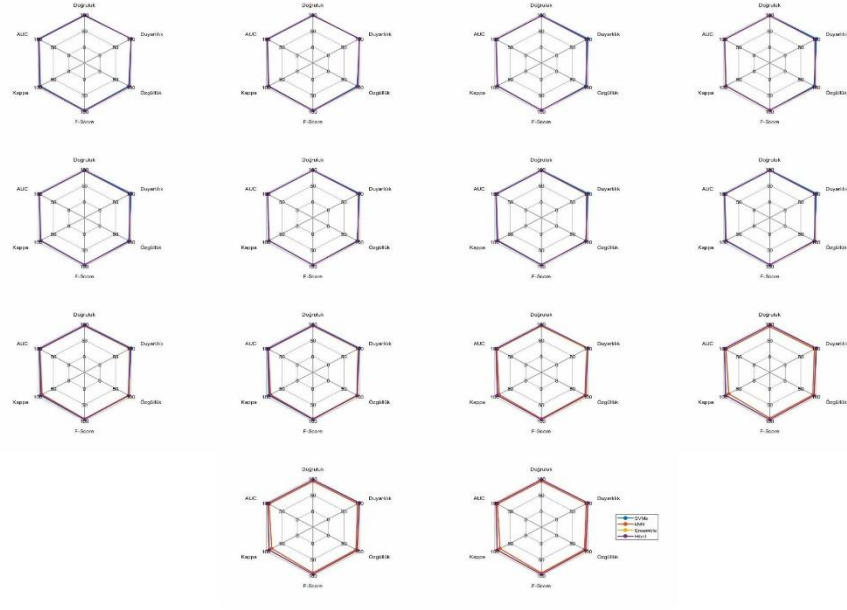
Şekil 4.2. B Veri Seti Sınıflandırma Modellerinin Performansı.



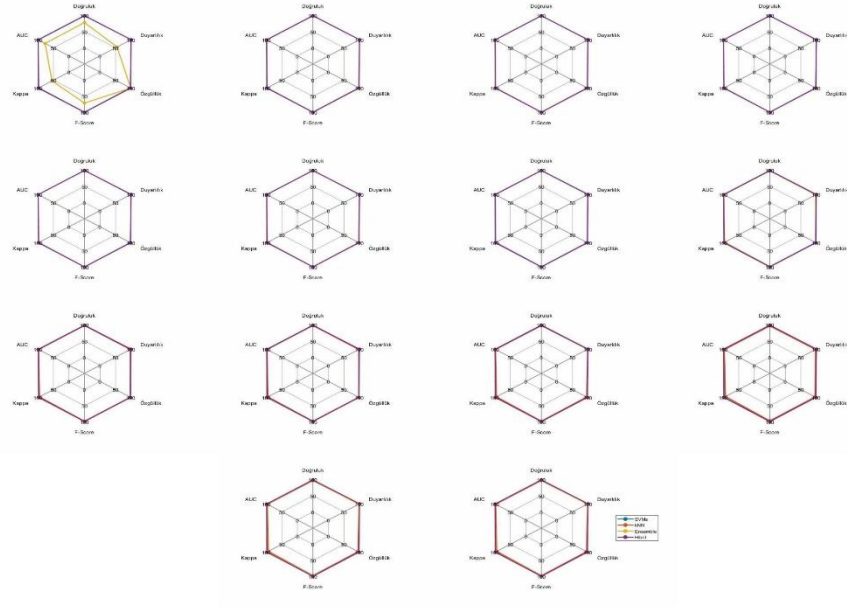
Şekil 4.3. C Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.4. D Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.5. E Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.6. F Veri Seti Sınıflandırma Modellerinin Performansı.

4.2. 2. Veri Seti (Breast Cancer) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması farklı bir veri seti üzerinde geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (9) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanmış, en iyi özellikten başlayarak 9 özellik sıralanmıştır. Bu sıralamaya göre 9 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf” a göre dengesiz bir veri olduğu için oluşturulan alt veri setlerinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 27 başka bir alt veri seti oluşturulmuştur. 27 alt veri seti, ham sınıf veri setinden 3 farklı veri seti elde edilerek 3 farklı veri setinin en iyi özellikten başlayarak 9 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.7’ye göre, Tablo 3.37’de yer alan A Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar artmış sonrasında 9 özelliğe kadar azalmıştır. 9 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “SVMs” olduğu görülmüştür. SVMs sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 5, 6 ve 7 özelliklerine sahip veri setlerinde görülmüştür. Şekil 4.7’de ise en iyi 5, 6 ve 7 özelliklerine sahip şekillerde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “SVMs” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.8’e göre, Tablo 3.37’de yer alan B Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar sıralandığında genel olarak en iyi 4 özelliğe kadar artmış sonrasında 9 özelliğe kadar azalmalar ve artmalar olmuştur. 9 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “SVMs” olduğu görülmüştür. SVMs sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 4 özelliğe sahip veri setinde görülmüştür. Şekil 4.8’de ise en iyi 4

özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “SVMs” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.9’a göre, Tablo 3.37’de yer alan C Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar sıralandığında genel olarak en iyi 7 özelliğe kadar artmış sonrasında 9 özelliğe kadar azalmalar ve artmalar olmuştur. 9 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 7 özelliğe sahip veri setinde görülmüştür. Şekil 4.9’da ise en iyi 7 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 3.37’ye göre, tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre Tablo 4.7’de görüldüğü üzere “Duyarlılık” değeri “Doğruluk” değerinden yüksek olduğu için hastalıkların teşhis oranı daha yüksek olduğu gözlemlenmiştir. Böylece sistemin genel olarak meme kanseri olan kişilerin bulunması için etkili olduğu görülmüştür. Sistem, meme kanseri olan kişilerin kanser teşhisinin hemen bulunmasında etkilidir. Tablo 4.7 ve Şekil 4.7’ye göre; Tablo 3.37’deki A Veri Setindeki en iyi 1 özelliğe sahip “SVMs, kNN, Ensemble ve Hibrit” yöntemleri, en iyi 2 özelliğe sahip “SVMs” ve “Hibrit” yöntemi, en iyi 4, 5, 6 ve 7 özelliğe sahip “SVMs” sınıflandırma yönteminin kullanıldığı veri setleri %100 ile en yüksek ve en çok duyarlılık oranına sahiptir. Tablo 4.9 ve Şekil 4.9’a göre; Tablo 3.37’deki C Veri Setindeki en iyi 1 özelliğe sahip “SVMs, Ensemble ve Hibrit” sınıflandırma yönteminin kullanıldığı veri setleri %100 ile en yüksek ve en çok duyarlılık oranına sahiptir. Ancak tüm veri setlerinde uygulanan yöntemlere göre sistemin genel performansında özgüllük değeri duyarlılık değerinden daha düşük olduğu için sistemin sağlıklı kişileri bulması kanserli kişileri bulmasına göre daha düşüktür. Böylece meme kanseri olan bireyin sistemde sağlıklı gözükme olasılığı düşük olup, kanserli hastaların çoğu teşhis edilmeye çalışılmıştır.

Yapılan tüm bu uygulamada doğruluk oranı, Tablo 3.37’de göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı genel olarak değişkenlik

göstermektedir. Özellikle, Tablo 4.7 ve Şekil 4.7'deki A Veri Seti ile Tablo 4.8 ve Şekil 4.8'deki B Veri Seti çoğunluğu “SVMs “ sınıflandırma yöntemi içerisinde en yüksek doğruluk oranlarına sahip veri setleridir.

Tablo 4.7. A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	67,11	100,00	40,48	57,63	37,83	70,24
			kNN	67,11	100,00	40,48	57,63	37,83	70,24
			Ensemble	67,11	100,00	40,48	57,63	37,83	70,24
			Hibrit	67,11	100,00	40,48	57,63	37,83	70,24
2	2	22	SVMs	68,42	100,00	42,86	60,00	40,16	71,43
			kNN	67,11	100,00	40,48	57,63	37,83	70,24
			Ensemble	63,16	85,29	45,24	59,12	29,07	65,27
			Hibrit	68,42	100,00	42,86	60,00	40,16	71,43
3	3	33	SVMs	69,74	44,12	90,48	59,31	36,11	67,30
			kNN	63,16	94,12	38,10	54,24	30,18	66,11
			Ensemble	71,05	76,47	66,67	71,23	42,42	71,57
			Hibrit	71,05	76,47	66,67	71,23	42,42	71,57
4	4	44	SVMs	75,00	100,00	54,76	70,77	51,99	77,38
			kNN	67,11	94,12	45,24	61,11	37,17	69,68
			Ensemble	68,42	82,35	57,14	67,47	38,21	69,75
			Hibrit	72,37	94,12	54,76	69,24	46,66	74,44
5	5	56	SVMs	77,63	100,00	59,52	74,63	56,82	79,76
			kNN	72,37	94,12	54,76	69,24	46,66	74,44
			Ensemble	73,68	91,18	59,52	72,03	48,79	75,35
			Hibrit	75,00	94,12	59,52	72,93	51,48	76,82
6	6	67	SVMs	77,63	100,00	59,52	74,63	56,82	79,76
			kNN	65,79	79,41	54,76	64,82	33,06	67,09
			Ensemble	68,42	73,53	64,29	68,60	37,19	68,91
			Hibrit	72,37	85,29	61,90	71,74	45,79	73,60
7	7	78	SVMs	77,63	100,00	59,52	74,63	56,82	79,76
			kNN	69,74	85,29	57,14	68,44	40,95	71,22
			Ensemble	67,11	64,71	69,05	66,81	33,66	66,88
			Hibrit	75,00	88,24	64,29	74,38	50,95	76,26
8	8	89	SVMs	77,63	79,41	76,19	77,77	55,14	77,80
			kNN	64,47	73,53	57,14	64,31	29,92	65,34
			Ensemble	67,11	61,76	71,43	66,25	33,29	66,60
			Hibrit	72,37	73,53	71,43	72,46	44,58	72,48

Tablo 4.7. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
9	9	100	SVMs	76,32	76,47	76,19	76,33	52,37	76,33
			kNN	63,16	73,53	54,76	62,77	27,52	64,15
			Ensemble	68,42	67,65	69,05	68,34	36,49	68,35
			Hibrit	73,68	76,47	71,43	73,86	47,37	73,95

Tablo 4.8. B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	63,16	91,18	40,48	56,06	29,82	65,83
			kNN	46,05	100,00	2,38	4,65	2,14	51,19
			Ensemble	63,16	91,18	40,48	56,06	29,82	65,83
			Hibrit	63,16	91,18	40,48	56,06	29,82	65,83
2	2	22	SVMs	64,47	91,18	42,86	58,31	32,14	67,02
			kNN	59,21	100,00	26,19	41,51	24,10	63,10
			Ensemble	52,63	41,18	61,90	49,46	3,12	51,54
			Hibrit	64,47	91,18	42,86	58,31	32,14	67,02
3	3	33	SVMs	76,32	64,71	85,71	73,74	51,28	75,21
			kNN	56,58	82,35	35,71	49,82	17,06	59,03
			Ensemble	59,21	55,88	61,90	58,74	17,74	58,89
			Hibrit	64,47	67,65	61,90	64,65	29,14	64,78
4	4	44	SVMs	77,63	73,53	80,95	77,06	54,63	77,24
			kNN	60,53	79,41	45,24	57,64	23,59	62,32
			Ensemble	57,89	64,71	52,38	57,89	16,71	58,54
			Hibrit	65,79	76,47	57,14	65,41	32,70	66,81
5	5	56	SVMs	68,42	91,18	50,00	64,58	39,20	70,59
			kNN	63,16	76,47	52,38	62,17	27,91	64,43
			Ensemble	61,84	70,59	54,76	61,68	24,73	62,68
			Hibrit	65,79	76,47	57,14	65,41	32,70	66,81
6	6	67	SVMs	72,37	79,41	66,67	72,48	45,19	73,04
			kNN	60,53	67,65	54,76	60,53	21,92	61,20
			Ensemble	61,84	61,76	61,90	61,83	23,47	61,83
			Hibrit	67,11	67,65	66,67	67,15	34,03	67,16
7	7	78	SVMs	75,00	73,53	76,19	74,84	49,58	74,86
			kNN	63,16	58,82	66,67	62,50	25,49	62,75
			Ensemble	60,53	55,88	64,29	59,79	20,17	60,08
			Hibrit	69,74	61,76	76,19	68,22	38,28	68,98

Tablo 4.8. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

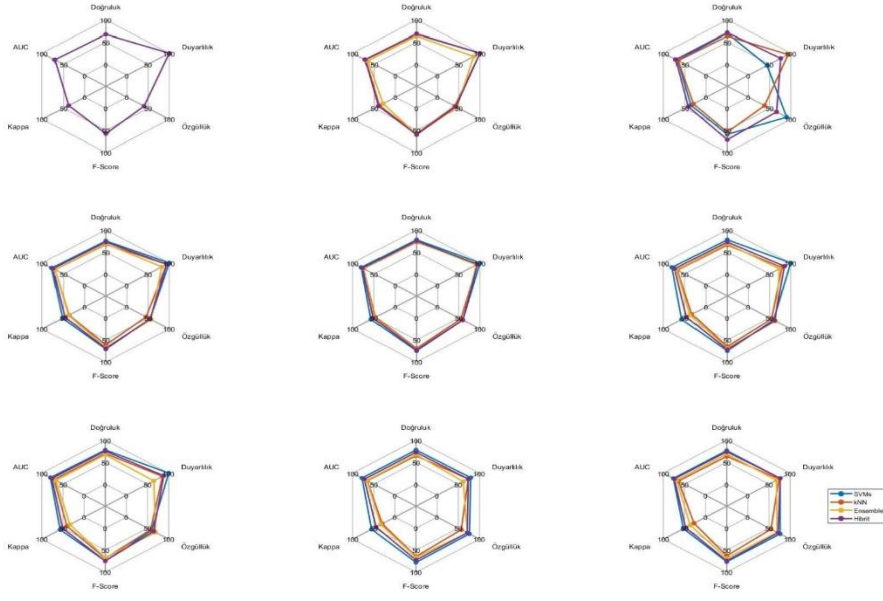
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
8	8	89	SVMs	71,05	73,53	69,05	71,22	42,11	71,29
			kNN	64,47	67,65	61,90	64,65	29,14	64,78
			Ensemble	60,53	61,76	59,52	60,62	21,05	60,64
			Hibrit	72,37	73,53	71,43	72,46	44,58	72,48
9	9	100	SVMs	71,05	70,59	71,43	71,01	41,78	71,01
			kNN	65,79	70,59	61,90	65,96	31,96	66,25
			Ensemble	61,84	61,76	61,90	61,83	23,47	61,83
			Hibrit	69,74	70,59	69,05	69,81	39,31	69,82

Tablo 4.9. C Veri Seti Özet Tablo - En İyi Yöntemler.

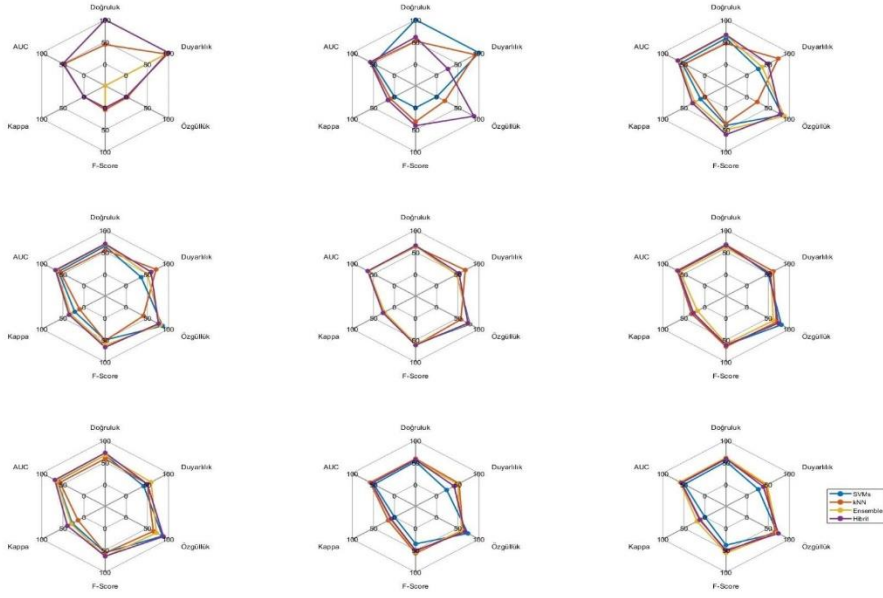
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	100,00	100,00	0,00	0,00	0,00	50,00
			kNN	43,42	94,12	2,38	4,64	0,00	48,25
			Ensemble	100,00	100,00	0,00	0,00	0,00	50,00
			Hibrit	100,00	100,00	0,00	0,00	0,00	50,00
2	2	22	SVMs	100,00	100,00	0,00	0,00	0,00	50,00
			kNN	51,32	91,18	19,05	31,51	9,41	55,11
			Ensemble	60,53	26,47	88,10	40,71	15,43	57,28
			Hibrit	60,53	26,47	88,10	40,71	15,43	57,28
3	3	33	SVMs	57,89	26,47	83,33	40,18	10,32	54,90
			kNN	46,05	73,53	23,81	35,97	0,00	48,67
			Ensemble	64,47	35,29	88,10	50,40	24,56	61,69
			Hibrit	65,79	50,00	78,57	61,11	29,23	64,29
4	4	44	SVMs	63,16	35,29	85,71	50,00	21,99	60,50
			kNN	53,95	70,59	40,48	51,45	10,62	55,53
			Ensemble	67,11	50,00	80,95	61,82	31,75	65,48
			Hibrit	68,42	58,82	76,19	66,39	35,41	67,51
5	5	56	SVMs	64,47	47,06	78,57	58,86	26,29	62,82
			kNN	61,84	67,65	57,14	61,95	24,31	62,39
			Ensemble	63,16	47,06	76,19	58,18	23,78	61,62
			Hibrit	64,47	52,94	73,81	61,66	27,13	63,38

Tablo 4.9. (Devamı) C Veri Seti Özet Tablo - En İyi Yöntemler.

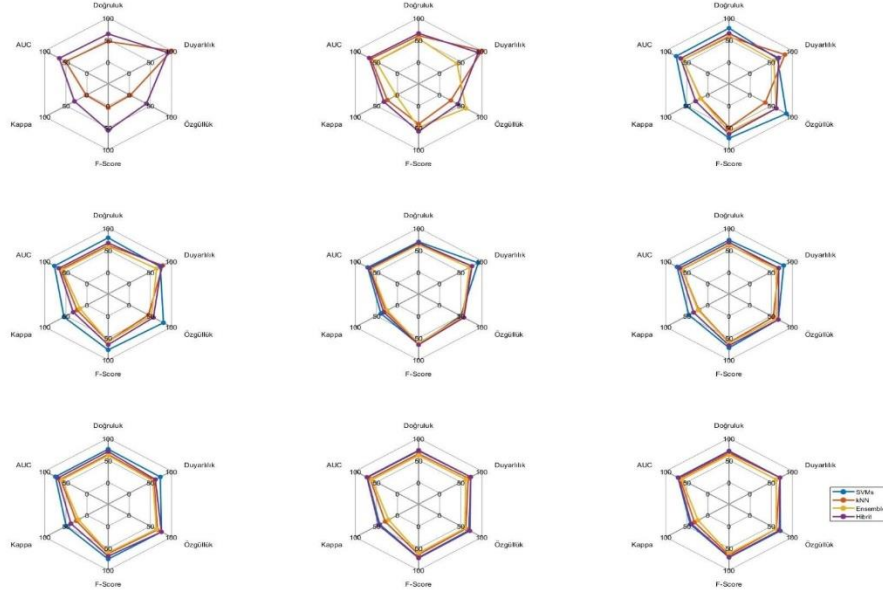
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
6	6	67	SVMs	67,11	50,00	80,95	61,82	31,75	65,48
			kNN	65,79	61,76	69,05	65,20	30,81	65,41
			Ensemble	59,21	55,88	61,90	58,74	17,74	58,89
			Hibrit	64,47	52,94	73,81	61,66	27,13	63,38
7	7	78	SVMs	65,79	41,18	85,71	55,63	27,99	63,45
			kNN	57,89	50,00	64,29	56,25	14,37	57,14
			Ensemble	65,79	58,82	71,43	64,52	30,42	65,13
			Hibrit	71,05	50,00	88,10	63,79	39,42	69,05
8	8	89	SVMs	51,32	23,53	73,81	35,68	0,00	48,67
			kNN	57,89	52,94	61,90	57,07	14,85	57,42
			Ensemble	55,26	50,00	59,52	54,35	9,52	54,76
			Hibrit	55,26	41,18	66,67	50,91	7,98	53,92
9	9	100	SVMs	51,32	26,47	71,43	38,63	0,00	48,95
			kNN	56,58	44,12	66,67	53,10	10,94	55,39
			Ensemble	60,53	47,06	71,43	56,74	18,80	59,24
			Hibrit	57,89	38,24	73,81	50,38	12,39	56,02



Şekil 4.7. A Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.8. B Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.9. C Veri Seti Sınıflandırma Modellerinin Performansı.

4.3. 3. Veri Seti (Coimbra) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması farklı bir veri seti üzerinde geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (9) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanmış, en iyi özellikten başlayarak 9 özellik sıralanmıştır. Bu sıralamaya göre 9 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf”a göre dengesiz bir veri olduğu için oluşturulan alt veri setinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 9 başka bir alt veri seti oluşturulmuştur. 9 alt veri seti, ham sınıf veri setinden 1 veri seti elde edilerek bu veri setinin en iyi özellikten başlayarak 9 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.10’a göre, Tablo 3.38’de yer alan Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar

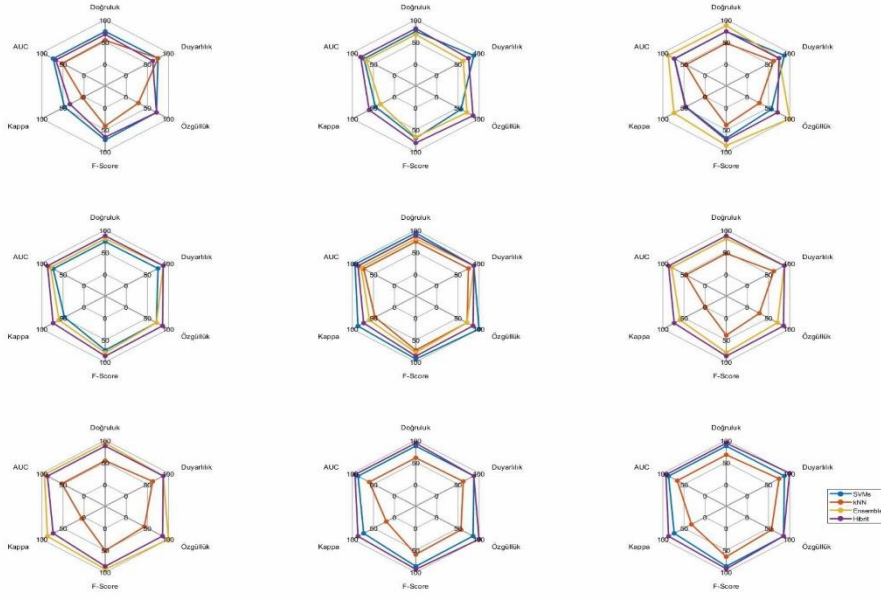
sıralandığında genel olarak artmalar ve azalmalar olduğu görülmüştür. 9 özelliğin sıralaması göz önünde bulundurulduğunda genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Hibrit” olduğu görülmüştür. Hibrit sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 8 özelliğe sahip veri setinde sahip olduğu görülmüştür. Şekil 4.10’da ise en iyi 8 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın özellikle “Hibrit” sınıflandırma yöntemlerinin kapladığı görülmektedir.

Tablo 3.38’e göre, veri setinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre “Duyarlılık” değeri “Doğruluk” değerinden yüksek olduğu için hastalıkların teşhis oranı daha yüksek olduğu gözlemlenmiştir. Böylece sistemin meme kanseri olan kişilerin bulunması için etkili olduğu görülmüştür. Sistem, meme kanseri olan kişilerin kanser teşhisinin hemen bulunmasında etkilidir. Tablo 4.10 ve Şekil 4.9’a göre, Tablo 3.38’de bulunan veri setindeki en iyi 9 özelliğe sahip “Ensemble ve Hibrit” yöntemi %100 ile en çok duyarlılık oranına sahiptir. Tablo 4.10 ve Şekil 4.9’a göre, Tablo 3.38’de bulunan veri setindeki en iyi 3, 5, 7 ve 8 özelliklerine sahip “SVMs, Ensemble ve Hibrit” yöntemleri %100 ile en çok özgüllük oranına sahiptir. Ancak tüm veri setinde uygulanan yöntemlere göre sistemin genel performansında özgüllük değeri duyarlılık değerinden daha düşük olduğu için sistemin sağlıklı kişileri bulması kanserli kişileri bulmasına göre daha düşüktür. Böylece meme kanseri olan bireyin sistemde sağlıklı gözükme olasılığı düşük olup, kanserli hastaların çoğu teşhis edilmeye çalışılmıştır.

Yapılan tüm bu uygulamada tüm parametreler doğrultusunda değerlendirildiğinde doğruluk oranı, Tablo 3.38’de göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı genel olarak hem artıp hem de azalarak değişkenlik gösterdiği görülmektedir. Bu yüzden tutarlı bir özelliğe ait doğruluk oranı söylenememektedir. Özellikle, Tablo 4.10 ve Şekil 4.9’da yer alan veri seti çoğunluğu “SVMs, Ensemble ve Hibrit” sınıflandırma yöntemlerinin olduğu veri setleri en yüksek doğruluk oranlarına sahip veri setleri olarak görülmüştür. En yüksek doğruluk oranı Tablo 4.10 ve Şekil 4.9’daki en iyi 5, 7, 8 ve 9 özelliklerine sahip olan “SVMs, Ensemble ve Hibrit” sınıflandırma yöntemlerinde ise %93,33 ile en yüksek doğruluk oranına sahip olduğu görülmektedir.

Tablo 4.10. Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	73,33	75	71,43	73,17	46,43	73,21
			kNN	53,33	75	28,57	41,38	3,67	51,79
			Ensemble	66,67	62,5	71,43	66,67	33,63	66,96
			Hibrit	66,67	62,5	71,43	66,67	33,63	66,96
2	2	22	SVMs	73,33	87,5	57,14	69,14	45,45	72,32
			kNN	66,67	62,5	71,43	66,67	33,63	66,96
			Ensemble	66,67	62,5	71,43	66,67	33,63	66,96
			Hibrit	80	75	85,71	80	60,18	80,36
3	3	33	SVMs	73,33	87,5	57,14	69,14	45,45	72,32
			kNN	46,67	62,5	28,57	39,22	0	45,54
			Ensemble	86,67	75	100	85,71	73,68	87,5
			Hibrit	73,33	75	71,43	73,17	46,43	73,21
4	4	44	SVMs	73,33	75	71,43	73,17	46,43	73,21
			kNN	80	87,5	71,43	78,65	59,46	79,46
			Ensemble	80	87,5	71,43	78,65	59,46	79,46
			Hibrit	86,67	87,5	85,71	86,6	73,21	86,61
5	5	56	SVMs	93,33	87,5	100	93,33	86,73	93,75
			kNN	73,33	75	71,43	73,17	46,43	73,21
			Ensemble	80	87,5	71,43	78,65	59,46	79,46
			Hibrit	86,67	87,5	85,71	86,6	73,21	86,61
6	6	67	SVMs	86,67	87,5	85,71	86,6	73,21	86,61
			kNN	46,67	62,5	28,57	39,22	0	45,54
			Ensemble	80	87,5	71,43	78,65	59,46	79,46
			Hibrit	86,67	87,5	85,71	86,6	73,21	86,61
7	7	78	SVMs	86,67	87,5	85,71	86,6	73,21	86,61
			kNN	53,33	62,5	42,86	50,85	5,41	52,68
			Ensemble	93,33	87,5	100	93,33	86,73	93,75
			Hibrit	86,67	87,5	85,71	86,6	73,21	86,61
8	8	89	SVMs	86,67	87,5	85,71	86,6	73,21	86,61
			kNN	60	62,5	57,14	59,7	19,64	59,82
			Ensemble	93,33	87,5	100	93,33	86,73	93,75
			Hibrit	93,33	87,5	100	93,33	86,73	93,75
9	9	100	SVMs	86,67	87,5	85,71	86,6	73,21	86,61
			kNN	66,67	75	57,14	64,86	32,43	66,07
			Ensemble	93,33	100	85,71	92,31	86,49	92,86
			Hibrit	93,33	100	85,71	92,31	86,49	92,86



Şekil 4.10. Kullanılan Veri Seti Sınıflandırma Modellerinin Performansı.

4.4. 4. Veri Seti (WPBC) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (33) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanılmış, en iyi özellikten başlayarak 33 özellik sıralanmıştır. Bu sıralamaya göre 33 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf”a göre dengesiz bir veri olduğu için oluşturulan alt veri setlerinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 99 başka bir alt veri seti oluşturulmuştur. 99 alt veri seti, ham sınıf veri setinden 3 farklı veri seti elde edilerek 3 farklı veri setinin en iyi özellikten başlayarak 33 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.11’e göre, Tablo 3.39’da yer alan A Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 33 özelliğe kadar

sıralandığında genel olarak en iyi 3 özelliğe kadar artmış sonrasında genel olarak 33 özelliğe kadar artmalar ve azalmalar olmuştur. 33 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Hibrit” olduğu görülmüştür. Hibrit sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 11 ve 14 özelliklerine sahip veri setlerinde görülmüştür. Şekil 4.11’de ise en iyi 11 ve 14 özelliklerine sahip şekillerde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Hibrit” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.12’ye göre, Tablo 3.39’da yer alan B Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 33 özelliğe kadar sıralandığında genel olarak en iyi 4 özelliğe kadar artmış sonrasında genel olarak 33 özelliğe kadar artmalar ve azalmalar olmuştur. 33 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “SVMs” olduğu görülmüştür. SVMs sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 15 özelliğe sahip veri setinde görülmüştür. Şekil 4.12’de ise en iyi 15 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “SVMs” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.13’e göre, Tablo 3.39’da yer alan C Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 33 özelliğe kadar sıralandığında genel olarak 33 özelliğe kadar artmalar ve azalmalar olmuştur. 33 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Hibrit” olduğu görülmüştür. Hibrit sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 24 özelliğe sahip veri setlerinde görülmüştür. Şekil 4.13’de ise en iyi 24 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Hibrit” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 3.39’a göre, tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre “Duyarlılık” değeri “Doğruluk” değerinden yüksek olduğu için hastalıkların teşhis oranı daha yüksek olduğu gözlemlenmiştir. Böylece sistemin meme kanseri olan kişilerin bulunması için etkili olduğu görülmüştür. Sistem, meme

kanseri olan kişilerin kanser teşhisinin hemen bulunmasında etkilidir. Tablo 4.11, Şekil 4.11'e göre; Tablo 3.39'daki A Veri Setinde en iyi 1 özelliğe sahip "Ensemble" yöntemi, Tablo 4.12, Şekil 4.12'ye göre Tablo 3.39'daki B Veri Setinde en iyi 6 özelliğe sahip "Ensemble" yöntemi, Tablo 4.13, Şekil 4.13'e göre; Tablo 3.39'daki C Veri Setinde en iyi 1, 2, 3, 4 ve 10 özelliklerine sahip "SVMs" sınıflandırma yönteminin kullanıldığı veri setleri %88 ile en yüksek ve en çok duyarlılık oranına sahiptir. Ancak tüm veri setlerinde uygulanan yöntemlere göre sistemin genel performansında sistemin sağlıklı ya da hasta kişileri bulması konusunda kesin yorum yapılamamaktadır.

Yapılan tüm bu uygulamada doğruluk oranı, Tablo 3.39'a göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı genel olarak değişkenlik göstermektedir. Özellikle Tablo 4.11, Şekil 4.11'deki A Veri Setinde bulunan "SVMs" sınıflandırma yöntemi %85,42 ile en yüksek doğruluk oranlarına sahip veri setidir.

Çalışmada Tablo 4.11, Şekil 4.11'de en iyi 3 özellik kullanılarak, Tablo 4.12; Şekil 4.12'de en iyi 4 özellik kullanılarak özellik çıkarımı yapılmıştır. Sistemi 33 özelliğe göre kıyasladığımda bu durum sistemin genel olarak performansı açısından az özellik çıkartılarak daha az iş yükü ile verimli sonuca ulaşılabilceği kanıtlanmış ve görselleştirilmiştir.

Tablo 4.11. A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	3	SVMs	62,50	88,00	34,78	49,86	23,27	61,39
			kNN	60,42	96,00	21,74	35,45	18,28	58,87
			Ensemble	64,58	88,00	39,13	54,17	27,66	63,57
			Hibrit	60,42	92,00	26,09	40,65	18,57	59,04
2	2	6	SVMs	81,25	80,00	82,61	81,28	62,50	81,30
			kNN	66,67	72,00	60,87	65,97	32,98	66,43
			Ensemble	66,67	76,00	56,52	64,83	32,75	66,26
			Hibrit	68,75	72,00	65,22	68,44	37,28	68,61

Tablo 4.11. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
3	3	9	SVMs	85,42	80,00	91,30	85,28	70,93	85,65
			kNN	72,92	76,00	69,57	72,64	45,64	72,78
			Ensemble	75,00	76,00	73,91	74,94	49,91	74,96
			Hibrit	77,08	80,00	73,91	76,84	54,01	76,96
4	4	12	SVMs	83,33	80,00	86,96	83,33	66,72	83,48
			kNN	72,92	76,00	69,57	72,64	45,64	72,78
			Ensemble	77,08	80,00	73,91	76,84	54,01	76,96
			Hibrit	79,17	80,00	78,26	79,12	58,26	79,13
5	5	15	SVMs	81,25	76,00	86,96	81,11	62,63	81,48
			kNN	72,92	76,00	69,57	72,64	45,64	72,78
			Ensemble	68,75	72,00	65,22	68,44	37,28	68,61
			Hibrit	77,08	76,00	78,26	77,11	54,17	77,13
6	6	18	SVMs	81,25	76,00	86,96	81,11	62,63	81,48
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	68,75	76,00	60,87	67,60	37,06	68,43
			Hibrit	79,17	80,00	78,26	79,12	58,26	79,13
7	7	21	SVMs	79,17	76,00	82,61	79,17	58,41	79,30
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	72,92	72,00	73,91	72,94	45,83	72,96
			Hibrit	79,17	80,00	78,26	79,12	58,26	79,13
8	8	24	SVMs	81,25	80,00	82,61	81,28	62,50	81,30
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	66,67	76,00	56,52	64,83	32,75	66,26
			Hibrit	75,00	80,00	69,57	74,42	49,74	74,78
9	9	27	SVMs	81,25	80,00	82,61	81,28	62,50	81,30
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	68,75	68,00	69,57	68,77	37,50	68,78
			Hibrit	77,08	80,00	73,91	76,84	54,01	76,96
10	10	30	SVMs	72,92	72,00	73,91	72,94	45,83	72,96
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	68,75	80,00	56,52	66,24	36,84	68,26
			Hibrit	75,00	76,00	73,91	74,94	49,91	74,96
11	11	33	SVMs	75,00	76,00	73,91	74,94	49,91	74,96
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	77,08	84,00	69,57	76,10	53,85	76,78
			Hibrit	81,25	80,00	82,61	81,28	62,50	81,30
12	12	36	SVMs	75,00	76,00	73,91	74,94	49,91	74,96
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	75,00	80,00	69,57	74,42	49,74	74,78
			Hibrit	77,08	80,00	73,91	76,84	54,01	76,96

Tablo 4.11. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
13	13	39	SVMs	75,00	76,00	73,91	74,94	49,91	74,96
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	75,00	80,00	69,57	74,42	49,74	74,78
			Hibrit	79,17	84,00	73,91	78,63	58,12	78,96
14	14	42	SVMs	75,00	72,00	78,26	75,00	50,09	75,13
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	72,92	68,00	78,26	72,77	46,02	73,13
			Hibrit	81,25	80,00	82,61	81,28	62,50	81,30
15	15	45	SVMs	77,08	76,00	78,26	77,11	54,17	77,13
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	75,00	72,00	78,26	75,00	50,09	75,13
			Hibrit	81,25	84,00	78,26	81,03	62,37	81,13
16	16	48	SVMs	68,75	72,00	65,22	68,44	37,28	68,61
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	64,58	72,00	56,52	63,33	28,67	64,26
			Hibrit	79,17	80,00	78,26	79,12	58,26	79,13
17	17	52	SVMs	66,67	72,00	60,87	65,97	32,98	66,43
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	64,58	72,00	56,52	63,33	28,67	64,26
			Hibrit	79,17	80,00	78,26	79,12	58,26	79,13
18	18	55	SVMs	68,75	72,00	65,22	68,44	37,28	68,61
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	62,50	64,00	60,87	62,40	24,87	62,43
			Hibrit	79,17	84,00	73,91	78,63	58,12	78,96
19	19	58	SVMs	70,83	76,00	65,22	70,20	41,36	70,61
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	72,92	76,00	69,57	72,64	45,64	72,78
			Hibrit	79,17	84,00	73,91	78,63	58,12	78,96
20	20	61	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	68,75	84,00	52,17	64,37	36,62	68,09
			Hibrit	75,00	80,00	69,57	74,42	49,74	74,78
21	21	64	SVMs	68,75	72,00	65,22	68,44	37,28	68,61
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	58,33	56,00	60,87	58,33	16,81	58,43
			Hibrit	72,92	80,00	65,22	71,86	45,45	72,61
22	22	67	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	64,58	68,00	60,87	64,24	28,92	64,43
			Hibrit	77,08	80,00	73,91	76,84	54,01	76,96

Tablo 4.11. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
23	23	70	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	72,92	72,00	73,91	72,94	45,83	72,96
			Ensemble	70,83	80,00	60,87	69,14	41,16	70,43
			Hibrit	79,17	84,00	73,91	78,63	58,12	78,96
24	24	73	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	60,42	64,00	56,52	60,03	20,56	60,26
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78
25	25	76	SVMs	72,92	72,00	73,91	72,94	45,83	72,96
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	58,33	56,00	60,87	58,33	16,81	58,43
			Hibrit	75,00	80,00	69,57	74,42	49,74	74,78
26	26	79	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	68,75	84,00	52,17	64,37	36,62	68,09
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78
27	27	82	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	77,08	76,00	78,26	77,11	54,17	77,13
			Hibrit	77,08	80,00	73,91	76,84	54,01	76,96
28	28	85	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	68,75	84,00	52,17	64,37	36,62	68,09
			Hibrit	72,92	80,00	65,22	71,86	45,45	72,61
29	29	88	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	68,75	84,00	52,17	64,37	36,62	68,09
			Hibrit	72,92	80,00	65,22	71,86	45,45	72,61
30	30	91	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	66,67	80,00	52,17	63,16	32,51	66,09
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78
31	31	94	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	66,67	80,00	52,17	63,16	32,51	66,09
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78
32	32	97	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	66,67	80,00	52,17	63,16	32,51	66,09
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78
33	33	100	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	66,67	80,00	52,17	63,16	32,51	66,09
			Hibrit	77,08	84,00	69,57	76,10	53,85	76,78

Tablo 4.12. B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	3	SVMs	58,33	88,00	26,09	40,24	14,44	57,04
			kNN	58,33	88,00	26,09	40,24	14,44	57,04
			Ensemble	50,00	32,00	69,57	43,84	1,54	50,78
			Hibrit	62,50	84,00	39,13	53,39	23,54	61,57
2	2	6	SVMs	72,92	72,00	73,91	72,94	45,83	72,96
			kNN	70,83	72,00	69,57	70,76	41,57	70,78
			Ensemble	66,67	72,00	60,87	65,97	32,98	66,43
			Hibrit	72,92	68,00	78,26	72,77	46,02	73,13
3	3	9	SVMs	70,83	80,00	60,87	69,14	41,16	70,43
			kNN	54,17	44,00	65,22	52,55	9,12	54,61
			Ensemble	66,67	64,00	69,57	66,67	33,45	66,78
			Hibrit	72,92	72,00	73,91	72,94	45,83	72,96
4	4	12	SVMs	75,00	80,00	69,57	74,42	49,74	74,78
			kNN	54,17	44,00	65,22	52,55	9,12	54,61
			Ensemble	68,75	80,00	56,52	66,24	36,84	68,26
			Hibrit	79,17	84,00	73,91	78,63	58,12	78,96
5	5	15	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	54,17	44,00	65,22	52,55	9,12	54,61
			Ensemble	68,75	80,00	56,52	66,24	36,84	68,26
			Hibrit	75,00	76,00	73,91	74,94	49,91	74,96
6	6	18	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	62,50	68,00	56,52	61,73	24,61	62,26
			Ensemble	75,00	88,00	60,87	71,96	49,38	74,43
			Hibrit	72,92	84,00	60,87	70,59	45,26	72,43
7	7	21	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	62,50	68,00	56,52	61,73	24,61	62,26
			Ensemble	68,75	84,00	52,17	64,37	36,62	68,09
			Hibrit	68,75	84,00	52,17	64,37	36,62	68,09
8	8	24	SVMs	68,75	64,00	73,91	68,60	37,72	68,96
			kNN	62,50	68,00	56,52	61,73	24,61	62,26
			Ensemble	70,83	84,00	56,52	67,57	40,95	70,26
			Hibrit	66,67	76,00	56,52	64,83	32,75	66,26
9	9	27	SVMs	72,92	76,00	69,57	72,64	45,64	72,78
			kNN	62,50	68,00	56,52	61,73	24,61	62,26
			Ensemble	70,83	84,00	56,52	67,57	40,95	70,26
			Hibrit	72,92	88,00	56,52	68,83	45,07	72,26
10	10	30	SVMs	62,50	68,00	56,52	61,73	24,61	62,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	64,58	72,00	56,52	63,33	28,67	64,26
			Hibrit	64,58	72,00	56,52	63,33	28,67	64,26

Tablo 4.12. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
11	11	33	SVMs	62,50	64,00	60,87	62,40	24,87	62,43
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	60,42	72,00	47,83	57,47	20,00	59,91
			Hibrit	60,42	68,00	52,17	59,04	20,28	60,09
12	12	36	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	64,00	47,83	54,74	11,89	55,91
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
13	13	39	SVMs	72,92	76,00	69,57	72,64	45,64	72,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	64,58	68,00	60,87	64,24	28,92	64,43
			Hibrit	70,83	76,00	65,22	70,20	41,36	70,61
14	14	42	SVMs	64,58	60,00	69,57	64,43	29,41	64,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	56,00	56,52	56,26	12,50	56,26
			Hibrit	60,42	60,00	60,87	60,43	20,83	60,43
15	15	45	SVMs	75,00	76,00	73,91	74,94	49,91	74,96
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	54,17	56,00	52,17	54,02	8,17	54,09
			Hibrit	64,58	68,00	60,87	64,24	28,92	64,43
16	16	48	SVMs	72,92	72,00	73,91	72,94	45,83	72,96
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	68,00	43,48	53,04	11,58	55,74
			Hibrit	64,58	72,00	56,52	63,33	28,67	64,26
17	17	52	SVMs	72,92	76,00	69,57	72,64	45,64	72,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	60,42	72,00	47,83	57,47	20,00	59,91
			Hibrit	70,83	80,00	60,87	69,14	41,16	70,43
18	18	55	SVMs	70,83	76,00	65,22	70,20	41,36	70,61
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	58,33	64,00	52,17	57,49	16,23	58,09
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
19	19	58	SVMs	64,58	68,00	60,87	64,24	28,92	64,43
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	60,42	68,00	52,17	59,04	20,28	60,09
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
20	20	61	SVMs	68,75	76,00	60,87	67,60	37,06	68,43
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	52,00	60,87	56,09	12,80	56,43
			Hibrit	68,75	72,00	65,22	68,44	37,28	68,61

Tablo 4.12. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
21	21	64	SVMs	62,50	68,00	56,52	61,73	24,61	62,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	54,17	52,00	56,52	54,17	8,49	54,26
			Hibrit	60,42	60,00	60,87	60,43	20,83	60,43
22	22	67	SVMs	64,58	68,00	60,87	64,24	28,92	64,43
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	58,33	56,00	60,87	58,33	16,81	58,43
			Hibrit	64,58	68,00	60,87	64,24	28,92	64,43
23	23	70	SVMs	68,75	72,00	65,22	68,44	37,28	68,61
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	56,00	56,52	56,26	12,50	56,26
			Hibrit	64,58	68,00	60,87	64,24	28,92	64,43
24	24	73	SVMs	62,50	68,00	56,52	61,73	24,61	62,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	54,17	48,00	60,87	53,67	8,81	54,43
			Hibrit	62,50	64,00	60,87	62,40	24,87	62,43
25	25	76	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	60,42	64,00	56,52	60,03	20,56	60,26
			Hibrit	64,58	72,00	56,52	63,33	28,67	64,26
26	26	79	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	36,00	78,26	49,32	13,99	57,13
			Hibrit	62,50	60,00	65,22	62,50	25,13	62,61
27	27	82	SVMs	66,67	68,00	65,22	66,58	33,22	66,61
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	48,00	65,22	55,30	13,10	56,61
			Hibrit	56,25	52,00	60,87	56,09	12,80	56,43
28	28	85	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	58,33	64,00	52,17	57,49	16,23	58,09
			Hibrit	66,67	76,00	56,52	64,83	32,75	66,26
29	29	88	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	56,25	56,00	56,52	56,26	12,50	56,26
			Hibrit	56,25	60,00	52,17	55,81	12,20	56,09
30	30	91	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	64,58	60,00	69,57	64,43	29,41	64,78
			Hibrit	58,33	60,00	56,52	58,21	16,52	58,26

Tablo 4.12. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
31	31	94	SVMs	68,75	68,00	69,57	68,77	37,50	68,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	62,50	60,00	65,22	62,50	25,13	62,61
			Hibrit	68,75	72,00	65,22	68,44	37,28	68,61
32	32	97	SVMs	68,75	68,00	69,57	68,77	37,50	68,78
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	60,42	60,00	60,87	60,43	20,83	60,43
			Hibrit	62,50	68,00	56,52	61,73	24,61	62,26
33	33	100	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	60,42	64,00	56,52	60,03	20,56	60,26
			Ensemble	66,67	60,00	73,91	66,23	33,68	66,96
			Hibrit	66,67	68,00	65,22	66,58	33,22	66,61

Tablo 4.13. C Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	3	SVMs	64,58	88,00	39,13	54,17	27,66	63,57
			kNN	64,58	96,00	30,43	46,22	27,14	63,22
			Ensemble	58,33	64,00	52,17	57,49	16,23	58,09
			Hibrit	62,50	92,00	30,43	45,74	22,99	61,22
2	2	6	SVMs	64,58	88,00	39,13	54,17	27,66	63,57
			kNN	52,08	56,00	47,83	51,59	3,83	51,91
			Ensemble	52,08	60,00	43,48	50,42	3,50	51,74
			Hibrit	56,25	60,00	52,17	55,81	12,20	56,09
3	3	9	SVMs	60,42	88,00	30,43	45,23	18,86	59,22
			kNN	37,50	48,00	26,09	33,80	0,00	37,04
			Ensemble	58,33	52,00	65,22	57,86	17,10	58,61
			Hibrit	54,17	64,00	43,48	51,78	7,53	53,74
4	4	12	SVMs	66,67	88,00	43,48	58,20	32,04	65,74
			kNN	37,50	48,00	26,09	33,80	0,00	37,04
			Ensemble	58,33	52,00	65,22	57,86	17,10	58,61
			Hibrit	58,33	64,00	52,17	57,49	16,23	58,09
5	5	15	SVMs	60,42	84,00	34,78	49,19	19,15	59,39
			kNN	37,50	48,00	26,09	33,80	0,00	37,04
			Ensemble	56,25	52,00	60,87	56,09	12,80	56,43
			Hibrit	54,17	60,00	47,83	53,23	7,85	53,91
6	6	18	SVMs	60,42	76,00	43,48	55,31	19,72	59,74
			kNN	54,17	68,00	39,13	49,68	7,21	53,57
			Ensemble	47,92	44,00	52,17	47,74	0,00	48,09
			Hibrit	56,25	60,00	52,17	55,81	12,20	56,09

Tablo 4.13. (Devamı) C Veri Seti Özet Tablo - En İyi Yöntemler.

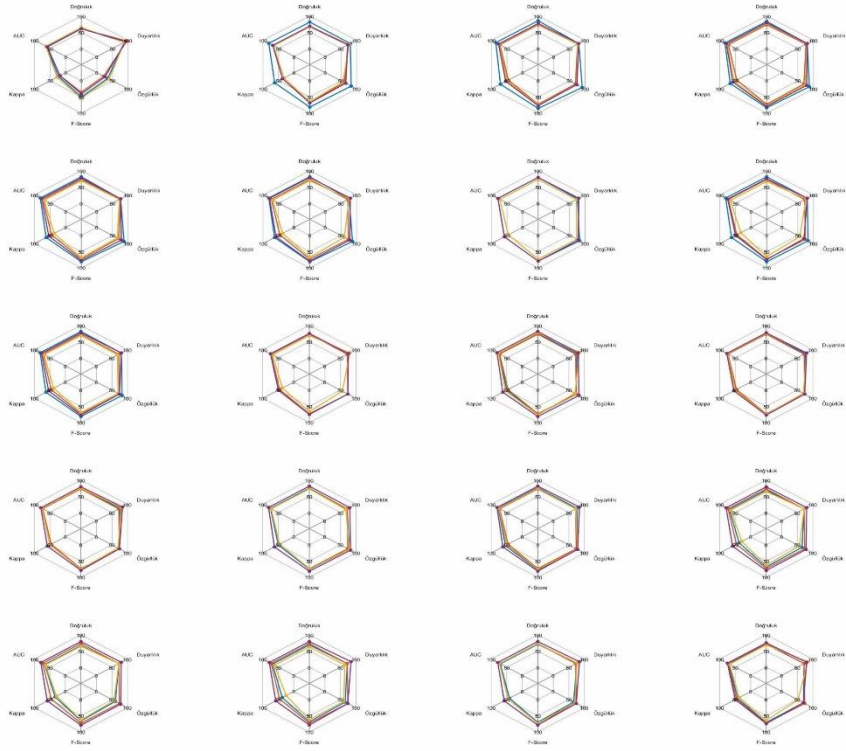
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
7	7	21	SVMs	62,50	72,00	52,17	60,50	24,34	62,09
			kNN	54,17	68,00	39,13	49,68	7,21	53,57
			Ensemble	58,33	56,00	60,87	58,33	16,81	58,43
			Hibrit	68,75	72,00	65,22	68,44	37,28	68,61
8	8	24	SVMs	64,58	68,00	60,87	64,24	28,92	64,43
			kNN	54,17	68,00	39,13	49,68	7,21	53,57
			Ensemble	58,33	56,00	60,87	58,33	16,81	58,43
			Hibrit	66,67	68,00	65,22	66,58	33,22	66,61
9	9	27	SVMs	68,75	76,00	60,87	67,60	37,06	68,43
			kNN	54,17	68,00	39,13	49,68	7,21	53,57
			Ensemble	58,33	52,00	65,22	57,86	17,10	58,61
			Hibrit	68,75	68,00	69,57	68,77	37,50	68,78
10	10	30	SVMs	70,83	88,00	52,17	65,51	40,74	70,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	56,25	52,00	60,87	56,09	12,80	56,43
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
11	11	33	SVMs	64,58	80,00	47,83	59,86	28,17	63,91
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	60,42	56,00	65,22	60,26	21,11	60,61
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
12	12	36	SVMs	64,58	80,00	47,83	59,86	28,17	63,91
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	60,42	56,00	65,22	60,26	21,11	60,61
			Hibrit	64,58	72,00	56,52	63,33	28,67	64,26
13	13	39	SVMs	66,67	84,00	47,83	60,95	32,28	65,91
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	52,08	48,00	56,52	51,91	4,50	52,26
			Hibrit	64,58	72,00	56,52	63,33	28,67	64,26
14	14	42	SVMs	50,00	40,00	60,87	48,28	0,86	50,43
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	58,33	44,00	73,91	55,16	17,67	58,96
			Hibrit	56,25	44,00	69,57	53,91	13,40	56,78
15	15	45	SVMs	64,58	76,00	52,17	61,87	28,42	64,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	60,42	52,00	69,57	59,51	21,38	60,78
			Hibrit	68,75	68,00	69,57	68,77	37,50	68,78
16	16	48	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	50,00	48,00	52,17	50,00	0,17	50,09
			Hibrit	56,25	60,00	52,17	55,81	12,20	56,09

Tablo 4.13. (Devamı) C Veri Seti Özet Tablo - En İyi Yöntemler.

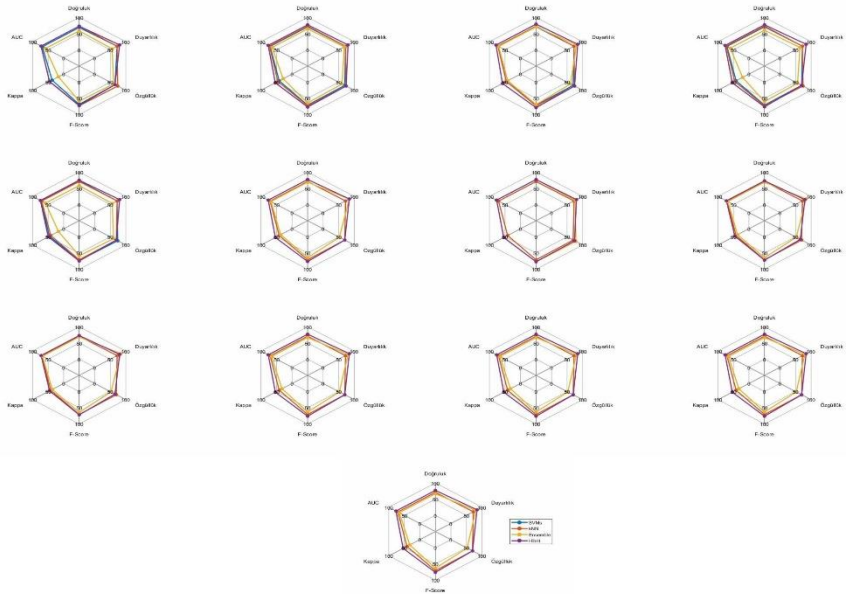
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
17	17	52	SVMs	64,58	76,00	52,17	61,87	28,42	64,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	52,08	52,00	52,17	52,09	4,17	52,09
			Hibrit	60,42	68,00	52,17	59,04	20,28	60,09
18	18	55	SVMs	62,50	72,00	52,17	60,50	24,34	62,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	43,75	36,00	52,17	42,60	0,00	44,09
			Hibrit	60,42	68,00	52,17	59,04	20,28	60,09
19	19	58	SVMs	64,58	72,00	56,52	63,33	28,67	64,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	54,17	52,00	56,52	54,17	8,49	54,26
			Hibrit	58,33	68,00	47,83	56,16	15,94	57,91
20	20	61	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	58,33	44,00	73,91	55,16	17,67	58,96
			Hibrit	60,42	56,00	65,22	60,26	21,11	60,61
21	21	64	SVMs	72,92	76,00	69,57	72,64	45,64	72,78
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	60,42	52,00	69,57	59,51	21,38	60,78
			Hibrit	72,92	76,00	69,57	72,64	45,64	72,78
22	22	67	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	68,75	56,00	82,61	66,75	38,14	69,30
			Hibrit	68,75	64,00	73,91	68,60	37,72	68,96
23	23	70	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	62,50	52,00	73,91	61,05	25,65	62,96
			Hibrit	66,67	64,00	69,57	66,67	33,45	66,78
24	24	73	SVMs	70,83	72,00	69,57	70,76	41,57	70,78
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	70,83	64,00	78,26	70,42	41,97	71,13
			Hibrit	75,00	76,00	73,91	74,94	49,91	74,96
25	25	76	SVMs	72,92	80,00	65,22	71,86	45,45	72,61
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	56,25	48,00	65,22	55,30	13,10	56,61
			Hibrit	66,67	72,00	60,87	65,97	32,98	66,43
26	26	79	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	52,08	32,00	73,91	44,66	5,80	52,96
			Hibrit	64,58	56,00	73,91	63,72	29,66	64,96

Tablo 4.13. (Devamı) C Veri Seti Özet Tablo - En İyi Yöntemler.

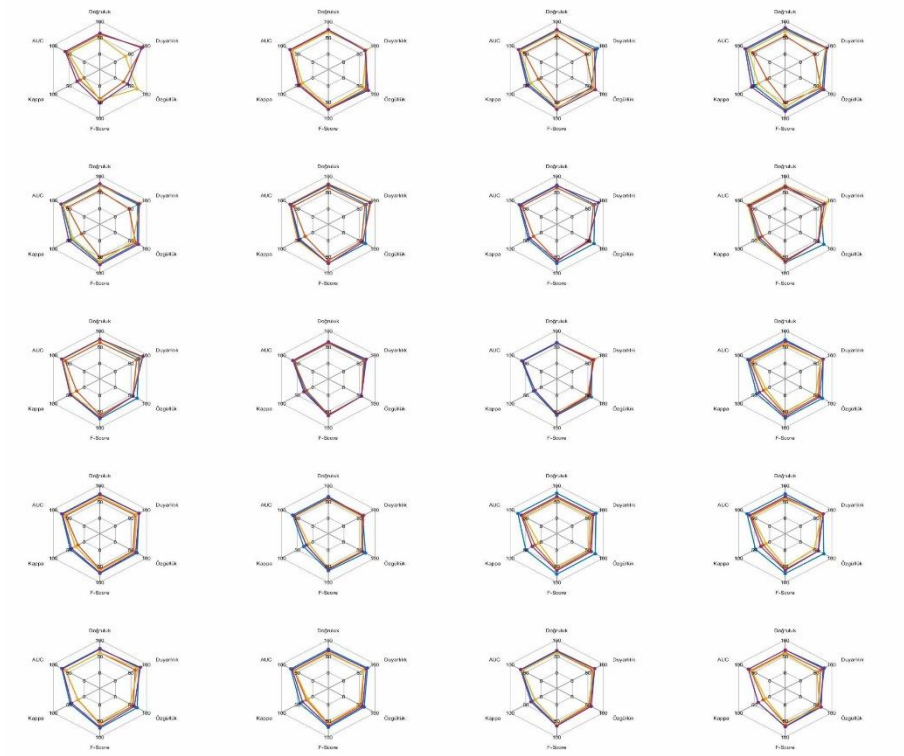
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
27	27	82	SVMs	62,50	72,00	52,17	60,50	24,34	62,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	54,17	48,00	60,87	53,67	8,81	54,43
			Hibrit	66,67	68,00	65,22	66,58	33,22	66,61
28	28	85	SVMs	66,67	76,00	56,52	64,83	32,75	66,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	64,58	56,00	73,91	63,72	29,66	64,96
			Hibrit	66,67	68,00	65,22	66,58	33,22	66,61
29	29	88	SVMs	62,50	72,00	52,17	60,50	24,34	62,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	50,00	60,00	39,13	47,37	0,00	49,57
			Hibrit	62,50	80,00	43,48	56,34	23,81	61,74
30	30	91	SVMs	58,33	64,00	52,17	57,49	16,23	58,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	58,33	48,00	69,57	56,80	17,38	58,78
			Hibrit	56,25	56,00	56,52	56,26	12,50	56,26
31	31	94	SVMs	68,75	72,00	65,22	68,44	37,28	68,61
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	54,17	48,00	60,87	53,67	8,81	54,43
			Hibrit	68,75	68,00	69,57	68,77	37,50	68,78
32	32	97	SVMs	58,33	64,00	52,17	57,49	16,23	58,09
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	62,50	56,00	69,57	62,05	25,39	62,78
			Hibrit	60,42	68,00	52,17	59,04	20,28	60,09
33	33	100	SVMs	60,42	64,00	56,52	60,03	20,56	60,26
			kNN	58,33	68,00	47,83	56,16	15,94	57,91
			Ensemble	54,17	56,00	52,17	54,02	8,17	54,09
			Hibrit	60,42	64,00	56,52	60,03	20,56	60,26



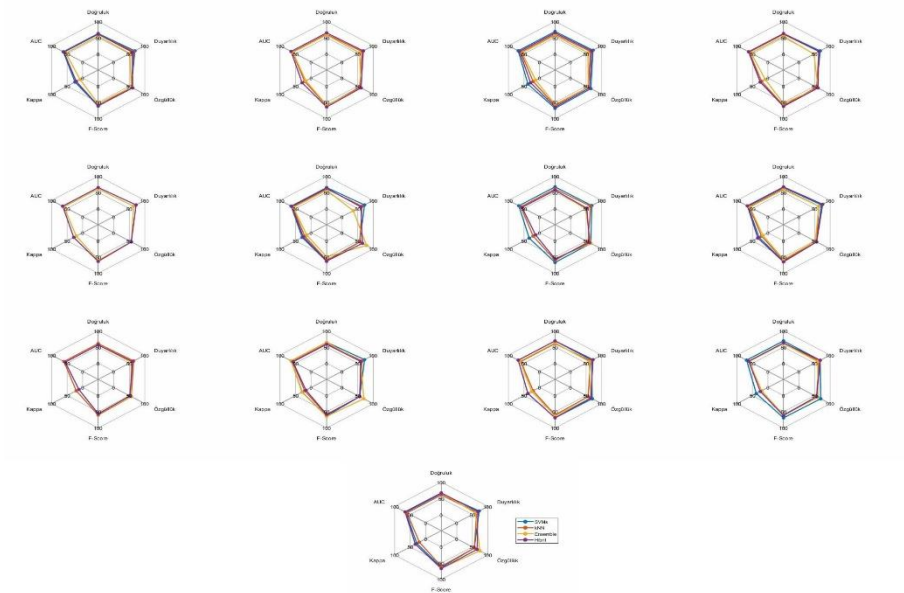
Şekil 4.11. A Veri Seti Sınıflandırma Modellerinin Performansı.



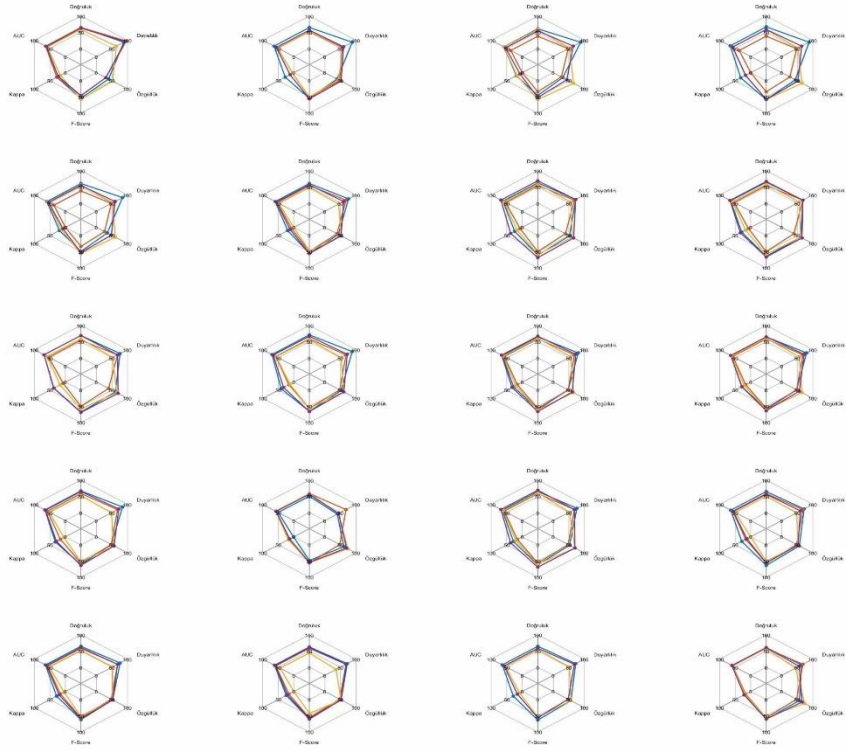
Şekil 4.11. (Devamı) A Veri Seti Sınıflandırma Modellerinin Performansı.



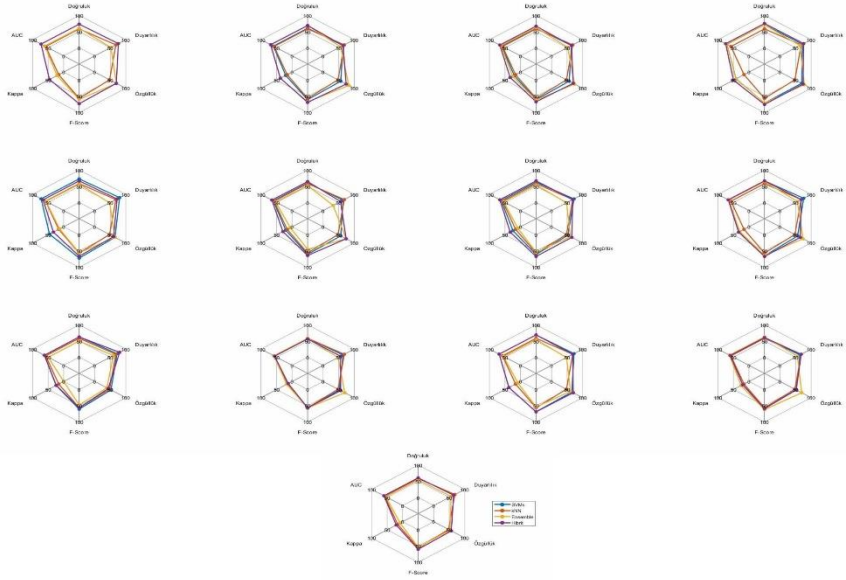
Şekil 4.12. B Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.12. (Devamı) B Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.13. C Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.13. (Devamı) C Veri Seti Sınıflandırma Modellerinin Performansı.

4.5. 5. Veri Seti (WBC) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (9) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanmış, en iyi özellikten başlayarak 9 özellik sıralanmıştır. Bu sıralamaya göre 9 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf”a göre dengesiz bir veri olduğu için oluşturulan alt veri setlerinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 18 başka bir alt veri seti oluşturulmuştur. 18 alt veri seti, ham sınıf veri setinden 2 farklı veri seti elde edilerek 2 farklı veri setinin en iyi özellikten başlayarak 9 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.14’e göre; Tablo 3.40’de yer alan A Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar sıralandığında genel olarak artmalar ve azalmalar olmuştur. 9 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 8 özelliğe sahip veri setinde görülmüştür. Şekil 4.14’de ise en iyi 8 özelliğe sahip şekilde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 4.15’e göre; Tablo 3.40’da yer alan B Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 9 özelliğe kadar sıralandığında genel olarak artmalar ve azalmalar olmuştur. 9 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 6 özelliğe sahip veri setinde görülmüştür. Şekil 4.15’de ise en iyi 6 özelliğe sahip şekilde diğer özelliklere sahip

şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 3.40’a göre, tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre “Duyarlılık” değeri “Doğruluk” değerine göre hastalıkların teşhis oranı hakkında kesin yorum yapılamamaktadır. Tablo 4.15 ve Şekil 4.15’e göre; Tablo 3.40’daki B Veri Setinde en iyi 3, 5, 6, 7, 8 ve 9 özelliklerine sahip “Ensemble” sınıflandırma yöntemlerinin kullanıldığı veri setleri %97,83 ile en yüksek ve en çok duyarlılık oranına sahiptir. Tablo 4.14 ve Şekil 4.14’e göre; Tablo 3.40’daki A Veri Setinde en iyi 6 özelliğe sahip “SVMs” sınıflandırma yönteminin kullanıldığı veri seti ile Tablo 4.15 ve Şekil 4.15’e göre; Tablo 3.40’daki B Veri Setinde en iyi 5 özelliğe sahip “SVMs”, en iyi 6 özelliğe sahip “SVMs, kNN, Ensemble ve Hibrit”, en iyi 7 özelliğe sahip “SVMs”, en iyi 8 özelliğe sahip “SVMs” ve en iyi 9 özelliğe sahip “SVMs” sınıflandırma yöntemlerinin kullanıldığı veri setleri %100 ile en yüksek ve en çok özgüllük oranına sahiptir. Tüm veri setlerinde uygulanan yöntemlere göre sistemin genel performansında özgüllük değeri duyarlılık değerinden daha yüksek olduğu için sistemin sağlıklı kişileri bulması kanserli kişileri bulmasına göre daha yüksektir. Böylece meme kanseri olmayan sağlıklı hastaların çoğu teşhis edilmeye çalışılmıştır.

Yapılan tüm bu uygulamada doğruluk oranı, Tablo 3.40’da göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı genel olarak yüksek olup, sağlık alanında bu sistemin kullanılması oldukça elverişli olduğunu göstermektedir. Özellikle Tablo 4.15 ve Şekil 4.15’de yer alan B veri setinde en iyi 6 özelliğe sahip “Ensemble” sınıflandırma yönteminde %98,94 ile en yüksek doğruluk oranına sahiptir.

Tablo 4.14. A Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	88,09	87,83	88,33	88,08	76,16	88,08
			kNN	51,49	94,78	10,00	18,09	4,70	52,39
			Ensemble	88,09	87,83	88,33	88,08	76,16	88,08
			Hibrit	88,09	87,83	88,33	88,08	76,16	88,08
2	2	22	SVMs	92,77	93,04	92,50	92,77	85,53	92,77
			kNN	80,85	93,04	69,17	79,35	61,89	81,11
			Ensemble	94,89	94,78	95,00	94,89	89,78	94,89
			Hibrit	92,77	93,04	92,50	92,77	85,53	92,77

Tablo 4.14. (Devamı) A Veri Seti Özet Tablo - En İyi Yöntemler.

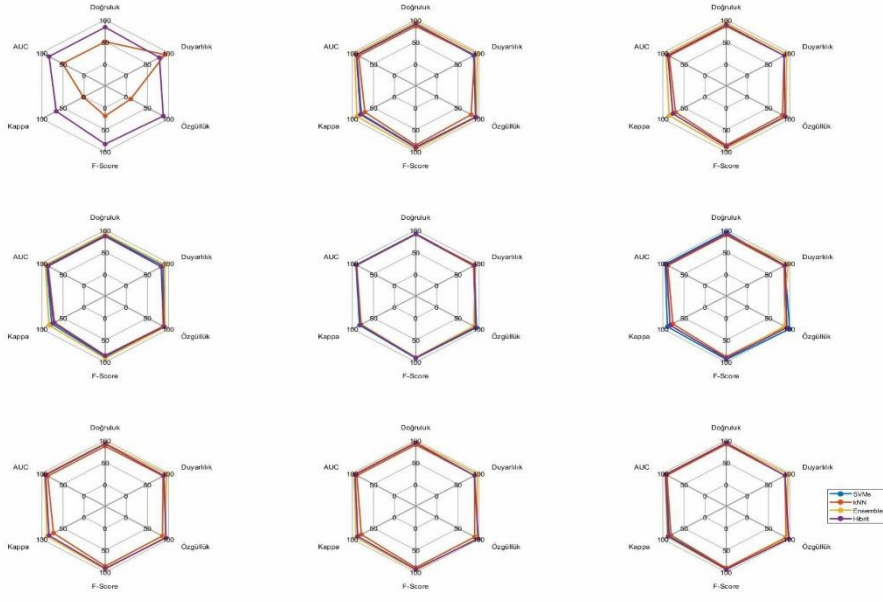
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
3	3	33	SVMs	93,62	93,04	94,17	93,60	87,23	93,61
			kNN	85,96	93,04	79,17	85,55	71,99	86,11
			Ensemble	92,34	97,39	87,50	92,18	84,71	92,45
			Hibrit	91,06	93,04	89,17	91,06	82,14	91,11
4	4	44	SVMs	93,62	93,04	94,17	93,60	87,23	93,61
			kNN	88,09	91,30	85,00	88,04	76,19	88,15
			Ensemble	92,77	94,78	90,83	92,77	85,54	92,81
			Hibrit	91,49	92,17	90,83	91,50	82,98	91,50
5	5	56	SVMs	92,77	92,17	93,33	92,75	85,52	92,75
			kNN	88,94	93,04	85,00	88,84	77,90	89,02
			Ensemble	94,04	96,52	91,67	94,03	88,09	94,09
			Hibrit	92,77	92,17	93,33	92,75	85,52	92,75
6	6	67	SVMs	93,19	92,17	94,17	93,16	86,37	93,17
			kNN	88,51	93,04	84,17	88,38	77,06	88,61
			Ensemble	92,77	95,65	90,00	92,74	85,54	92,83
			Hibrit	92,77	92,17	93,33	92,75	85,52	92,75
7	7	78	SVMs	91,91	92,17	91,67	91,92	83,83	91,92
			kNN	91,49	93,04	90,00	91,50	82,98	91,52
			Ensemble	94,89	95,65	94,17	94,90	89,79	94,91
			Hibrit	93,19	93,04	93,33	93,19	86,38	93,19
8	8	89	SVMs	91,06	91,30	90,83	91,07	82,12	91,07
			kNN	92,77	92,17	93,33	92,75	85,52	92,75
			Ensemble	95,32	95,65	95,00	95,32	90,64	95,33
			Hibrit	93,19	92,17	94,17	93,16	86,37	93,17
9	9	100	SVMs	91,49	92,17	90,83	91,50	82,98	91,50
			kNN	93,19	93,04	93,33	93,19	86,38	93,19
			Ensemble	94,89	96,52	93,33	94,90	89,79	94,93
			Hibrit	93,19	93,04	93,33	93,19	86,38	93,19

Tablo 4.15. B Veri Seti Özet Tablo - En İyi Yöntemler.

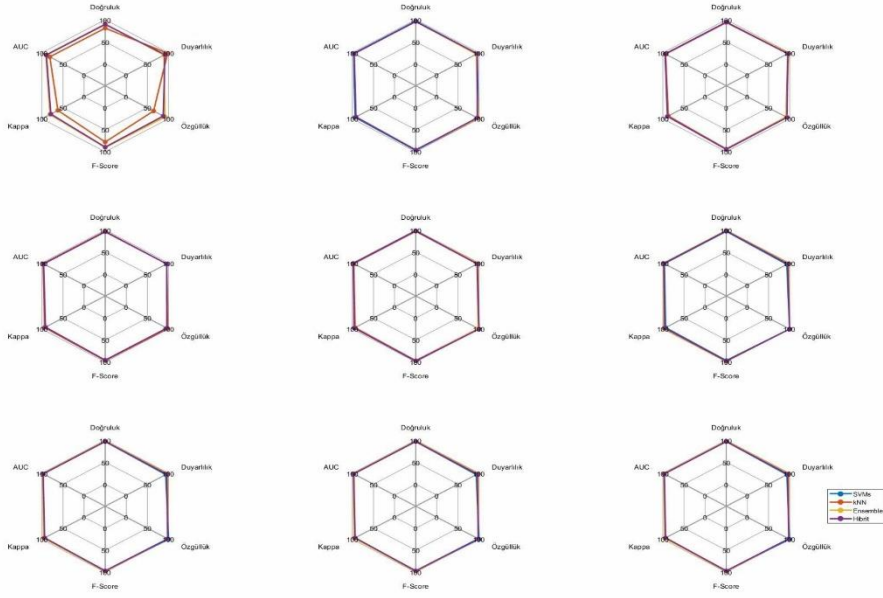
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	11	SVMs	88,94	89,57	88,33	88,95	77,87	88,95
			kNN	83,83	95,65	72,50	82,48	67,81	84,08
			Ensemble	88,94	89,57	88,33	88,95	77,87	88,95
			Hibrit	88,94	89,57	88,33	88,95	77,87	88,95
2	2	22	SVMs	96,17	93,91	98,33	96,07	92,33	96,12
			kNN	94,47	98,26	90,83	94,40	88,95	94,55
			Ensemble	94,89	93,91	95,83	94,86	89,78	94,87
			Hibrit	96,17	95,65	96,67	96,16	92,34	96,16

Tablo 4.15. (Devamı) B Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
3	3	33	SVMs	97,02	93,91	100,00	96,86	94,03	96,96
			kNN	96,17	95,65	96,67	96,16	92,34	96,16
			Ensemble	96,17	95,65	96,67	96,16	92,34	96,16
			Hibrit	96,60	94,78	98,33	96,53	93,18	96,56
4	4	44	SVMs	97,02	98,26	95,83	97,03	94,04	97,05
			kNN	96,60	98,26	95,00	96,60	93,19	96,63
			Ensemble	96,60	96,52	96,67	96,59	93,19	96,59
			Hibrit	97,45	98,26	96,67	97,46	94,89	97,46
5	5	56	SVMs	97,02	98,26	95,83	97,03	94,04	97,05
			kNN	95,74	98,26	93,33	95,73	91,49	95,80
			Ensemble	94,47	97,39	91,67	94,44	88,95	94,53
			Hibrit	97,02	99,13	95,00	97,02	94,05	97,07
6	6	67	SVMs	95,74	97,39	94,17	95,75	91,49	95,78
			kNN	96,17	98,26	94,17	96,17	92,34	96,21
			Ensemble	96,60	97,39	95,83	96,61	93,19	96,61
			Hibrit	97,02	98,26	95,83	97,03	94,04	97,05
7	7	78	SVMs	94,47	96,52	92,50	94,47	88,94	94,51
			kNN	97,45	97,39	97,50	97,45	94,89	97,45
			Ensemble	91,91	96,52	87,50	91,79	83,85	92,01
			Hibrit	96,17	97,39	95,00	96,18	92,34	96,20
8	8	89	SVMs	95,74	96,52	95,00	95,75	91,49	95,76
			kNN	97,02	97,39	96,67	97,03	94,04	97,03
			Ensemble	91,91	96,52	87,50	91,79	83,85	92,01
			Hibrit	95,74	97,39	94,17	95,75	91,49	95,78
9	9	100	SVMs	94,47	95,65	93,33	94,48	88,94	94,49
			kNN	96,60	97,39	95,83	96,61	93,19	96,61
			Ensemble	94,47	97,39	91,67	94,44	88,95	94,53
			Hibrit	95,32	98,26	92,50	95,29	90,65	95,38



Şekil 4.14. A Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.15. B Veri Seti Sınıflandırma Modellerinin Performansı.

4.6. 6. Veri Seti (WDBC) için Sonuçlar

Bu çalışmadaki amaç meme kanseri teşhisinde yapay zeka yöntemleri ile kural tabanlı teşhis algoritması farklı bir veri seti üzerinde geliştirmektir. Çalışmada bulunan veri setinde bireylere ait olan teşhis sonuçları konulurken klinik bulguların (30) tamamından yararlanılmıştır. Çalışmada “Eta” özellik seçme algoritması veri setine uygulanmış, en iyi özellikten başlayarak 30 özellik sıralanmıştır. Bu sıralamaya göre 30 farklı alt veri seti oluşturulmuştur. Veri setinde bulunan “Sınıf”a göre dengesiz bir veri olduğu için oluşturulan alt veri setinde veri dengeleme işlemi yapılmıştır. Veri dengeleme işlemi sonucu 30 başka bir alt veri seti oluşturulmuştur. 30 alt veri seti, ham sınıf veri setinden 1 veri seti elde edilerek bu veri setinin en iyi özellikten başlayarak 30 özelliğe göre ayrı ayrı değerlendirilmesi sonucu oluşturulmuştur. Dengelenen tüm veri setleri “Ensemble, kNN, SVMs ve Hibrit Yapay Zeka” yöntemleri kullanılarak sınıflandırılmıştır. Elde edilen sonuçlara göre Meme Kanseri için teşhis algoritması oluşturulmuştur.

Tablo 4.16’ya göre, Tablo 3.41’de yer alan Veri Seti tüm sınıflandırma modellerine (SVMs, kNN, Ensemble ve Hibrit) göre “Doğruluk, Duyarlılık, Özgüllük, F-Score, Kappa ve AUC” performans değerlendirme kriterleri en iyi özellikten 30 özelliğe kadar sıralandığında genel olarak artmalar ve azalmalar olmuştur. 30 özelliğin sıralamasına göre genel olarak bakılırsa; en başarılı sınıflandırma yönteminin “Ensemble” olduğu görülmüştür. Ensemble sınıflandırma yöntemi içerisinde en yüksek performans değerlendirme kriterlerine sahip değerler en iyi 16, 17, 22, 28 ve 29 özelliklerine sahip veri setlerinde görülmüştür. Şekil 4.16’da ise en iyi 16, 17, 22, 28 ve 29 özelliklerine sahip şekillerde diğer özelliklere sahip şekiller göz önünde bulundurulduğunda görsel olarak en fazla alanın “Ensemble” sınıflandırma yönteminin kapladığı görülmektedir.

Tablo 3.41’e göre, tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin genel performansına göre “Duyarlılık” değeri “Doğruluk” değerinden yüksek olduğu için hastalıkların teşhis oranı daha yüksek olduğu gözlemlenmiştir. Böylece sistemin meme kanseri olan kişilerin bulunması için etkili olduğu görülmüştür. Sistem, meme kanseri olan kişilerin kanser teşhisinin hemen bulunmasında etkilidir. Tablo 4.16, Şekil 4.16’da bulunan en iyi 16, 17, 22, 28 ve 29 özelliklerine sahip veri setlerinde “Ensemble” yöntemi %100 ile en çok duyarlılık oranını sahiptir. Tablo 4.16, Şekil 4.16’da bulunan en iyi 16, 17, 22, 28 ve 29 özelliklerine sahip veri setlerinde “Ensemble” yöntemi %100 ile en çok özgüllük oranını sahiptir. Tüm veri setinde

uygulanan yöntemlere göre sistemin genel performansında duyarlılık değeri özgüllük değerinden daha yüksek olduğu için sistemin kanserli kişileri bulması sağlıklı kişileri bulmasına göre daha yüksektir. Böylece meme kanserinde hasta kişilerin çoğu teşhis edilmeye çalışılmıştır.

Yapılan tüm bu uygulamada doğruluk oranı, Tablo 3.41’de göre tüm veri setlerinde uygulanan yöntemler incelendiğinde sistemin doğruluk oranı “Ensemble” yöntemi %100 ile en çok doğruluk oranını sahiptir. Bu durum da sistemin sağlık alanında kullanılmasında oldukça elverişli yapmaktadır. Özellikle Tablo 4.16, Şekil 4.16’daki en iyi 16, 17, 22, 28 ve 29 özelliklerine sahip “Ensemble” sınıflandırma yöntemlerinde %100 ile en yüksek doğruluk oranına sahiptir.

Tablo 4.16. Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	1	3	SVMs	91,67	93,02	90,24	91,61	83,31	91,63
			kNN	88,1	88,37	87,8	88,09	76,18	88,09
			Ensemble	86,9	83,72	90,24	86,86	73,84	86,98
			Hibrit	89,29	88,37	90,24	89,3	78,57	89,31
2	2	7	SVMs	92,86	95,35	90,24	92,73	85,69	92,8
			kNN	89,29	88,37	90,24	89,3	78,57	89,31
			Ensemble	90,48	88,37	92,68	90,48	80,96	90,53
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85
3	3	10	SVMs	92,86	95,35	90,24	92,73	85,69	92,8
			kNN	89,29	88,37	90,24	89,3	78,57	89,31
			Ensemble	92,86	90,7	95,12	92,86	85,72	92,91
			Hibrit	94,05	95,35	92,68	94	88,08	94,02
4	4	13	SVMs	92,86	95,35	90,24	92,73	85,69	92,8
			kNN	84,52	88,37	80,49	84,25	68,98	84,43
			Ensemble	91,67	90,7	92,68	91,68	83,33	91,69
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85
5	5	17	SVMs	92,86	93,02	92,68	92,85	85,71	92,85
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	94,05	93,02	95,12	94,06	88,1	94,07
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85
6	6	20	SVMs	91,67	93,02	90,24	91,61	83,31	91,63
			kNN	86,9	90,7	82,93	86,64	73,75	86,81
			Ensemble	95,24	93,02	97,56	95,24	90,48	95,29
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85

Tablo 4.16. (Devamı) Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.

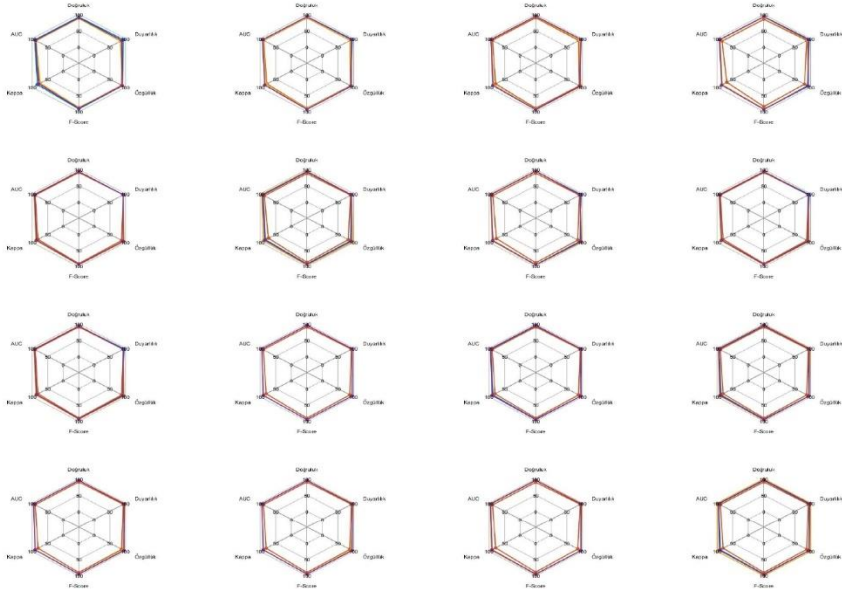
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
7	7	23	SVMs	94,05	95,35	92,68	94	88,08	94,02
			kNN	88,1	90,7	85,37	87,95	76,15	88,03
			Ensemble	95,24	93,02	97,56	95,24	90,48	95,29
			Hibrit	94,05	93,02	95,12	94,06	88,1	94,07
8	8	27	SVMs	92,86	95,35	90,24	92,73	85,69	92,8
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	94,05	93,02	95,12	94,06	88,1	94,07
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85
9	9	30	SVMs	92,86	95,35	90,24	92,73	85,69	92,8
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	94,05	93,02	95,12	94,06	88,1	94,07
			Hibrit	92,86	93,02	92,68	92,85	85,71	92,85
10	10	33	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	95,24	95,35	95,12	95,24	90,47	95,24
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
11	11	37	SVMs	94,05	93,02	95,12	94,06	88,1	94,07
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	95,24	95,35	95,12	95,24	90,47	95,24
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
12	12	40	SVMs	94,05	93,02	95,12	94,06	88,1	94,07
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	96,43	97,67	95,12	96,38	92,85	96,4
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
13	13	43	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	95,24	97,67	92,68	95,11	90,46	95,18
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
14	14	47	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	95,24	97,67	92,68	95,11	90,46	95,18
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
15	15	50	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	89,29	93,02	85,37	89,03	78,52	89,19
			Ensemble	96,43	97,67	95,12	96,38	92,85	96,4
			Hibrit	95,24	95,35	95,12	95,24	90,47	95,24
16	16	53	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	100	100	100	100	100	100
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4

Tablo 4.16. (Devamı) Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.

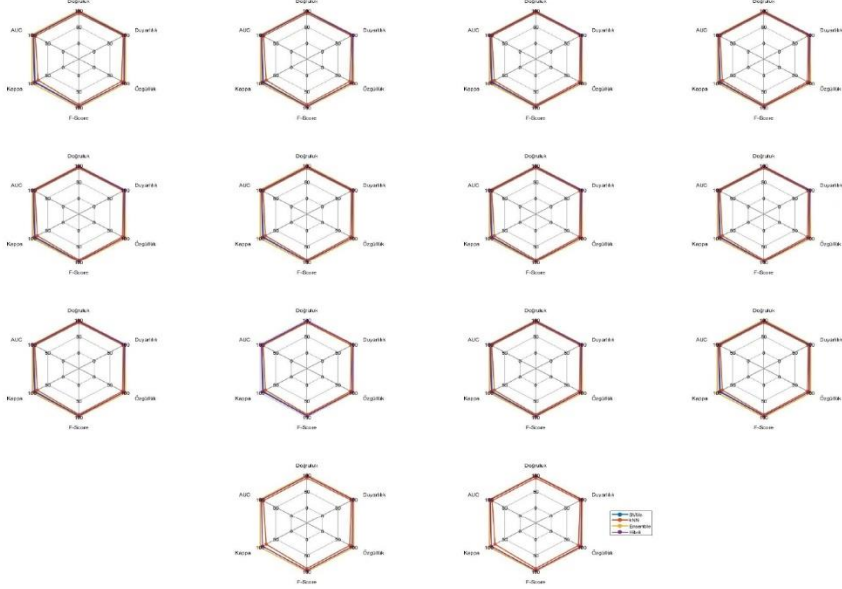
L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
17	17	57	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	100	100	100	100	100	100
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
18	18	60	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
19	19	63	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
20	20	67	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
21	21	70	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
22	22	73	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	100	100	100	100	100	100
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
23	23	77	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
24	24	80	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
25	25	83	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
26	26	87	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	96,43	95,35	97,56	96,44	92,86	96,45
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4

Tablo 4.16. (Devamı) Kullanılan Veri Seti Özet Tablo - En İyi Yöntemler.

L	FN	FP	Model	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
27	27	90	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	98,81	97,67	100	98,82	97,62	98,84
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
28	28	93	SVMs	95,24	95,35	95,12	95,24	90,47	95,24
			kNN	91,67	93,02	90,24	91,61	83,31	91,63
			Ensemble	100	100	100	100	100	100
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
29	29	97	SVMs	96,43	97,67	95,12	96,38	92,85	96,4
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	100	100	100	100	100	100
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4
30	30	100	SVMs	96,43	97,67	95,12	96,38	92,85	96,4
			kNN	90,48	93,02	87,8	90,34	80,92	90,41
			Ensemble	98,81	100	97,56	98,77	97,62	98,78
			Hibrit	96,43	97,67	95,12	96,38	92,85	96,4



Şekil 4.16. Kullanılan Veri Seti Sınıflandırma Modellerinin Performansı.



Şekil 4.16. (Devamı) Kullanılan Veri Seti Sınıflandırma Modellerinin Performansı.

5. TARTIŞMA VE KARAR

Bu çalışmada makine öğrenmesi doğrultusunda Ensemble, kNN, SVMs ve Hibrit Yapay Zeka gibi sınıflandırma işlemleri uygulanmıştır. Elde edilen sonuçlar ile Meme Kanseri teşhis algoritması oluşturulmuştur. Literatürde yapılan çalışmalar Meme Kanseri teşhisinde farklı makine öğrenmesi ve özellik seçme algoritmaları kullanarak farklı özelliklerdeki veri setleri üzerinde çalışıldığı görülmüştür.

Literatürde makine öğrenmesi teknikleri olarak Yapay Sinir Ağları (ANN), Destek Vektör Makineleri (SVMs), Naive Bayes (NB), Sınıflandırma ve Regresyon Ağaçları (CART), k-En yakın Komşu (kNN), Doğrusal Regresyon (LR) ve Çok Katmanlı Algılayıcı (MLP), Rastgele Orman (Random Forest), Aşırı Güçlendirme (Extreme Boost), Karar Ağaçları (C4.5), Lojistik Regresyon, Doğrusal Ayrımcı Analiz (Linear Discriminant Analysis), Boosting ve AdaBoost, Bagging Algoritması, IBk (Bazı Parametrelerle Öznitelik Tabanlı Öğrenme Bazı), Rastgele Komite Algoritması kullanılmıştır.

Literatürde makine öğrenmesi modelinin performans değerlendirme için F-Score, AUC, ROC eğrisi altındaki alan, doğruluk (accuracy), duyarlılık değeri (recall), duyarlılık (sensitivity), özgüllük (specificity), kappa istatistiği, TP Oranı (Doğru Pozitif Oran), FP Oranı (Yanlış Pozitif Oran), MCC, Alıcı İşletim Karakteristikleri Alanı (AUC), zaman karmaşıklığı (Time Complexity), Lift Eğrisi (Lift Curve) ve Kalibrasyon Grafiği (Calibration Plot), Rekürsif Özellik İyileştirme Tekniği (Recursive Feature Elimination) gibi farklı performans metrikleri kullanılmıştır.

Çalışmamızda da literatürdeki gibi doğruluk oranı, duyarlılık, özgüllük, kappa sayısı, F-Score, AUC gibi performans değerlendirme kriterleri uygulanmıştır (Uçar ve ark., 2017, 2018).

Çalışmamızda meme kanseri ile ilgili 6 farklı veri seti kullanılmıştır.

Çalışmamızda elde edilen sonuçlarda tüm veri setleri üzerindeki değerler kıyaslandığında 1. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, en yüksek özgüllük oranı Ensemble yöntemi ile %98,7 ve en yüksek duyarlılık oranı birçok yöntem ile %100 olarak bulunmuştur. 2. veri setine göre en yüksek

doğruluk oranı SVMs yöntemi ile 77,63, en yüksek özgüllük oranı Hibrit ve Ensemble yöntemi ile %88,10 ve en yüksek duyarlılık oranı tüm yöntemler ile %100 olarak bulunmuştur. 3. veri setine göre en yüksek doğruluk oranı SVMs, Ensemble ve Hibrit yöntemi ile %99,3, en yüksek özgüllük oranı SVMs, Ensemble ve Hibrit yöntemi ile %100 ve en yüksek duyarlılık oranı Ensemble ve Hibrit yöntemi ile %100 olarak bulunmuştur. 4. veri setine göre en yüksek doğruluk oranı SVMs yöntemi ile %85,42, en yüksek özgüllük oranı SVMs ile %91.3 ve en yüksek duyarlılık oranı SVMs ve Ensemble yöntemi ile %88 olarak bulunmuştur. 5. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %98,94, en yüksek özgüllük oranı tüm yöntemler ile %100 ve en yüksek duyarlılık oranı yöntemi ile Ensemble %97,83 olarak bulunmuştur. 6. veri setine göre en yüksek doğruluk oranı Ensemble yöntemi ile %100, en yüksek özgüllük oranı Ensemble yöntemi ile %100 ve en yüksek duyarlılık oranı Ensemble yöntemi ile %100 olarak bulunmuştur.

Literatürdeki çalışmalarda R programlama, Weka , Spark ve Python gibi farklı platformlarda makine öğrenmesi algoritmaları geliştirilmiştir (Bharati ve ark., 2018; Keleş, 2019; Sharma ve ark., 2018). Bu platformlarda uygulanan makine öğrenmesi algoritmaları bu çalışmadakine göre daha az etkili olduğu görülmüştür.

Literatürde yapılan çalışmaların büyük çoğunluğu çalışmamızda kullanılan 6. Veri Seti ile aynı olan Wisconsin Diagnostic Breast Cancer (WDBC) veri seti üzerinde gerçekleşmiştir (Abdar ve ark., 2020; Bharat ve ark., 2018; Fatima ve ark., 2020).

Diğer çalışmalar ise WDBC'den farklı WPBC, WBC, Coimbra, UMMC, Using three machine learning techniques for predicting breast cancer, Antenna, Gene Expression vb. veri setleri üzerinde gerçekleşmiştir (LG ve AT, 2013; Maliha ve ark., 2019).

Çalışmamızda olduğu gibi literatürdeki bazı çalışmalarda özellik azaltılarak alt veri kümeleri oluşturma ve veri temizleme, özellik seçimi, özellik çıkarımı gibi ortak aşamalar uygulanmış ancak yüksek doğruluk oranlarına ulaşamamıştır (Israni, 2019; Memon ve ark., 2019; Said ve ark., 2018).

Literatürde bazı çalışmalarda veri seti %80 eğitim, %20 test olarak bölünmüştür. Literatürdeki bu çalışmalarda olduğu gibi çoğu çalışmada özellik azaltılarak alt veri kümeleri oluşturulmuş ve özellik seçimi yapılmıştır (Chaurasia ve Pal, 2020). Literatürde bir diğer çalışmada eğitim ve test veri setleri %66'ya %33 olarak bölünmüştür. Böylece deneysel sonuçlarda SVMs %96,9957 ile en iyi doğruluk

performansını göstermiştir (Bayrak ve ark., 2019). Ancak literatürdeki bu çalışmaların çoğu modelinde doğruluk %90'nın üzerinde olmasına rağmen çalışmamızdaki 1. veri seti, 2. veri seti ve 4. veri setinde %50 eğitim - %50 test veri seti, 3. veri setinde %85 eğitim - %15 test veri seti, 5. veri seti ve 6. veri setinde %80 eğitim - %20 test veri seti uygulaması yapılmıştır. Sonuç olarak literatür ile tüm veri setlerinde bulunan doğruluk, duyarlılık, özgüllük, F-Score, Kappa ve AUC değerleri kıyaslandığında 4. veri seti hariç aynı veri setleri ile yapılan literatürdeki çalışmalar ile uygulamamızdaki çalışmamız kıyaslandığında literatürdeki çalışmaların hepsinin değerlerinin düşük çıktığı görülmüştür. Dolayısı ile 1. veri seti, 2. veri seti, 3. veri seti, 5. veri seti ve 6. veri setinde yapmış olduğumuz makine öğrenmesi uygulaması meme kanseri teşhisinde çok üstün sonuçlara ulaştığı görülmüştür.

Literatürde en yüksek doğruluk oranı WDBC veri setinde SVMs yöntemi ile %99,68, WPBC veri setinde %99,68 bulunmuştur (Abdel ve Eldeib, 2016; Fatima ve ark., 2020). Bu değerler çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 6. veri seti ismi altında WDBC veri seti kullanarak en yüksek doğruluk oranı Ensemble yöntemi ile %100, 4. veri seti ismi altında WPBC veri seti kullanarak en yüksek doğruluk oranı SVMs yöntemi ile %85,42 bulunmuştur. Çalışmamızda en yüksek doğruluk oranı 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100 bulunmuştur.

Literatürde en yüksek duyarlılık değeri WPBC veri setinde %99,68 bulunmuştur (Abdel ve Eldeib, 2016). Bu değer çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 4. veri seti ismi altında WPBC veri seti kullanarak en yüksek duyarlılık oranı SVMs ve Ensemble yöntemleri ile %88 bulunmuştur. Çalışmamızda en yüksek duyarlılık değeri 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100 bulunmuştur.

Literatürde en yüksek özgüllük değeri WPBC veri setinde %99,47 bulunmuştur (Abdel ve Eldeib, 2016). Bu değer çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 4. veri seti ismi altında WPBC veri seti kullanarak en yüksek özgüllük oranı SVMs yöntemi ile %91,3 bulunmuştur. Çalışmamızda en yüksek duyarlılık değeri 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100, 3. veri seti ismi altında Coimbra veri seti kullanarak SVMs, Ensemble ve Hibrit yöntemleri ile %100, 1. veri seti ismi altında UMMC veri seti kullanarak %100 bulunmuştur.

Literatürde en yüksek F-Score değeri WDBC ve WBC veri setinde %98,1 bulunmuştur (Abdar ve ark., 2020; Israni, 2019). Bu değerler çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 6. veri seti ismi altında WDBC veri seti kullanarak en yüksek F-Score değeri Ensemble yöntemi ile %100, 5. veri seti ismi altında WBC veri seti kullanarak en yüksek F-Score değeri Ensemble yöntemi ile %98,91 bulunmuştur. Çalışmamızda en yüksek F-Score değeri 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100 bulunmuştur.

Literatürde en yüksek Kappa değeri Coimbra veri setinde %70,39 bulunmuştur (Sen, 2021). Bu değer çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 3. veri seti ismi altında Coimbra veri seti kullanarak en yüksek Kappa değeri SVMs, Ensemble ve Hibrit yöntemleri ile %86,73 bulunmuştur. Çalışmamızda en yüksek Kappa değeri 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100 bulunmuştur.

Literatürde en yüksek AUC değeri WBC veri setinde %99,5 bulunmuştur (Israni, 2019). Bu değer çalışmamızda ise aynı veri seti kullanarak kıyaslandığında 5. veri seti ismi altında WBC veri seti kullanarak en yüksek AUC değeri Ensemble yöntemi ile %98,91 bulunmuştur. Çalışmamızda en yüksek AUC değeri 6. veri seti ismi altında WDBC veri seti kullanarak Ensemble yöntemi ile %100 bulunmuştur.

Literatürde en yüksek doğruluk oranının bulunduğu çalışma ile çalışmamız kıyaslandığında ise literatürdeki çalışmada Spark ortamında SVMs algoritması ile %99,68 bulunan doğruluk oranı çalışmamızda 6. veri seti ile yapılan uygulama hariç tüm veri setlerinin sahip olduğu doğruluk oranlarından yüksek çıktığı görülmüştür. Ancak literatürdeki CNN, SAE ve SSAE gibi derin öğrenme tekniklerinin kullanıldığı bu çalışmada doğruluk oranı çalışmamızda 1. veri seti ve 5. veri seti adı altında kullanılan en yüksek değere sahip 4 makine öğrenme algoritmasının sadece 1'inin üzerinde çıkmıştır. Çalışmamızda 1. veri seti ve 5. veri seti adı altında en yüksek değerde geriye kalan makine öğrenmesi algoritmalarının 3'ü derin öğrenme teknikleri ile kullanılan algoritmaların doğruluk oranından daha fazla çıktığı görülmüştür (Fatima ve ark., 2020).

Meme kanseri için ANN, SVMs ve NB üzerinde özellik seçme ve özellik çıkarma teknikleri gibi bu çalışmadaki ortak aşamalar ve ortak veri seti (WDBC) kullanılan literatürdeki bir çalışmada en yüksek özgüllük oranı %99,07 ve en yüksek duyarlılık

oranı %98,41 bulunmuştur. Bu değerler araştırma yapılan literatürdeki en yüksek özgüllük ve duyarlılık değerler olduğu görülmüştür (Omondiagbe ve ark., 2019). Literatürde kullanılan veri seti, çalışmamızda 6. veri seti (WDBC) olarak yer almıştır. Çalışmamızda aynı veri seti kullanarak uygulanan yöntemler ile en yüksek özgüllük oranı %100 ve en yüksek duyarlılık oranı %100 bulunmuştur. Aynı zamanda çalışmamızda farklı veri seti üzerinde, 1. veri seti (UMMC) olarak kullanılan veri setinde en yüksek özgüllük oranı %98,7 ve duyarlılık oranı %100 bulunmuştur. Böylece çalışmamızda hem aynı hem de farklı veri setleri üzerinde meme kanseri teşhisinde hastalıklı ve sağlıklı kişilerin bulunmasında güvenilir bir çalışma elde edilebileceği ortaya koyulmuştur.

Meme kanseri teşhisi için 6 farklı makine öğrenmesi tekniği olan; CART, SVMs, NB, kNN, LR and MLP algoritmalar karşılaştırılmıştır. Çalışmada veri seti %80 eğitim, %20 test olarak bölünmüştür. Bu çalışmada olduğu gibi bazı çalışmalarda özellik azaltılarak alt veri kümeleri oluşturulmuş ve özellik seçimi yapılmıştır (Chaurasia ve Pal, 2020). Ancak literatürdeki bu çalışmanın çoğu modelinde doğruluk %90'nın üzerinde değildir. Çalışmamızda da %80 eğitim, %20 test veri seti uygulanarak literatürdeki aynı veri seti ile 6. veri setindeki (WDBC) uygulamamızda Ensemble yöntemi ile bulunan en yüksek doğruluk oranı %100 bulunmuştur. Sonuç olarak literatürdeki bu çalışma yapılan uygulamamızın doğruluk oranlarını geçememiştir. Bu durum da çalışmamızın sağlık alanında daha uygulanabilir ve elverişli olduğunu göstermektedir.

Bazı çalışmalarda uygulamamızda kullanılan veri setiyle aynı veri seti kullanılmasıyla beraber çoğunluğu benzer makine öğrenmesi algoritmaları kullanılmış ve en yüksek doğruluk oranı Random Forest yöntemi ile %82,70, en yüksek özgüllük oranı SVMs yöntemi ile %84 ve en yüksek duyarlılık oranı Extreme Boost yöntemi ile %84 olarak bulunmuştur (Ganggayah ve ark., 2019). Çalışmamızda, literatürde kullanılan veri seti ile aynı veri seti kullanılarak 1. veri seti (UMMC) adı altında uygulaması gerçekleştirilmiştir. Yapılan çalışmamızda elde edilen sonuçlarda ise en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, en yüksek özgüllük oranı Ensemble yöntemi ile %98,7 ve en yüksek duyarlılık oranı birçok yöntem ile %100 olarak bulunmuştur. Bu durum da uygulamamızın farklı makine öğrenmesi teknikleri ile daha yüksek yüzdelerde doğruluk, özgüllük ve duyarlılık oranları bulunabileceğini ispatlamaktadır. Böylece çalışmamızın meme kanseri teşhisinde hastalıklı ve sağlıklı

kişilerin bulunmasında güvenilir bir çalışma olduğunu tekrar ortaya koyarak sağlık alanında daha uygulanabilir ve elverişli bir çalışma olduğunu göstermektedir.

Çalışmamızdaki 6. veri seti (WDBC) ile aynı veri seti kullanılarak yapılan başka bir çalışmada ise meme kanseri sınıflandırma tahmini için ANN ve SVMs kullanılmıştır. Bu yöntemler çeşitli aşamalarla WEKA kullanılarak gerçekleştirilmiştir. Eğitim ve test veri setleri %66'ya %33 olarak bölünmüştür. Deneysel sonuçlarda SVMs %96,9957 ile en iyi doğruluk performansını göstermiştir (Bayrak ve ark., 2019). Uygulamamızda literatürdeki bu çalışmadan farklı platformda, daha az makine öğrenmesi algoritması uygulayarak, test ve eğitim yüzdeleri farklı olmasına rağmen %80 eğitim ve %20 test veri seti uygulanmıştır. Aynı veri seti kullanılarak 6. Veri seti adı altında yapılan uygulamamızda en yüksek doğruluk oranı %100 bulunmuştur. Aynı zamanda uygulamamızda farklı veri setleri kullanarak yapılmış diğer çalışmalar 5. Veri seti (WBC) adı altında yapılan çalışmada en yüksek doğruluk değeri %97,45, 1. Veri seti (UMMC) adı altında yapılan çalışmada en yüksek doğruluk değeri %99,3'e yaklaşamamıştır. Bu durum da hem aynı hem de farklı veri setleri üzerinde uygulamasını yapmış olduğumuz çalışmalarımızın sağlık alanında daha üstün sonuçlara ulaşabileceğini göstermektedir.

Çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak yapılan farklı bir çalışmada Random Forest (RF), NB, SVMs ve kNN gibi makine öğrenmesi algoritmaları kullanılmıştır. Özellik seçme ve çıkarma uygulamasından sonra makine öğrenmesi algoritmaları kullanılmıştır (kNN, SVMs, NB and Random Forest). WEKA programı ile gerçekleştirilen uygulamada veri setindeki etiket ise iyi huylu ve kötü huylu olarak sınıflandırılmıştır. Ancak bazı değerlerin olmamasından dolayı veri seti sayısı düşürülmüştür. Literatürdeki bu çalışmada farklı platformda yapmış olduğumuz çalışmaya benzer şekilde veri dengesizliğinden dolayı veri azalma tekniği uygulayarak, veri seti üzerinde benzer makine öğrenmesi algoritmaları ve uygulamaları kullanılmıştır. Ancak en yüksek doğruluk oranı SVMs yöntemi ile %97,9, RF yöntemi ile %96, Native Bayes yöntemi ile %92,6, kNN yöntemi ile %96,1 olarak bulunmuştur (Khourdifi ve Bahaj, 2018). Çalışmamızda, literatürde kullanılan veri seti ile aynı veri seti kullanılarak 6. Veri seti (WDBC) adı altında uygulaması gerçekleştirilmiştir. Aynı veri seti ile yapılan çalışmamızda elde edilen sonuçlarda ise en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Aynı zamanda

uygulamamızdaki farklı veri setleri üzerinde gerçekleştirilen 1. Veri seti (UMMC) adı altında uygulması yapılan çalışmamızda en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 ve 5. Veri seti (WBC) adı altında uygulması yapılan çalışmamızda en yüksek doğruluk oranı SVM, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94 bulunmuştur. Bu durum da aynı veri seti ile yapılan çalışmalarımızın bir kısmı, farklı veri setleri üzerinde yapılan çalışmalarımızın çoğu literatürdeki çalışmalardan sağlık alanında daha iyi olduğunu göstermektedir.

Çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak yapılan başka bir çalışmada SVMs, C4.5, NB ve kNN gibi makine öğrenmesi algoritmaları kullanılmıştır. Her bir algoritma doğruluk, kararlılık, duyarlılık ve özgüllük bakımından incelenmiştir. En yüksek doğruluk oranı %97,13 ile SVMs'ye ait olduğu görülmüştür. C4.5, NB ve kNN doğruluk oranları %95,12 ve %95,28 arasında olduğu görülmüştür. Literatürdeki bu çalışmada tüm uygulamalar WEKA veri madenciliği gerçekleştirilmiştir (Asri ve ark., 2016). Benzer algoritmaların incelendiği, aynı veri seti kullanarak ve Matlab üzerinden yapılan çalışmamızda elde edilen sonuçlarda ise en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Aynı zamanda uygulamamızdaki farklı veri setleri üzerinde gerçekleştirilen, 1. Veri seti (UMMC) adı altında uygulması yapılan çalışmamızda en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 ve 5. Veri seti (WBC) adı altında uygulması yapılan çalışmamızda en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94 bulunmuştur. Bu durumda çalışmamızın literatürdeki benzer algoritmalar ve veri setleri kullanılmasına rağmen farklı platformda gerçekleştirilmiştir. Sonuç olarak çalışmamızdaki aynı veri setleri üzerinde gerçekleştirilen uygulamaların bir kısmı, farklı veri setleri üzerinde gerçekleştirilen uygulamaların çoğu yüksek doğruluk değeri oranlarına sahip olduğunu göstermektedir. Aynı zamanda hem aynı hem de farklı veri setleri kullanılarak uygulaması yapılan çalışmamızın sağlık alanında literatürdeki bu çalışmadan kısmen daha iyi olduğu görülmüştür.

Çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak yapılan literatürdeki farklı bir çalışmada SVMs, RF ve Bayesian Networks (BN) isimli 3 makine öğrenmesi tekniği uygulanıp, karşılaştırılmıştır. Bu teknikler doğruluk, duyarlılık değeri (recall),

kesinlik (precision) ve Area of ROC parametreleri bakımından incelenmiştir. Tüm çalışma WEKA ortamında yapılmıştır. Sonuçlara göre en yüksek doğruluk değeri %97 olduğu bulunmuştur (Shinde ve ark., 2022). Uygulamamızda literatürden farklı olarak Matlab ortamında yapılan fakat aynı veri seti kullanılan çalışmamızda ise en yüksek doğruluk değeri %100 bulunmuştur. Aynı zamanda uygulamamızdaki farklı veri setleri üzerinde gerçekleştirilen, 1. Veri seti (UMMC) adı altında yapılan çalışmada en yüksek doğruluk değeri %99,3 ve 5. Veri seti (WBC) adı altında yapılan çalışmada en yüksek doğruluk değeri %97,45'e yaklaşamamıştır. Bu sonuç da çalışmamızda yer alan aynı alandaki hem aynı hem de farklı veri setleri üzerinde yapılan uygulamaların meme kanseri konusunda daha güvenilir sonuçlar elde edebileceğini göstermektedir.

Literatürde çalışmamızdan farklı bir seti (Using three machine learning techniques for predicting breast veri seti) kullanılarak elde edilen farklı bir çalışmada C4.5, SVMs ve ANN makine öğrenmesi teknikleri uygulanmıştır. Bu 3 makine öğrenmesi tekniği doğruluk, özgüllük ve duyarlılık açısından incelenmiştir. Analiz sonuçları göz önünde bulundurulduğunda doğruluk değeri DT, ANN ve SVMs makine öğrenmesi teknikleri %93,6, %94,7 ve %95,7 sıralı şekilde sonuçlanmıştır (LG ve AT, 2013). Literatürdeki bu çalışma uygulamamızda farklı veri setleri üzerinde yapılan çalışmalarla kıyaslandığında; benzer algoritmaların incelendiği çalışmamızda 6. Veri setine (WDBC) göre en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43, 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Bu durum literatürdeki bütün makine öğrenmesi tekniklerine rağmen çalışmamızda uygulanan makine öğrenmesi tekniklerinin çoğunun altında değerler elde etmiştir. Bu durum da çalışmamızda farklı veri setleri üzerinde uygulanan makine öğrenmesi tekniklerinin daha güvenilir sonuçlar elde edebileceğini göstermektedir.

Çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak yapılan literatürde bulunan bir diğer çalışmada RF, kNN ve NB gibi makine öğrenme algoritmaları kullanılmıştır. Python ortamında yapılan çalışmada doğruluk, kesinlik (precision) ve F-Score parametreleri açısından makine öğrenmesi algoritmaları üzerinde karşılaştırmalar yapılmıştır. Tüm sonuçlar incelendiğinde en yüksek doğruluk değeri

kNN yöntemi ile %95,9, en yüksek kesinlik (precision) değeri kNN yöntemi ile %98,27, en yüksek recall değeri RF yöntemi ile %93,65, en yüksek F-Score değeri kNN yöntemi ile %94,2 bulunmuştur. Diğer RF ve NB gibi yöntemler ile doğruluk değeri sırasıyla %94,74 ve %94,47 bulunmuştur (Sharma ve ark., 2018). Çalışmamızda ise literatürdeki bu çalışmada kullanılan veri seti, yapılan makine öğrenmesi algoritmaları ve algoritmalar üzerinde uygulanan yöntemler benzerlik göstermektedir. Ancak tüm bu benzerliklere rağmen uygulamamızdaki çalışma farklı platformda uygulanmıştır. Tüm bunların sonucunda matlab ortamında uygulaması yapılan ve benzer veri seti kullanılan uygulamamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 ve en yüksek F-Score değeri Ensemble yöntemi ile %100 bulunmuştur. Böylece çalışmamızın literatürdeki bahsedilen çalışmaya kıyasla ortak kullanılan yöntemlerden daha verimli sonuçlar elde ettiğini göstermektedir. Bu durum da meme kanseri teşhisinde uygulamamızdaki çalışmanın daha üstün sonuçlar elde edebileceğini göstermektedir.

Çalışmamızda 5. Veri Seti (WBC) ile aynı veri seti kullanılarak yapılan başka bir çalışmada SVMs, C4.5, NB ve kNN gibi makine öğrenme algoritmaları karşılaştırılmıştır. Her bir algoritma doğruluk, kesinlik, duyarlılık ve özgüllük bakımından incelenmiştir. En yüksek doğruluk oranı %97,13 ile SVMs'ye ait olduğu görülmüştür. C4.5, Naïve Bayes ve kNN gibi makine öğrenmesi algoritmalarının da doğruluk değeri %95,12 ve %95,28 arasında değişkenlik göstermiştir. WEKA veri madenciliği ile tüm uygulama gerçekleştirilmiştir (Bharat ve ark., 2018). Matlab ortamında aynı veri seti kullanılarak gerçekleştirilen uygulamamızda elde edilen sonuçlarda ise en yüksek doğruluk oranı Ensemble yöntemi ile %98,94, en yüksek özgüllük oranı SVMs, kNN, Ensemble ve Hibrit yöntemleri ile %100 ve en yüksek duyarlılık oranı Ensemble yöntemi ile %97,83 olarak bulunmuştur. Bu durum da uygulamamızın benzer makine öğrenmesi teknikleri ve aynı veri seti ile daha yüksek yüzdelerde doğruluk, özgüllük ve duyarlılık oranları bulunabileceğini ispatlamaktadır. Böylece çalışmamızın meme kanseri teşhisinde en az hata oranına sahip olduğu görülmekte olup en yüksek doğru sınıflandırma yapabilme kapasitesine sahip olduğu görülmüştür.

Çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak literatürde yer alan bu çalışmadaki temel amaç meme kanseri için kullanılan farklı veri madenciliği ve

makine öğrenmesi algoritmalarını gözden geçirmektir. Temel odak noktası uygun algoritma ve en yüksek doğruluk değerine sahip meme kanseri tahminini bulmaktır. Linear (LR, Logistic Regression, Linear Discriminant Analysis), nonlinear (CART, NB, kNN, SVMs) ve Ensemble Algorithms (Ctree, RF, Boosting ve AdaBoost) gibi algoritmalara sahip çalışmalar incelenmiştir. Literatürdeki bu çalışmada da doğruluk değerleri dikkate alınarak her bir algoritmanın karşılaştırmalı analizi yapılmıştır. Daha sonra ise meme kanseri teşhisi için en uygun makine öğrenmesi algoritmaları belirtilmiştir. Her tekniğin farklı koşullar altında ve farklı veri kümeleri üzerinde uygun olduğu görülmüştür. Tüm bu tekniklerin karşılaştırmalı analizden sonra makine öğrenmesi algoritmalarından SVMs'nin meme kanseri tahmini için en uygun algoritma olduğunu öğrenilmiştir. Aynı zamanda SVMs'nin doğruluk değeri diğer makine öğrenmesi algoritmalarından yüksek çıktığı bulunmuştur. Tüm makine öğrenmesi algoritmaları ile kıyaslandığında SVMs'nin en yüksek doğruluk oranı Weka ortamında %98.03 iken Spark ortamında %99,68 bulunmuştur. Meme kanseri teşhisi için farklı bir veri tabanından toplanılan başka bir veri seti üzerinde CNN, SAE ve SSAE gibi derin öğrenme teknikleri uygulandığında en yüksek doğruluk oranı %98.9 bulunmuştur (Fatima ve ark., 2020). Uygulamamızdaki çalışma ise literatürdeki bu çalışmaya benzer ortam olan Matlab üzerinde yapılmış, aynı veri seti ve benzer algoritmalar olan Ensemble, kNN, SVMs ve Hibrit gibi makine öğrenmesi algoritmaları kullanılmıştır. Uygulamamızdaki çalışmada kullanılan makine öğrenmesi algortimaları açısından doğruluk oranları kıyaslandığında benzer veri seti kullanılan uygulamamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Aynı konuda farklı veri seti kullanılarak 1. Veri seti (UMMC) adı altında uygulması yapılan çalışmamızda ise en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Literatür ile uygulamamızda yer alan aynı veri setleri üzerinde en yüksek doğruluk oranları kıyaslandığında ise literatürdeki çalışmada Weka ortamında SVMs algoritması ile %99,68 bulunan doğruluk oranı çalışmamızda Matlab ortamında Ensemble algoritması ile %100 bulunan doğruluk oranından düşük çıkmıştır. Aynı zamanda literatürde CNN, SAE ve SSAE gibi derin öğrenme tekniklerinin kullanıldığı çalışmadaki doğruluk oranı farklı veri setleri açısından kıyaslandığında; 1. Veri seti (UMMC) adı altında uygulması yapılan çalışmamızda kullanılan 4 makine öğrenme algoritmasının sadece 1'inin üzerinde çıkmıştır. 1. Veri seti kullanılarak yapılan

çalışmamızda geriye kalan makine öğrenmesi algoritmalarının 3'ü derin öğrenme teknikleri ile kullanılan algoritmaların doğruluk oranından daha fazla çıktığı görülmüştür. Meme kanseri teşhisinde literatürdeki bu çalışma uygulamamızdaki çalışmadan başarısız olduğu söylenebilir. Çünkü meme kanseri teşhisinde çalışmamızın, hem aynı ortamlarda en yüksek doğruluk değeri açısından hem de farklı ortamlarda CNN, SAE ve SSAE gibi farklı algoritmalar olan derin öğrenme teknikleri açısından kıyaslandığında çalışmamızın daha üstün olduğu görülmüştür. Bu durumda meme kanseri teşhisinde uygulamamızdaki çalışmanın doğruluk oranlarına bakıldığında güvenilir sonuçlar vermektedir.

Literatürde çalışmamızdan farklı bir seti (Antenna Veri Seti) kullanılarak elde edilen bir diğer çalışmada veri madenciliği algoritmaları olan Bagging Algorithm, IBk, Random Committee Algorithm, Random Forest Algorithm, Simple Classification ve Regression Tree (Simple CART Algorithm) algoritmaları meme kanserini teşhis ve tahmininde kullanılmıştır. Sonuçlar WEKA programında Bayes, Function, Meta, Lazy, Trees vb. açısından analiz edilmiştir. Analiz sonuçlarından sonra Random Forest Algorithm'ın en yüksek doğruluk seviyesine sahip olduğu ve meme kanseri için en uygun algoritma olduğu görülmüştür. Random Forest algoritmasının doğruluk oranı %92,2 iken Bagging, IBk and Random Committee Algoritmalarının doğruluk oranları sırasıyla %90,9, %90 and %90,9 şeklinde bulunmuştur (Keleş, 2019). Matlab ortamında uygulaması yapılan çalışmamızda ise literatürdeki bu çalışma ile makine öğrenmesi algoritmaları ve algoritmalar üzerinde uygulanan yöntemler farklılık göstermektedir. Literatürdeki bu çalışma uygulamamızda farklı veri setleri üzerinde yapılan çalışmalarla kıyaslandığında; tüm bunların sonucunda çalışmamızdaki 6. Veri setine (WDBC) göre en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43, 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVM, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Böylece çalışmamızın literatürdeki bahsedilen çalışma ile kıyaslandığında farklı veri setleri üzerinde daha uygun makine öğrenmesi algoritmaları kullanılıp daha verimli sonuçlar elde edilebileceği görülmüştür. Bu durum da farklı veri setleri üzerinde meme kanseri hastalığı

konusunda çalışmamızın uygulanan algoritmalar ve yöntem konusunda daha üstün olduğunu göstermektedir.

Literatürde çalışmamızdan farklı bir seti (Gene Expression Veri Seti) kullanılarak elde edilen bir diğer çalışmada SVMs, Ctree ve Random Forest algoritmaları meme kanserinin teşhisi için 9 modelin sınıflandırılmasında kullanılmıştır. Çalışmada veriseti Weka ve Spark olan iki farklı veri madenciliği aracında algoritmaların doğruluk (accuracy) ve hata oranları gösterilmiştir. Çalışmada iki veri seti (GE ve DM) ana amacı tümörlerin varlığını belirlemek olan genleri elde etmek için filtrelenmiştir. İki farklı araçta algoritmaların karşılaştırmaları sonucunda SVMs'nin diğer algoritmalarından daha yüksek olduğu ortaya çıkmıştır. Bu doğruluk değeri Spark'ta %99,68, Weka'da ise %98,03 olarak bulunmuştur (Alghunaim ve Al-Baity, 2019). Çalışmamızdaki veri setlerinde veri madenciliğinden farklı bir yöntem olarak Matlab ortamında makine öğrenmesi algoritmaları kullanılmıştır. Literatürde yer alan bu çalışma uygulamamızdaki tüm veri setleri ile yapılan çalışmalarla kıyaslandığında sadece 1. Veri setinde (UMMC) bulunan uygulama ile 6. Veri setinde (WDBC) bulunan uygulama literatürdeki bu çalışma ile kıyaslanabilecek düzeyde doğruluk oranlarına sahip olabilmıştır. Çalışmamızda SVMs gibi benzer makine öğrenme algoritması kullanılmasına rağmen tüm veri setlerinde en yüksek doğruluk oranına sahip 1. Veri setinde Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2, 6. Veri seti (WDBC) adı altında uygulaması yapılan çalışmamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Literatürde en yüksek doğruluk oranına sahip bu çalışmada doğruluk oranları kıyaslandığında çalışmamızda bulunan bazı algoritmaların doğruluk oranlarını geçse de genel olarak çalışmamızda farklı veri seti üzerinde meme kanserin alanında daha yüksek doğruluk değerleri ile teşhis edilebileceği görülmüştür.

Literatürde çalışmamızdan farklı bir seti kullanılarak uygulamamızdaki çalışmayla uyumlu bir çalışmada NB, kNN ve J48 algoritmaları ile meme kanserinin olduğu 9 farklı türdeki kanserin tahmini için kullanılmıştır. Eğitim veri seti temel alınarak çalışmada başlangıçta sonuçların doğru ve yanlışlığını test etmek için semptomları karşılaştırılmış ve eşleşen semptomların doğru olduğu anlamına gelmiştir. Bu süreç boyunca meme kanserinin farklı tipleri tahmin edilmiştir ve doğruluk oranları temel alınarak her bir algoritmada sınıflandırılmıştır. Meme kanserinin teşhisi süreci boyunca

NB ve KNN doğruluk (accuracy) oranları J48 Decision Tree Classifier'den daha yüksek olduğu bulunmuş ve sırasıyla doğruluk değerleri %98,2, %98,8 ve %98,5 olarak bulunmuştur (Maliha ve ark., 2019). Literatürde yer alan bu çalışma uygulamamızdaki tüm veri setleri ile yapılan çalışmalarla kıyaslandığında 1. Veri setinde (UMMC) ve 6. Veri setinde (WDBC) bulunan uygulamalar literatürdeki bu çalışma ile kıyaslanabilecek düzeyde doğruluk oranlarına sahip olabilmektedir. Matlab ortamında uygulaması yapılan 1. Veri setindeki (UMMC) uygulamamızda kNN gibi benzer makine öğrenme algoritması kullanılmasına rağmen en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Gene aynı şekilde 6. Veri setindeki (WDBC) uygulamamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Bu durum da farklı veri seti üzerinde sağlık alanında meme kanseri bulgularının uygulamamızdaki çalışmada daha yüksek oranlarda tespit edilebileceği görülmüştür.

Meme kanseri teşhisi için çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak literatürde yer alan bu çalışmada SVMs olan tahmine dayalı makine öğrenme modeli tekrarlanan özellik eleme tekniği kullanılmıştır. Çalışmanın amacı iyi huylu ve kötü huylu tümörlü insanların veri setinden doğru özellikleri seçmektir. Recursive feature elimination technique, SVMs algoritmasının değerlendirilmesi için kullanılmıştır. Performans matrisi farklı kernel türleri üzerinde tahmine dayalı SVMs modelinin doğruluk oranının kontrolü için tasarlanmıştır. SVMs Linear Kernel'de %99 doğruluk oranına RBF Kernel'de %98 doğruluk oranına, Polynomial Kernel'de %97 doğruluk oranına, Sigmoid Kernel'de %84 doğruluk oranına sahiptir (Memon ve ark., 2019). Uygulamamızda Matlab üzerinde yapılan çalışmamızda, aynı veri seti ve benzer algoritma olan SVMs makine öğrenmesi algoritması kullanılmıştır. Uygulamamızdaki çalışmada kullanılan makine öğrenmesi algortimaları açısından doğruluk oranları kıyaslandığında benzer veri seti kullanılan uygulamamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. Aynı konuda farklı veri seti kullanılarak 1. Veri seti (UMMC) adı altında uygulması yapılan çalışmamızda ise en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. SVMs gibi aynı makine öğrenme algoritması üzerinde çalışmalar kıyaslandığında literatürdeki çalışmada

SVMs yöntemi ile daha yüksek doğruluk değeri bulunmuştur. Ancak genel olarak sistem kıyaslandığında çalışmamızdaki farklı veri setleri üzerindeki diğer makine öğrenme algortimalarının tümü literatürdeki bu çalışmanın üzerinde çıkmıştır. Bu durum da çalışmamızın aynı ve farklı veri setleri göz önüne alınarak sağlık alanında güvenilirliği açısından daha iyi tespit elde edilebileceğini göstermiştir.

Meme kanseri teşhisi için çalışmamızda 5. Veri Seti (WBC) ile aynı veri seti kullanılarak literatürde yer alan bu çalışmada erken aşamada meme kanseri tespiti için önerilen modelin etkili olduğu savunulmuştur. Literatürde önerilen BCD modelinde veri ile ilgili problemlerin üstesinden gelmek ve sınıflandırıcının performansını arttırmak için 10 kat çapraz doğrulama tekniği ile SVMs kullanılmıştır. BCD F-measure, class accuracy, ROC Curve, AUC, Lift Curve ve Calibration Plot gibi çeşitli ölçüm parametreleri kullanarak AdaBoost, Random Forest, Naïve Bayes gibi modeller uygulanmıştır. Önerilen BCD modeli uygulanan tüm modeller arasında en yüksek doğruluk oranı %98,1 ve en yüksek AUC değeri ise %99,5 bulunmuştur (Israni, 2019). Matlab ortamında uygulanmış çalışmamızda literatür ile aynı veri seti üzerinde farklı makine öğrenmesi algortimaları kullanılmasına rağmen doğruluk, F-Score ve AUC gibi benzer ölçüm parametreleri kullanılmıştır. Benzer ölçüm parametreleri kıyaslandığında literatürde en yüksek doğruluk oranı SVMs yöntemi ile %98,1 iken , 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. AUC değerinde ise literatürdeki çalışmada en yüksek oran SVMs yöntemi ile %99,5 iken, uygulamamızda 5. Veri setine göre (WBC) Ensemble yöntem ile %98,91, 1. Veri setine göre (UMMC) Ensemble yöntemi ile %99,4 bulunmuştur. F-Score değerinde ise literatürdeki çalışmada en yüksek oran SVMs yöntemi ile %98,1 iken, uygulamamızda 5. Veri setine göre (WBC) Ensemble yöntemi ile %98,9, 1. Veri setine göre (UMMC) Ensemble yöntemi ile %99,3 bulunmuştur. Literatürdeki çalışmanın AUC değeri uygulamamızdaki AUC değerlerinden yüksek olsa da F-Score ve doğruluk değeri açısından uygulamamızdaki çalışma aynı ve farklı veri setleri üzerinde daha yüksek bulunmuştur. Bu durumda literatürdeki çalışma AUC değeri açısından meme kanseri teşhisinin tahmininde güvenilir olduğunu gösterse de F-Score ve doğruluk değeri bakımından bu durum değişkenlik gösterebilmektedir. AUC, F-Score ve doğruluk gibi ortak ölçüm

parametreleri incelenip kıyaslandığında ve hem farklı hem de aynı veri setleri göz önünde bulundurulduğunda, uygulamamızdaki çalışmanın meme kanseri teşhisinde performansın daha yüksek ve daha güvenilir sonuçlar elde edilebileceği görülmüştür.

Meme kanseri teşhisi için çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak literatürde yer alan literatürdeki bu çalışmada meme kanseri tespiti için yeni bir İç İçe Topluluk (NE) tekniğini kullanmıştır. Literatürdeki bu çalışmada, oylama ve istifleme tekniklerine dayalı dört adet iki katmanlı İç İçe Topluluk Sınıflandırıcısı oluşturulmuştur. Bunlar; SV-BayesNet-2-MetaClassifier, SV-Naive Bayes-2-MetaClassifier, SVBayesNet-3-MetaClassifier, SV-Naive Bayes-3-MetaClassifier olarak isimlendirilmiştir. Bu dört adet iki katmanlı sınıflandırıcılara ek olarak BayesNet ve NB sınıflandırıcıları da değerlendirilmiştir. Sınıflandırıcıların performansları veri sınıflandırması için doğruluk, duyarlılık, kesinlik, ROC Curve gibi tipik ölçütler kullanılarak değerlendirilmiştir. Tüm bu literatürdeki uygulama açık kaynaklı bir makine öğrenimi yazılımı olan WEKA 3.9.1 ortamında yapılmıştır. Uygulama sonucunda ise BayesNet algoritması için en yüksek doğruluk %95,25 iken en yüksek F-Score değeri %95,30, NB algoritması için en yüksek doğruluk %93,32 iken en yüksek F-Score değeri %93,30, SV-BayesNet-2-MetaClassifier algoritması için en yüksek doğruluk %97,72 iken en yüksek F-Score değeri %97,70, SV-Naive Bayes-2-MetaClassifier algoritması için en yüksek doğruluk %97,72 iken en yüksek F-Score değeri %97,70, SV-BayesNet-3-MetaClassifier algoritması için en yüksek doğruluk %98,07 iken en yüksek F-Score değeri %98,10, SV-Naive Bayes-3-MetaClassifier algoritması için en yüksek doğruluk %98,07 iken en yüksek F-Score değeri %98,10 bulunmuştur (Abdar ve ark., 2020). Matlab ortamında uygulanmış çalışmamızda literatürden farklı makine öğrenmesi algoritmaları kullanılmasına rağmen doğruluk ve F-Score gibi benzer ölçüm parametreleri ve aynı veri seti kullanılmıştır. Benzer ölçüm parametreleri kıyaslandığında literatürde en yüksek doğruluk oranı %98,07 iken, 6. Veri setine (WDBC) göre en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. F-Score değerinde ise literatürdeki çalışmada en yüksek oran %98,1 iken, uygulamamızda 6. Veri setine göre (WDBC) Ensemble

yöntemi ile %98,9, 1. Veri setine göre (UMMC) Ensemble yöntemi ile %99,3 bulunmuştur. Bu durumda literatürdeki çalışmada bulunan F-Score değeri açısından hem aynı hem de farklı veri setleri üzerinde yapılan uygulamamıza göre düşük kaldığı görülmüştür. Literatürdeki çalışmada doğruluk değerleri aynı veri seti ile yapılan çalışmamıza göre çoğu değerde oldukça düşük çıktığı görülmüştür. Literatürdeki çalışmada doğruluk değeri farklı veri seti ile yapılan çalışmamıza göre ise tüm değerler düşük çıktığı görülmüştür. F-Score ve doğruluk gibi ortak ölçüm parametreleri incelendiğinde meme kanseri teşhisinde uygulamamızdaki çalışmanın doğruluk oranı açısından daha emin ve güvenilir sonuçlar elde edebileceği görülmüştür.

Meme kanseri teşhisi için çalışmamızda 5. Veri Seti (WBC) ile aynı veri seti kullanılarak literatürde bulunan farklı bir çalışmada meme kanseri tahmininde veri madenciliği araçları kullanılmıştır. Literatürdeki çalışmadaki temel uygulama NB algoritması, Bayes Logistic Regression, Simple CART ve J48 gibi algoritmaları doğruluk ve time complexity yaparak sınıflandırılmasıdır. WEKA ortamında uygulaması yapılan bu analizin en yüksek doğruluk oranı NB yöntemi ile %95,27, Bayesian logistic Regression yöntemi ile %65,42, Simple CART yöntemi ile %98,13, J48 yöntemi ile %97,27 bulunmuştur (Singh ve Thakral, 2018). Matlab ile uygulaması yapılan uygulamamızdaki çalışmada literatürle aynı veri seti ile tamamen farklı algoritmalar kullanılarak 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Sonuç olarak aynı veri seti kullanılarak uygulaması yapılan çalışmamızın çoğu algoritmalarının, farklı veri seti kullanılarak uygulaması yapılan çalışmamızın tüm algoritmalarının doğruluk oranı literatürdeki çalışmanın üzerinde çıkmıştır. Bu durum da sağlık alanında meme kanseri teşhisinde çalışmamızın daha uygun bir çalışma ortaya koyduğu görülmüştür.

Meme kanseri teşhisi için çalışmamızda 6. Veri Seti (WDBC) ile aynı veri seti kullanılarak literatürde yer alan farklı bir çalışmada meme kanseri tespiti için MLP, kNN, CART, SVMs ve Gaussian NB dahil olmak üzere beş doğrusal olmayan makine öğrenme algoritmasının karşılaştırılması yapıldı. Çalışmanın temel amacında meme kanseri tespiti için algoritmaların doğruluk, kesinlik ve duyarlılık açısından etkinliği karşılaştırılmıştır. Tüm algoritmaların doğruluk oranları incelendiğinde MLP'nin

doğruluk değeri kNN, CART, NB ve SVMs algoritmalarından daha yüksek olarak %99,12 bulunmuştur. Diğer algoritmalarda ise doğruluk oranı kNN yöntemi ile %95,61, CART yöntemi ile %93,85, NB yöntemi ile %94,73, SVMs yöntemi ile %98,24 bulunmuştur (Bataineh, 2019). Matlab üzerinde yapılan uygulamamızdaki çalışmada kNN ve SVMs gibi literatüre benzer algoritmalar ve aynı veri seti kullanarak 6. Veri setine (WDBC) göre en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43, 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVM, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. Literatürde MLP ile yapılan çalışmada benzer makine öğrenmesi teknikleri olan kNN ve SVMs algoritmalarında benzer veri seti ile uygulamamızdaki çalışmaya göre daha yüksek çıktığı görülmüştür. Ancak farklı veri seti olan 1. Veri setine göre bu algoritmaların değerleri çalışmamızda daha yüksek çıkmıştı görülmüştür. Diğer açıdan, literatürdeki tüm makine öğrenmesi teknikleri açısından en yüksek doğruluk oranı çalışmamızdaki benzer veri seti ile yapılan uygulamamızla kıyaslandığında çalışmamızdaki uygulamanın daha yüksek olduğu görülmüştür. Ayrıca diğer benzer makine öğrenmesi teknikleri olan kNN ve SVMs algoritmalarında farklı veri setleri (1. Veri seti ve 5. Veri seti) ile kıyaslandığında da çalışmamızdaki uygulamalar gene daha yüksek olduğu görülmüştür. Bu durum da uygulamamızdaki çalışmanın tıbbi uygulamada meme kanseri teşhisinde farklı veri setlerinde ve aynı veri setlerinde çoğunlukla daha kesin sonuçlar elde edilebileceği görülmüştür.

Meme kanseri teşhisi için çalışmamızda 2. Veri Seti ile aynı veri seti kullanılarak literatürde başka bir çalışmada araştırmacılar meme kanseri tahmini için NB, Random Forest, Logistics Regresyon , MLP ve kNN karşılaştırmalı analizini sağladı. Tüm bu algoritmaların değerlendirilmesi Kappa Statistics, TP Rate, FP Rate, kesinlik (Precision), duyarlılık (Recall), F-Measure Lining, MCC, Receiver Operating Characteristics Area or ROC Area üzerinde algoritmanın kesinliği açısından yapılmıştır. Her bir algoritmanın doğruluk değerini analiz etmek için her algoritma veri kümesine uygulanmıştır. kNN, NB ve Random Forest'ın doğruluk değeri sırasıyla %72,3, %71,6 ve %69,5 iken Logistics Regresyon ve MLP örneklerin doğruluk değerleri sırasıyla %68,8 ve %64,6 bulunmuştur. F-Measure açısından kNN,

NB ve Random Forest'in doğruluk değerleri sırasıyla %69,7, %71,7 ve %66,9 iken Logistics Regresyon ve MLP örneklerin doğruluk değerleri sırasıyla %67,5 ve %64,7 bulunmuştur (Bharati ve ark., 2018). Matlab üzerinde aynı veri seti ile yapılmış uygulamamızdaki çalışmada literatürdeki kNN gibi benzer makine öğrenmesi algoritmaları ile birlikte doğruluk ve F-Score gibi benzer ölçüm parametreleri de kullanılmıştır. 2. Veri setine göre en yüksek doğruluk oranı SVMs yöntemi ile %77,63, kNN yöntemi ile %72,37, Ensemble yöntemi ile %73,68 ve Hibrit yöntemi ile %75, 6. Veri setine (WDBC) göre en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43, 5. Veri setine göre (WBC) en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94, 1. Veri setine göre (UMMC) en yüksek doğruluk oranı Ensemble yöntemi ile %99,3, kNN yöntemi ile %99,2, SVMs yöntemi ile %98,8, Hibrit yöntemi ile %99,2 bulunmuştur. F-Score değerinde ise literatürdeki çalışmada en yüksek oran %71,7 iken uygulamamızda 2. Veri setine göre en yüksek oran SVMs yöntemi ile %77,77, 6. Veri setine göre (WDBC) Ensemble yöntemi ile %100, 5. Veri setine göre (WBC) Ensemble yöntemi ile %98,9, 1. Veri setine göre (UMMC) Ensemble yöntemi ile %99,3 bulunmuştur. Bu durumda çalışmamızda ve literatürde ortak kullanılan doğruluk ve F-Score değerleri kıyaslandığında çalışmamızda yer alan hem aynı hem farklı veri setlerindeki değerlerin oldukça yüksek olduğu görülmektedir. Bu durum da uygulamamızdaki çalışmanın meme kanseri tahmininde daha yüksek performans elde edilebileceği görülmektedir.

Meme kanseri teşhisi için çalışmamızda 3. Veri Seti (Coimbra) ile aynı veri seti kullanılarak literatürde başka bir çalışmada araştırmacılar meme kanseri tahmini için doğruluk, özgüllük ve duyarlılık gibi uygulamamız ile ortak ölçüm parametreleri kullanılmıştır. Literatürdeki bu çalışmada DNN-1 modeline göre doğruluk değeri %91,4, özgüllük %88,2 ve duyarlılık %94,4, DNN-2 modeline göre doğruluk değeri %91,4, özgüllük %94,1 ve duyarlılık %88,8, DNN-1 modeline göre doğruluk değeri %88,5, özgüllük %88,2 ve duyarlılık %88,8 bulunmuştur (Artificial Intelligence and Applied Mathematics in Engineering Problems, 2024). Matlab üzerinde aynı veri seti ile yapılmış uygulamamızdaki çalışmada literatürdeki doğruluk, özgüllük ve duyarlılık gibi benzer ölçüm parametreleri kullanılmıştır. Literatür ile aynı veri seti kullanılarak yapılan 3. Veri seti (Coimbra) çalışmamızda en yüksek doğruluk oranı %93,33, en yüksek özgüllük oranı %100 ve en yüksek duyarlılık oranı %100 bulunmuştur. 6. Veri

seti (WDBC) çalışmamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. 5. Veri seti (WDBC) çalışmamızda en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94 bulunmuştur. 1. Veri seti (UMMC) çalışmamızda en yüksek doğruluk oranı %99,3, en yüksek özgüllük oranı %98,7 ve en yüksek duyarlılık oranı %100 bulunmuştur. Bu durumda çalışmamızda ve literatürde ortak kullanılan doğruluk, özgüllük ve duyarlılık değerleri kıyaslandığında çalışmamızda yer alan hem aynı veri hem de farklı veri setlerindeki tüm değerlerin oldukça yüksek olduğu görülmektedir. Bu durum uygulamamızdaki hem aynı hem de farklı veri seti ile yapılan çalışmaların literatürdeki çalışmaya göre tıbbi uygulamada meme kanseri teşhisinde daha iyi sonuçlar elde edilebileceği de görülmüştür.

Meme kanseri teşhisi için çalışmamızda 3. Veri Seti (Coimbra) ile aynı veri seti kullanılarak literatürde başka bir çalışmada araştırmacılar meme kanseri tahmini için doğruluk ve kapa istatistiği gibi çalışmamızla ortak ölçüm parametreleri kullanılmıştır. Literatürdeki çalışmada en yüksek doğruluk değeri Node Harvest sınıflandırma algoritması ile %86,95 bulunmuştur. Literatürde bu çalışmada diğer sınıflandırma algoritmaları ile bulunan doğruluk oranları %56,52'ye kadar en yüksek doğruluk oranından aşağı doğru düşerek çalışma sonuçlandırılmıştır. Literatürdeki çalışmaya kappa istatistiği açısından bakıldığında ise en yüksek kappa değeri %70,39 bulunmuştur. Kappa istatistik değeri de tıpkı doğruluk değeri gibi %42,98'e kadar en yüksek kappa değerinden aşağı doğru düşerek çalışma sonuçlandırılmıştır (Sen, 2021). Literatür ile aynı veri seti kullanılarak yapılan 3. Veri seti (Coimbra) çalışmamızda en yüksek doğruluk oranı %93,33 ve en yüksek kappa oranı %86,73 bulunmuştur. 6. Veri seti (WDBC) çalışmamızda en yüksek doğruluk oranı %100 ve en yüksek kappa oranı %100 bulunmuştur. 5. Veri seti (WDBC) çalışmamızda en yüksek doğruluk oranı %98,94 ve en yüksek kappa oranı %97,87 bulunmuştur. 1. Veri seti (UMMC) çalışmamızda en yüksek doğruluk oranı %99,3 en yüksek kappa oranı %98,3 bulunmuştur. Bu durumda çalışmamızda ve literatürde ortak kullanılan doğruluk ve kappa değerleri kıyaslandığında çalışmamızda yer alan hem aynı hem de farklı veri setlerindeki değerlerin oldukça yüksek olduğu görülmektedir. Bu durum da uygulamamızdaki çalışmanın meme kanseri teşhisinde hem aynı hem farklı veri setlerinde daha iyi sonuçlar elde edilebileceği görülmüştür.

Meme kanseri teşhisi için çalışmamızda 4. Veri Seti (WPBC) ile aynı veri seti kullanılarak literatürdeki bir çalışmada araştırmacılar duyarlılık ve özgüllük değerleri gibi çalışmamızdala ortak ölçüm parametreleri kullanılmıştır. Literatürdeki çalışmaya göre en yüksek doğruluk değeri %99,68, en yüksek duyarlılık değeri %100, en yüksek özgüllük değeri %99,47 bulunmuştur (Abdel ve Eldeib, 2016). Literatür ile aynı veri seti kullanılarak yapılan 4. Veri seti (WPBC) çalışmamızda en yüksek doğruluk oranı %85,42 en yüksek duyarlılık oranı %88, en yüksek özgüllük oranı %91,30 bulunmuştur. 6. Veri seti (WDBC) çalışmamızda en yüksek doğruluk oranı SVMs yöntemi ile %96,43, kNN yöntemi ile %91,67, Ensemble yöntemi ile %100, Hibrit yöntem ile %96,43 bulunmuştur. 5. Veri seti (WDBC) çalışmamızda en yüksek doğruluk oranı SVMs, Hibrit ve kNN yöntemleri ile %97,9 ve Ensemble yöntemi ile %98,94 bulunmuştur. 1. Veri seti (UMMC) çalışmamızda en yüksek doğruluk oranı %99,3, en yüksek özgüllük oranı %98,7 ve en yüksek duyarlılık oranı %100 bulunmuştur. Gene, bu durumda çalışmamızda ve literatürde ortak kullanılan doğruluk, özgüllük ve duyarlılık değerleri kıyaslandığında çalışmamızda yer alan aynı veri setindeki değerlerin düşük olduğu ancak farklı veri setlerindeki değerlerin oldukça yüksek olduğu görülmektedir. Bu durum da çalışmamızın tıbbi uygulamada meme kanseri teşhisinde farklı veri setlerinde daha net ve iyi sonuçlar elde edilebileceği görülmüştür.

Tüm bu sonuçlarda literatürde yer alıp çalışmamızda yer almayan veri setleri, çalışmamızda en başarılı değerlere sahip veri setleri ile kıyaslama yapılmıştır. Ayrıca literatürde benzer veri setleri kullanılan çalışmalarda gene çalışmamızda farklı veri setleri kullanılan uygulamalar ile kıyaslama yapılarak meme kanseri konusunda uygulaması yapılan çalışmamızın çeşitliliğinin ispatlanması istenilmiştir.

Tüm bu sonuçlar kıyaslandığında uygulamamızın farklı makine öğrenmesi teknikleri ve farklı veri setleri ile daha yüksek yüzdelerde doğruluk, özgüllük, duyarlılık, F-Score, Kappa ve AUC oranları bulunabileceğini ispatlamaktadır. Bu durum da meme kanseri teşhisinde hastalıklı ve sağlıklı kişilerin bulunmasında güvenilir bir çalışma olduğunu tekrar ortaya koyarak sağlık alanında daha uygulanabilir ve elverişli bir çalışma olduğunu göstermektedir.

Tablo 5.1. Çalışmamızda Yer Alan Tüm Veri Setleri ve Değerlerinin Yüzdesel Performans Sonuçları.

[illegible]

Tablo 5.2. Literatürde Araştırması Yapılan Tüm Veri Setleri ve Değerlerinin Yüzdesel Performans Sonuçları.

Sayı	Literatürdeki Veri Seti	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
1	WDBC (Omondiagbe ve ark., 2019).		98,41	99,07			
2	WDBC (Bayrak ve ark., 2019).	96,9957					
3	WDBC (Khourdifi ve Bahaj, 2018).	97,9 (SVMs)					
4	WDBC (Asri ve ark., 2016).	97,13 (SVMs)					
5	WDBC (Shinde ve ark., 2022).	97					
6	WDBC (Sharma ve ark., 2018).	95,9 (kNN)			94,2 (kNN)		
7	WDBC (Fatima ve ark., 2020).	99,68 (SVMs)					
8	WDBC (Memon ve ark., 2019).	99 (SVMs)					
9	WDBC (Abdar ve ark., 2020).	98,07			98,1		
10	WDBC (Bataineh, 2019).	99,12					
11	UMMC (Ganggayah ve ark., 2019).	82,70 (Random Forest)	84 (Extreme Boost)	84 (SVMs)			
12	WBC (Bharat ve ark., 2018).	97,13 (SVMs)					

Tablo 5.2. (Devamı) Literatürde Araştırması Yapılan Tüm Veri Setleri ve Değerlerinin Yüzdesel Performans Sonuçları.

Sayı	Literatürdeki Veri Seti	Doğruluk	Duyarlılık	Özgüllük	F-Score	Kappa	AUC
13	WBC (Israni, 2019).	98,1			98,1		99,5
14	WBC (Singh ve Thakral, 2018).	98,13 (NB)					
15	Breast Cancer Veri Seti (Bharati ve ark., 2018).	72,3 (kNN)			71,7 (NB)		
16	Coimbra veri seti (Artificial Intelligence and Applied Mathematics in Engineering Problems. Lecture Notes on Data Engineering and Communications Technologies, 2024).	91,4	94,4	94,1			
17	Coimbra veri seti (Sen, 2021).	86,95 (Node Harvest)				70,39	
18	WPBC (Abdel ve Eldeib, 2016).	99,68	100	99,47			
19	Using three machine learning techniques for predicting breast veri seti (LG ve AT, 2013).	95,7 (SVMs)					
20	Antenna Veri Seti (Keleş, 2019).	92,2 (Random Forest)					
21	Gene Expression Veri Seti (Alghunaim ve Al-Baity, 2019).	99,68 (SVMs)					
22	Literatür Veri Seti (Maliha ve ark., 2019).	98,80					



KAYNAKLAR

- 80/20 Principle, The: Richard Koch, Richard Aspel: 9781486213429: Amazon.com: Books. (2024). 23 Aralık 2024 tarihinde <https://www.amazon.com/80-20-Principle-Secret-Achieving/dp/1486213421> adresinden erişildi.
- Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D. ve Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*, 132, 123–131. doi:10.1016/j.patrec.2018.11.004
- Abdel-Zaher, A. M. ve Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139–144. doi:10.1016/j.eswa.2015.10.015
- Agarap, A. F. M. (2018). On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset. *ACM International Conference Proceeding Series* içinde (ss. 5–9). Association for Computing Machinery. doi:10.1145/3184066.3184080
- Ak, M. F. (2020). A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. *Healthcare (Switzerland)*, 8(2). doi:10.3390/healthcare8020111
- Alghunaim, S. ve Al-Baity, H. H. (2019). On the Scalability of Machine-Learning Algorithms for Breast Cancer Prediction in Big Data Context. *IEEE Access*, 7, 91535–91546. doi:10.1109/ACCESS.2019.2927080
- Alpar, R. (2010). Alpar, R., (2020). Uygulamalı İstatistik ve Geçerlik-Güvenirlik, 6. Baskı. Ankara: Detay Yayıncılık. <https://www.seckin.com.tr/kitap/746711289> adresinden erişildi.
- American Cancer Society. (2021). What is breast cancer? | american cancer society. *Www.Cancer.Org*. 15 Ocak 2024 tarihinde <https://www.cancer.org/cancer/types/breast-cancer/about/what-is-breast-cancer.html> adresinden erişildi.
- Asri, H., Mousannif, H., Al Moatassime, H. ve Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Computer Science*, 83(Fams), 1064–1069. doi:10.1016/j.procs.2016.04.224
- Aymen Fathalla Alhasadi, B. H. (2016). *Predicting Breast Cancer By Using Artificial Neural Network a Master'S Thesis*.
- Ayon, S. I., Islam, M. M. ve Hossain, M. R. (2022). Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques. *IETE Journal of Research*, 68(4), 2488–2507. doi:10.1080/03772063.2020.1713916
- Bataineh, A. Al. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. *International Journal of Machine Learning and Computing*, 9(3), 248–254. doi:10.18178/ijmlc.2019.9.3.794

- Bayrak, E. A., Kirci, P. ve Ensari, T. (2019). Comparison of machine learning methods for breast cancer diagnosis. *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, 4–6. doi:10.1109/EBBT.2019.8741990
- Bharat, A., Pooja, N. ve Reddy, R. A. (2018). Using Machine Learning algorithms for breast cancer risk prediction and diagnosis. *2018 IEEE 3rd International Conference on Circuits, Control, Communication and Computing, I4C 2018*, (x), 1–4. doi:10.1109/CIMCA.2018.8739696
- Bharati, S., Rahman, M. A. ve Podder, P. (2018). Breast cancer prediction applying different classification algorithm with comparative analysis using WEKA. *4th International Conference on Electrical Engineering and Information and Communication Technology, iCEEiCT 2018*, 581–584. doi:10.1109/CEEICT.2018.8628084
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. ve Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. doi:10.3322/caac.21492
- Breast Cancer - Symptoms, Causes and Treatment | MedPark Hospital. (2024). 7 Aralık 2024 tarihinde <https://www.medparkhospital.com/en-US/disease-and-treatment/breast-cancer> adresinden erişildi.
- Breast cancer (female) | NHS inform. (2024). 7 Aralık 2024 tarihinde <https://www.nhsinform.scot/illnesses-and-conditions/cancer/cancer-types-in-adults/breast-cancer-female/> adresinden erişildi.
- Breast Cancer Risk Factors You Can't Change | American Cancer Society. (2024). 7 Aralık 2024 tarihinde <https://www.cancer.org/cancer/types/breast-cancer/risk-and-prevention/breast-cancer-risk-factors-you-cannot-change.html> adresinden erişildi.
- Breast Cancer Signs and Symptoms | Most Common Symptoms | American Cancer Society. (2024). 7 Aralık 2024 tarihinde <https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/breast-cancer-signs-and-symptoms.html> adresinden erişildi.
- Breast Cancer Treatment Options - National Breast Cancer Foundation. (2020). 15 Ocak 2024 tarihinde <https://www.nationalbreastcancer.org/breast-cancer-treatment/> adresinden erişildi.
- Breastcancer. (2023). Breast Cancer Facts and Statistics. *Breastcancer.org*. 15 Ocak 2024 tarihinde https://www.breastcancer.org/facts-statistics?gad_source=1&gclid=CjwKCAiAzJOtBhALEiwAtwj8tlbQuo59n0mvpqVNs4YuzG07eSYQa53w4PbnkQQYEyqnfQyC5Nq41hoCSMIQAvD_BwE adresinden erişildi.
- Centers for Disease Control and Prevention, D. of C. P. and C. (2020). What Is Breast Cancer? | CDC. *September 14*. 14 Ocak 2024 tarihinde https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm adresinden erişildi.

- Chaurasia, V. ve Pal, S. (2020). Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer. *SN Computer Science*, 1(5). doi:10.1007/s42979-020-00296-8
- Chaurasia, V., Pal, S. ve Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms and Computational Technology*, 12(2), 119–126. doi:10.1177/1748301818756225
- Cheng, K., Wang, J., Liu, J., Zhang, X., Shen, Y. ve Su, H. (2023). Public health implications of computer-aided diagnosis and treatment technologies in breast cancer care. *AIMS Public Health*, 10(4), 867–895. doi:10.3934/publichealth.2023057
- Conceição, R. C., Medeiros, H., Godinho, D. M., O'Halloran, M., Rodriguez-Herrera, D., Flores-Tapia, D. ve Pistorius, S. (2020). Classification of breast tumor models with a prototype microwave imaging system. *Medical Physics*, 47(4), 1860–1870. doi:10.1002/mp.14064
- Content on Early Breast Cancer. (2024). 26 Ocak 2024 tarihinde <https://www.webmd.com/breast-cancer/toc-early-breast-cancer> adresinden erişildi.
- “Ensemble Karar Ağaçları-Ensemble Decision Trees (DT) / Regresyon Algoritması” by İLYAS BERK FIRSAT on Prezi. (2024). 28 Mart 2024 tarihinde <https://prezi.com/p/vzfhtellvaz4/ensemble-karar-agaclar-ensemble-decision-trees-dt-regresyon-algoritmas/> adresinden erişildi.
- Fan, M., Yuan, W., Zhao, W., Xu, M., Wang, S., Gao, X. ve Li, L. (2020). Joint Prediction of Breast Cancer Histological Grade and Ki-67 Expression Level Based on DCE-MRI and DWI Radiomics. *IEEE Journal of Biomedical and Health Informatics*, 24(6), 1632–1642. doi:10.1109/JBHI.2019.2956351
- Fatima, N., Liu, L., Hong, S. ve Ahmed, H. (2020). Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis. *IEEE Access*, 8, 150360–150376. doi:10.1109/ACCESS.2020.3016715
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A. ve Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4), 778–789. doi:10.1002/ijc.33588
- Ganggayah, M. D., Taib, N. A., Har, Y. C., Lio, P. ve Dhillon, S. K. (2019). Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Medical Informatics and Decision Making*, 19(1), 1–17. doi:10.1186/s12911-019-0801-4
- Haq, A. U., Li, J. P., Saboor, A., Khan, J., Wali, S., Ahmad, S., ... Zhou, W. (2021). Detection of Breast Cancer through Clinical Data Using Supervised and Unsupervised Feature Selection Techniques. *IEEE Access*, 9, 22090–22105. doi:10.1109/ACCESS.2021.3055806
- Haque, M. R., Islam, M. M., Iqbal, H., Reza, M. S. ve Hasan, M. K. (2018). Performance Evaluation of Random Forests and Artificial Neural Networks for the Classification of Liver Disorder. *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering, IC4ME2 2018*. doi:10.1109/IC4ME2.2018.8465658

- Home - UCI Machine Learning Repository. (2024). 8 Aralık 2024 tarihinde <https://archive.ics.uci.edu/> adresinden erişildi.
- Houssein, E. H., Emam, M. M., Ali, A. A. ve Suganthan, P. N. (2021, 1 Nisan). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*. Elsevier Ltd. doi:10.1016/j.eswa.2020.114161
- Huang, Z. ve Chen, D. (2022). A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm. *IEEE Access*, 10, 3284–3293. doi:10.1109/ACCESS.2021.3139595
- Islam Ayon, S. ve Milon Islam, M. (2019). Diabetes Prediction: A Deep Learning Approach. *International Journal of Information Engineering and Electronic Business*, 11(2), 21–27. doi:10.5815/ijieeb.2019.02.03
- Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M. ve Kabir, M. N. (2020). Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. *SN Computer Science*, 1(5). doi:10.1007/s42979-020-00305-w
- Islam, M. Z., Islam, M. M. ve Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, 20. doi:10.1016/j.imu.2020.100412
- Israni, P. (2019). Breast cancer diagnosis (BCD) model using machine learning. *International Journal of Innovative Technology and Exploring Engineering*, 8(10), 4456–4463. doi:10.35940/ijitee.J9973.0881019
- k-en Yakın Komşu Algoritması ve Bir Uygulama (Kredi Riskini Sınıflandırma) - PDF Free Download. (2024). 27 Mart 2024 tarihinde <https://docplayer.biz.tr/15448102-K-en-yakin-komsu-algoritmasi-ve-bir-uygulama-kredi-riskini-siniflandirma.html> adresinden erişildi.
- Kaan Ülgen. (2017). Makine Öğrenimi Bölüm-5 (Karar Ağaçları). *Makine Öğrenimi Bölüm-5 (Karar Ağaçları)*. 27 Mart 2024 tarihinde <https://www.linkedin.com/pulse/makine-ogrenimi-bolum-5-karar-agac-lari-eyup-kaan-ulgen/?originalSubdomain=tr> adresinden erişildi.
- Kadam, V. J., Jadhav, S. M. ve Vijayakumar, K. (2019). Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax Regression. *Journal of Medical Systems*, 43(8). doi:10.1007/s10916-019-1397-z
- Kaggle: Your Machine Learning and Data Science Community. (2024). 8 Aralık 2024 tarihinde <https://www.kaggle.com/> adresinden erişildi.
- Kartal, E. (2015). Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama, (January).
- Kaya Keleş, M. (2019). Breast cancer prediction and detection using data mining classification algorithms: A comparative study. *Tehnicki Vjesnik*, 26(1), 149–155. doi:10.17559/TV-20180417102943
- Khourdifi, Y. ve Bahaj, M. (2018). Applying best machine learning algorithms for breast cancer prediction and classification. *2018 International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2018*, 1–5. doi:10.1109/ICECOCS.2018.8610632

- KNN (K-En Yakın Komsu). KNN algoritması, son derece basit... | by ABDULLAH ATCILI | Machine Learning Türkiye | Medium. (2024). 28 Mart 2024 tarihinde <https://medium.com/machine-learning-türkiye/knn-k-en-yakın-komşu-7a037f056116> adresinden erişildi.
- Krag, D. (2020). Breast Cancer. *Surgical Oncology*. doi:10.1201/9781498712651-4
- Kumar Jakhar, A., Singh, M. ve Gupta, A. (2022). SELF: A Stacked-based Ensemble Learning Framework for Breast Cancer Classification SELF: A Stacked-based Ensemble Learning Framework for Breast Cancer Classification. doi:10.21203/rs.3.rs-2013877/v1
- LG, A. ve AT, E. (2013). Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. *Journal of Health & Medical Informatics*, 04(02). doi:10.4172/2157-7420.1000124
- Maliha, S. K., Ema, R. R., Ghosh, S. K., Ahmed, H., Mollick, M. R. J. ve Islam, T. (2019). Cancer Disease Prediction Using Naive Bayes, K-Nearest Neighbor and J48 algorithm. *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, 1–7. doi:10.1109/ICCCNT45670.2019.8944686
- Meme kanseri - Belirtiler ve nedenler - Mayo Clinic. (2024). 7 Aralık 2024 tarihinde <https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470> adresinden erişildi.
- Meme Kanseri Belirtileri ve İşaretleri | Umut Şehri. (2024). 7 Aralık 2024 tarihinde <https://www.cancercenter.com/cancer-types/breast-cancer/symptoms> adresinden erişildi.
- Memon, M. H., Li, J. P., Haq, A. U., Memon, M. H., Zhou, W. ve Lacuesta, R. (2019). Breast Cancer Detection in the IOT Health Environment Using Modified Recursive Feature Selection. *Wireless Communications and Mobile Computing*, 2019. doi:10.1155/2019/5176705
- Muduli, D., Dash, R. ve Majhi, B. (2020). Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach. *Biomedical Signal Processing and Control*, 59. doi:10.1016/j.bspc.2020.101912
- Muhammad, L. J., Islam, M. M., Usman, S. S. ve Ayon, S. I. (2020). Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN Computer Science*, 1(4). doi:10.1007/s42979-020-00216-w
- Naseem, U., Rashid, J., Ali, L., Kim, J., Haq, Q. E. U., Awan, M. J. ve Imran, M. (2022). An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers. *IEEE Access*, 10, 78242–78252. doi:10.1109/ACCESS.2022.3174599
- Omondigbe, D. A., Veeramani, S. ve Sidhu, A. S. (2019). Machine Learning Classification Techniques for Breast Cancer Diagnosis. *IOP Conference Series: Materials Science and Engineering*, 495(1). doi:10.1088/1757-899X/495/1/012033
- Polat, K. ve Güneş, S. (2009). A new feature selection method on classification of medical datasets: Kernel F-score feature selection. *Expert Systems with Applications*, 36(7), 10367–10373. doi:10.1016/j.eswa.2009.01.041

- Said, A. A., Abd-Elmegid, L. A., Kholeif, S. ve Gaber, A. A. (2018). Classification based on clustering model for predicting main outcomes of breast cancer using Hyper-Parameters Optimization. *International Journal of Advanced Computer Science and Applications*, 9(12), 268–273. doi:10.14569/IJACSA.2018.091239
- Sci-Hub | Artificial Intelligence and Applied Mathematics in Engineering Problems. Lecture Notes on Data Engineering and Communications Technologies | 10.1007/978-3-030-36178-5. (2024). 17 Aralık 2024 tarihinde https://sci-hub.se/10.1007/978-3-030-36178-5_88 adresinden erişildi.
- Sen, A. (2021). Data Mining and Principal Component Analysis on Coimbra Breast Cancer Dataset. *Proceedings of Intelligent Computing and Technologies Conference*, 42–49. doi:10.21467/proceedings.115.5
- Sharma, S., Aggarwal, A. ve Choudhury, T. (2018). Breast Cancer Detection Using Machine Learning Algorithms. *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018*, (MI), 114–118. doi:10.1109/CTEMS.2018.8769187
- Shinde, Y., Kenchappagol, A. ve Mishra, S. (2022). Comparative Study of Machine Learning Algorithms for Breast Cancer Classification. *Smart Innovation, Systems and Technologies*, 286, 545–554. doi:10.1007/978-981-16-9873-6_49
- Singh, S. N. ve Thakral, S. (2018). Using data mining tools for breast cancer prediction and analysis. *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*, 1–4. doi:10.1109/CCAA.2018.8777713
- Symptoms of Breast Cancer | Breast Cancer | CDC. (2024). 7 Aralık 2024 tarihinde <https://www.cdc.gov/breast-cancer/symptoms/index.html> adresinden erişildi.
- Uçar, M. K. (2017). *OBSTRÜKTİF Uyku ApnTeşhisiİçinMakineÖğrenmesi TabanlıYeniBir YönGeliştirilmesi*.
- UÇAR, M. K. (2019). Eta Correlation Coefficient Based Feature Selection Algorithm for Machine Learning: E-Score Feature Selection Algorithm. *Journal of Intelligent Systems: Theory and Applications*, 2(1), 7–12. doi:10.38016/jista.498799
- Uçar, M. K., Bozkurt, M. R., Bilgin, C. ve Polat, K. (2017). Automatic detection of respiratory arrests in OSA patients using PPG and machine learning techniques. *Neural Computing and Applications*, 28(10), 2931–2945. doi:10.1007/s00521-016-2617-9
- Uçar, M. K., Bozkurt, M. R., Bilgin, C. ve Polat, K. (2018). Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques. *Neural Computing and Applications*, 29(8), 1–16. doi:10.1007/s00521-016-2365-x
- Wang, Y., Chu, Y. M., Thaljaoui, A., Khan, Y. A., Chammam, W. ve Abbas, S. Z. (2021). A multi-feature hybrid classification data mining technique for human-emotion. *BioData Mining*, 14(1). doi:10.1186/s13040-021-00254-x

ÖZGEÇMİŞ

Ad-Soyad : İlker ÇAKAR

ÖĞRENİM DURUMU:

- **Lisans** : 2020, Düzce Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2021 yılından beri İstanbul Büyükşehir Belediyesi bünyesinde çalışmakta.