

**KOCAELİ ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ**

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

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
MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
YÜKSEK LİSANS TEZİ**

**ENDOMETRİAL KANSERİN AMELİYAT ÖNCESİ BT
GÖRÜNTÜLERİNDE ELDE EDİLEN RADIOMİCS 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 3.6.	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 sonuçları	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 adımıyla 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.

