

**KOCAELİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**YÜKSEK LİSANS TEZİ**

**ENDOMETRİAL KANSERİN AMELİYAT ÖNCESİ BT  
GÖRÜNTÜLERİNDE ELDE EDİLEN RADİOMİCS VERİLERİ  
KULLANILARAK MAKİNE ÖĞRENMESİ TEKNİKLERİYLE  
KİTLE İÇİN YÜKSEK RİSK DEĞERLENDİRMESİ YÖNTEMİ**

**FEHİME YİĞİT**

**KOCAELİ 2021**

**KOCAELİ ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ**  
**ANABİLİM DALI**

**YÜKSEK LİSANS TEZİ**

**ENDOMETRİAL KANSERİN AMELİYAT ÖNCESİ BT**  
**GÖRÜNTÜLERİNDE ELDE EDİLEN RADİOMİCS VERİLERİ**  
**KULLANILARAK MAKİNE ÖĞRENMESİ TEKNİKLERİYLE**  
**KİTLE İÇİN YÜKSEK RİSK DEĞERLENDİRMESİ YÖNTEMİ**

**FEHİME YİĞİT**

**Dr.Öğr. Üyesi Alpaslan Burak İNNER**

**Danışman, Kocaeli Üniversitesi**

.....

**Dr.Öğr. Üyesi Özgür ÇAKIR**

**Jüri Üyesi, Kocaeli Üniversitesi**

.....

**Dr.Öğr. Üyesi Ersin KAYA**

**Jüri Üyesi, Konya Teknik Üniversitesi**

.....

**Tezin Savunulduğu Tarih: 28.01.2021**

## ÖNSÖZ VE TEŞEKKÜR

Bu tez çalışmasında, Bilgisayarlı Tomografi görüntülerinden elde edilen doku analizi verilerinde makine öğrenmesiyle Endometrial Kanser kitle alt-tipi tespiti yapılmıştır. Yapılan çalışmaların makine öğrenmesi ve hastalık tespiti gibi alanlarda çalışan kişilere faydalı olmasını ümit ederim.

Tez çalışmam süresince büyük bir özen ve özveriyle, bilgi ve tecrübelerini benimle paylaşan, beni motive edip desteklerini esirgemeyen değerli tez danışmanım, Dr. Öğr. Üyesi Alpaslan Burak İNNER Hocama çok teşekkür ederim. Tez çalışmalarına çok büyük destek ve emek veren Dr. Öğr. Üyesi Özgür ÇAKIR'a değerli katkıları için teşekkür ederim.

Çalışmalarım ve eğitim hayatım boyunca maddi ve manevi destekleriyle beni hiçbir zaman yalnız bırakmayan ve beni motive eden canım anneme, babama, kardeşlerime ve teyzeme çok teşekkür ederim. Zorlandığımda hep yanımda olan, benden sevgi ve desteklerini esirgemeyen kıymetli arkadaşlarıma çok teşekkür ederim.

Ocak – 2021

Fehime YİĞİT

## İÇİNDEKİLER

ÖNSÖZ VE TEŞEKKÜR .....	i
İÇİNDEKİLER .....	ii
ŞEKİLLER DİZİNİ.....	iv
TABLolar DİZİNİ .....	v
SİMGELER VE KISALTMALAR DİZİNİ .....	vii
ÖZET.....	viii
ABSTRACT.....	ix
GİRİŞ .....	1
1. GENEL BİLGİLER .....	4
1.1. Endometrial Kanser .....	4
1.2. Radyolojik Tanısal Yöntemler.....	4
1.2.1. Manyetik rezonans görüntüleme.....	4
1.2.2. Bilgisayarlı tomografi.....	5
1.3. Radiomics .....	5
1.3.1. ROI ve segmentasyon .....	6
1.3.2. 3D Slicer .....	6
1.3.3. Doku(Texture) analizi.....	6
1.3.4. Pyradiomics .....	9
1.4. Makine Öğrenmesi.....	10
1.4.1. Destek vektör makineleri.....	10
1.4.2. K-En yakın komşu algoritması .....	12
1.4.3. Karar ağaçları (Decision Tree) .....	13
1.4.4. Rasgele orman (Random forest) .....	14
1.4.5. Çok katmanlı algılayıcı.....	14
1.4.6. Gradient boosting machines .....	15
1.4.7. XGBoost .....	16
1.4.8. LightGBM .....	16
1.4.9. CatBoost .....	16
1.5. Öznitelik Seçimi .....	17
1.5.1. Chi square test (Ki-kare yöntemi) .....	17
1.5.2. Mutual information (Karşılıklı bilgi).....	18
1.5.3. MRMR.....	18
1.5.4. ReliefF .....	19
1.5.5. Step forward selection .....	19
1.5.6. Step backward selection .....	19
1.6. Sınıflandırma Performans Metrikleri.....	20
1.6.1. Karmaşıklık matrisi .....	20
1.6.2. Doğruluk.....	20
1.6.3. Recall (Sensitivity) .....	21
1.6.4. Specificity (Özgüllük) .....	21
1.6.5. Eğri altında kalan alan (AUC) .....	21
1.6.6. Precision (Kesinlik) .....	22
1.6.7. F-Score.....	22

1.6.8. Matthews correlation coefficient .....	22
2. MATERYAL VE YÖNTEM.....	24
2.1. Veri Seti .....	24
2.2. Hasta Seçimi .....	24
2.3. BT Parametreleri.....	25
2.4. Radiomics Verilerinin Çıkarılması .....	25
2.5. Veri Ön İşleme.....	26
2.6. Öznitelik Seçimi .....	27
2.7. Makine Öğrenmesinin Uygulanması .....	28
3. BULGULAR VE TARTIŞMA.....	30
3.1. Endometrioid- Seröz Alt-Tip İkili Sınıflandırma Sonuçları.....	30
3.2. Myom- NonMyom İkili Sınıflandırma Sonuçları.....	38
3.3. Myom-Endometrioid-Seröz Çok Sınıf Sınıflandırma Sonuçları .....	46
4. SONUÇLAR VE ÖNERİLER.....	57
KAYNAKLAR .....	58
KİŞİSEL YAYIN VE ESERLER .....	64
ÖZGEÇMİŞ .....	65

## ŞEKİLLER DİZİNİ

Şekil 1.1. GLCM analizinin şematik çizimi, a) Gri Seviye Görüntü, b)Nümerik Gri Seviye Görüntü, c) Co-occurence Matris .....	8
Şekil 1.2. Destek vektör makineleri .....	11
Şekil 1.3. KNN algoritması için örnek veri dağılımı .....	13
Şekil 1.4. Karar Ağacı yapısı .....	14
Şekil 1.5. Çok Katmanlı Algılayıcı Modeli .....	15
Şekil 1.6. Karmaşıklık matrisi.....	20
Şekil 2.1. Üç kesitli BT görüntüsü üzerinde segmentasyon işlemi, a) Axial Plan, b) Sagittal Plan, c)Koronal Plan.....	26
Şekil 3.1. a) Decision Tree kullanarak yapılan sınıflandırma sonucu elde edilen karmaşıklık matrisi b) CatBoost karmaşıklık matrisi .....	47
Şekil 3.2. Karmaşıklık matrisleri a) GBM b) CatBoost .....	49
Şekil 3.3. Karmaşıklık matrisleri, a) SVM, b)GBM, c)LightGBM, d) CatBoost .....	50
Şekil 3.4. MLP'ye ait karmaşıklık matrisi .....	52
Şekil 3.5. Karmaşıklık matrisleri, a) KNN, b) Random Forest.....	53
Şekil 3.6. Karmaşıklık matrisleri, a) SVM, b)MLP .....	55
Şekil 3.7. Karmaşıklık matrisleri, a)SVM, b)XGBoost .....	56

## TABLolar DİZİNİ

Tablo 3.1.	Tüm öznitelikler kullanılarak yapılan Endometrioid-Seröz sınıflandırma sonuçları.....	30
Tablo 3.2.	Endometrioid- Seröz sınıflandırması için kullanılan yöntemler ve elde edilen öznitelikler .....	31
Tablo 3.3.	Chi-Square Test ile öznitelik seçimi sonrası Endometrioid-Seröz sınıflandırma sonuçları.....	32
Tablo 3.4.	Mutual Information ile seçilen özniteliklerle elde edilen Endometrioid-Seröz sınıflandırma sonuçları .....	33
Tablo 3.5.	ReliefF ile seçilen özniteliklerle, K=4 ve K=10 için elde edilen Endometrioid-Seröz sınıflandırma sonuçları .....	33
Tablo 3.6.	MRMR ile seçilen özniteliklerle elde edilen Endometrioid Seröz sınıflandırma sonuçları.....	34
Tablo 3.7.	Endometrioid-Seröz sınıflandırması için SFS ve SBS algoritmalarıyla seçilen öznitelikler .....	36
Tablo 3.8.	SFS ile elde edilen Endometrioid-Seröz sınıflandırma sonuçları .....	37
Tablo 3.9.	SBS ile elde edilen Endometrioid-Seröz sınıflandırma sonuçları .....	37
Tablo 3.10.	Tüm Öznitelikler kullanılarak yapılan Myom-NonMyom sınıflandırma sonuçları.....	38
Tablo 3.11.	Myom-NonMyom Sınıflandırması için kullanılan öznitelik seçim yöntemleri ve elde edilen öznitelikler.....	39
Tablo 3.12.	Chi Square Test ile öznitelik seçimi sonrası Myom-NonMyom sınıflandırma sonuçları.....	40
Tablo 3.13.	Mutual Information ile seçilen özniteliklerle elde edilen Myom-Nonmyom sınıflandırma sonuçları.....	41
Tablo 3.14.	ReliefF ile seçilen özniteliklerle K=4 ve K=10 için yapılan Myom-NonMyom sınıflandırma sonuçları .....	42
Tablo 3.15.	MRMR ile seçilen özniteliklerle elde edilen Myom-NonMyom sınıflandırma sonuçları .....	43
Tablo 3.16.	Myom-NonMyom sınıflandırması için SFS ve SBS algoritmalarıyla seçilen öznitelikler .....	44
Tablo 3.17.	SFS kullanılarak elde edilen Myom-NonMyom sınıflandırma sonuçları .....	45
Tablo 3.18.	SBS ile elde edilen Myom-NonMyom sınıflandırma sonuçları .....	45
Tablo 3.19.	Tüm öznitelikler kullanılarak yapılan çoklu sınıflandırma sonuçları .....	46
Tablo 3.20.	Çok sınıflı sınıflandırma için seçilen öznitelikler.....	48
Tablo 3.21.	Chi Square Test ile öznitelik seçimi sonrası yapılan çoklu sınıflandırma sonuçları.....	48
Tablo 3.22.	Mutual Information ile seçilen özniteliklerle elde edilen çoklu sınıflandırma sonuçları.....	49
Tablo 3.23.	ReliefF ile seçilen özniteliklerle K=4 ve K=10 için elde edilen çoklu sınıflandırma sonuçlar .....	51

Tablo 3.24.	MRMR yöntemiyle elde edilen çoklu sınıflandırma sonuçları.....	52
Tablo 3.25.	Çok sınıflı sınıflandırma için SFS ve SBS ile seçilen öznitelikler.....	53
Tablo 3.26.	SFS algoritması kullanılarak elde edilen çoklu sınıflandırma sonuçları .....	54
Tablo 3.27.	SBS ile elde edilen çoklu sınıflandırma sonuçları .....	55





## SİMGELER VE KISALTMALAR DİZİNİ

### Kısaltmalar

AUC	: Area Under the Curve (Eğri Altında Kalan Alan)
BT	: Bilgisayarlı Tomografi
CPTAC	: Clinical Proteomic Tumor Analysis Consortium (Klinik Proteomik Tümör Analiz Konsorsiyumu)
DICOM	: Digital Imaging and Communications in Medicine (Tıpta Dijital Görüntüleme ve İletişim)
DMI	: Depth of Myometrial Invasion (Miyometriyal invazyon derinliği)
EFB	: Exclusive Feature Bundling (Özel Değişken Paketi)
GBM	: Gradient Boosting Machines (Gradyan Artırma Makineleri)
GLCM	: Grey Level Co-occurrence Matrix (Gri Seviye Eş Oluşum Matrisi)
GLDM	: Grey Level Dependence Matrix (Gri Seviye Bağımlılık Matrisi)
GLRLM	: Grey Level Run Length Matrix (Gri Seviye Dizi Uzunluğu Matrisi)
GLSZM	: Grey Level Size Zone Matrix (Gri Seviye Boyutu Bölge Matrisi)
KNN	: K-Nearest Neighbours (K-En Yakın Komşu )
MCC	: Matthews Correlation Coefficient (Matthews Korelasyon Katsayısı)
MLP	: Multi Layer Perceptrons (Çok Katmanlı Algılayıcılar)
MRI	: Magnetic Resonance Imaging (Manyetik Rezonans Görüntüleme)
MRMR	: Minimum Redundancy Maximum Relevance (Minimum Fazlalık Maksimum Alaka)
NGTDM	: Neighborhood Grey Tone Difference Matrix (Komşuluk Gri Ton Fark Matrisi)
PET	: Positron Emission Tomography (Pozitron Emisyon Tomografi)
ROC	: Receiver Operating Characteristic (Alıcı İşletim Karakteristiği)
ROI	: Region of Interest (İlgili Bölge)
SBS	: Step Backward Selection (Geri Yönlü Arama Seçimi)
SFS	: Step Forward Selection (İleri Yönlü Arama Seçimi)
SVM	: Support Vector Machines (Destek Vektör Makineleri)
TCGA	: The Cancer Genome Atlas (Kanser Genom Atlası)
TCIA	: The Cancer Imaging Archive (Kanser Görüntüleme Arşivi)
UCEC	: Uterine Corpus Endometrial Carcinoma (Rahim Yapısı Endometrial Karsinom)

# ENDOMETRİAL KANSERİN AMELİYAT ÖNCESİ BT GÖRÜNTÜLERİNDE ELDE EDİLEN RADIOMİCS VERİLERİ KULLANILARAK MAKİNE ÖĞRENMESİ TEKNİKLERİYLE KİTLE İÇİN YÜKSEK RİSK DEĞERLENDİRMESİ YÖNTEMİ

## ÖZET

Endometrial kanser; günümüzde giderek artan obezite vakaları, fiziksel aktivitelerin azalması ve ortalama yaşam süresinin artmasıyla birlikte gün geçtikçe yaygınlaşan bir jinekolojik kanser türüdür. Endometrial kanserde hızlı ve yüksek doğrulukla tanı koymak çok önemlidir. Makine öğrenmesiyle Endometrial kanser için yapılan çalışma sayısı oldukça az olduğundan yapılan çalışma bu alana katkı sağlamayı hedefledi. Bu tez çalışmasında açık kaynaklı veri kümesi koleksiyonu olan The Cancer Genome Atlas (TCGA) 'dan alınan Bilgisayarlı Tomografi (BT) verileri kullanıldı. 135 adet hastaya ait olan BT verileri uzman radyolog tarafından 3D Slicer yazılımı ve Pyradiomics eklentisiyle doku analizi işlemi yapılarak elde edilen 130 öznitelik içeren radiomics verileri kullanıldı. Radiomics radyografik tıbbi görüntülerden çok sayıda sayısal parametrik değerler çıkarılarak elde edilen veridir.

Veri ön işleme adımında radiomics sürüm bilgilerini içeren öznitelikler çıkarıldı ve aynı gruba ait olan kitle tipleri etiketlendi. Özniteliklerin sayısının azaltılması için Chi-Square Test, Mutual Information, ReliefF, MRMR, SBS ve SFS algoritmaları kullanıldı. Bu yöntemlerle elde edilen özniteliklerle; KNN, SVM, MLP, Decision Tree, Random Forest, GBM, LightGBM, XGBoost ve CatBoost sınıflandırma algoritmaları kullanılarak üç farklı sınıflandırma yapıldı. Yapılan ilk sınıflandırma Endometrioid ve Seröz alt-tiplerini kendi aralarında ayırabilmek için yapıldı ve bu sınıflandırma sonucunda elde edilen başarı %92 oldu. İkinci sınıflandırmada Myom tipini diğer tiplerden ayırmak istendi ve Myom-NonMyom sınıflandırması sonucunda %95 başarıya ulaşıldı. Son olarak çok sınıflı sınıflandırma yapıldı ve kitleler kendi aralarında sınıflandırıldı. Genel olarak yüksek sonuçlar veren bu sınıflandırmada %92 başarıya ulaşıldı. Bu çalışma sonucunda; radiomics verilerinin makine öğrenmesiyle kanser alt-tiplerini tespit etmek için kullanılmasının uygun olduğu görüldü.

**Anahtar Kelimeler:** Doku Analizi, Endometrial Kanser, Makine Öğrenmesi, Radiomics, Yapay Zeka.

# **HIGH RISK ASSESSMENT METHOD OF ENDOMETRIAL CANCER WITH MACHINE LEARNING TECHNIQUES USING RADIOMICS DATA OBTAINED IN PRE-OPERATIONAL CT IMAGES**

## **ABSTRACT**

Endometrial cancer is a type of gynecological cancer that is spreading day by day with increasing obesity cases, decrease in physical activities and increase in average life expectancy. It is very important to diagnose endometrial cancer quickly and with high accuracy. There are many studies conducted in the field of disease detection with machine learning. However, since the number of studies for endometrial cancer is quite low, the thesis study aimed to contribute to this field. In this thesis, CT data from The Cancer Genome Atlas (TCGA), an open source data set collection, were used. The CT data belonging to 135 patients were used by an expert radiologist, using 3D Slicer and Pyradiomics software, using tissue analysis including filtering and segmentation steps, and 130-feature radiomics data. Radiomics is the data obtained by extracting a large number of numerical parametric values from radiographic medical images.

After the data pre-processing steps, masses types were tagged. Chi-Square Test, Mutual Information, ReliefF, MRMR, SBS and SFS algorithms were used to reduce the number of features. With the attributes obtained by these methods; Three different classifications were made using KNN, SVM, MLP, Decision Tree, Random Forest, GBM, LightGBM, XGBoost and CatBoost classification algorithms. The first classification was made in order to distinguish between Endometrioid and Serous subtypes and the success achieved as a result of this classification was 92%. In the second classification, it was wanted to distinguish the Myom type from other types, and as a result of the Myom-NonMyom classification, 95% success was achieved. Finally, multi-class classification was made and masses were classified among themselves. In this classification, which generally gives high results, 92% success was achieved. As a result of this study, it was seen that it was appropriate to use radiomics data to detect cancer subtypes with machine learning.

**Keywords:** Texture Analysis, Endometrial Cancer, Machine Learning, Radiomics, Artificial Intelligence.

## GİRİŞ

Endometrial kanser; günümüzde giderek artan obezite vakaları, fiziksel aktivitelerin azalması ve ortalama yaşam süresinin artmasıyla birlikte gün geçtikçe yaygınlaşan bir jinekolojik kanser türüdür [1]. Dünya çapında her yıl yaklaşık 400.000 yeni vakanın görüldüğü ve 90.000 civarı hastanın endometriyal kanser ve ilişkili nedenlerden hayatını kaybettiği düşünülmektedir [2]. Ayrıca, bu yaşam koşullarıyla ve hastalık gelişimine sebep olabilecek diğer faktörler sebebiyle 2030 yılına kadar endometriyal kanser sıklığında %55'e varan oranlarda artış beklenmektedir [3].

Endometriyal kanser, diğer adıyla rahim kanseri; her yıl yaklaşık 382.100 yeni vaka ve 89.900 kanser ilişkili ölümlerle ciddi bir kanser yükü oluşturmaktadır [4]. 2018 yılı verilerine göre dünyada kadınları etkileyen en sık yedinci ve jinekolojik kanserler arasında servikal kanserden sonra en sık ikinci kanserdir [2, 4].

Endometriyal kanser, Avrupa'da kanser ilişkili ölümlere en sık sebep olan sekizinci kanserken; Kuzey Amerika'da ise her yıl yaklaşık 55.000 yeni vaka ve 11.000 ölümlerle en sık görülen altıncı kanser olarak yer almaktadır [5,6].

Doku analizi, bir tümör görüntüsündeki pikseller ve voksellerin değerlendirilmesi yoluyla heterojenliğin kapsamlı, kantitatif bir analizini sağlamaya çalışır [7].

Görüntü dokusu özellikleri, daha doğru bir teşhis, gelişmiş preoperatif risk sınıflandırması ve beyin [8], akciğer [9], meme [10], prostat [11] gibi çeşitli kanser türlerinde tedavi yanıtının daha iyi değerlendirilmesini sağlayan görüntüleme belirteçleri olarak önerilmiştir.

Doku analizi ile yapılan son çalışmalar, endometriyal kanserde klinik davranış, tedavi ve hastalık sonucunu tahmin etmek için MRI (Magnetic Resonance Imaging) [12] ve PET (Positron Emission Tomography) [13] 'ün faydasını bildirmiştir.

Son yıllarda tıp alanında hastalık tespiti için makine öğrenmesi yöntemlerinin kullanımını oldukça popüler olmuştur. Hastalık, tümör ve tümör alt-tiplerinin tespiti için radiomics verilerinin kullanımı da yaygınlaşmıştır.

Asadullayeva [14] tezinde; BT (Bilgisayarlı Tomografi) görüntülerine doku analizi yaparak akciğer kanserinin tiplerinin makine öğrenmesiyle sınıflandırılmasını incelemiştir. Taşlıçay [15] ise; MRI görüntülerini kullanarak, radiomics tabanlı makine öğrenmesinin menenjiom hastalarında tümör alt-tip ve derecelerini sınıflandırma üzerindeki katkısını incelemiştir.

Hsiang-Chun Dong ve arkadaşları [16] MRI görüntüleri üzerinde; Endometriyum Kanserinin derinliğini tanımlamak için Evrişimli sinir ağı ile derin öğrenmeyi kullanmışlardır. Akazawa ve arkadaşları [17] ise; rahim kanserinin erken tanısı için makine öğrenmesi yöntemlerini kullanmışlardır. Son olarak Stanzione ve arkadaşları [18]; MRI görüntülerinden elde edilen radiomics verileriyle Endometrial kanser tanısı konmuş hastalarda miyometriyal invazyon derinliğinin (DMI) tespit etmek için makine öğrenmesi yöntemlerini kullanmışlardır.

Bununla birlikte, yapılan son araştırmalara göre bilgimiz dahilinde daha önceki hiçbir çalışma endometrial kanserin klinik evreleme ve alt grup tespitinde makine öğrenmesiyle BT doku analizinin olası değerini değerlendirmemiştir.

Bu çalışmada riskli ve yaygın bir kanser türü olan Endometrial kanser alt-tipleri incelendi. Bu çalışmanın amacı, makine öğrenmesi yöntemleriyle bilgimiz dahilinde daha önce literatürde kullanılmamış olan endometrial kanserin klinik evreleme ve alt grup tespitinde BT doku analizi verileri kullanılarak endometriyal karsinomun yüksek riskli alt tiplerini ve riskli olmayan iyi huylu tiplerini birbirlerinden ayırt edebilmektir.

Tez çalışmasının ilk bölümünde; endometrial kanser alt-tipleri, radyolojik tanısal yöntemler, radiomics verileri, doku analizi, makine öğrenmesi yöntemleri, öznelik seçim yöntemleri ve sınıflandırma metrikleriyle ilgili genel bilgiler verildi.

İkinci bölümde; veri seti, radiomics verilerinin ve özniteliklerinin elde edilmesi, öznitelik seçim işlemlerinin ve makine öğrenmesi yöntemlerinin uygulanması ele alındı.

Üçüncü bölümde ise yapılan sınıflandırma sonuçları verildi. Yapılan ilk sınıflandırma endometrial kanserin iki alt-tipi olan Endometrioid ve Seröz kötü huylu kitlelerin sınıflandırılmasıdır. Daha sonra iyi huylu kitle türü olan Myom kitlesi ile iyi huylu olmayan Endometrioid ve Seröz kitleleri sınıflandırıldı. Bu sınıflandırma ikili bir sınıflandırma olup Myom-NonMyom olarak geçecektir. Yapılan son sınıflandırma ise bu üç kitle türünün birbirinden çoklu olarak sınıflandırılmasıdır. Yapılan çoklu sınıflandırma kitlenin iyi veya kötü huylu olmasını tespit etmek yerine hangi alt-tip grubuna ait olduğunu tespit etmeyi amaçlar.

## 1. GENEL BİLGİLER

### 1.1. Endometrial Kanser

Dünya Sağlık Örgütü'nün sınıflandırmasına göre endometrial kanser; endometrioid karsinom, müsinöz adenokarsinom, seröz adenokarsinom, şeffaf hücreli adenokarsinom, ayrışmamış karsinom, nöroendokrin tümörler ve miks karsinom olarak sınıflandırılmaktadır [4,19]. Endometrial karsinom, hücre tipine göre iki alt tipe [tip 1 ve tip 2] ayrılır. En yaygın alt tip, vakaların % 75 ile % 80'ini oluşturan tip 1 [endometrioid] 'dir [20]. Çoğunlukla 55-65 yaş arası kadınlarda görülür ve nispeten yavaş ilerleyen ve daha olumlu sonuç veren kötü huylu bir tümördür [20][21]. Seröz karsinom ise tip 2 endometriyal karsinomun en yaygın alt tipidir ve vakaların% 5-10'unu oluşturur [22]. Tip 2 karsinomlar ise yüksek evrede tanı alırken, agresif yayılım ve kötü seyirle gider. Tip 2 karsinomlar, 65-75 yaş arasındaki kadınlarda görülebilen kötü huylu tümörlerdir.

Üreme çağındaki kadınlarda en sık görülen iyi huylu tümörlere de myom denir [23]. Myomlar ilk Matthew Baille tarafından 1793 yılında tanımlanmıştır. Myomlar düz kas hücrelerinden oluşmaktadır ve değişik oranlarda fibröz doku içermektedir [24]. Çalışmaların gerçekleştirildiği popülasyona ve kullanılan tanısal tekniğe göre %5,4 ile %77 kadında myom olduğu literatürde belirtilmiştir [25]. Yaş ilerledikçe görülme oranı artar fakat kadınların menopoz dönemi başladıktan sonra gerilemeye başlar [26].

### 1.2. Radyolojik Tanısal Yöntemler

#### 1.2.1. Manyetik rezonans görüntüleme

Manyetik Rezonans Görüntüleme (Magnetic Resonance Imaging, MRI), üç boyutlu detaylı anatomik görüntüler üreten bir görüntüleme teknolojisidir. Genellikle hastalık tespiti, teşhisi ve tedavi takibi için kullanılır. MRI tarayıcıları, vücudun kemikli olmayan kısımlarını veya yumuşak dokularını görüntülemek için özellikle uygundur. MRI, röntgen veya diğer radyasyonu kullanmadığı için, özellikle beyinde teşhis veya tedavi için sık görüntüleme gerektiğinde tercih edilen görüntüleme yöntemidir.

### 1.2.2. Bilgisayarlı tomografi

BT, vücudun farklı bölgelerindeki birçok hastalığın teşhisi için tercih edilen bir yöntemdir ve yüksek çözünürlüklü görüntüler sağlar [27]. Bu yüksek çözünürlüklü görüntüler sayesinde daha kaliteli radiomics özelliklerin çıkarılmasına yardımcı olur.

Bununla birlikte, BT görüntüleme performansı, aşağıdaki üç ana özellik dahil olmak üzere kullanılan protokolün farklı bileşenlerine bağlıdır [28]:

- İki ardışık dilim arasındaki milimetre (mm) cinsinden uzaklık olan dilim kalınlığı;
- Yoğunluk değişimlerini görüntü yoğunluklarına yansıtma yeteneği;
- Tomografik ölçümleri kesitsel görüntülere dönüştürmeyi amaçlayan yeniden inşa algoritması.

Spesifik klinik endikasyonlar için BT protokolleri genellikle farklı kurumlar arasında benzer olsa da, radiomics özellikleri aynı ayarlara sahip farklı tarayıcılar arasında bile farklılık gösterebilir [29].

### 1.3. Radiomics

Radiomics, radyografik tıbbi görüntülerden çok sayıda sayısal parametrik değerler (öznitelikler) çıkarılarak elde edilen veridir. Bu sayısal radiomics verileri çıplak gözle görülmeyen hastalık özniteliklerini ortaya çıkarabilir.

Radyolojik görüntülerde aynı hastalığa sahipmiş gibi görünen iki hastaya ait radiomics verileri çok farklı olabilir. Gözlemlenen ve kaydedilen verideki bu farklılıklarla daha doğru hastalık tespiti yapılır [27].

Radiomics'in amacı hastalık tespitinde kullanılmak üzere mümkün olduğu kadar çok ve anlamlı gizli sayısal görüntüleme verilerini çıkarmak ve analiz etmektir.

Radiomics BT, MRI, PET, X-ışını ve ultrasonografi gibi çeşitli görüntüleme tekniklerine uygulanabilir. Radiomics onkoloji çalışmalarında başlatılmıştır, ancak potansiyel olarak tüm hastalıklara uygulanabilir [30].

Radiomics verilerine ulaşabilmek için elde edilen görüntünün bazı ön işlemlerden geçirilmesi gerekir. Kesit kalınlığı; görüntülerin çözünürlüğünü ifade eden önemli bir



parametredir. Görüntülerdeki sinyal-gürültü oranının değişkenliği, görüntülerin benzer protokoller ve kesit kalınlıklarıyla elde edilip edilmemesi sonuçları etkileyen faktörlerdir [31]. Bu sebeple görüntü elde edilirken kullanılan protokol farklılıklarının oluşturabileceği sorunları azaltmak için analiz öncesi görüntülerin standardize edilmesi gerekir [15]. Bu işlem sonrasındaki adım ise segmentasyon olacaktır. Bütün bu adımları bünyesinde barındıran bazı yazılımlar yardımıyla radiomics ön işleme adımları rahatlıkla yapılabilir. ImageJ, MIPAV (Medical Image Processing, Analysis, and Visualization), LifeX, Mazda ve 3DSlicer bu yardımcı yazılımlar arasındadır. Bölütleme (segmentation) adımını ise özellik çıkarımı adımı takip eder.

### **1.3.1. ROI ve segmentasyon**

Hastalık tespiti için yapılacak en önemli adımlardan biri de ilgili bölgeyi tespit etmektir. Yani elde olan görüntülerden istenen alanı istenmeyen alanlardan ayırıştırıp, gereksiz olan kısımların çıkarılmasıdır.

Çalışılmak istenen bölgeye; ilgili bölge yani ROI (Region of Interest) denir. ROI'nin diğer bölgelerden çıkarılmasına ise bölütleme denir. Görüntü verileri üzerinde bölütleme yapıp, radiomics verilerini elde etmek için 3D Slicer yazılımı ve Pyradiomics eklentisi kullanılmıştır.

### **1.3.2. 3D Slicer**

3D Slicer, tıp bilişimi, görüntü işleme ve üç boyutlu görselleştirmede kullanılmak üzere tasarlanmış açık kaynaklı bir yazılım platformudur [32]. 3D Slicer sahip olduğu araçlar(tools) sayesinde segmentasyon ve radiomics verilerini oluşturmada çok kolaylık sağlayan bir yazılımdır. Radiomics veri ve öznitelikleri, 3D Slicer'ın sahip olduğu kullanışlı arayüz sayesinde rahatlıkla çıkarılabilir.

### **1.3.3. Doku(Texture) analizi**

Doku analizi işlemi; segmentasyon adımından sonra, kullanılacak olan verinin değerlendirilmesi için yapılan çok önemli bir adımdır. Radiomics özellikleri bu adımda elde edilir. Radiomics özellikleri temelde semantik ve agnostik olarak iki gruba ayrılır [30].

Semantik özellikler, radyoloji alanında lezyon tanımlamak için sıklıkla kullanılan boyut, şekil, lokasyon gibi özelliklerin nicel karşılığıdır [30].

Agnostik radiomics özellikleri ise; genellikle radyoloji sözlüğü içinde barınmayan ve matematiksel olarak çıkarılan nicel değerlerdir. Agnostik özelliklerden istatistiksel doku özellikleri kendi içinde birinci, ikinci ve yüksek dereceli istatistiksel doku özellikleri olarak üçe ayrılır [30].

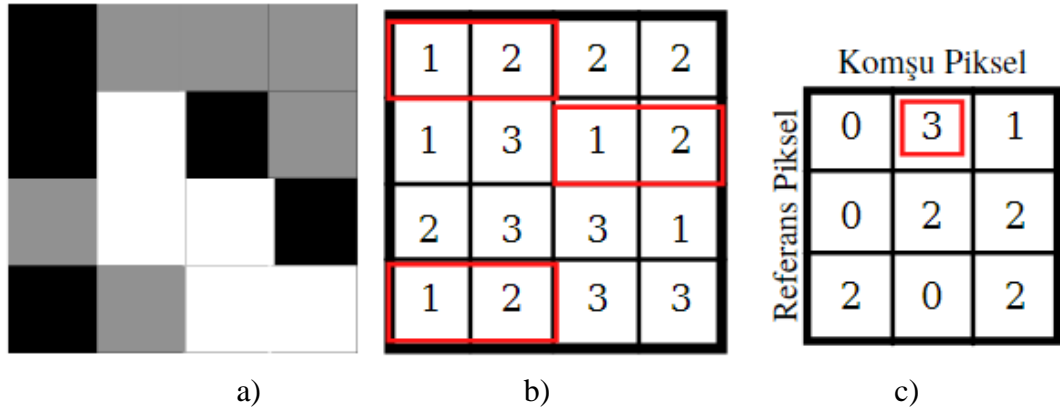
İstatistiksel doku analizinde doku öznitelikleri, görüntüde birbirine göre belirlenen konumlarda gözlenen yoğunluk kombinasyonlarının istatistiksel dağılımından hesaplanır.

Her kombinasyondaki yoğunluk noktalarının (piksel) sayısına göre, birinci dereceden istatistiksel öznitelikler, ikinci dereceden istatistiksel öznitelikler ve daha yüksek dereceden istatistiksel öznitelikler olarak sınıflandırılır [33].

Birinci dereceden istatistiksel öznitelikler; histogram tabanlı özniteliklerdir. Görüntüdeki gri düzey frekans dağılımını ölçer. Bu öznitelikler görüntüdeki piksellerin konumu ve birbirleriyle olan ilişkileriyle ilgilenmezler [34]. Görüntünün minimum, maksimum, ortalama, standart sapma, enerji, varyans, entropy (düzensizlik, rastgelelik), uniformity (tekillik), kurtosis (piksel histogramının yassılığı), skewness (çarpıklık, asimetri) değerleri gibi gri seviyelerin dağılımı ile ilgili nicel parametreleri ölçer [30].

İkinci dereceden istatistiksel öznitelikler; doku öznitelikleri olarak tanımlanır [30]. İkinci dereceden istatistiksel öznitelikler; görüntüde bulunan noktalar arasındaki ilişkileri ve gri seviye geçiş öznitelikler değerlendirir. Bu çıkarımları GLCM (Grey Level Co-occurrence Matrix) matrisini kullanarak elde eder [33].

GLCM, bir görüntüdeki noktalar arasındaki ilişkinin değerlendirmesi sonucunda oluşturulan matristir [34]. GLCM matris kullanılarak elde edilen ölçümler iki boyutlu görüntülerde 0, 45, 90 ve 135 derece olmak üzere dört farklı açıda, 3 boyutlu görüntülerde 13 farklı açıda değerlendirme yapılır [35].



Şekil 1.1. GLCM analizinin şematik çizimi, a) Gri Seviye Görüntü, b) Nümerik Gri Seviye Görüntü, c) Co-occurrence Matrisi

Şekil 1.1.'de GLCM analizinin şematik çizimi verildi. Şekil 1.1'deki ilk görüntü(a) gri seviye görüntüyü belirtir. İkinci görselde(b) ise gri seviye görüntünün rakamsal karşılığını ifade eder. Son görüntüdeki(c) matris ise soldan sağa horizontal yani 0 derece ile yapılan hesaplama ile elde edilen 'co-occurrence' matrisi göstermektedir. Nümerik gri seviye görüntüye bakıldığında, soldan sağa doğru 1 numaralı pikselden 2 numaralı piksele geçişler işaretlenmiştir. 1 numaralı pikselden 2 numaralı piksele geçiş üç defa görülmüştür. 'Co-occurrence' matriste dikey sütun referans piksel değerlerini, yatay sütun komşu piksel değerlerini göstermektedir. Referans olarak 1, komşu piksel rakamsal değeri olarak 2 alındığında çember içine alınmış toplam geçiş sayısı 3 olarak görülmektedir [35].

Yüksek düzey istatistiksel öznitelikler ise; GLRLM (Grey Level Run Length Matrix), GLSZM (Grey Level Size Zone Matrix) ve NGTDM (Neighborhood Grey Tone Difference Matrix) matrislerini kullanarak üç veya daha fazla nokta arasındaki ilişkileri değerlendirerek elde edilirler [34, 35].

GLRLM, aynı gri seviye yoğunluğuna sahip olan komşu noktaların sayısı üzerinden oluşturulan matristir [35]. GLSZM aynı gri seviye yoğunluğuna sahip noktaların bulunduğu alanların boyutlarını değerlendiren matristir [36].

NGTDM ise; görüntüdeki her noktanın gri seviye yoğunluk değeri ile komşu noktaların gri seviye yoğunluk değerlerinin ortalaması arasındaki farkını alarak işlem yapan matristir [37].

### 1.3.4. Pyradiomics

Pyradiomics; özellik algoritmaları kullanarak tıbbi görüntü verilerinden radiomics özelliklerin işlenmesini ve çıkarılmasını sağlayan açık kaynak kodlu bir platformdur [38]. PyRadiomics, 3D Slicer yazılımında bulunan bir eklentidir. Çok basit fakat oldukça kullanışlı olan pyradiomics kullanıcıya zaman kazandırmakla birlikte esnek bir analiz platformu da sağlar. PyRadiomics, sadece 3D Slicer platformunda değil, herhangi bir sistem üzerinde de kullanılabilir. PyRadiomics platformunda, radyolojik veriler segmente edilebilir, görüntülere daha iyi sonuç almak adına filtre uygulanabilir ve özellik çıkarımı yapılır; radiomics verileri elde edilebilir. Öznitelikler, orijinal görüntü üzerinde veya filtre seçeneği kullanılarak önceden işlenmiş görüntüler üzerinden hesaplanabilir. Bu filtreler; LoG (Gaussian of Laplacian) filtrelerinin yanı sıra kare, karekök, logaritma ve üstel filtrelerdir. LoG filtresinin uygulanması için Pyradiomics platformu, sırasıyla PyWavelets ve SimpleITK modüllerinden yararlanır. Diğer filtreler NumPy kütüphanesi kullanılarak uygulanır [38]. Her öznitelik, uygulanan filtre, öznitelik sınıfı ve öznitelik adından oluşan benzersiz bir isimle tanımlanır. PyRadiomics yazılımıyla elde edilen öznitelik sınıfları şunlardır:

- Birinci dereceden istatistiksel özellikler (First order statistics),
- 3 boyutlu şekil tabanlı özellikler (Shape-based 3D),
- 2 boyutlu şekil tabanlı özellikler (Shape-based 2D),
- Gri seviye eş oluşum matrisi (Grey level co-occurrence matrix, GLCM) kullanılarak elde edilen özellikler,
- Gri seviye bağımlılık matrisi (Grey level dependence matrix, GLDM) kullanılarak elde edilen özellikler,
- Gri seviye dizi uzunluğu matrisi (Grey level run length matrix, GLRLM) kullanılarak elde edilen özellikler,
- Gri seviye boyutu bölge matrisi (Grey level size zone matrix, GLSZM) kullanılarak elde edilen özellikler),
- Komşuluk gri ton fark matrisi (Neighborhood grey tone difference matrix, NGTDM) kullanılarak elde edilen özellikler).

Pyradiomics, hesaplanan radiomics özniteliklerinin yanı sıra; mevcut sürüm, uygulanan filtreler, ayarlar ve orijinal görüntü aralığı dahil olmak üzere çıkarma hakkında ek bilgiler içerir.

#### **1.4. Makine Öğrenmesi**

Son zamanlarda adından sıklıkça bahsettiren makine öğrenmesi kavramı, bilgisayarların herhangi bir şekilde programlanmadan öğrenmelerini sağlayan yapay zekâ alanını ifade eder [39, 40].

Makine öğrenmesi, daha önceki istatistiklere dayanarak gerçek zamanlı problemi tanıma becerisini geliştiren ve kesin talimatlar ve kuralların altındaki gerçek zamanlı problemi çözmek için performans gösteren algoritmaları planlamak ve geliştirmek için bilimsel bir alan olarak tanımlanabilir [40, 41].

Makine öğrenmesi denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olmak üzere üç kategoriye ayrılır [39].

Denetimsiz öğrenmede; etiketli veri yoktur. Tüm veri setine odaklanarak çalışılır. Veri setine bakılarak bir örüntü çıkarmak hedeflenir. Herhangi bir tahminleme yapılmaz. Kümeleme problemleri bu kategoridedir.

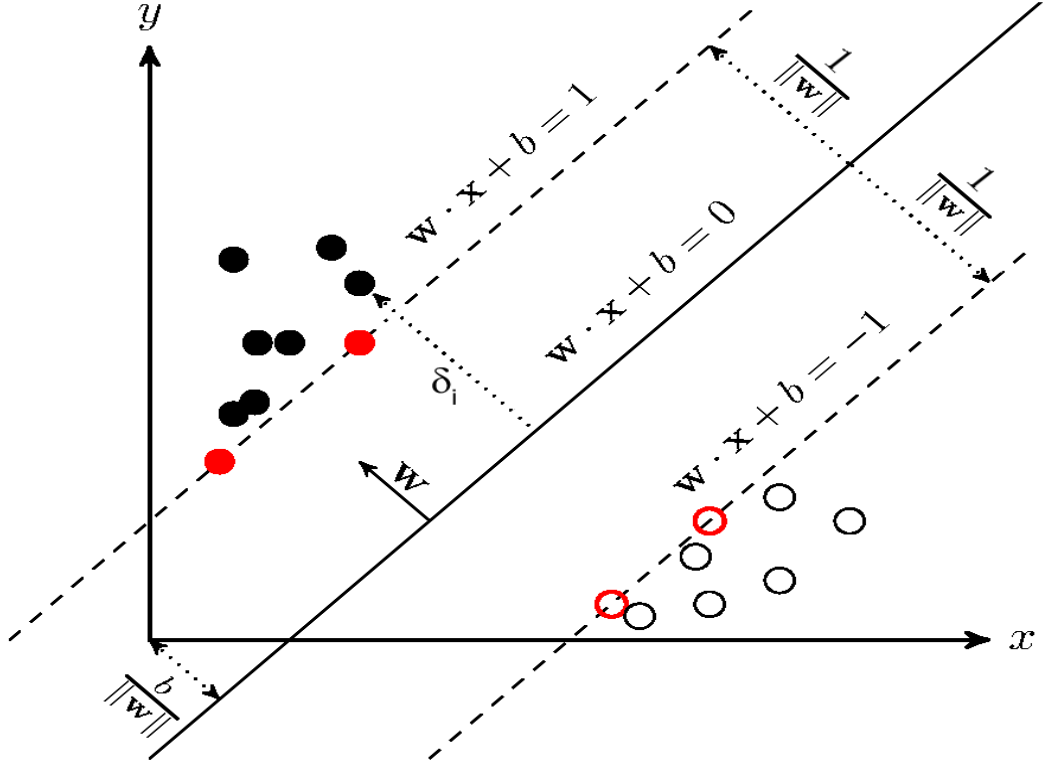
Denetimli öğrenmede ise; etiketli veri üzerinde çalışılır ve bir tahmin modeli geliştirilir. Sınıflandırma ve regresyon problemleri bu kategoridedir.

Sınıflandırma da kendi içinde ikili sınıflandırma ve çok sınıflı sınıflandırma olarak ikiye ayrılır. İkili sınıflandırmada veriler 0 ve 1 olarak etiketlenir ve iki kategoriye göre sınıflandırılır [42]. Çok sınıflı sınıflandırmada ise sınıf sayısı artar. Kullanılan her sınıf için bir sınıf etiketi atanır ve sınıflandırma buna göre yapılır.

##### **1.4.1. Destek vektör makineleri**

Destek vektör makineleri (Support Vector Machine-SVM) iki ve daha fazla sınıfa sahip verileri en iyi şekilde sınıflandırmak için kullanılan makine öğrenmesi algoritmasıdır.

Basit fakat etkili bir algoritma olan SVM; bu sınıflandırma işlemini hiper düzlemlerle yapar. İyi bir sınıflandırma yapmak için de en iyi hiper düzlemi bulmaya çalışır. En iyi hiper düzlemi ise destek vektörleri yardımıyla bulur. Sınıflandırmayı yapabilmek için; her sınıfa en yakın olacak şekilde sınır çizgileri çekilir ve bu sınır çizgileri yardımıyla en iyi hiper düzlem bulunur.



Şekil 1.2. Destek vektör makineleri [43]

Şekil 1.2.'de örnek bir destek vektör makinesinin grafiği verildi. Örnek sayısının  $N$  adet olduğu, doğrusal ayrılabilen bir eğitim kümesinin  $\{x_i|y_i\}$ ,  $i=1,2,\dots,N$  olduğunu varsayalım. Bu eğitim kümesinde  $x$ ; giriş vektörünü  $y$  ise; iki sınıfa ait etiket vektörünü temsil eder  $y \in \{-1,1\}$ . En iyi hiper düzlemi bulmak için  $b$  eşik değerine ve  $w$  ağırlık vektörüne ihtiyaç duyulur [44].

En iyi hiper düzlemi Denklem (1.1)'de verilen;

$$w \cdot x_i + b = 0 \quad (1.1)$$

ifadesi temsil eder. Bu denklemin amacı; iki sınıfı en iyi şekilde ayırmak ve iki farklı sınıfa ait olan veri noktalarını birbirinden olabildiğince uzak tutmaktır.

Destek vektör makinelerinin veri kümesinin dağılımı hakkında herhangi bir varsayımı yoktur. Dolayısıyla sınıflandırma için karar fonksiyonlarını en iyi şekilde kullanır. Doğrusal olarak ayrılabilen ve ayrılamayan veriler için kullanılan karar fonksiyonları farklıdır. Kullanılmak istenilen veriye göre en doğru karar fonksiyonu seçilmelidir.

Sınıflandırma aşamasında, ele alınan veri noktasının hangi sınıflara ait olduğunu belirleyen destek vektörlerinin denklemleri ise  $y_i \in \{1\}$  için Denklem (1.2)'de,  $y_i \in \{-1\}$  için Denklem (1.3)'de;

$$w \cdot x_i + b \geq 1; \quad y_i = 1 \quad (1.2)$$

$$w \cdot x_i + b \leq -1; \quad y_i = -1 \quad (1.3)$$

verildiği gibidir.

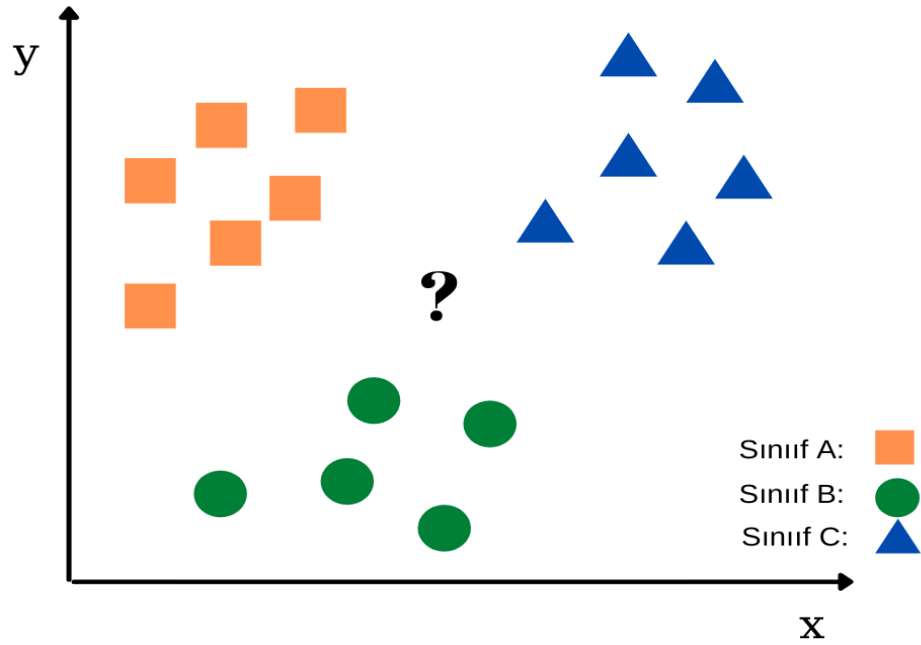
Destek vektör makineleri ikili sınıflandırma için oldukça başarılı bir sınıflandırma yöntemi olmakla birlikte çoklu sınıfa sahip olan verileri sınıflandırma konusunda aynı başarıya sahip olamamıştır. Bu yüzden Bire-bir ve Bire-Hepsi yöntemleri geliştirilmiştir.

Bire-Bir yöntemi seçtiği her sınıfı kalan sınıflarla birer birer sınıflandırarak çalışırken, Bire-Hepsi yönteminde; seçilen sınıf kalan sınıfların tümüyle karşılaştırılır. Böylelikle destek vektörlerinin çalışma mantığına uygun olarak ikili sınıflandırma yapılmış olur [45].

#### **1.4.2. K-En yakın komşu algoritması**

K-En Yakın Komşu Algoritması (K-Nearest Neighbors, KNN); sınıflandırma ve regresyonda kullanılan, kullanımı kolay bir makine öğrenmesi algoritmasıdır. Herhangi bir öğrenme algoritması kullanmaz ve model oluşturmaz.

Çalışma mantığı basittir. Elimizde olan veriye yeni bir veri geldiğinde, o verinin hangi sınıfa ait olduğunu bulmak için önce bir k değeri belirlenir. Belirlenen k sayısı en yakın komşu sayısıdır. Belirlenen bu k değerine göre verinin çoğunluk olarak hangi en yakın komşu grubuna dahil olduğuna karar verir. Yani KNN algoritmasında çıktısı bir sınıf üyeliğidir.



Şekil 1.3. KNN algoritması için örnek veri dağılımı

Şekil 1.3’ de üç farklı veri kümesi şematize edildi. KNN algoritmasının çalışma prensibine göre eldeki verilere yeni bir üye geldiğinde; yeni üyeyi belirlenen komşu sayısına göre en yakın veri kümesine dahil eder.

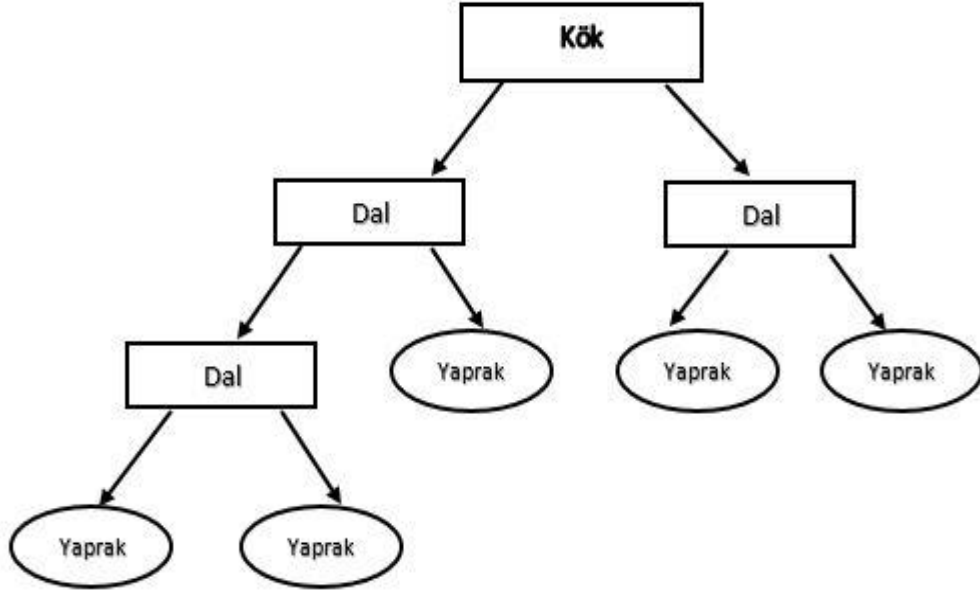
#### 1.4.3. Karar ağaçları (Decision Tree)

Karar ağaçları sınıflandırma ve regresyonda kullanılan gözetimli makine öğrenmesi algoritmalarındandır. Sayısal ve kategorik değişkenler üzerinde işlem yaparlar. Karar Ağaçları: Karar Ağaçları (DT), örnekleri özellik değerlerine göre sıralayarak sınıflandıran ağaçlardır [46, 47].

Karar ağaçlarındaki düğümler verideki her örneğin özelliklerini temsil eder. Dallar ise düğümlerin alabileceği değerleri temsil eder. En iyi sonucu vermek için karar düğümlerini en iyi özelliklerden seçerler.

Birçok farklı veri üzerinde işlem yapabilme özelliğine sahiptirler. Gürültülü ve eksik verileri tolere edebilirler. Kullanımı kolay olmasına karşı çok boyutlu verilerde kullanıma uygun değildirler. Random Forest, Gradient Boosting Machines ve GBM (Gradient Boosting Machines) tabanlı XGBoost, CatBoost gibi algoritmaların temelini oluştururlar.





Şekil 1.4. Karar Ağacı yapısı

Şekil 1.4.'de örnek bir karar ağacı yapısı sunuldu. Bu karar ağacı da kök, dall ve yapraklardan oluşur.

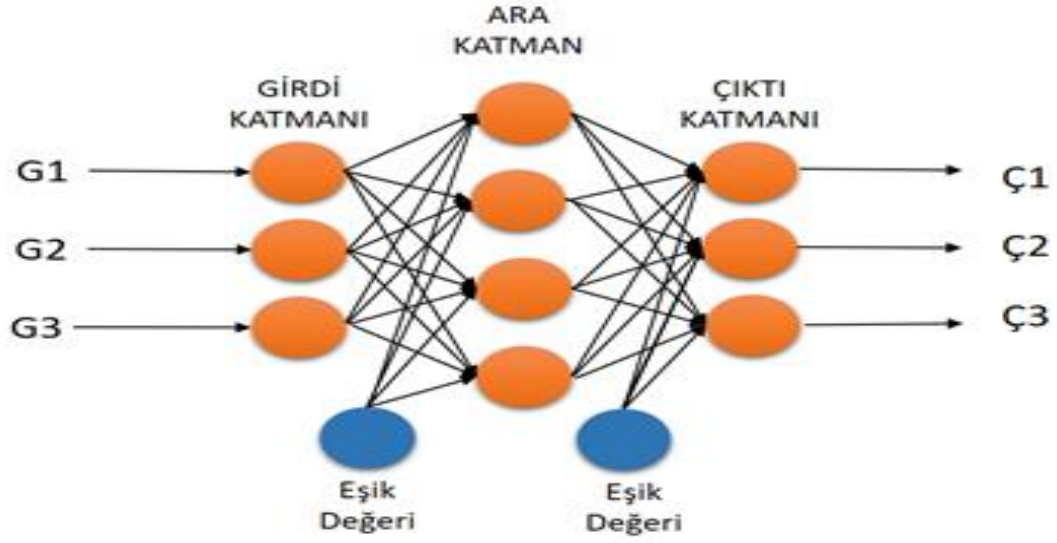
#### 1.4.4. Rasgele orman (Random forest)

Random Forest; Breiman [48] tarafından 2001 yılında önerilen bir makine öğrenmesi algoritmasıdır. Kullandığı veri setinden rastgele seçilen birden fazla alt kümenin oluşturduğu karar ağaçlarıyla sınıflandırma ve regresyon problemlerine göre özel olarak kullanılır [48]. Hangi şart ve durumda algoritmanın daha başarılı sonuç vereceğini söyler.

Çok hızlı çalışır ve ölçeklenebilir bir algoritmadır. Büyük veriler üzerinde de kullanılabilir. Kayıp ve eksik değerleri tolere eder, problem tanımına göre özelliklerin önem düzeyini verir. Karar ağacı sayısı arttıkça başarı da artar.

#### 1.4.5. Çok katmanlı algılayıcı

Çok katmanlı algılayıcı (Multi Layer Perceptrons, MLP); doğrusal olmayan bir problem olan XOR problemini çözmek için ortaya atılmıştır. Sınıflandırma için de kullanılan multi layer perceptrons, yapay sinir ağlarının bir sınıfıdır [49]. Temel amacı ağın beklenen çıktısı ile ürettiği çıktı arasındaki hatayı en aza indirmektir.



Şekil 1.5. Çok Katmanlı Algılayıcı Modeli [49]

Şekil 1.5.' de çok katmanlı algılayıcı modeli verildi. Çok katmanlı algılayıcı Şekil 1.5.'de görüldüğü gibi; girdi katmanı, ara katman ve çıktı katmanı olmak üzere üç katmandan oluşan bir ağıdır. Girdi katmanı ağa giren verileri ara katmana iletir. Ara katman ihtiyaca göre birçok gizli katmandan oluşabilir. Veriler ara katmanda işlenir ve daha sonra çıktı katmanına gönderilir. Çıktı katmanı ise sonuçları veren katmandır.

Lineer olmayan problemler için çok önemli olan bu sinir ağı, kolay uygulanabilmektedir fakat büyük boyutlu verilerde yapılan işlemler çok zaman almaktadır [49].

#### 1.4.6. Gradient boosting machines

Gradient Boosting Machines; 2001 yılında Jerome H. Friedman [50] tarafından önerilen denetimli makine öğrenmesi tekniğidir. Sınıflandırma ve regresyon için kullanılan bu algoritma; karar ağaçlarına benzer şekilde çalışır.

Gradient Boosting Machine algoritması yinelemeli olarak çalışır. Bu algoritmanın çalışma mantığı; zayıf tahmin fonksiyonlarını iterasyonlarla bir araya getirerek güçlü bir model oluşturup, kayıp fonksiyonunu en aza indirmektir [50].

GBM; bir optimizasyon algoritması olarak Gradient Descent Algoritmasını kullanarak kayıp fonksiyonunu en aza indirmeyi amaçlar.

#### **1.4.7. XGBoost**

XGBoost algoritması; 2016 yılında Tianqi Chen [51] tarafından önerilmiştir. Karar ağaçları temelli Gradient Boosting Machine algoritmasının geliştirilmiş ve daha yüksek performansa sahip versiyonudur. XGBoost algoritması diğer algoritmalarından 10 kat daha hızlı çalışır ve bunu ölçeklenebilir olması sayesinde yapar [51]. Sahip olduğu ölçeklenebilirlik, bazı algoritmik optimizasyonlardan kaynaklanmaktadır [51]. XGBoost paralel çalışabilmesi ve donanımsal optimizasyonları sayesinde de çok daha hızlı ve yüksek performanslı çalışır.

#### **1.4.8. LightGBM**

LightGBM; 2017 yılında Ke ve arkadaşları [52] tarafından önerilen sınıflandırma ve regresyon için kullanılan bir Gradient Boosting Machine algoritmasıdır [53]. Gradient Boosting Machines gibi karar ağaçlarına benzer şekilde çalışır. XGBoost modelinden farkı, eğitim sürecini hızlandırmak, bellek tüketimini azaltmak ve derinlik kısıtlamaları olan yaprak odaklı bir büyüme stratejisi uygulamak için histogram tabanlı algoritmalar kullanmasıdır [54]. Bu sayede daha hızlı sonuç verir ve yüksek performans sağlar. Algoritma yüksek hız ve performansa sahip olmasını sağlayan iki farklı tekniğe sahiptir. Bunlardan ilki tüm veri içinde en önemli verileri kullanmasını sağlayan Gradyan Tabanlı Tek Yönlü Örnekleme (Gradient-based One-Side Sampling); diğeri ise seyrek özellikleri birleştirip daha yoğun özellik oluşturan Özel Değişken Paketi (Exclusive Feature Bundling – EFB) dir [52].

#### **1.4.9. CatBoost**

CatBoost algoritması 2017 yılında Liudmila Prokhorenkova ve arkadaşları [55] tarafından önerilmiştir. Temel olarak Gradient Boosting Machine algoritması baz alınmıştır. CatBoost algoritması; karar ağaçlarını simetrik olarak oluşturur. Bu özelliği sayesinde ağaç derinliğini kontrol altında tutar. Kontrollü şekilde oluşturduğu ağaç yapısı daha hızlı ve yüksek sonuçlarla geri dönüş sağlar.

Diğer Gradient Boosting tabanlı algoritmalarından farklı olarak kategorik özellikleri direkt olarak kullanır. İsmi de bu özelliğinden; "Category" ve "Boosting" kelimelerinden almıştır [56].

## 1.5. Öznitelik Seçimi

Öznitelik seçimi, veri ile ilgilenen topluluklara uzun yıllar araştırma konusu olmuştur. Veri madenciliği üzerine yapılan araştırmaların da artmasıyla daha fazla ilgi gören bir alan olmuştur.

Çok boyutlu verilerde; sınıflandırma başarısına ve modelin öğrenme sürecine katkı sağlamayacak öznitelikler bulunabilir. Verideki alaka düzeyi düşük, gereksiz veya modeli yanıltabilecek özniteliklerin ayıklanması gerekir. Veri öznitelikleri içinde, sınıflandırma başarısı üzerinde önemli rol oynayan, değerli özniteliklerin seçilmesi işlemine öznitelik seçimi denir. Öznitelik seçimi için birçok yöntem vardır. Öznitelik seçiminde veriyi tanıyıp, yöntemi veriye göre seçmek başarıyı doğrudan etkileyeceği için çok önemli bir adımdır. Öznitelik seçim yöntemleri; filtreleme, sarmalayıcı ve gömülü yöntemler olmak üzere üçe ayrılır.

Filtreleme yöntemleri; öznitelikleri puanlama sistemiyle puanlar, düşük puana sahip olan öznitelikleri ve özniteliklerin birbiriyle olan alakalarını belirler. Ölçeklenebilir, kolay hesaplanabilir yöntemlerdir. Dolayısıyla maliyeti düşük ve uygulama açısından da kolaylardır. Bu çalışmada bir tek değişkenli filtre metodu olan Ki-Kare Testi (Chi Square Test), Karşılıklı Bilgi (Mutual Information), MRMR (Minimum Redundancy Maximum Relevance) ve çok değişkenli filtre metodu olan ReliefF kullanıldı.

Sarmalayıcı Yöntemler de filtre yöntemleri gibi çalışır. Filtre yöntemlerinden farklı olarak öznitelik seçimi için sınıflandırma algoritmalarını kullanır. Bu çalışma için kullanılan sarmalayıcı yöntem tabanlı öznitelik seçim algoritmaları ise Step Forward Selection ve Step Backward Selection algoritmaları olarak belirlendi.

Gömülü öznitelik seçim yöntemleri ise hem filtre hem sarmalayıcı yöntemlerin özelliklerine sahiptir

### 1.5.1. Chi square test (Ki-kare yöntemi)

Chi Square Test; tek değişkenli filtre yöntemidir. Chi-Square yöntemi, kategorik değişkenler üzerinde çalışır. Kategorik değişkenlerin ilişkilerini ve bağımlılıklarını tespit eder. Ki kare testi iki adımlı bir testtir. İlk adımda gözlenen değerlerin beklenen değerlere göre ki kare istatistiği hesaplanır. İkinci adımda ise elde edilen ki kare

istatistiđi belirlenen eşik değeriyle kıyaslanır ve buna göre karar verilir. Öznitelikler ki-kare istatistiđine göre puanlanır ve en iyi puana sahip olan öznitelikler kullanılır. Ki-kare istatistiđi ise Denklem (1.4), Denklem (1.5) ve Denklem (1.6) kullanılarak [57];

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \quad (1.4)$$

$$E_{ij} = N_i N_j / N \quad (1.5)$$

$$d = (I-1)(J-1) \quad (1.6)$$

elde edilir. Denklem (1.4)'de verilen I; aralıkların sayısıdır, J ise sınıf sayısıdır.  $N_{ij}$ ; i. aralık, j. sınıftaki örnek sayısıdır. Denklem (1.5)'de verilen  $E_{ij}$  iki özellik bağımsız iken; i. aralık j. sınıfta beklenen birim sayısıdır. Son olarak denklem (1.6)'da verilen d test istatistiđi için kullanılacak Ki-Kare dağılımının serbestlik derecesini göstermektedir [58].

### 1.5.2. Mutual information (Karşılıklı bilgi)

Mutual Information (Karşılıklı Bilgi) ile öznitelik seçme yaklaşımı ilk olarak Batiti [59] tarafından 1994 yılında ortaya atılmıştır. Karşılıklı bilgi; iki değişken arasında doğrusal ve doğrusal olmayan bağımlılıkları tespit eden bir ölçüttür [60]. Bu ölçüt; bir değişkenin diğer değişkenden elde edebileceđi bilgi miktarını da ölçer [61]. Entropi tabanlı çalışan bu yöntem için temel problem ikiden fazla özniteliđin birlikte karşılıklı bilgi değeriinin ölçülmesidir [62].

### 1.5.3. MRMR

MRMR algoritması Peng ve arkadaşları [63] tarafından 2005 yılında önerilen entropi tabanlı özellik seçim algoritmasıdır. MRMR algoritması sınıflandırma için kullanılacak olan veride bulunan sınıfların etiketleriyle en fazla ilişkili olan özellikleri seçerek çalışan bir filtreleme algoritmasıdır [64]. Bu algoritma iki öznitelik arasındaki veya özniteliklerle sınıf etiketleri arasındaki benzerlik oranını ölçmek için Mutual Information kullanır [64].

MRMR algoritması özünde tüm öznitelikleri en değerliden değersize doğru sıralamaya çalışır ve sınıflandırma problemi için kaç adet öznitelik kullanılması gerektiği kararını kullanıcıya bırakır. Bu nedenle MRMR algoritması bir öznitelik seçme algoritmasından ziyade öznitelik sıralama algoritması olarak değerlendirilmelidir [65].

#### **1.5.4. ReliefF**

Kira ve Rendell [66] tarafından 1992'de önerilen verimliliği yüksek bir öznitelik seçme algoritması olan Relief algoritması sadece iki sınıflı verileri sınıflandırabilmektedir. Bu özelliğiyle algoritma; çok sınıflı verileri sınıflandırmakta yetersiz kalmıştır. Bu nedenle Kononenko [67], 1994 yılında ReliefF algoritmasını önermiştir.

ReliefF, çok sınıflı verilerin sınıflandırılmasını sağlayabilir. Ayrıca, eksik ve gürültülü veri durumunun üstesinden de gelebilmektedir [68]. Öznitelikler arasındaki karşılıklı ilişkiyi ortaya çıkarır. Tüm özniteliklerin ağırlıklarını sıfır kabul ederek işleme başlar. Daha sonra veriden seçilen bir örneğin diğer örneklerle alakalarına bakar ve özniteliklerin ağırlıklarını günceller. Kurulan bu model için yeterlilik sağlamayan öznitelikleri veriden çıkarır. Bu şekilde öznitelik seçme işlemine son verir.

#### **1.5.5. Step forward selection**

SFS (Step Forward Selection) algoritması öznitelik seçme işlemine boş bir kümeyle başlar. Her adımda öznitelikleri teker teker ekleyerek ilerler. Bu işlem sınıflandırma kalitesindeki yükselme durana kadar devam eder. Fakat bu aşamaya kadar eklediği önemsiz öznitelikleri; seçtiği özniteliklerin arasından çıkarmayışı sıralı ileri seçim algoritmasının dezavantajı olarak bilinir. En iyi performansı az sayıda özneliğe sahip olan veriler üzerinde çalışınca elde eder [69].

#### **1.5.6. Step backward selection**

SBS (Step Backward Selection) algoritması; SFS algoritmasından farklı olarak öznitelik işlemini tersten yapar. En iyi öznitelik grubunu bulmak için verinin tüm özniteliklerini kullanarak başlar. Her adımda en kötü performansı gösteren özneliği gruptan çıkararak ilerler. Bu işlem de sınıflandırma başarısındaki yükseliş durana kadar devam eder. SFS algoritmasındaki gibi, önceden çıkarmış olduğu öznitelikleri tekrar öznitelik grubuna ekleyemez [70].

## 1.6. Sınıflandırma Performans Metrikleri

Kullanılan sınıflandırma algoritmalarının performanslarını değerlendirebilmek amacıyla bazı performans metrikleri kullanıldı.

### 1.6.1. Karmaşıklık matrisi

Karmaşıklık matrisi; yapılan sınıflandırmanın performansını görsel olarak sunan matristir [71]. Bu matris; sınıflandırmanın sonuçlarını, Doğru Pozitif (True Positive, TP), Doğru Negatif (True Negative, TN), Yanlış Pozitif (False Positive, FP) ve Yanlış negatif (False Negative, FN) değerleriyle oluşturur.

	Gerçek Sınıf Değeri (True Value)		
		Pozitif	Negatif
Modelin Tahmin Ettiği Sınıf Değeri (Predicted Value)	Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
	Negatif	Yanlış Negatif (FN)	Doğru Negatif (TN)

Şekil 1.6. Karmaşıklık matrisi

Şekil 1.6.' da bir karmaşıklık matrisi verildi. Burada doğru pozitif değeri; gerçek değeri pozitif ve tahmin ettiğimiz değeri de pozitif olan örneklerdir. Doğru negatif değerlerini ise; gerçek değeri negatif olup, negatif olarak tahmin ettiğimiz örnekler oluşturur. Yanlış pozitif değerler; pozitif tahmin ettiğimiz fakat gerçekte negatif olan değerlerdir. Yanlış negatif değerler ise; gerçek değeri pozitif olduğu halde, negatif olarak tahmin ettiğimiz değerlerdir.

### 1.6.2. Doğruluk

Doğruluk; sınıflandırıcının genel performansını gösteren metriktir. Eğitim verisi kullanılarak eğitilen modelin test verisi üzerindeki performansını gösterir. Denklem (1.7) ' de;

$$\text{Doğruluk} = \frac{TP+TN+FP+FN}{TP+TN} \quad (1.7)$$

doğruluk değeri formülize edilmiştir.

İkili sınıflandırma için Denklem (1.7) kullanılır. Çok sınıflı sınıflandırma yapılırken, her sınıf için doğruluk değeri hesaplandıktan sonra ortalaması alınır.

### 1.6.3. Recall (Sensitivity)

Recall değeri etiketli veri için pozitif olan değerleri doğru olarak pozitif tahmin yapma yetkinliğini gösteren değerdir. Tıp alanında sensitivity olarak kullanılan recall değeri Denklem (1.8)'de, çok sınıflı sınıflandırma için hesaplanan makro recall değeri de Denklem (1.9)'da;

$$\text{Recall} = \frac{TP}{TP+FN} \quad (1.8)$$

$$\text{Recall}_M = \frac{\sum_{i=1}^l \frac{TP_i}{TP_i+FN_i}}{l} \quad (1.9)$$

verildi. Makro recall değeri; sınıf etiketlerini belirlemek için bir sınıflandırıcının sınıf başına ortalama etkinliğini gösterir. Denklem (1.9)'da verilen l; sınıf sayısıdır.

### 1.6.4. Specificity (Özgüllük)

Specificity; etiketli veri içinde negatif olarak etiketlenmiş veriyi doğru tespit etme yetkinliğini gösteren değerdir.

Specificity değeri Denklem (1.10)'da;

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (1.10)$$

verildiği gibidir.

### 1.6.5. Eğri altında kalan alan (AUC)

ROC (Receiver Operating Characteristic) eğrisi sınıflandırma performansını görsel olarak sunan önemli bir performans metriğidir. ROC eğrisinin altında kalan alan ise AUC (Area Under the Curve-Eğri altında kalan alan) değeri olarak belirlenir. ROC eğrisi; y eksenine doğru pozitif oranını (sensitivity, recall); x eksenine ise yanlış pozitif oranını (1-specificity) alır. Burada AUC değerinin büyük olması istenilen durumdur. AUC değeri ne kadar büyük olursa; sınıflandırma için kurulan modelin sınıflandırmada ve veri üzerinde kullanımındaki yetkinliği artıyor demektir



### 1.6.6. Precision (Kesinlik)

Precision değeri; doğru olarak sınıflandırılmış pozitif örneklerin pozitif olarak sınıflandırılmış toplam örneklerin sayısına oranıdır. Diğer bir deyişle pozitif olarak tahmin edilen verinin gerçekten pozitif değere sahip olma ölçütüdür. Precision değeri Denklem (1.11)'de;

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1.11)$$

verildi. Çok sınıflı sınıflandırmada ise makro precision değeri kullanılmıştır. Bu değer; verideki tüm sınıflar için hesaplanan precision değerinin ortalamasıdır. Makro precision değeri Denklem (1.12)'de;

$$\text{Precision}_M = \frac{\sum_{i=1}^l \frac{TP_i}{TP_i+FP_i}}{l} \quad (1.12)$$

verildi ve Denklem (1.12)'de l; sınıf sayısıdır.

### 1.6.7. F-Score

Precision ve recall değerlerinin harmonik ortalamasını alarak hesaplama yapan metriktir. Dengesiz veriler üzerinde oldukça faydalıdır. F-score ne kadar büyükse o kadar doğru tahminleme yapılıyor denebilir. İkili sınıflandırmada kullanılan F-score değeri Denklem (1.13)'de;

$$\text{F-Score} = \frac{2TP}{2TP+FP+FN} \quad (1.13)$$

verildiği gibidir. Çok sınıf sınıflandırmada sistem performansını değerlendirmek için kullanılan makro F-Score ise Denklem (1.14)'de;

$$\text{F-Score}_M = \frac{2\text{Precision}_M\text{Recall}_M}{\text{Precision}_M+\text{Recall}_M} \quad (1.14)$$

olarak verilmiştir.

### 1.6.8. Matthews correlation coefficient

Matthew Correlation Coefficient(MCC) ise ikili ve çoklu sınıflandırma için kullanılan bir ölçüttür. MCC; gözlemlenen(observation) ve tahmin edilen değer arasındaki

korelasyonu belirten bir sınıflandırma metriğidir. MCC değeri; -1 ile +1 arasında değer alır. Sınıflandırıcı +1'e yakın değerler almışsa doğru tahminlerde bulunduğunu, 0 değerini almışsa rastgele tahminlerde bulunduğunu ve -1'e yakın değerler almışsa yanlış tahminlerde bulunduğunu göstermektedir [72, 73]. MCC metriği Denklem (1.15)'de;

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1.15)$$

verildi.



## **2. MATERYAL VE YÖNTEM**

### **2.1. Veri Seti**

Bu çalışmada kullanılan 135 adet hastanın klinik, histopatolojik ve görüntüleme verileri halka açık veri tabanları olan TCGA-UCEC (The Cancer Genome Atlas-Uterine Corpus Endometrial Carcinoma) [74, 75, 76] ve CPTAC-UCEC (Ulusal Kanser Enstitüsü Klinik Proteomik Tümör Analizi Konsorsiyumu)'dan [77] alınmıştır. Bu yayında kullanılan veriler, Ulusal Kanser Enstitüsü Klinik Proteomik Tümör Analizi Konsorsiyumu (CPTAC) tarafından oluşturulmuştur. Burada gösterilen sonuçlar tamamen TCGA Araştırma Ağı [78] tarafından oluşturulan verilere dayanmaktadır.

Bu çalışmada kullanılan veriler; kontrastlı bilgisayarlı tomografiye sahip olan hasta görüntüleri, kesit kalınlığı 5 mm'den az olan görüntüler ve piksel aralığı 1 mm'den az olan görüntüler kullanılarak elde edilmiştir. Önemli yapay nesnelere (metal, halka vs.), görüntü paraziti içeren veri kümeleri, tip I olmayan ve tip II olmayan endometriyal kanserli hastalara ait görüntüler çalışmaya dahil edilmemiştir. Verilerde homojen dağılımın sağlanması için; çok nadir görülen bir alt-tip olan Seröz kitle türüne ait veriler farklı veri tabanlarından, TCGA-UCEC (The Cancer Genome Atlas-Uterine Corpus Endometrial Carcinoma) [74, 75, 76] ve CPTAC-UCEC (Ulusal Kanser Enstitüsü Klinik Proteomik Tümör Analizi Konsorsiyumu)'dan [77] alınmıştır.

Kullanılan veri setinde; 45 adet Myom, 44 adet Seröz ve 46 adet Endometrioid kitle türüne sahip hasta görüntüleri bulunmaktadır.

### **2.2. Hasta Seçimi**

Çalışmada patolojik olarak kanıtlanmış endometriyal kanserli 90 hastanın ve iyi huylu kitle türü olan Myom kitlesine sahip 45 hastanın klinik görüntüleri kullanıldı. Toplamda 46 hastada Endometrioid ve 44 hastada Endometriyal kanserin Seröz alt tipleri vardı.

Hastaların ortalama yaşı  $54.6 \pm 15.3$  idi. Kullanılan hasta verilerinde; Endometrioid alt tip kitleli hastaların ortalama yaşı, Seröz alt tip kitleli hastalara göre çok az farkla daha düşüktür. ( $54,2 \pm 12,6$  yıl ve  $57,9 \pm 15,5$  yıl;  $p = 0,017$ ). Kitle boyutu da Endometrioid alt tip kitlesi olan katılımcılarda benign kitlelere göre anlamlı olarak daha büyüktür ( $58.41 \pm 33.01$  mm'ye karşı  $36.04 \pm 13.96$  mm;  $p = 0.012$ ).

### **2.3. BT Parametreleri**

BT edinim parametreleri şunlardır: dilim kalınlığı, 3 mm; tüp voltajı, 100–130 kV; tüp akımı, 85–575 mA; ve piksel boyutu,  $0,582 \times 0,582$  mm<sup>2</sup> ila  $0,906 \times 0,906$  mm<sup>2</sup>. Arteriyel faz görüntüleri vücut ağırlıklarına göre (2 mL / kg; maksimum = 150 mL) mekanik güç enjektörleri ile noniyonik intravenöz kontrast madde ile 25 saniye gecikmeyle elde edilmiştir.

CPTAC-UCEC, TCGA-UCEC veritabanları çeşitli veri kaynaklarını [75, 78] içerir. Bu nedenle, tamamen tek tip bir görüntüleme protokolüne ulaşmak imkansızdır. Tüm görüntüler, protokol heterojenliğinin olası olumsuz sonuçlarını en aza indirmek için normalizasyon, piksel yeniden örnekleme, yeniden ölçekleme ve gri seviye ayrıklaştırma gibi ön işlemlerden geçirilmiştir.

### **2.4. Radiomics Verilerinin Çıkarılması**

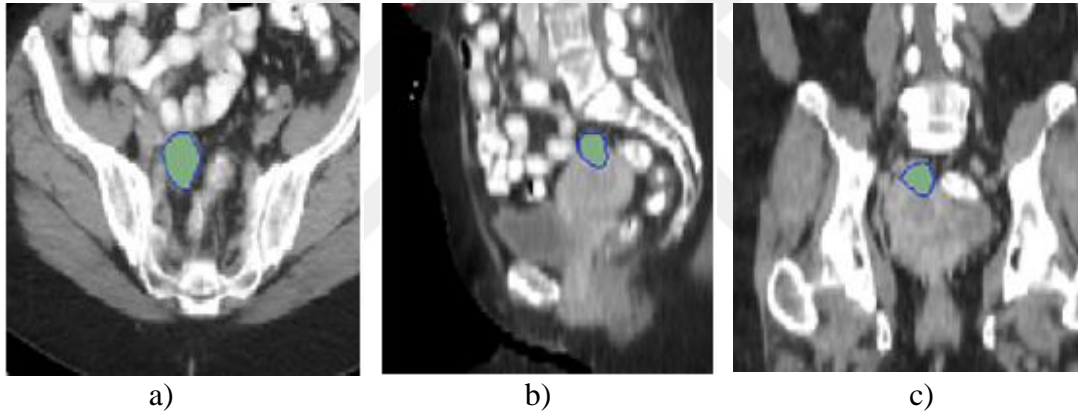
Veri setleri için aynı kesit kalınlığı 2 mm olan görüntüleri analiz etmek için DICOM (Digital Imaging and Communications in Medicine) formatı kullanılmıştır. Doku özelliğinin çıkarılmasından önce, tüm görüntülere bazı görüntü işleme adımları uygulanmıştır. Bu görüntü işleme adımları; gri düzey normalleştirme ( $\pm 3$  sigma tekniği), piksel yeniden örnekleme ve kübik B-spline enterpolasyonu kullanarak yeniden ölçekleme (elde edilen piksel boyutu,  $1 \times 1$  mm<sup>2</sup>) ve gri-seviye ayrıklaştırma (bölme genişliği, 0.3) olarak 4 adımdan oluşmaktadır.

Tümörler, açık kaynaklı yazılım olan 3D Slicer yazılımının 4.10.2. sürümü [80] kullanılarak geliştirilmiş BT görüntülerinden manuel olarak segmentlere ayrılmıştır. Segmentasyon işlemi Kocaeli Üniversitesi Radyoloji Bölümü Öğretim Üyesi Dr. Özgür ÇAKIR tarafından yapılmıştır.

Tümörlerin en büyük kesit alanı, Şekil 2.1.'de verildiği gibi tümör segmentasyonu için seçilmiştir. Segmentasyon gerçekleştirilirken, tüm durumlarda, mevcut kontrastlı BT veya kontrastlı MRI çalışmaları da tümör sınırlarını belirlemek için kullanılmıştır.

Radiomics doku öznitelikleri; PyRadiomics yazılımı (sürüm, 3.9.0.) [81] kullanılarak orijinal, filtrelenmiş ve dalgacıkla dönüştürülmüş görüntülerden çıkarıldı. Sırasıyla ince, orta ve kaba görüntüleri temsil eden 2-mm, 4-mm ve 6-mm değerleriyle görüntü filtrasyonu için LoG filtresi kullanılmıştır.

Görünür lezyon konturunu dikkate alarak ROI'ler dikkatlice çizilmiştir ve daha sonra çevre dokulardan kısmi hacim etkisini azaltmak için lezyon konturundan 1-2 mm'ye kadar küçültülmüştür. Bir BT görüntüsünün üç farklı kesit alanında yapılan segmentasyon işlemi Şekil 2.1.'de verildi.



Şekil 2.1. Üç kesitli BT görüntüsü üzerinde segmentasyon işlemi, a) Axial Plan, b) Sagittal Plan, c)Koronal Plan

Çıkarılan doku öznitelikleri; histogram öznitelikleri, gradyan özellikleri, gri düzey birlikte oluşma matrisi öznitelikleri, sayı matrisi öznitelikleri ve otopregresif model öznitelikleridir. Çıkarılan özniteliklerin toplam sayısı, her aşama için hasta başına 130'dur.

## 2.5. Veri Ön İşleme

Elde edilen radiomics verileri kullanılmadan önce bazı ön işleme adımlarından geçti. Öncelikle; veri içinde anlamsız veya gürültü sayılabilecek olan değerlerin varlığının kontrolü yapıldı. Veriler içerisinde eksik veya yanlış veri olmadığı için bu adımda herhangi bir düzeltme yapılmadı. Bu adımdan sonra veri entegrasyonu adımına

geçildi. Farklı dosyalarda olan veriler birleştirilerek ortak bir veri tabanı oluşturuldu. Aynı kitle grupları etiketlendi. Bu etiketlenen veriler daha sonra sınıflandırma aşamasında hedef veri olarak kullanıldı.

Her hasta için 130 adet radiomics verisi özniteliği bulunmaktadır. 130 adet özniteliğin içerisinde 22 adet öznitelik; kullanılan cihaz, yazılım ve sürüm bilgisi içerdiği için özniteliklerin arasından çıkarıldı. Böylelikle elde edilen öznitelik sayısı hasta başına 108 adet oldu. Son olarak verilere; minimum-maximum normalizasyonu yapıldı. Bu normalizasyonda verilerin içindeki en büyük ve en küçük değerleri alınarak diğer verilerin bu değerlere göre normalleştirme işlemi yapılır.

## 2.6. Öznitelik Seçimi

Veri setinde bulunan 108 adet özniteliğin içerisinde en yararlı olan özniteliklerin bulunması için öznitelik seçimine gidildi. 108 adet öznitelikle yapılan sınıflandırma başarısına, daha az öznitelik kullanarak ulaşılmaya çalışıldı. Öznitelik seçimleri Chi-Square Test, Mutual Information, ReliefF, MRMR, Step Forward Selection ve Step Backward Selection kullanarak yapıldı. Öznitelik seçim işlemleri Python programlama dili kullanılarak gerçekleştirildi.

Chi-Square Test kullanarak seçilen öznitelik sayısı optimum 20 olarak belirlendi. Chi-Square Test kullanarak öznitelikler puanlandı. Puanlanan öznitelikler içinde en yüksek puana sahip 20 öznitelik seçildi. Daha az sayıda öznitelikle de sınıflandırma yapılmıştır fakat ikili sınıflandırmada en iyi sonuçlara ulaşılabilen öznitelik sayısı Chi-Square Test için 20 oldu. Çoklu sınıflandırma yapılırken de 20 adet öznitelik seçmiş ve sınıflandırma yapıldı. Daha sonra sınıflandırma işlemi 10 adet öznitelikle tekrarlandı. Sınıflandırma sonuçlarında çok bariz bir fark olmadığından öznitelik sayısı 10 olarak alındı. Chi-Square Test kullanarak öznitelik seçme işlemi Python dilinde, matematiksel ve istatistiksel işlemler için kullanılan Scipy modülü ile yapıldı.

Mutual Information öznitelik seçimi ise özellikle veri ön işleme ve makine öğrenmesi alanında çok faydalı olan Sklearn kütüphanesinin öznitelik seçme modülü kullanılarak yapıldı. Mutual Information ile yüksek performanslı özniteliklerin %10' u alınarak, 11 adet öznitelik seçildi.

MRRM algoritması ise [64] tarafından geliştirilen Pymrr yazılımı kullanılarak yapıldı. Öznitelik sayısı (n); n=3 ve n=5 seçilerek elde edilen sonuçlar karşılaştırıldı.

ReliefF algoritmasıyla öznitelik seçimi için Randy Olson [82] tarafından geliştirilen ReliefF modülü kullanılarak gerçekleştirildi. Öznitelik sayısı sabit tutuldu ve 5 olarak belirlendi. Bu algoritmayla iki farklı öznitelik seçimi gerçekleştirildi. Algoritmanın parametrelerinden biri olan en yakın komşu sayısı (K); öncelikle K=4 olarak seçildi. Daha sonra K=10 olarak seçildi. İki farklı en yakın komşu sayısı ile aynı sayıda öznitelik kullanarak sınıflandırma yapıldı ve sonuçlar karşılaştırıldı.

Step Forward Selection ve Step Backward Selection algoritmaları için Sebastian Raschka tarafından geliştirilen MLXTEND modülü [83] kullanılarak yapıldı. Bu iki algoritma için her sınıflandırıcı kullanacağı öznitelikleri kendine özel olarak seçti.

## **2.7. Makine Öğrenmesinin Uygulanması**

Radiomics verileriyle kitle tespiti; sınıflandırma algoritmaları kullanılarak yapıldı. Bu algoritmalar, KNN, SVM, Decision Tree, Random Forest, MLP, GBM, XGBoost, LightGBM ve CatBoost'tur. Bu algoritmalar kullanılarak Endometrioid-Seröz ikili sınıflandırması, Myom-NonMyom ikili sınıflandırması ve Endometrioid-Seröz-Myom sınıflandırması yapıldı. Her sınıflandırma için önce verideki tüm öznitelikler kullanıldı. Daha sonra öznitelik seçim yöntemleriyle seçilen öznitelikler kullanılarak sınıflandırma yapıldı. Tüm sınıflandırmalar için bu işlemler tekrarlandı.

Sınıflandırma algoritmalarının alacağı parametreleri belirlenen başarı metriğine göre en iyi şekilde seçilme işlemine hiperparametre optimizasyonu denir Sınıflandırıcı olarak kullanılan algoritmalara hiperparametre optimizasyonu GridSearch Cross Validation kullanılarak uygulandı. Bunun için Phyton programlama dili ve Sklearn kütüphanesi kullanıldı. Sınıflandırma modelinde kullanılmak istenen parametreler modele verilmiş ve verilen parametrelerin hepsi teker teker tüm kombinasyonlar denenerek en iyi sonucu verecek hiperparametreler alındı. Böylelikle modellerde meydana gelebilecek aşırı öğrenme ve öğrenememe gibi problemlerin önüne geçilmek istendi. Ayrıca hiperparametre optimizasyonu ile kurulan modeller kullanılan veri setine göre özelleştirildi.

Sınıflandırma işlemleri Python programlama dili kullanılarak gerçekleştirildi. Parametre optimizasyonu ile birlikte yapılan çok aşamalı sınıflandırmalar işlemlerin çok uzun sürmesinden dolayı yüksek GPU ve CPU'ya sahip bilgisayarlar kullanılarak yapıldı.





### 3. BULGULAR VE TARTIŞMA

Bu kısımda yapılan sınıflandırma sonuçları verildi. Üç farklı sınıflandırma yapıldı. Bunların ilki iki kötü huylu kitle türü olan endometrioid ve seröz sınıflandırmasıdır. Diğeri; sahip olunan kitlenin iyi huylu mu kötü huylu mu olduğunu belirten Myom-NonMyom sınıflandırmasıdır. Son olarak bu üç farklı kitlenin çoklu sınıflandırma yapılarak birbirlerinden ayrılmaları ele alındı.

#### 3.1. Endometrioid- Seröz Alt-Tip İkili Sınıflandırma Sonuçları

Endometrioid- Seröz sınıflandırması; elimizde bulunan 90 adet hastaya ait radiomics verileri kullanılarak yapıldı. Bu veriler ön işleme adımlarına tabi tutulduktan sonra etiketlendi. Kitleleri endometrioid grubuna dahil olan hastalara ait veriler 1; seröz grubuna dahil olan hastalara ait veriler 0 olarak etiketlendi. Etiketleme işlemi sonucunda 46 adet endometrioid, 44 adet seröz grubuna ait hastaya sahip olduğu gözlemlendi. Verilerin %30'u test verisi, %70'i ise eğitim verisi olarak ayrıldı. Bu verilerle; farklı makine öğrenmesi algoritmaları kullanılarak hastanın sahip olduğu kitlenin hangi kötü huylu kitle türüne ait olduğu tespit edilmek istendi. Aşağıda verilen Tablo 3.1.'de, verinin 108 adet özniteliği kullanılarak makine öğrenmesi algoritmalarından elde edilen sonuçlar verildi. Tüm tablolarda sonuçlar yüzde cinsinden verildi.

Tablo 3.1. Tüm öznitelikler kullanılarak yapılan Endometrioid-Seröz sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
KNN	0,85	1,0	0,69	0,82	1,0	0,85	0,73	9	0	4	14
SVM	0,85	1,0	0,69	0,82	1,0	0,85	0,73	9	0	4	14
DecisionTree	0,81	0,83	0,77	0,80	0,85	0,81	0,62	10	2	3	12
RandomForest	0,62	0,64	0,54	0,58	0,71	0,63	0,25	7	4	6	10
MLP	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
GBM	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
XGBOOST	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
LightGBM	0,77	0,73	0,85	0,79	0,71	0,78	0,56	11	4	2	10
CatBoost	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12

Doğruluk ve Auc değerleri en yüksek algoritmalar %92 doğruluk ve %92 Auc değeriyle MLP ve % 85 doğruluk ve %85 Auc değeriyle KNN, SVM, XGBoost ve CatBoost olarak belirlendi. KNN, SVM ve MLP %100 specificity değerine sahip olmakla; seröz grubuna ait hastaların hepsini doğru tahmin edebileceğini göstermektedir. Ayrıca %100 precision değeriyle de endometrioid grubundaki hastaların gerçekten endometrioid grubuna ait olma durumunu gösterir. SVM ve KNN algoritmalarının Recall değerlerinin %69 olması, endometrioid grubuna ait olan hastaların %69 oranla doğru tespit edildiğini gösterir. Dolayısıyla bu sınıflandırma sonuçlarına sadece precision ve recall değerleriyle bakılması yeterli olamayacağı için F-Score değerleri ele alınır. MLP'nin %92 F-Score değerleriyle KNN, SVM, XGBoost ve CatBoost'a göre daha dengeli ve doğru şekilde sınıflandırma yaptıklarını gösterdi. MLP, XGBoost, Catboost algoritmaları 13 adet endometrioid grubundaki hastanın 11 tanesini doğru tahmin ederek KNN ve SVM algoritmalarına göre endometrioid grubunu daha iyi tahminlemede daha başarılı oldu. Fakat %85 specificity değeriyle seröz grubuna ait hastaları KNN, SVM ve MLP algoritmalarına göre daha iyi tespit edemediğini gösterdi. Bu tablo için en kötü sonucu Random Forest algoritması verdi. Özellikle MCC değerinin %25 olması, algoritmanın rastgele tahminlerde bulunduğunu göstermektedir.

Tüm özellikleri kullanarak yapılan sınıflandırma sonrası öznelik seçim işlemlerine gidildi. Yapılan öznelik seçim yöntemleri ve elde edilen öznelikler Tablo 3.2'de verildi.

Tablo 3.2. Endometrioid- Seröz Sınıflandırması İçin Kullanılan Yöntemler ve Elde Edilen Öznelikler

Öznelik Seçim Yöntemleri	Seçilen Öznelikler
<b>Chi-Square Test</b>	gldm_LargeDependenceHighGrayLevelEmphasis glszm_SizeZoneNonUniformity glszm_GrayLevelNonUniformity shape_SurfaceArea shape_VoxelVolume shape_MeshVolume gldm_DependenceNonUniformity gldm_DependenceNonUniformityNormalized gldm_GrayLevelNonUniformity firstorder_Energy firstorder_TotalEnergy glszm_LargeAreaLowGrayLevelEmphasis glszm_ZoneVariance glszm_LargeAreaEmphasis gllrm_LongRunLowGrayLevelEmphasis ngtdm_Busyness gllrm_LongRunEmphasis gllrm_RunVariance gldm_LargeDependenceLowGrayLevelEmphasis glcm_ClusterProminence

Tablo 3.2. (Devam) Endometrioid- Seröz Sınıflandırması İçin Kullanılan Yöntemler ve Elde Edilen Öznitelikler

Öznitelik Seçim Yöntemleri	Seçilen Öznitelikler
<b>Mutual Information</b>	gldm_LargeDependenceHighGrayLevelEmphasis glszm_SmallAreaEmphasis gldm_SmallDependenceLowGrayLevelEmphasis glcm_Contrast glszm_LowGrayLevelZoneEmphasis glcm_DifferenceEntropy glcm_Imc2 glcm_DifferenceAverage glszm_LargeAreaLowGrayLevelEmphasis glrlm_LongRunEmphasis glszm_SizeZoneNonUniformityNormalized
<b>ReliefF; K=4</b>	shape_SurfaceVolumeRatio gldm_LargeDependenceHighGrayLevelEmphasis ngtdm_Coarseness glszm_ZoneEntropy ngtdm_Contrast
<b>ReliefF; K=10</b>	gldm_LargeDependenceHighGrayLevelEmphasis shape_SurfaceVolumeRatio glcm_Idm ngtdm_Contrast glcm_DifferenceVariance
<b>MRMR; n=3</b>	shape_MeshVolume glcm_ClusterProminence firstorder_RootMeanSquared
<b>MRMR; n=5</b>	shape_MeshVolume glcm_ClusterProminence firstorder_RootMeanSquared firstorder_TotalEnergy firstorder_Energy

Bu aşamada kullanılan ilk yöntem Chi-Square yöntemidir. Aşağıda verilen Tablo 3.3’de; Chi Square Test ile seçilen 20 adet öznitelikle elde edilen sonuçlar verildi.

Tablo 3.3. Chi-Square Test ile Öznitelik Seçimi Sonrası Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>SVM</b>	0,77	0,73	0,85	0,79	0,71	0,78	0,56	11	4	2	10
<b>DecisionTree</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>RandomForest</b>	0,92	0,87	1,0	0,93	0,85	0,93	0,86	11	2	2	12
<b>MLP</b>	0,85	0,80	0,92	0,86	0,78	0,85	0,71	12	3	1	11
<b>GBM</b>	0,85	0,85	0,85	0,85	0,85	0,85	0,70	9	4	4	10
<b>XGBOOST</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>LightGBM</b>	0,85	0,85	0,85	0,85	0,85	0,78	0,70	11	2	2	12
<b>CatBoost</b>	0,85	0,76	1,0	0,87	0,71	0,86	0,73	13	4	0	10

Chi Square Test ile seçilen 20 özelliğin kullanılmasıyla en iyi sonuçları veren algoritmalar Decision Tree, Random Forest ve XGBoost oldu. Bu algoritmalar %92 doğruluk ve %86 MCC değerleriyle aynı sonuçları verdikleri görüldü. Random Forest %93 F-score değeriyle diğer algoritmalara göre daha iyi performans gösterirken, Decision Tree ve XGBoost algoritmaları seröz grubuna ait olan hastaları %100 doğru tahmin etti. MLP; endometrioid grubunun en iyi tahmin eden algoritma olsa da; seröz grubunda aynı başarıyı sağlayamadı. En kötü sonuçları ise KNN algoritması verdi.

Tablo 3.4. Mutual Information ile seçilen özniteliklerle elde edilen Endometrioid-Seröz sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,77	1,0	0,54	0,70	1,0	0,77	0,61	7	0	6	14
<b>SVM</b>	0,77	1,0	0,54	0,70	1,0	0,77	0,61	7	0	6	14
<b>DecisionTree</b>	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>RandomForest</b>	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>MLP</b>	0,77	0,73	0,85	0,79	0,78	0,78	0,56	11	4	2	10
<b>GBM</b>	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>XGBOOST</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>LightGBM</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>CatBoost</b>	0,77	0,73	0,85	0,79	0,71	0,78	0,56	11	4	2	10

Yukarıda verilen Tablo 3.4.'de ise Mutual Information ile seçilen 11 özellekle elde edilen sonuçlar verildi. Tabloya göre en iyi sonuçları XGBoost ve LightGBM algoritmaları verdi. Bu iki algoritma; %100 specificity değeriyle seröz grubunda olan hastaların tamamını doğru tahmin etti. Ayrıca yüksek MCC değerleriyle de yapılan tahminlemelerinin doğru olduğu söylenebilir. Yüksek F-Score değeriyle de sınıflandırma performanslarının iyi olduğu gözlemlenebilir. Bu sınıflandırma sonucunda Chi Square ile elde edilen başarıya, Mutual Information kullanarak 11 özellekle ulaşılabildiği söylenebilir.

Tablo 3.5. ReliefF ile Seçilen Özniteliklerle, K=4 ve K=10 için Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

K=4 için											
Algoritmalar	Doğruluk	Precision	Recall	Fscore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>SVM</b>	0,70	0,78	0,54	0,64	0,85	0,70	0,41	7	2	6	12
<b>DecisionTree</b>	0,70	0,73	0,62	0,67	0,78	0,70	0,40	8	3	5	11
<b>RandomForest</b>	0,70	0,73	0,62	0,67	0,78	0,70	0,40	8	3	5	11

Tablo 3.5. (Devam) ReliefF ile Seçilen Özniteliklerle, K=4 ve K=10 İçin Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	Fscore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>MLP</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>GBM</b>	0,85	0,82	0,69	0,75	0,85	0,77	0,55	9	2	4	12
<b>XGBOOST</b>	0,62	0,64	0,54	0,58	0,71	0,63	0,25	7	4	6	10
<b>LightGBM</b>	0,85	1,0	0,69	0,82	1,0	0,85	0,73	9	0	4	14
<b>CatBoost</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>K=10 için</b>											
<b>KNN</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>SVM</b>	0,85	1,0	0,69	0,82	1,0	0,85	0,73	9	0	4	14
<b>DecisionTree</b>	0,92	1,0	0,85	0,92	1,0	0,92	0,86	11	0	2	14
<b>RandomForest</b>	0,81	1,0	0,62	0,76	1,0	0,81	0,67	8	0	5	14
<b>MLP</b>	0,62	0,57	1,0	0,72	0,28	0,64	0,40	13	10	0	4
<b>GBM</b>	0,85	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>XGBOOST</b>	0,77	0,82	0,69	0,75	0,71	0,77	0,55	9	2	4	12
<b>LightGBM</b>	0,77	0,82	0,69	0,75	0,85	0,77	0,55	9	2	4	12
<b>CatBoost</b>	0,77	0,82	0,69	0,75	0,85	0,77	0,55	9	2	4	12

Tablo 3.5. ise ReliefF özellik seçim algoritmasıyla seçilen özellikler ve bu özelliklerle elde edilen sonuçları göstermektedir. Bu algoritma için K; en yakın komşu sayısı önce 4, daha sonra 10 olarak seçildi. K=4 ve K=10 için özellik sayısı sabit tutulup 5 olarak belirlendi. K'nın değişmesiyle 3 ortak özellik seçildi. Seçilen özniteliklerle elde edilen sonuçlar karşılaştırılınca; Decision tree, random forest ve GBM algoritmalarının doğruluk ve sınıflandırma metriklerinde artış gözlemlendi. K=4 seçildiğinde en iyi sonuçları veren algoritma KNN oldu. K=10 seçildiğinde; KNN başarısını devam ettirirken, Decision Tree algoritması da KNN ile aynı sonuçları verdi. Böylelikle Mutual Information ile 11 adet öznitelik kullanılarak elde edilen sonuçlara ReliefF ile sadece 5 adet öznitelik kullanarak ulaşıldı.

Tablo 3.6. MRMR ile Seçilen Özniteliklerle Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

<b>N=3 için</b>											
Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,77	0,82	0,69	0,75	0,85	0,77	0,55	9	2	4	12
<b>SVM</b>	0,55	0,52	1,0	0,68	0,14	0,57	0,27	13	12	0	2
<b>DecisionTree</b>	0,62	0,64	0,54	0,58	0,71	0,63	0,25	7	4	6	10
<b>RandomForest</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>MLP</b>	0,48	0,47	0,54	0,50	0,42	0,48	-0,3	7	8	6	6
<b>GBM</b>	0,62	0,64	0,54	0,58	0,85	0,63	0,25	7	4	6	10

Tablo 3.6. (Devam) MRMR ile Seçilen Özniteliklerle Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>XGBOOST</b>	0,66	0,64	0,69	0,67	0,64	0,67	0,33	9	5	4	9
<b>LightGBM</b>	0,62	0,64	0,54	0,58	0,85	0,63	0,25	7	4	6	10
<b>CatBoost</b>	0,77	0,73	0,85	0,79	0,71	0,78	0,56	11	4	2	10
<b>N=5 için</b>											
<b>KNN</b>	0,77	0,82	0,69	0,75	0,85	0,77	0,55	9	2	4	12
<b>SVM</b>	0,55	0,52	1,0	0,68	0,14	0,57	0,27	13	12	0	2
<b>DecisionTree</b>	0,74	0,71	0,76	0,73	0,71	0,78	0,48	11	4	2	10
<b>RandomForest</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>MLP</b>	0,48	0,46	0,46	0,46	0,50	0,48	-0,3	6	7	7	7
<b>GBM</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>XGBOOST</b>	0,59	0,57	0,62	0,59	0,57	0,59	0,18	8	6	5	8
<b>LightGBM</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10
<b>CatBoost</b>	0,70	0,69	0,69	0,69	0,71	0,70	0,40	9	4	4	10

Tablo 3.6.'de ise; MRMR özellik seçim yönteminden elde edilen özelliklerle yapılan sınıflandırma sonuçları yer almaktadır. Özellik sayısı(n); n=3 seçildiğinde; %77 doğrulukla KNN ve CatBoost algoritmaları en iyi sonuçları veren algoritmalar oldu. KNN; %85 specificity değeriyle Seröz grubuna ait hastalara; CatBoost algoritmasından daha iyi teşhis koyabileceğini göstermektedir. CatBoost algoritması ise sınıflandırma performansı ile KNN'e göre daha iyi çalıştığını göstermektedir. N=5 seçildiğinde ise; en iyi sonuçları KNN verdi. Öznitelik sayısının artmasıyla KNN sınıflandırma sonuçlarında değişiklik olmasa da diğer algoritmaların üzerinde ufak değişimlere sebep oldu. Örneğin Catboost algoritmasının sonuçları düşerken, GBM ve LightGBM algoritmalarının sonuçlarında artış gözlemlendi. MRMR öznitelik seçim metoduyla seçilen öznitelikler kullanılarak kurulan modellerin öğrenmelerinde sıkıntı olduğu söylenebilir. Özellikle Seröz kitle grubuna ait hasta verilerinin tahminlerinde çok büyük sıkıntılar olduğu ve MCC değerlerinin 1'e yakınsaması gerekirken, çok düşük değerlerde ve hatta negatif değerlerde seyretmesi modellerin rasgele tahmin yürüttüğünü göstermektedir.

Aşağıda verilen Tablo 3.7.'de ise SFS ve SBS algoritmalarıyla her sınıflandırıcının kendisine özel olarak seçtiği öznitelikler verildi.

Tablo 3.7. Endometrioid-Seröz Sınıflandırması için SFS ve SBS Algoritmalarıyla Seçilen Öznitelikler

	SFS	SBS
<b>KNN</b>	glcm_DifferenceAverage glszm_ZonePercentage	firstorder_TotalEnergy
<b>SVM</b>	glcm_Idmn	shape_VoxelVolume hape_Maximum3DDiameter shape_MeshVolume shape_MajorAxisLength shape_Sphericity shape_LeastAxisLength shape_Elongation firstorder_Energy
<b>Decision Tree</b>	shape_Maximum3DDiameter shape_MajorAxisLength shape_Maximum2DDiameterSlice gldm_LargeDependenceHighGrayLevelEmphasis gldm_SmallDependenceLowGrayLevelEmphasis firstorder_Range glszm_LargeAreaLowGrayLevelEmphasis glszm_ZoneEntropy	shape_MeshVolume shape_LeastAxisLength shape_SurfaceArea gldm_LargeDependenceHighGrayLevelEmphasis gldm_SmallDependenceLowGrayLevelEmphasis firstorder_Range
<b>Random Forest</b>	shape_MeshVolume shape_Maximum2DDiameterColumn gldm_GrayLevelNonUniformity glcm_Contrast glszm_GrayLevelNonUniformity glszm_LargeAreaLowGrayLevelEmphasis	shape_Maximum3DDiameter shape_Flatness shape_Maximum2DDiameterColumn gldm_LargeDependenceHighGrayLevelEmphasis glrlm_GrayLevelNonUniformity
<b>MLP</b>	shape_Elongation shape_SurfaceVolumeRatio gldm_DependenceNonUniformityNormalized glcm_MaximumProbability glcm_Id firstorder_Skewness glszm_LowGrayLevelZoneEmphasis ngtdm_Strength	shape_Flatness glcm_Idn glcm_Autocorrelation firstorder_TotalEnergy glszm_LargeAreaHighGrayLevelEmphasis
<b>GBM</b>	firstorder_Mean glszm_LargeAreaLowGrayLevelEmphasis	glcm_ClusterShade glcm_Idmn glszm_LargeAreaHighGrayLevelEmphasis
<b>XGBOOST</b>	shape_Maximum3DDiameter shape_MajorAxisLength glcm_ClusterShade firstorder_Skewness firstorder_TotalEnergy firstorder_Mean glrlm_GrayLevelNonUniformity ngtdm_Complexity	shape_MajorAxisLength shape_Flatness glcm_ClusterShade glcm_Idmn glszm_LargeAreaHighGrayLevelEmphasis
<b>LightGBM</b>	shape_Maximum2DDiameterSlice glcm_JointAverage glcm_Contrast glcm_Autocorrelation firstorder_InterquartileRange glrlm_GrayLevelNonUniformityNormalized glszm_ZoneVariance glszm_LargeAreaEmphasis	shape_Maximum2DDiameterSlice glcm_DifferenceVariance glcm_Idn glcm_MCC firstorder_Skewness glszm_LargeAreaLowGrayLevelEmphasis glszm_LargeAreaHighGrayLevelEmphasis glszm_HighGrayLevelZoneEmphasis

Tablo 3.7. (Devam) Endometrioid-Seröz Sınıflandırması için SFS ve SBS Algoritmalarıyla Seçilen Öznitelikler

	SFS	SBS
<b>CatBoost</b>	shape_Maximum2DDiameterSlice shape_Maximum2DDiameterRow gldm_LargeDependenceHighGrayLevelEmphasis glszm_LowGrayLevelZoneEmphasis	shape_Maximum2DDiameterSlice glcm_DifferenceVariance glcm_Idn glcm_MCC firstorder_Skewness glszm_LargeAreaLowGrayLevelEmphasis glszm_LargeAreaHighGrayLevelEmphasis glszm_HighGrayLevelZoneEmphasis

Tablo 3.8. SFS ile Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	MCC	TP	FP	FN	TN
<b>KNN</b>	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>SVM</b>	0,48	0,48	1,0	0,65	0,0	0,0	13	14	0	0
<b>DecisionTree</b>	0,77	0,82	0,69	0,75	0,85	0,55	9	2	4	12
<b>RandomForest</b>	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>MLP</b>	0,92	1,0	0,85	0,92	1,0	0,86	11	0	2	14
<b>GBM</b>	0,92	0,87	1,0	0,93	0,85	0,86	13	2	0	12
<b>XGBOOST</b>	0,77	0,82	0,69	0,75	0,85	0,55	9	2	4	12
<b>LightGBM</b>	0,77	0,82	0,69	0,75	0,85	0,55	9	2	4	12
<b>CatBoost</b>	0,81	1,0	0,61	0,75	1,0	0,67	8	0	5	14

Tablo 3.8.' da ise; Step forward selection algoritmasıyla elde edilen sonuçlar verildi. Bu tabloya göre; en başarılı olan algoritmalar MLP ve GBM oldu. Yüksek sınıflandırma performanslarına sahip olan bu iki algoritmadan, MLP'nin, GBM'e göre Seröz grubuna ait hasta tahmininde hiç hata yapmadığı görüldü. GBM ise Endometrioid grubuna ait hastaların tamamını doğru tahmin etti. SVM ise en kötü sınıflandırma performansına sahip olan sınıflandırıcı oldu. Endometrioid grubunun tamamını doğru tahminlerken, Seröz grubuna ait hastaları hiç doğru tahmin edemedi. MCC değerine bakarak tahminlerini rasgele yaptığı da söylenebilir.

Tablo 3.9. SBS ile Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	MCC	TP	FP	FN	TN
<b>KNN</b>	0,85	0,84	0,84	0,84	0,85	0,70	11	2	2	12
<b>SVM</b>	0,85	0,76	1,0	0,86	0,71	0,73	13	4	0	10
<b>DecisionTree</b>	0,77	0,81	0,69	0,74	0,85	0,55	9	2	4	12
<b>RandomForest</b>	0,92	1,0	0,84	0,91	1,0	0,86	11	0	2	14
<b>MLP</b>	0,85	0,85	0,85	0,85	0,85	0,70	11	2	2	12
<b>GBM</b>	0,77	0,73	0,84	0,78	0,71	0,56	11	4	2	10



Tablo 3.9. (Devam) SBS ile Elde Edilen Endometrioid-Seröz Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	MCC	TP	FP	FN	TN
<b>XGBOOST</b>	0,77	0,73	0,84	0,78	0,71	0,56	11	4	2	10
<b>LightGBM</b>	0,92	0,86	1,0	0,92	0,85	0,86	13	2	0	12
<b>CatBoost</b>	0,92	0,86	1,0	0,92	0,85	0,86	13	2	0	12

Endometrioid-Seröz kitlelerinin sınıflandırmaları için son olarak kullanılan öznelik seçim algoritması Step Backward Selection algoritması oldu. Bu tabloda elde edilen sonuçlar Tablo 3.9.' de verildi. Bu tabloya göre; en iyi sonuçları Random Forest, LightGBM ve CatBoost algoritmaları verdi. LightGBM ve CatBoost aynı sonuçları verdi. LightGBM ve CatBoost algoritmaları Endometrioid grubundaki hastaların tamamını doğru tahminlediler. Random Forest algoritması da bu iki algoritmaya göre hiç hata yapmadan %100 performansla Seröz grubuna ait hastaları doğru tahmin etti.

### 3.2. Myom- NonMyom İkili Sınıflandırma Sonuçları

Myom-NonMyom sınıflandırmasının amacı kitle türünü belirlemektir. Kitle türleri iyi ve kötü huylu olarak ikiye ayrılır. Yapılan ilk sınıflandırma iki kötü huylu kitlenin birbirinden ayrılmasını hedeflerken, bu kez iyi ve kötü huylu kitlelerin birbirinden ayrılması hedeflendi. Bu sınıflandırma elimizde bulunan 135 adet hastaya ait radiomics verileri kullanılarak yapıldı. Endometrioid ve Seröz kötü huylu kitleler grubuna dahil olan hastalara ait veriler 1; Myom, iyi huylu kitle grubuna dahil olan hastalara ait veriler 0 olarak etiketlendi. Etiketleme işlemi sonucunda 90 adet Myom olmayan, 45 adet Myom grubuna ait hastaya sahip olduğu gözlemlendi. Bu verilerle; bir önceki aşamadaki gibi farklı makine öğrenmesi algoritmaları ve öznelik seçim algoritmaları kullanılarak hastanın sahip olduğu kitlenin iyi huylu mu kötü huylu mu olduğu tespit edilmek istendi. Yapılan ilk işlem verinin tüm özellikleri kullanılarak yapılan sınıflandırma oldu. Sınıflandırma sonuçları aşağıdaki Tablo 3.10.'de verildi.

Tablo 3.10. Tüm Öznelikler Kullanılarak Yapılan Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,82	0,84	0,93	0,88	0,61	0,77	0,58	26	5	2	8
<b>SVM</b>	0,90	0,93	0,93	0,93	0,84	0,89	0,77	26	2	2	11
<b>DecisionTree</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>RandomForest</b>	0,92	0,90	1,0	0,95	0,76	0,88	0,71	28	3	0	10
<b>MLP</b>	0,90	0,92	0,92	0,92	0,84	0,89	0,71	26	2	2	11
<b>GBM</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8

Tablo 3.10. (Devam) Tüm Öznitelikler Kullanılarak Yapılan Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>XGBOOST</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>LightGBM</b>	0,90	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9
<b>CatBoost</b>	0,90	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9

Tablo 3.10.' a bakıldığında, tüm algoritmalarından iyi ve yüksek sonuçlar alınmış gibi görünse de aslında bu tablo için önemli olan metrikler specificity ve MCC değerleri olacaktır. Çünkü; Myom ve Myom olmayanları sınıflandırırken kullanılan verilerin sayıları eşit değildir. Myom olmayan veriler, Myom olan verilerin iki katı olduğu için modeller Myom tanısı koymada, Myom olmayan kitlelere tanı koymadaki başarılarını gösterememişlerdir. Dolayısıyla bu sınıflandırma yapılırken, iyi huylu kitle türü olan Myom sınıfına ait verilerin doğru tahmin edilme oranı olan specificity değerine ve MCC metriğine daha çok önem verildi. Tablo 3.10.'da, diğer metriklerin yanı sıra, specificity ve MCC değerleri en yüksek olan algoritmalar SVM, MLP ve Random Forest oldu. SVM ve MLP algoritmaları da myom türüne ait tahminleri, yani hasta olmayana hasta değil demeyi %84 oranında doğru yaparken, Random Forest algoritması da myom olmayan kitlelerin tamamını doğru tahmin etti.

Tüm özniteliklerle yapılan sınıflandırma sonrasında, öznitelik seçim adımlarına başlandı. Tablo 3.11'de Myom-NonMyom sınıflandırması için seçilen öznitelikler verildi.

Tablo 3.11. Myom-NonMyom Sınıflandırması için Kullanılan Öznitelik Seçim Yöntemleri ve Elde Edilen Öznitelikler

Öznitelik Seçim Yöntemleri	Seçilen Öznitelikler
<b>Chi-Square Test</b>	gldm_DependenceVariance glrlm_RunLengthNonUniformityNormalized firstorder_Kurtosis gldm_LargeDependenceHighGrayLevelEmphasis glszm_ZoneEntropy gldm_DependenceNonUniformityNormalized shape_SurfaceArea gldm_LargeDependenceEmphasis glszm_GrayLevelNonUniformityNormalized shape_VoxelVolume shape_MeshVolume ngtdm_Strength firstorder_Energy firstorder_TotalEnergy glszm_LargeAreaLowGrayLevelEmphasis glrlm_LongRunLowGrayLevelEmphasis glszm_ZoneVariance

Tablo 3.11. (Devam) Myom-NonMyom Sınıflandırması için Kullanılan Öznitelik Seçim Yöntemleri ve Elde Edilen Öznitelikler

Öznitelik Seçim Yöntemleri	Seçilen Öznitelikler
<b>Chi-Square Test</b>	glszm_LargeAreaEmphasis gldm_DependenceNonUniformity grlm_LongRunEmphasis
<b>Mutual Information</b>	gldm_DependenceNonUniformityNormalized gldm_DependenceVariance glcm_InverseVariance glcm_DifferenceAverage firstorder_Median grlm_RunVariance grlm_LongRunEmphasis grlm_ShortRunHighGrayLevelEmphasis grlm_ShortRunEmphasis grlm_RunPercentage grlm_RunLengthNonUniformityNormalized
<b>ReliefF; K=4</b>	glszm_SizeZoneNonUniformityNormalized glszm_SmallAreaEmphasis glszm_ZoneEntropy glszm_GrayLevelNonUniformityNormalized gldm_DependenceVariance
<b>ReliefF; K=10</b>	glszm_SizeZoneNonUniformityNormalized glszm_SmallAreaEmphasis glszm_ZoneEntropy gldm_DependenceVariance shape_Flatness
<b>MRMR; n=3</b>	firstorder_Uniformity glcm_Id firstorder_RootMeanSquared
<b>MRMR; n=5</b>	firstorder_Uniformity glcm_Id firstorder_RootMeanSquared firstorder_TotalEnergy firstorder_Energy

Tablo 3.12.'de Chi Square Test ile seçilen 20 öznitelikle yapılan sınıflandırma sonuçları verildi.

Tablo 3.12. Chi-Square Test ile Öznitelik Seçimi Sonrası Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,92	0,96	0,93	0,95	0,92	0,93	0,83	26	1	2	12
<b>SVM</b>	0,87	0,93	0,89	0,91	0,84	0,87	0,72	25	2	3	11
<b>DecisionTree</b>	0,85	0,82	1,0	0,90	0,53	0,92	0,66	28	6	0	7
<b>RandomForest</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>MLP</b>	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
<b>GBM</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8

Tablo 3.12. (Devam) Chi Square Test ile Öznitelik Seçimi Sonrası Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
<b>XGBOOST</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>LightGBM</b>	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
<b>CatBoost</b>	0,90	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9

Chi Square Test ile seçilen 20 özelliğin kullanılmasıyla en iyi sonuçları veren algoritma KNN ve CatBoost oldu. KNN'in; kitlesi iyi huylu myom olanlara, myom tanısını %92 specificity değeriyle diğer algoritmalarından çok daha iyi bir şekilde koyduğu söylenebilir. CatBoost ise kötü huylu kitlelerin tamamını doğru tahminlerken; diğer algoritmalara göre iyi huylu kitle tanısını KNN'den sonra en iyi yapan algoritma oldu. Decision Tree, MLP ve LightGBM ise iyi huylu kitleye sahip hastaların yarısına kötü huylu teşhisi koydu. Dolayısıyla en başarısız algoritmalar Decision Tree, MLP ve LightGBM oldu.

Tablo 3.13. Mutual Information ile Seçilen Özniteliklerle Elde Edilen Myom-Nonmyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
<b>KNN</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>SVM</b>	0,85	0,84	0,96	0,90	0,61	0,79	0,65	27	5	1	8
<b>DecisionTree</b>	0,85	0,84	0,96	0,90	0,61	0,79	0,65	27	5	1	8
<b>RandomForest</b>	0,92	0,90	1,0	0,95	0,76	0,88	0,83	28	3	0	10
<b>MLP</b>	0,90	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9
<b>GBM</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
<b>XGBOOST</b>	0,92	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9
<b>LightGBM</b>	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
<b>CatBoost</b>	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8

Yukarıda verilen Tablo 3.13.'de, Mutual Information ile seçilen 11 özellekle elde edilen sonuçlar verilmiştir. Tabloya göre en iyi sonucu Random Forest algoritması vermiştir. Random Forest algoritması, kötü huylu kitleye sahip hastaların tümünü doğru tahmin etti. İyi huylu myom kitlesine sahip olan hastalara ise %76 oranında doğru teşhis koydu. XGBoost algoritması da, Random Forest algoritmasına yakın sonuçlar vermesine rağmen, iyi huylu kitleye sahip hastaların yalnızca %69'unu doğru tahmin ettiği için Random Forest algoritması daha başarılı sonuç verdi denilebilir.

Tablo 3.14. ReliefF ile seçilen özniteliklerle K=4 ve K=10 için yapılan Myom-NonMyom sınıflandırma sonuçları

K=4 için											
Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
KNN	0,80	0,86	0,86	0,86	0,69	0,77	0,54	24	4	4	9
SVM	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
DecisionTree	0,78	0,79	0,93	0,85	0,46	0,70	0,45	26	7	2	6
RandomForest	0,81	0,84	0,96	0,90	0,61	0,79	0,65	27	5	1	8
MLP	0,95	0,93	1,0	0,97	0,84	0,92	0,88	28	2	0	11
GBM	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
XGBOOST	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
LightGBM	0,85	0,84	0,96	0,90	0,61	0,79	0,65	27	5	1	8
CatBoost	0,90	0,88	1,0	0,93	0,69	0,85	0,77	28	4	0	9
K=10 için											
Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	AUC	MCC	TP	FP	FN	TN
KNN	0,85	0,87	0,93	0,90	0,69	0,81	0,65	26	4	2	9
SVM	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
DecisionTree	0,82	0,82	0,96	0,89	0,53	0,75	0,59	27	6	1	7
RandomForest	0,80	0,84	0,96	0,90	0,53	0,79	0,52	27	5	1	8
MLP	0,92	0,90	1,0	0,95	0,76	0,88	0,83	28	3	0	10
GBM	0,87	0,81	0,93	0,87	0,61	0,73	0,72	26	6	2	7
XGBOOST	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
LightGBM	0,78	0,79	0,93	0,85	0,46	0,70	0,45	26	7	2	6
CatBoost	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8

Tablo 3.14. ise ReliefF öznitelik seçim algoritmasıyla seçilen özniteliklerle elde edilen sonuçları göstermektedir. Bu algoritma için yine K; en yakın komşu sayısı önce 4, daha sonra 10 olarak seçildi. K=4 ve K=10 için özellik sayısı sabit tutulup 5 olarak belirlendi. K'nın değişmesiyle K=4 için seçilen özniteliklerin sadece 1 tanesi değişti. K=4 için en iyi sonuçları çok büyük farkla MLP algoritması verdi. Kötü huylu kitleye sahip hastaların tamamını, iyi huylu olan hastaların ise %84'ünü doğru tahmin etti. Ayrıca MCC değerinin %88 olması da, yaptığı tahminlerin doğru olduğunu göstermektedir. K=10 için yapılan sınıflandırma sonucunda ise yine MLP iyi ve kötü huylu kitle tahminlerinde en iyi performansı verdi. Başarı metriklerinde düşüş olduğu gözlemlendi. Böylelikle K=4 için seçilen özniteliklerin daha iyi sonuçlar verdiği söylenebilir.

Tablo 3.15. MRMR ile seçilen özneliklerle elde edilen Myom-NonMyom sınıflandırma sonuçları

N=3 için											
Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
KNN	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
SVM	0,75	0,75	0,96	0,84	0,30	0,64	0,38	27	9	0	4
DecisionTree	0,80	0,78	1,0	0,88	0,38	0,69	0,54	28	8	0	5
RandomForest	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
MLP	0,78	0,76	1,0	0,86	0,30	0,65	0,48	28	9	0	4
GBM	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
XGBOOST	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
LightGBM	0,80	0,78	1,0	0,88	0,38	0,69	0,54	28	8	0	5
CatBoost	0,78	0,77	0,96	0,86	0,38	0,67	0,45	27	8	1	5
N=5 için											
Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	AUC	MCC	TP	FP	FN	TN
KNN	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
SVM	0,75	0,75	0,96	0,84	0,30	0,64	0,38	27	9	1	4
DecisionTree	0,78	0,78	0,78	0,76	0,46	0,70	0,45	26	7	2	6
RandomForest	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
MLP	0,75	0,74	1,0	0,85	0,23	0,62	0,41	28	10	0	3
GBM	0,87	0,85	1,0	0,92	0,61	0,81	0,72	28	5	0	8
XGBOOST	0,85	0,82	1,0	0,90	0,53	0,77	0,66	28	6	0	7
LightGBM	0,82	0,80	1,0	0,89	0,46	0,73	0,60	28	7	0	6
CatBoost	0,82	0,82	0,96	0,89	0,53	0,75	0,59	27	6	1	7

Tablo 3.15.'de ise; MRMR özellik seçim yönteminden elde edilen özelliklerle yapılan sınıflandırma sonuçları yer almaktadır. Özellik sayısı(n); n=3 seçildiğinde; SVM ve CatBoost kötü huylu kitle tespitini %96 oranında doğru tahmin ederken, diğer algoritmaların tümü bu tahmini %100 oranında doğru yaptılar. Fakat aynı başarıyı iyi huylu kitle tespiti için yapılan tahminler için söylemek mümkün değildir. En yüksek specificity değerine sahip olan KNN %53 oranında iyi huylu kitle tahminini doğru yapmaktadır. Bu da iyi huylu kitleye sahip hastaların yaklaşık olarak yarısına kötü huylu kitle tanısı konmuş anlamına gelir. N=5 seçildiğinde ise sonuçlar çok büyük bir değişime uğramadı. Genel olarak MRMR yöntemiyle seçilen öznelikler iyi huylu kitle tespiti için yeterli olmadı.

Tablo 3.16'da Myom-NonMyom sınıflandırması için SFS ve SBS algoritmalarıyla her sınıflandırıcının özel olarak seçtiği öznelikler verildi.

Tablo 3.16. Myom-NonMyom Sınıflandırması için SFS ve SBS Algoritmalarıyla Seçilen Öznitelikler

	SFS	SBS
<b>KNN</b>	gldm_DependenceVariance firstorder_RootMeanSquared glszm_ZoneEntropy	shape_VoxelVolume shape_Maximum3DDiameter shape_MeshVolume shape_MajorAxisLength shape_Sphericity shape_LeastAxisLength firstorder_Energy firstorder_TotalEnergy
<b>SVM</b>	gldm_DependenceVariance	shape_VoxelVolume shape_Maximum3DDiameter shape_MeshVolume shape_MajorAxisLength shape_Sphericity shape_LeastAxisLength shape_Elongation shape_SurfaceVolumeRatio
<b>Decision Tree</b>	gldm_DependenceNonUniformityNormalized gldm_DependenceVariance glcm_DifferenceEntropy glcm_InverseVariance glcm_DifferenceVariance glcm_SumEntropy glrlm_RunVariance glszm_LargeAreaHighGrayLevelEmphasis	shape_Maximum3DDiameter shape_SurfaceArea gldm_GrayLevelVariance gldm_DependenceNonUniformityNormalized glcm_JointEnergy glcm_Idm glcm_Imc1 glszm_LowGrayLevelZoneEmphasis
<b>Random Forest</b>	shape_Maximum3DDiameter gldm_DependenceNonUniformityNormalized glcm_SumEntropy glrlm_LongRunEmphasis glszm_ZoneEntropy	shape_Maximum2DDiameterSlice shape_SurfaceArea gldm_DependenceVariance glcm_ClusterProminence firstorder_Median glrlm_ShortRunLowGrayLevelEmphasis glrlm_RunPercentage glszm_LowGrayLevelZoneEmphasis
<b>MLP</b>	firstorder_InterquartileRange firstorder_Skewness firstorder_10Percentile glrlm_LowGrayLevelRunEmphasis	shape_Elongation gldm_SmallDependenceHighGrayLevelEmphasis firstorder_Energy firstorder_TotalEnergy firstorder_90Percentile glrlm_GrayLevelNonUniformity glrlm_LongRunEmphasis glrlm_RunPercentage
<b>GBM</b>	shape_Maximum2DDiameterSlice gldm_GrayLevelVariance gldm_DependenceNonUniformityNormalized glcm_Idm glcm_Correlation glszm_GrayLevelNonUniformityNormalized	shape_Maximum3DDiameter shape_MeshVolume shape_LeastAxisLength shape_SurfaceArea gldm_GrayLevelNonUniformity gldm_DependenceVariance firstorder_RobustMeanAbsoluteDeviation glszm_LowGrayLevelZoneEmphasis
<b>XGBOOST</b>	gldm_DependenceNonUniformityNormalized glcm_InverseVariance glszm_LargeAreaLowGrayLevelEmphasis glszm_ZoneEntropy	shape_Maximum3DDiameter shape_SurfaceArea shape_Maximum2DDiameterRow gldm_GrayLevelNonUniformity gldm_DependenceVariance glcm_MaximumProbability

Tablo 3.16. (Devam) Myom-NonMyom Sınıflandırması için SFS ve SBS Algoritmalarıyla Seçilen Öznitelikler

SFS		SBS
<b>XGBoost</b>		firstorder_Median glszm_LowGrayLevelZoneEmphasis
<b>Lightgbm</b>	gldm_GrayLevelVariance gldm_DependenceVariance glcm_Contrast glcm_InverseVariance	shape_MeshVolume shape_MajorAxisLength gldm_DependenceNonUniformity gldm_DependenceVariance gldm_LargeDependenceHighGrayLevelEmphasis glcm_JointEnergy glcm_InverseVariance
<b>CatBoost</b>	gldm_GrayLevelVariance gldm_HighGrayLevelEmphasis gldm_SmallDependenceHighGrayLevelEmphasis gldm_DependenceNonUniformityNormalized gldm_DependenceVariance glcm_JointEnergy glcm_InverseVariance glszm_GrayLevelNonUniformityNormalized	shape_Elongation shape_Maximum2DDiameterSlice gldm_DependenceNonUniformity gldm_LargeDependenceEmphasis gldm_DependenceVariance

Tablo 3.17. SFS kullanılarak elde edilen Myom-NonMyom sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	MCC	TP	FP	FN	TN
<b>KNN</b>	0,78	0,76	1,0	0,86	0,30	0,48	28	9	0	4
<b>SVM</b>	0,80	0,79	0,96	0,87	0,46	0,52	27	7	1	6
<b>DecisionTree</b>	0,90	0,88	1,0	0,93	0,69	0,77	28	4	0	9
<b>RandomForest</b>	0,87	0,90	0,93	0,91	0,76	0,71	26	3	2	10
<b>MLP</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8
<b>GBM</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8
<b>XGBOOST</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8
<b>LightGBM</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8
<b>CatBoost</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8

Tablo 3.17.'de ise; özellik seçimini Step forward selection algoritmasıyla; her sınıflandırıcı özel olarak seçildi. Bu tabloya göre; Decision Tree kötü huylu kitle tahminini %100 doğru yaparak Random Forest algoritmasına göre daha iyi bir performans sergilerken, Random Forest ise iyi huylu myom kitlesi tahminini Decision Tree algoritmasına göre daha yüksek oranda yaptı.

Tablo 3.18. SBS ile Elde Edilen Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	F-Score	Specificity	MCC	TP	FP	FN	TN
<b>KNN</b>	0,87	0,72	1,0	0,84	0,61	0,72	28	5	0	8
<b>SVM</b>	0,73	0,76	1,0	0,86	0,15	0,33	27	11	0	2
<b>DecisionTree</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8
<b>RandomForest</b>	0,85	0,82	1,0	0,90	0,53	0,66	28	6	0	7
<b>MLP</b>	0,82	0,84	0,93	0,88	0,61	0,58	26	5	2	8



Tablo 3.18. (Devam) SBS ile Elde Edilen Myom-NonMyom Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision	Recall	FScore	Specificity	MCC	TP	FP	FN	TN
<b>GBM</b>	0,80	0,81	0,93	0,87	0,53	0,52	26	6	2	7
<b>XGBOOST</b>	0,90	0,88	1,0	0,93	0,69	0,77	28	4	0	9
<b>LightGBM</b>	0,82	0,84	0,93	0,88	0,61	0,58	26	5	2	8
<b>CatBoost</b>	0,87	0,85	1,0	0,92	0,61	0,72	28	5	0	8

Son olarak Tablo 3.18.'de Step Backward Selection algoritması kullanılarak yapılan öznelik seçimi ve elde edilen sonuçlar verilmektedir. Bu tabloda verilen sonuçlar için seçilen öznelikleri Step backward selection algoritmasıyla; her sınıflandırıcı özel olarak seçti. Sonuçlara göre; kötü huylu kitleye sahip hastaların tamamına, iyi huylu kitleye sahip hastaların %69'una doğru tanı koyan XGBoost algoritması en iyi performansı gösteren algoritma oldu. Kötü huylu kitleye sahip hastaların tamamına doğru tanı koymasına rağmen, iyi huylu kitleye sahip hastaların neredeyse hepsine yanlış teşhis koyan SVM ise en kötü performans gösteren algoritma oldu.

### 3.3. Myom-Endometrioid-Seröz Çok Sınıf Sınıflandırma Sonuçları

Myom-Endometrioid- Seröz sınıflandırması; elimizde bulunan 135 adet hastaya ait radiomics verileri kullanılarak yapıldı. Kitlesi Myom grubuna dahil olan hastalara ait veriler 0 olarak etiketlenirken, endometrioid grubuna dahil olan hastalara ait veriler 1; seröz grubuna dahil olan hastalara ait veriler 2 olarak etiketlendi. Etiketleme işlemi sonucunda 45 adet myom, 46 adet endometrioid, 44 adet seröz grubuna ait hastaya sahip olduğu gözlemlendi. Bu verilere de önceki aşamalarda kullanılan sınıflandırma ve öznelik seçim algoritmaları uygulandı. Çoklu sınıflandırma yapılmasındaki amaç; kitlenin iyi huylu veya kötü huylu olmasının tespit edilmesi değil, hangi kitle sınıfına ait olduğunun belirlenmesidir. Dolayısıyla bu üç sınıfa ait veriler kullanılarak sınıflandırma yapıldı. Aşağıda verilen Tablo 3.19.'da, verinin 108 adet özneliği kullanılarak makine öğrenmesi algoritmalarından elde edilen sonuçlar verildi.

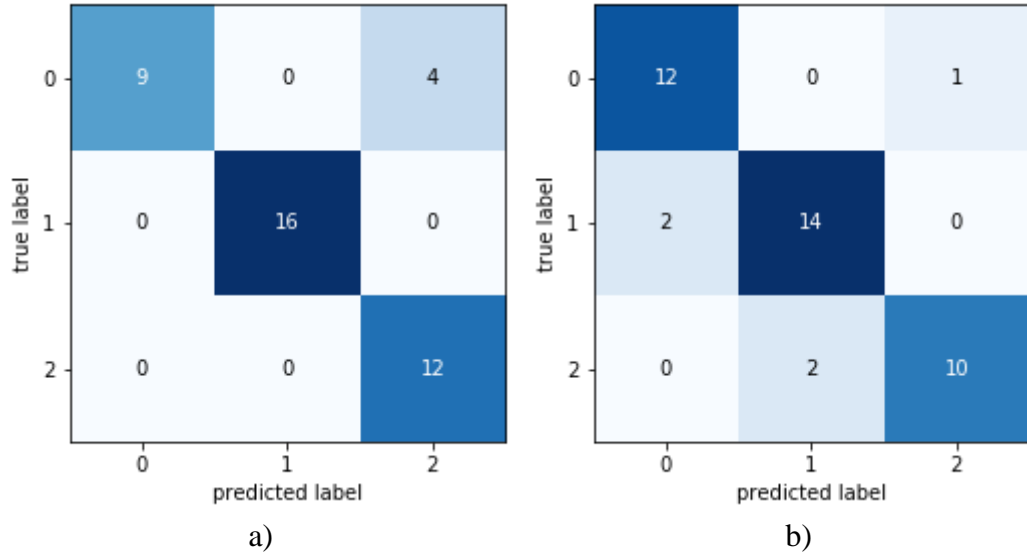
Tablo 3.19. Tüm Özneliklerle Yapılan Çoklu Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
<b>KNN</b>	0,82	0,80	0,77	0,78	0,67
<b>SVM</b>	0,81	0,80	0,80	0,80	0,70
<b>DecisionTree</b>	0,92	0,91	0,89	0,89	0,86

Tablo 3.19. (Devam) Tüm Özniteliklerle Yapılan Çoklu Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
RandomForest	0,85	0,82	0,79	0,80	0,70
MLP	0,89	0,88	0,88	0,88	0,81
GBM	0,87	0,84	0,82	0,82	0,74
XGBOOST	0,87	0,84	0,82	0,82	0,74
LightGBM	0,86	0,81	0,81	0,81	0,74
CatBoost	0,90	0,88	0,87	0,87	0,81

Tablo 3.19’da verilen sonuçlara bakıldığında; en iyi sınıflandırma performansı sunan algoritmalar Decision Tree ve CatBoost olduğu görüldü. Decision Tree algoritması makro ölçekte %89 oranında performans gösterirken, CatBoost algoritmasında ise bu oran %87’dir. En yüksek MCC değerine sahip olan Decision Tree algoritmasının doğru tahminleme yaptığı söylenebilir.



Şekil 3.1. a) Decision Tree kullanılarak yapılan sınıflandırma sonucu elde edilen karmaşıklık matrisi b) CatBoost karmaşıklık matrisi

Şekil 3.1.’de Decision Tree ve CatBoost algoritmalarına ait karmaşıklık matrisleri bulunmaktadır. Decision Tree algoritması; kötü huylu endometrioid ve seröz kitlelerine ait hastaların tamamını doğru tahminlerken; CatBoost algoritması ise iyi huylu kitle tespitinde daha başarılı sonuçlar verdi.

Tablo 3.20.’de ise Endometrioid-Seröz-Myom sınıflandırması için seçilen öznitelikler verilmiştir.

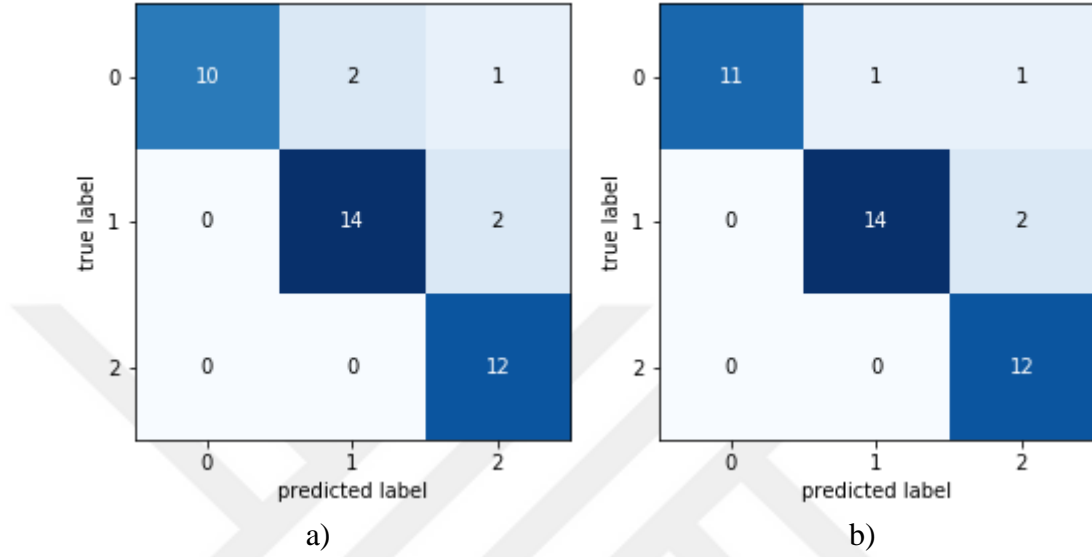
Tablo 3.20. Çok sınıflı Sınıflandırma için Seçilen Öznitelikler

Öznitelik Seçim Yöntemleri	Seçilen Öznitelikler
<b>Chi-Square Test</b>	glszm_SizeZoneNonUniformity gldm_DependenceVariance shape_SurfaceArea shape_VoxelVolume shape_MeshVolume gldm_DependenceNonUniformity gldm_DependenceNonUniformityNormalized firstorder_Energy firstorder_TotalEnergy glszm_LargeAreaLowGrayLevelEmphasis
<b>Mutual Information</b>	gldm_DependenceNonUniformityNormalized gldm_DependenceVariance glcm_InverseVariance glcm_DifferenceAverage firstorder_Median grlm_RunVariance grlm_LongRunEmphasis grlm_ShortRunHighGrayLevelEmphasis grlm_ShortRunEmphasis grlm_RunPercentage grlm_RunLengthNonUniformityNormalized
<b>ReliefF; K=4</b>	glszm_SizeZoneNonUniformityNormalized glszm_SmallAreaEmphasis glszm_ZoneEntropy glszm_GrayLevelNonUniformityNormalized gldm_DependenceVariance
<b>ReliefF; K=10</b>	glszm_SizeZoneNonUniformityNormalized glszm_SmallAreaEmphasis glszm_ZoneEntropy gldm_DependenceVariance shape_Flatness
<b>MRMR; n=3</b>	firstorder_Uniformity glcm_Id firstorder_RootMeanSquared
<b>MRMR; n=5</b>	firstorder_Uniformity glcm_Id firstorder_RootMeanSquared firstorder_TotalEnergy firstorder_Energy

Tablo 3.21. Chi Square Test ile öznitelik seçimi sonrası yapılan çoklu sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
<b>KNN</b>	0,86	0,85	0,83	0,82	0,75
<b>SVM</b>	0,77	0,69	0,68	0,68	0,53
<b>DecisionTree</b>	0,82	0,82	0,75	0,72	0,67
<b>RandomForest</b>	0,81	0,77	0,75	0,75	0,63
<b>MLP</b>	0,84	0,80	0,77	0,77	0,67
<b>GBM</b>	0,90	0,89	0,88	0,87	0,82
<b>XGBOOST</b>	0,88	0,87	0,85	0,85	0,79
<b>LightGBM</b>	0,89	0,86	0,85	0,85	0,78
<b>CatBoost</b>	0,92	0,91	0,90	0,90	0,85

Chi Square Test ile seçilen 10 özelliğin kullanılmasıyla elde edilen sonuçlar Tablo 3.21’ de verildi. En iyi sonuçları veren algoritmalar GBM ve CatBoost oldu. CatBoost algoritması %92 doğruluk ve %90 sınıflandırma performansı ile başarılı algoritmadır. Yüksek MCC değeri de tahminleme doğruluğunu kanıtlar niteliktedir.



Şekil 3.2. Karmaşıklık matrisleri a) GBM b) CatBoost

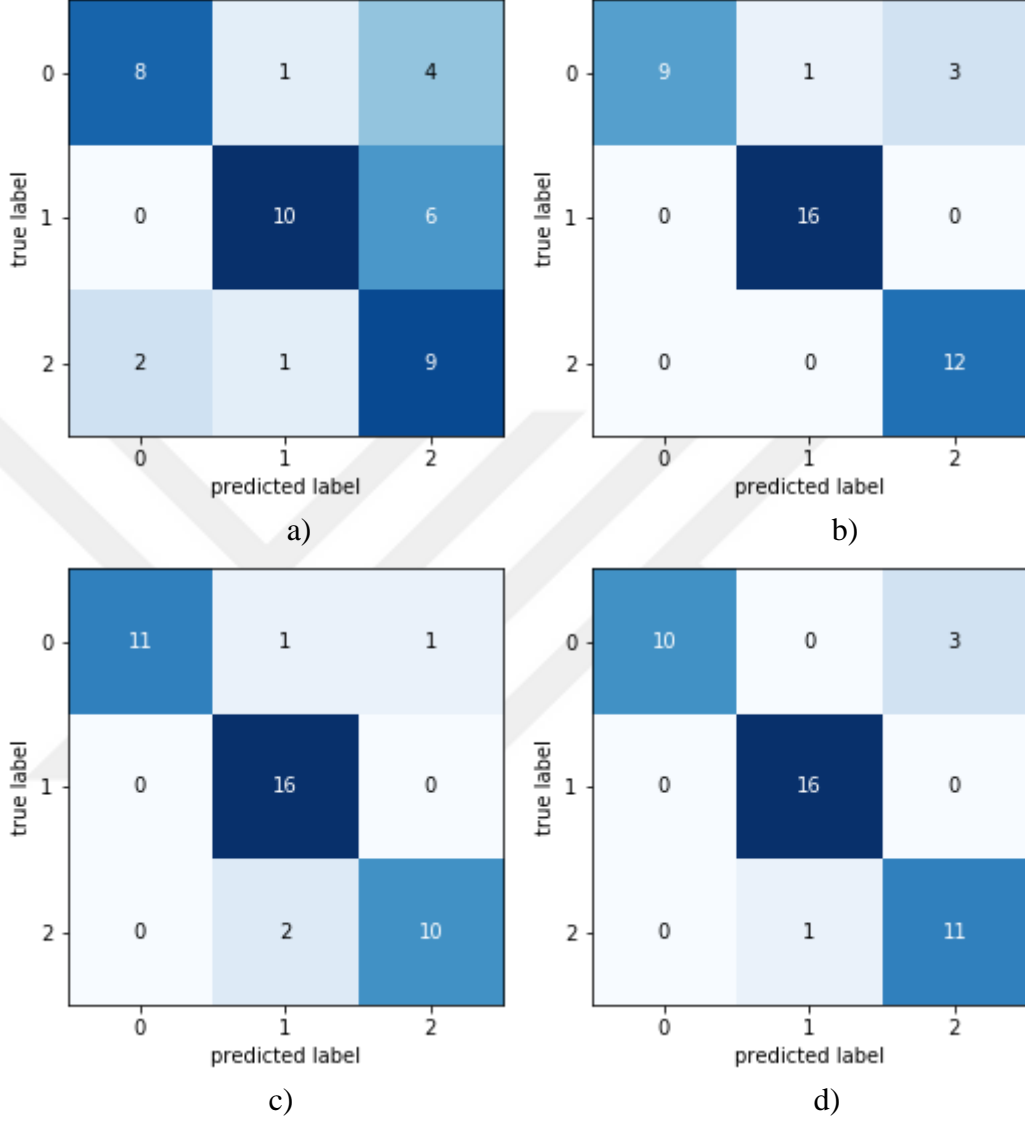
Şekil 3.2.’de GBM ve CatBoost algoritmalarına ait karmaşıklık matrisleri bulunmaktadır. Her iki algoritma da; kötü huylu endometrioid ve seröz kitlelerine ait hastaların eşit şekilde tahminlerken; CatBoost algoritması ise iyi huylu kitle tespitinde daha başarılı sonuçlar verdi

Tablo 3.22. Mutual Information ile seçilen özneliklerle elde edilen çoklu sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
KNN	0,90	0,89	0,86	0,87	0,82
SVM	0,76	0,70	0,66	0,66	0,50
DecisionTree	0,90	0,89	0,86	0,86	0,82
RandomForest	0,90	0,87	0,86	0,86	0,81
MLP	0,86	0,82	0,82	0,82	0,74
GBM	0,91	0,91	0,89	0,89	0,86
XGBOOST	0,89	0,86	0,85	0,85	0,78
LightGBM	0,92	0,91	0,89	0,90	0,85
CatBoost	0,92	0,90	0,89	0,89	0,85

Yukarıda verilen Tablo 3.22.’de ise Mutual Information ile seçilen 11 özellikle elde edilen sonuçlar verildi. Chi Square Test ile elde edilen sonuçlara göre çok daha iyi

sonular verildiđi gzlemlendi. Tabloya gre en iyi sonuları GBM, LightGBM ve CatBoost algoritmaları verildi. En kt sonu ise SVM ile elde edildi.



Őekil 3.3. KarmaŐıklık matrisleri, a) SVM, b)GBM, c)LightGBM, d) CatBoost

Őekil 3.3. 'deki karmaŐıklık matrislerine bakıldıđında; GBM, LightGBM ve CatBoost kullanarak yapılan sınıflandırmalarda; en iyi tahmin edilebilen grup endometrioid grubu oldu.

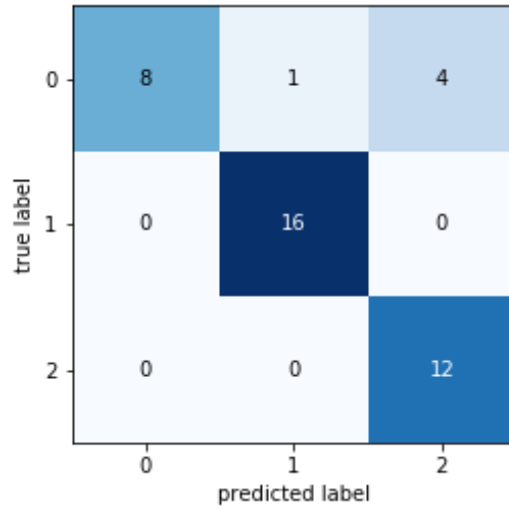
SVM e ait karmaŐıklık matrisine bakılırsa; zellikle myom ve endometrioid grubuna ait tahminlemelerde yanlıŐ tanı koyulduđu sylenbilir.

Tablo 3.23. ReliefF ile Seçilen Özniteliklerle K=4 ve K=10 için Elde Edilen Çoklu Sınıflandırma Sonuçları

K=4 için					
Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
KNN	0,76	0,77	0,66	0,67	0,52
SVM	0,86	0,82	0,83	0,81	0,74
DecisionTree	0,82	0,76	0,75	0,74	0,64
RandomForest	0,86	0,82	0,83	0,81	0,75
MLP	0,90	0,89	0,87	0,86	0,83
GBM	0,88	0,85	0,85	0,84	0,78
XGBOOST	0,87	0,85	0,85	0,85	0,78
LightGBM	0,88	0,85	0,85	0,84	0,78
CatBoost	0,88	0,85	0,85	0,84	0,78
K=10 için					
KNN	0,87	0,85	0,85	0,85	0,78
SVM	0,83	0,78	0,78	0,77	0,67
DecisionTree	0,80	0,74	0,74	0,73	0,61
RandomForest	0,86	0,82	0,83	0,82	0,74
MLP	0,90	0,89	0,87	0,86	0,83
GBM	0,84	0,80	0,80	0,78	0,71
XGBOOST	0,88	0,87	0,85	0,85	0,78
LightGBM	0,88	0,85	0,85	0,84	0,78
CatBoost	0,89	0,87	0,87	0,87	0,81

Tablo 3.23.'de ise ReliefF öznitelik seçim algoritmasıyla seçilen öznitelikler ve bu özniteliklerle elde edilen sonuçlar verildi. K; en yakın komşu sayısı önce 4, daha sonra 10 olarak seçildi.

K=4 ve K=10 için öznitelik sayısı sabit tutulup 5 olarak belirlendi. K'nın değişmesiyle 1 farklı özellik seçildi. Seçilen özniteliklerle elde edilen sonuçlar karşılaştırılınca; her iki tablo için de MLP başarılı oldu. %90 doğruluk, %87 sınıflandırma performansı ve %83 MCC değeriyle yaptığı tahminlerin doğru olduğu söylenebilir. K=4 için elde edilen sonuçlara bakıldığında; en kötü performansı KNN algoritması verdi. K, en yakın komşu sayısının artmasıyla, KNN algoritmasından elde edilen sonuçlar artış olduğu gözlemlendi. K=10 için; elde edilen sonuçlar genel olarak yüksek olsa da, diğer algoritmalara kıyasla en kötü sonuçları veren algoritmalar SVM ve GBM oldu.



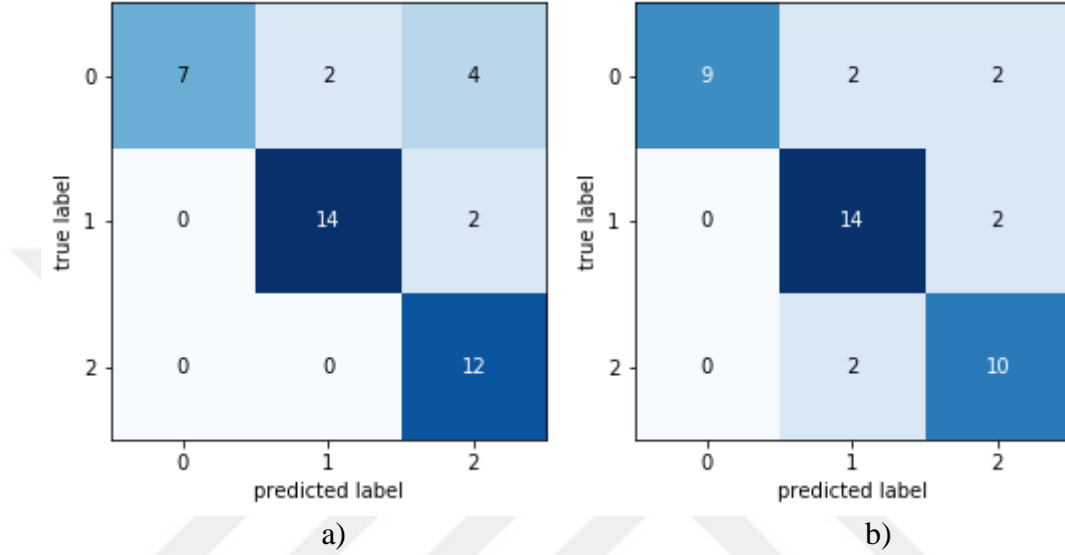
Şekil 3.4. MLP'ye ait karmaşıklık matrisi

Şekil 3.4.'deki karmaşıklık matrisine bakılırsa; MLP kullanarak yapılan sınıflandırmalarda; kötü huylu kitle türleri olan Endometrioid ve Seröz grubuna ait hastaların tamamı doğru tahmin edilirken; yapılan yanlış tahminlerin Myom sınıfına ait olduğu gözlemlendi.

Tablo 3.24. MRMR Yöntemiyle Elde Edilen Çoklu Sınıflandırma Sonuçları

N=3 için					
Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
KNN	0,85	0,84	0,80	0,79	0,73
SVM	0,76	0,67	0,66	0,65	0,50
DecisionTree	0,72	0,62	0,60	0,58	0,42
RandomForest	0,81	0,80	0,74	0,74	0,64
MLP	0,73	0,62	0,63	0,59	0,46
GBM	0,81	0,80	0,74	0,74	0,64
XGBOOST	0,80	0,78	0,72	0,71	0,61
LightGBM	0,78	0,77	0,69	0,68	0,58
CatBoost	0,69	0,56	0,55	0,56	0,34
N=5 için					
KNN	0,85	0,84	0,80	0,79	0,73
SVM	0,76	0,67	0,66	0,65	0,50
DecisionTree	0,77	0,69	0,67	0,68	0,52
RandomForest	0,85	0,83	0,80	0,80	0,71
MLP	0,77	0,69	0,68	0,66	0,54
GBM	0,87	0,84	0,82	0,82	0,74
XGBOOST	0,83	0,81	0,77	0,77	0,67
LightGBM	0,82	0,80	0,74	0,74	0,64
CatBoost	0,73	0,62	0,61	0,61	0,44

MRMR özellik seçim yönteminden elde edilen özelliklerle yapılan sınıflandırma sonuçları Tablo 3.24.'de yer almaktadır. Özellik sayısı(n); n=3 seçildiğinde; %85 doğrulukla KNN algoritması en iyi sonuçları veren algoritma oldu. N=5 seçildiğinde de KNN; yine aynı sonuçları verirken, Random Forest algoritmasının sonuçlarında artış gözlemlendi.



Şekil 3.5. Karmaşıklık matrisleri, a) KNN, b) Random Forest

Şekildeki karmaşıklık matrisine bakılırsa; myom grubuna ait hastalara daha fazla yanlış tanı koyulduğu, endometrioid grubunun ise daha doğru tahmin edilebildiği sonucuna varılabilir.

Tablo 3.25'de ise çok sınıflı sınıflandırma için SFS ve SBS algoritmalarıyla seçilen öznitelikler verildi.

Tablo 3.25. Çok Sınıflı Sınıflandırma için SFS ve SBS ile Seçilen Öznitelikler

Algoritmalar	SFS	SBS
<b>KNN</b>	gldm_DependenceVariance firstorder_RootMeanSquared glszm_ZoneEntropy	shape_VoxelVolume shape_Maximum3DDiameter shape_MeshVolume shape_MajorAxisLength shape_Sphericity shape_LeastAxisLength firstorder_Energy firstorder_TotalEnergy
<b>SVM</b>	shape_Maximum2DDiameterColumn gldm_DependenceEntropy gldm_DependenceVariance glcm_SumAverage glszm_ZoneEntropy	shape_VoxelVolume hape_Maximum3DDiameter shape_MeshVolume shape_MajorAxisLength firstorder_Energy



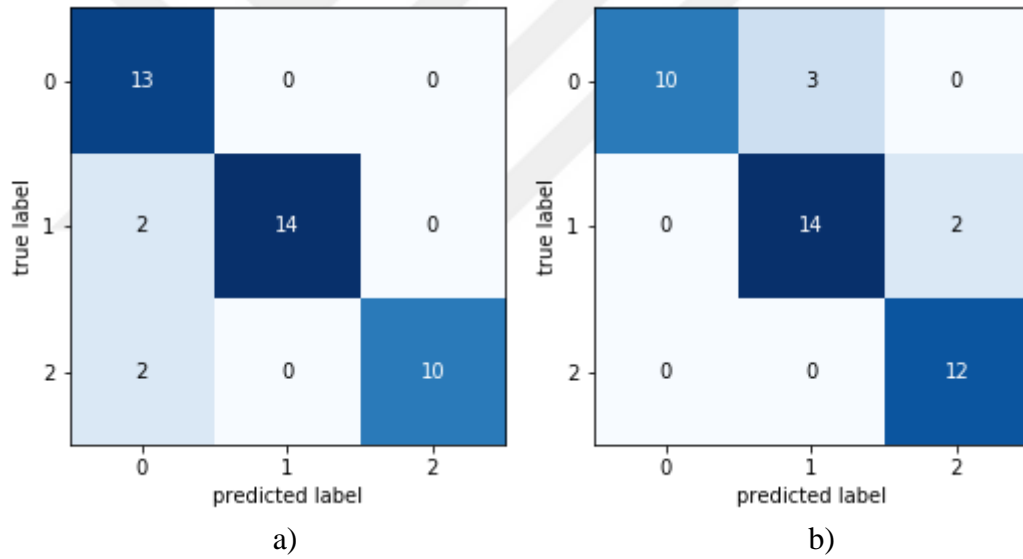
Tablo 3.25. (Devam) Çok Sınıflı Sınıflandırma için SFS ve SBS ile Seçilen Öznitelikler

Algoritmalar	SFS	SBS
<b>SVM</b>	glcm_JointEntropy glcm_SumSquares	shape_LeastAxisLength glszm_LargeAreaHighGrayLevelEmphasis shape_Sphericity
<b>Decision Tree</b>	shape_SurfaceVolumeRatio gldm_DependenceVariance glcm_DifferenceEntropy glcm_Idm glcm_Id firstorder_Skewness firstorder_RobustMeanAbsoluteDeviation glrlm_ShortRunEmphasis	shape_Elongation shape_Flatness shape_SurfaceArea gldm_DependenceVariance glcm_DifferenceEntropy glszm_LargeAreaHighGrayLevelEmphasis glszm_LowGrayLevelZoneEmphasis
<b>Random Forest</b>	shape_Sphericity gldm_DependenceVariance glcm_DifferenceEntropy glcm_Id firstorder_TotalEnergy glrlm_RunPercentage	shape_Maximum3DDiameter shape_Maximum2DDiameterSlice gldm_DependenceVariance glcm_MaximumProbability firstorder_Skewness glszm_LowGrayLevelZoneEmphasis
<b>MLP</b>	shape_Sphericity shape_Elongation gldm_LowGrayLevelEmphasis glcm_Id glrlm_ShortRunLowGrayLevelEmphasis glrlm_RunPercentage	shape_Sphericity shape_SurfaceVolumeRatio glcm_ClusterTendency firstorder_Energy firstorder_TotalEnergy glrlm_GrayLevelNonUniformity glrlm_RunLengthNonUniformity glszm_LargeAreaHighGrayLevelEmphasis
<b>GBM</b>	gldm_DependenceVariance glcm_DifferenceEntropy glcm_Id firstorder_Variance	shape_Maximum3DDiameter shape_Elongation shape_Maximum2DDiameterSlice shape_SurfaceArea gldm_LargeDependenceEmphasis gldm_DependenceVariance glcm_Idmn glszm_LowGrayLevelZoneEmphasis
<b>XGBOOST</b>	gldm_DependenceNonUniformityNormalized gldm_DependenceVariance glcm_DifferenceVariance glcm_Id	shape_Sphericity shape_Elongation shape_MinorAxisLength gldm_SmallDependenceEmphasis gldm_DependenceVariance gldm_LargeDependenceHighGrayLevelEmphasis glszm_LargeAreaHighGrayLevelEmphasis glszm_LowGrayLevelZoneEmphasis
<b>LightGBM</b>	gldm_GrayLevelNonUniformity gldm_DependenceVariance glcm_JointEntropy glcm_DifferenceEntropy glcm_SumEntropy firstorder_10Percentile glrlm_RunLengthNonUniformityNormalized glszm_ZonePercentage	shape_Sphericity shape_Maximum2DDiameterSlice gldm_DependenceVariance glcm_DifferenceEntropy firstorder_Skewness firstorder_Median firstorder_TotalEnergy glszm_LowGrayLevelZoneEmphasis
<b>CatBoost</b>	shape_Sphericity gldm_SmallDependenceEmphasis gldm_DependenceVariance glcm_Contrast glcm_DifferenceEntropy glcm_DifferenceVariance glcm_Idm glrlm_ShortRunEmphasis	gldm_GrayLevelVariance gldm_HighGrayLevelEmphasis shape_Sphericity gldm_SmallDependenceEmphasis gldm_DependenceVariance glcm_Contrast glcm_DifferenceEntropy glcm_DifferenceVariance glcm_Idm glrlm_ShortRunEmphasis

Tablo 3.26. SFS algoritması kullanılarak elde edilen çoklu sınıflandırma sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
KNN	0,80	0,80	0,72	0,68	0,63
SVM	0,91	0,92	0,90	0,90	0,86
DecisionTree	0,87	0,86	0,82	0,81	0,76
RandomForest	0,85	0,80	0,80	0,80	0,70
MLP	0,90	0,89	0,88	0,88	0,82
GBM	0,87	0,85	0,81	0,81	0,75
XGBOOST	0,88	0,90	0,84	0,85	0,79
LightGBM	0,88	0,86	0,84	0,84	0,78
CatBoost	0,86	0,82	0,79	0,80	0,70

Tablo 3.26.' da ise; özellik seçimini Step forward selection algoritmasıyla; her sınıflandırıcı özel olarak seçti. Tablo 20' ye göre %91 doğrulukla en iyi sonucu SVM verdi. MLP ise %90 doğrulukla en iyi sonuçları veren ikinci algoritma oldu.



Şekil 3.6. Karmaşıklık matrisleri, a) SVM, b)MLP

SVM ve MLP ye ait karmaşıklık matrisleri şekilde verildi. Matrislere göre; SVM Myom grubuna ait hastaların tamamına doğru teşhis koyarken, MLP ise Seröz grubundaki hastaların tamamını doğru tespit etti.

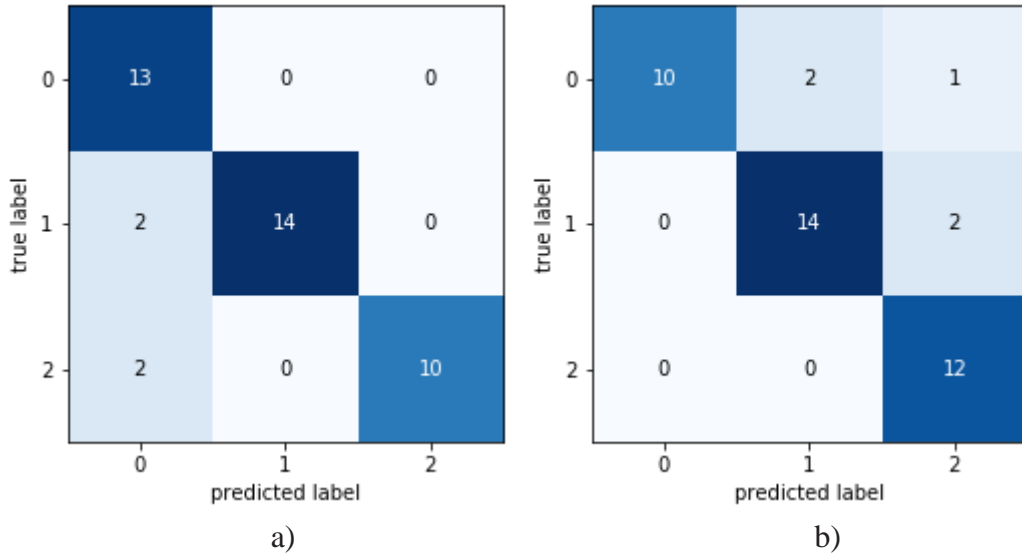
Tablo 3.27. SBS ile Elde Edilen Çoklu Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
KNN	0,89	0,89	0,87	0,86	0,82
SVM	0,91	0,92	0,90	0,90	0,86
DecisionTree	0,81	0,79	0,72	0,72	0,62
RandomForest	0,86	0,82	0,81	0,81	0,74

Tablo 3.27. (Devam) SBS ile Elde Edilen Çoklu Sınıflandırma Sonuçları

Algoritmalar	Doğruluk	Precision <sub>M</sub>	Recall <sub>M</sub>	F-Score <sub>M</sub>	MCC
MLP	0,76	0,72	0,69	0,69	0,57
GBM	0,84	0,83	0,79	0,78	0,72
XGBOOST	0,90	0,89	0,88	0,87	0,82
LightGBM	0,84	0,83	0,79	0,78	0,72
CatBoost	0,87	0,84	0,82	0,82	0,74

Endometriotid-Seröz-Myom kitlelerinin sınıflandırmaları için kullanılan son öznelik seçim algoritması Step Backward Selection algoritması oldu. Elde edilen sonuçlar Tablo 3.27’ de verildi. Bu tabloda verilen sonuçlar için seçilen öznelikleri Step backward selection algoritmasıyla; her sınıflandırıcı özel olarak seçti. Tabloya göre; en iyi sonuçları SVM ve XGBoost algoritmaları verdi. SVM ile elde edilen doğruluk %91 iken, XGBoost ile elde edilen doğruluk %90 oldu.



Şekil 3.7. Karmaşıklık matrisleri, a)SVM, b)XGBoost

Yukarıda verilen karmaşıklık matrislerine göre; SVM Myom grubuna teşhis koymada başarılıyken, XGBoost ise Seröz grubuna teşhis koymada daha başarılıdır. Endometrioid grubu için yapılan tahmin sayısı her iki algoritma için de aynıdır.

#### 4. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında, BT görüntülerinden elde edilen doku analizi verilerinde makine öğrenmesiyle Endometrial Kanser kitle alt-tipi tespiti yapıldı. Yapılan kitle alt-tip sınıflandırmalarında yüksek sonuçlar elde edildi.

Endometrioid-Seröz ikili sınıflandırmasında SFS öznitelik seçim yöntemiyle seçilen 2 adet öznitelikle %92 doğruluk elde eden GBM algoritması en az sayıda özellik kullanarak en iyi sonuca ulaşan algoritma oldu.

Myom-NonMyom sınıflandırmasında K=4 için; ReliefF algoritmasıyla seçilen 5 adet öznitelikle yapılan sınıflandırmada en iyi sonucu %95 doğrulukla MLP verdi. Myom-NonMyom sınıflandırmasında Endometrioid-Seröz ikili sınıflandırmasına göre genel olarak her grupta yüksek sonuçlar elde edildi. Ulaşılan bu sonucun sebebi; iki kötü huylu kitleyi birbirinden ayırmanın; iyi ve kötü huylu kitleleri birbirinden ayırmaktan daha zor olmasıdır.

Endometrioid-Seröz-Myom üçlü sınıflandırmasında da; sonuçlar genel olarak yüksek çıktı. Bu sınıflandırma için en iyi sonucu mutual information yöntemiyle seçilen özniteliklerle yapılan sınıflandırmada CatBoost algoritması verdi. Yapılan bu üçlü sınıflandırmada Seröz grubuna ait olan verilerin daha yüksek doğrulukla tahmin edildiği görüldü.

Yapılan sınıflandırmalara bakıldığında mutual information öznitelik seçme yöntemleri içerisinde genel olarak en iyi sonuçları veren yöntem oldu. MRMR yöntemiyle seçilen özniteliklerle yapılan sınıflandırma sonuçları ise bütün sınıflandırmalar için en kötü sonuçları verdi. Kullanılan sınıflandırıcılar içinde CatBoost algoritması en başarılı sonuç veren algoritmalarından oldu. Yapılan deneysel çalışmaların; ileride yapılacak olan çalışmalara yardımcı olabileceği düşünülmektedir.

## KAYNAKLAR

- [1] Renehan A.G., Tyson M., Egger M., Heller R., Body-Mass Index and Incidence of Cancer: A Systematic Review and Meta-analysis of Prospective Observational Studies, *The Lancet*, 2008, **371** (9612), 569–578.
- [2] P.Ferlay J., Colombet M., Soerjomataram I., Mathers C., Parkin M., Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods, *Int J Cancer*, 2019, **144**(8),1941–1953.
- [3] Sheikh M., Althouse A., Freese K., Soisson S., Edwards R., Endometrial Cancer Projections to 2030: Should We Be Concerned?, *Future Oncology*,2014, **10**(16),2561–2568.
- [4] D. Keleş, Endometriyum Kanserinde Manyetik Rezonans Görüntüleme ile Kantitatif Değerlendirmenin Prognostik Öneminin Araştırılması, Tıpta Uzmanlık Tezi, Süleyman Demirel Üniversitesi, Tıp Fakültesi, Isparta, 2020, 621399.
- [5] Amant F., Mirza M., Koskas M, Cancer of the Corpus Uteri, *Int J Gynecol Obs.*, 2018, **143**(2), 37–50.
- [6] Piulats J., Guerra E., Gil-Martín M., Roman-Canal B., Gatiús S., Molecular Approaches for Classifying Endometrial Carcinoma,” *Gynecol Oncol.*, 2017., **145**(1), 200–207.
- [7] Davnall F., Yip C., Ljungqvist G., Assessment of Tumor Heterogeneity: An Emerging Imaging Tool for Clinical Practice?, *Insights Imaging*, DOI: 10.1007/s13244-012-0196-6.
- [8] Rose C., Mills S., Quantifying Spatial Heterogeneity in Dynamic Contrast-Enhanced MRI Parameter Maps, *Magn. Reson. Med.*,2009, **62** (2), 488–499.
- [9] Ganeshan B., Goh V., Mandeville H., Non-small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT, *Radiology*, 2013, **266** (1), 326–336.
- [10] Eun E., Kang D., Texture Analysis with 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer, *Radiology*,2020, **294** (1), 31–41.
- [11] Wibmer A., Hricak H., Haralick Texture Analysis of Prostate MRI: Utility for Differentiating Non-Cancerous Prostate From Prostate Cancer and Differentiating Prostate Cancers with Different Gleason Scores, *Eur. Radiol.*, 2015, **25** (10), 2840–2850.

- [12] Ueno Y., Forghani B., Endometrial Carcinoma: MR Imaging-Based Texture Model for Preoperative Risk Stratification - A Preliminary Analysis, *Radiology*, 2017, **284** (3), 748–757.
- [13] Bernardi E., Radiomics of the Primary Tumour as a Tool to Improve 18F-FDG-PET Sensitivity in Detecting Nodal Metastases in Endometrial Cancer, *EJNMMI Res.*, DOI: 10.1186/s13550-018-0441-1.
- [14] Asedullayeva M., Küçük Hücreli Akciğer Kanseri ile Küçük Hücreli Dışı Akciğer Kanserinin Ayırıcı Tanısında Kontrastlı BT Görüntülemeye Radiomics ve Makine Öğrenmesinin Katkısı, Tıpta Uzmanlık Tezi, Kocaeli Üniversitesi, Tıp Fakültesi, Kocaeli, 2020, 635348.
- [15] Taşlıçay C., Konvansiyonel Manyetik Rezonans Görüntülemenin Ve Radiomiks Tabanlı Makina Öğrenmesinin Menenjiom Hastalarında Preoperatif Grade, Histolojik Subtip Ve Surveyansı Öngörmeye Katkısı, Tıpta Uzmanlık Tezi, Kocaeli Üniversitesi, Tıp Fakültesi, Kocaeli, 2020, 635722.
- [16] Dong C., Dong K., Yu H., Using Deep Learning with Convolutional Neural Network Approach to Identify the Invasion Depth of Endometrial Cancer in Myometrium Using MR Images: A Pilot Study, *Int. J. Environ. Res. Public Health*, 2020, **17** (16), 1–18.
- [17] Akazawa M., Hashimoto K., Noda K., The Application of Machine Learning for Predicting Recurrence in Patients with Early-Stage Endometrial Cancer : A Pilot Study, *Obstetrics & Gynecology Sci.*, DOI: 10.5468/ogs.20248.
- [18] Arnaldo S., Deep Myometrial Infiltration of Endometrial Cancer on MRI: A Radiomics-Powered Machine Learning Pilot Study, *Acad. Radiol.*, DOI: 10.1016/j.acra.2020.02.028.
- [19] Kurman R., Cargangiu M., Herrington C., *WHO Classification of Tumours of Female Reproductive Organs*, 4th ed. Lyon, 2014.
- [20] Fanning J., Evans M., Endometrial Adenocarcinoma Histologic Subtypes: Clinical and Pathologic Profile, *Gynecol. Oncol.*, 1989, **32** (3), 288–291.
- [21] Morice P., Leary A., Endometrial Cancer, *The Lancet*, 2016, **387**(10023), 1094–1108.
- [22] Van Mieghem I., Papillary Serous Adenocarcinoma of the Endometrium: CT–Pathologic Correlation, *Am. J. Roentgenol.*, 2004, **182** (6), 1534–1536.
- [23] Topçu E., Myoma Uteri Tanısı İle Opere Edilen Hastalarda Histopatolojik İncelemenin Uterin Sarkom, Stump Ve Benign Myoma Varyantları Açısından Araştırılması, Tıpta Uzmanlık Tezi, Sağlık Bilimleri Üniversitesi, Tıp Fakültesi, İstanbul, 2019, 608602.
- [24] Okolo S., Incidence, Aetiology and Epidemiology of Uterine Fibroids, *Best Practice and Research: Clinical Obstetrics and Gynaecology*, DOI:10.1016/j.bpobgyn.2008.04.002.

- [25] Evans P., Brunzell S., Uterine Fibroid Tumors: Diagnosis and Treatment, *American Family Physician*, 2007, **75** (10), 1503,1508.
- [26] Myom Nedir?, İstanbul Başkent Üniversitesi, <https://istanbulbaskentuniversitesi.com/saglik-rehberi/kadin-hastaliklari-ve-dogum/myom-nedir>,(Ziyaret Tarihi :03 Ocak 2021).
- [27] Lambin P., Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis, *Eur J Cancer*, 2012, **48**(4), 6188,441-446.
- [28] Afshar P., Mohammadi A., Plataniotis K., From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities, *IEEE Signal Process. Mag.*, 2019, **36** (4), 132–160.
- [29] Berenguer R., Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters, *Radiology*, 2018, **288** (2), 407–415.
- [30] Gillies R., Kinahan P., Hricak H., Radiomics: Images Are More Than Pictures, They Are Data, *Radiology*, 2016, **278** (2), 563–577.
- [31] Koçak B., Durmaz E., Radiomics with Artificial Intelligence: A Practical Guide for Beginners, *Diagnostic Interv. Radiol.*, 2019, **25** (6), 485–495.
- [32] Pieper S., Halle M., 3D Slicer, *2004 2nd IEEE Int. Symp. Biomed. Imaging Macro to Nano*, 2004, **1** (1), 632–635.
- [33] Albrechtsen F., Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices, Image Processing Laboratory, Department of Informaticss, University of Oslo, November 5, 2008.
- [34] Lumber G., Smith D., CT Texture Analysis: Definitions, Applications, Biologic Correlates and Challenges, *RadioGraphics*, DOI: 10.1148/rg.2017170056.
- [35] Larroza A., Bodí V., Moratal D., Texture analysis in magnetic resonance imaging: review and considerations for future applications, *In Assessment of Cellular and Organ Function and Dysfunction Using Direct and Derived MRI Methodologies*, InTech. Rijeka, DOI:10.5772/64641.
- [36] Thibault G., Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification, *Pattern Recognit. Inf. Process.*, November, 2009.
- [37] Peper J., Dahl R., Radiomics: A New Application From Established Techniques, *Expert Rev. Precis. Med. Drug Dev.*, 2016, **1** (2), 207–226.
- [38] Griethuysen J., Computational Radiomics System to Decode the Radiographic Phenotype, *Cancer Res.*, 2017, **77** (21), e104–e107.
- [39] Deo R., Machine Learning in Medicine, *Circulation*, 2015, **132** (20), 1920–1930.

- [40] Erdursun M., Kronik Böbrek Hastalığının Makine Öğrenmesi Yöntemleriyle Analizi, Yüksek Lisans Tezi, Kırıkkale Üniversitesi, Fen Bilimleri Enstitüsü, Kırıkkale, 2019,554000.
- [41] Chandna D., Diagnosis of Heart Disease Using Data Mining Algorithm, *IJCSIT*, 2014, **5** (2), 1678-1680.
- [42] Baran M., Makine Öğrenmesi Yöntemleriyle Çoklu Etiketli Verilerin Sınıflandırılması, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas, 2020, 619685.
- [43] Haltuf M., Support Vector Machines for Credit Scoring, Michal Haltuf, <https://svm.michalhaltuf.cz/>, (Ziyaret Tarihi: 03 Ocak 2021).
- [44] Osuna E., Freund R., Girosi F., Improved Training Algorithm for Support Vector Machines, *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, DOI: 10.1109/nnspp.1997.622408.
- [45] Canbay Y., Diyabet Verilerinin Dvm Kullanılarak Sınıflandırılması, Yüksek Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 2013, 335394.
- [46] Singh A., Thakur N., A Review of Supervised Machine Learning Algorithms, *3rd Int. Conf. Comput. Sustain. Glob. Dev. INDIACom*, 2016, 1310–1315.
- [47] Osisanwo F., Akinsola J., Supervised Machine Learning Algorithms: Classification and Comparison, *Int. J. Comput. Trends Technol.*, 2017, **48** (3), 128–138.
- [48] Breiman L., Random Forests, *Stat. Dep. Univ. Calif. Berkeley*, DOI :10.1023/A:1010933404324.
- [49] ÖZTEMEL E., *Yapay Sinir Ağları*, 2, Papatya Basımevi, İstanbul, 2006.
- [50] Friedman J., Greedy Function Approximation: A Gradient Boosting Machine, *Ann. Stat.*, 2001, **29**(5), 1189–1232.
- [51] Chen T., XGBoost: A Scalable Tree Boosting System, *KDD '16 Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, DOI :10.1145/2939672.2939785.
- [52] Ke G., LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Adv. Neural Inf. Process. Syst.*, 2017, 3147–3155.
- [53] Sun X., Liu M., A Novel Cryptocurrency Price Trend Forecasting Model Based on LightGBM, *Financ. Res. Lett.*, DOI:10.1016/j.frl.2018.12.032.
- [54] Fan J., Ma X., Wu L., Light Gradient Boosting Machine: An Efficient Soft Computing Model for Estimating Daily Reference Evapotranspiration with Local and External Meteorological Data, *Agric. Water Manag.*, DOI: /10.1016/j.agwat.2019.105758.



- [55] Dorogush A., Ershov V., CatBoost: Gradient Boosting with Categorical Features Support, 2018, arXiv:1810.11363.
- [56] Balaban E., Samur O., Boosting Algoritmaları , Data Science For The Earth, <https://www.datascienceearth.com/boosting-algoritmaları/>, (Ziyaret Tarihi: 03 Ocak 2021).
- [57] Liu H., Li J., A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns., *Genome Inform.*, 2002, **13** (January), 51–60.
- [58] BUDAK H., Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım, *Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg.*, DOI: 10.19113/sdufbed.01653.
- [59] Battiti R., Using Mutual Information for Selecting Features in Supervised Neural Net Learning, *IEEE Trans. Neural Networks*, 1994, **5**(4), 537–550.
- [60] Kule A., Feature Selection Using Different Mutual Information Estimation Methods, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen ve Teknoloji Enstitüsü, İstanbul, 2010, 295247.
- [61] Witten H., Frank E., Hall M., Pal C., *Data Mining: Practical Machine Learning Tools and Techniques*, 4 th, 2016.
- [62] ÇELİK C., BİLGE H., Ağırlıklandırılmış Koşullu Karşılıklı Bilgi ile Öznitelik Seçimi, *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Derg.*, 2015, **30** (4), 585–596.
- [63] Peng H., Long F., Ding C., Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27** (8), 1226–1238.
- [64] Ding C., Peng H., Minimum Redundancy Feature Selection from Microarray Gene Expression Data, *Proceedings of the 2003 IEEE Bioinformatics Conference*, 2003, 523–528.
- [65] Gülgezen G., Kararlı ve Başarımı Yüksek Öznitelik Seçimi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2009, 251774.
- [66] Kira K., Rendell L., The Feature Selection Problem: Traditional Methods and a New Algorithm, *Proceedings Ninth National Conference on Artificial Intelligence*, 1992, 129–134.
- [67] Kononenko I., Estimating Attributes: Analysis and Extensions of Relief, *Machine Learning: ECML-94*, 1994, 171–182.
- [68] Yang F., Cheng W., Dou R., An Improved Feature Selection Approach Based on ReliefF and Mutual Information, 2011, *Int. Conf. Inf. Sci. Technol.*, 246–250.

- [69] Wah Y., Ibrahim N., Hamid H., Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy, *Pertanika J. Sci. Technol.*, 2018, **26** (1), 329–340.
- [70] Ladha L., Deepa T., Feature Selection Methods And Algorithms, *Int. J. Comput. Sci. Eng.*, 2011, **3** (5), 1787–1797.
- [71] Stehman S., Selecting and Interpreting Measures of Thematic Classification Accuracy, *Remote Sens. Environ.*, 1997, **62** (1), 77–89.
- [72] Alan A., Karabatak M., Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi, *Fırat Üniversitesi Mühendislik Bilim. Derg.*, 2020, **32** (2), 531–540.
- [73] Akosa J., Predictive Accuracy: A Misleading Performance Measure for Highly Imbalanced Data, Oklahoma State University, 2017.
- [74] Genomic Data Commons Data Portal, <https://portal.gdc.cancer.gov/>. (Ziyaret Tarihi: 04 Ocak 2021).
- [75] Clark K., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, *J. Digit. Imaging*, 2013, **26** (6), 1045–1057.
- [76] TCGA-UCEC - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki, <https://wiki.cancerimagingarchive.net/display/Public/TCGA-UCEC> . (Ziyaret Tarihi: 04 Ocak 2021).
- [77] CPTAC-UCEC, [https://proteomics.cancer.gov/news\\_and\\_announcements/cptac-proteomics-data-ucsc-genome-browser](https://proteomics.cancer.gov/news_and_announcements/cptac-proteomics-data-ucsc-genome-browser). (Ziyaret Tarihi: 04 Ocak 2021)
- [78] The Cancer Genome Atlas Program - National Cancer Institute, <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>. (Ziyaret Tarihi: 04 Ocak 2021).
- [79] Erickson B. J., Mutch D., Radiology Data from The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma (TCGA-UCEC) collection. The Cancer Imaging Archive. DOI:10.7937/K9/TCIA.2016.GKJ0ZWAC.
- [80] 3D Slicer, <https://www.slicer.org/>. (Ziyaret Tarihi: 04 Ocak 2021)
- [81] Radiomics , <https://www.radiomics.io/pyradiomics.html>. (Ziyaret Tarihi: 04 Ocak 2021)
- [82] Olson R., ReliefF, PyPI, <https://pypi.org/project/ReliefF/>. (Ziyaret Tarihi: 04 Ocak 2021)
- [83] Raschka S., Mlxtend, PyPI, <https://pypi.org/project/mlxtend/>. (Ziyaret Tarihi: 04 Ocak 2021)

## KİŞİSEL YAYIN VE ESERLER

- [1] **Yigit F**, İner B.,Çakır Ö., Akciğer CT Doku Analizi Verilerinde Makine Öğrenmesi ile Covid-19 Tespiti, *Imascon*, Kocaeli, Türkiye, 4-5 Aralık 2020.



## ÖZGEÇMİŞ

İlköğretimi sırasıyla İnönü İlköğretim ve Cumhuriyet İlköğretim Okullarında okuduktan sonra, 2008 yılında kazandığı Adıyaman Fatih Anadolu Lisesi'nden 2012 yılında mezun oldu. 2013 yılında yerleştiği Kocaeli Üniversitesi Bilgisayar Mühendisliği Bölümü'nü, 2018 yılında, Bilgisayar Mühendisi olarak tamamladı. Şubat 2019'da Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda yüksek lisans eğitimine başladı ve eğitimine hala devam ediyor.

