

**T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**AÇIK KAYNAKLI KLİNİK KANSER VERİLERİNİN R-SHINY
UYGULAMASI İLE YAPAY ZEKA TABANLI WEB ARAYÜZÜ
DESTEK SİSTEMİNİN GELİŞTİRİLMESİ**

Hüseyin Koray MISIRLIOĞLU

0000-0002-3473-5721

ONKOLOJİ ANABİLİM DALI

Translasyonel Onkoloji

YÜKSEK LİSANS TEZİ

İZMİR

HAZİRAN 2021

TEZ KODU: DEU.HSI.MSc-2018970114

**T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**AÇIK KAYNAKLI KLİNİK KANSER VERİLERİNİN R-SHINY
UYGULAMASI İLE YAPAY ZEKA TABANLI WEB ARAYÜZÜ
DESTEK SİSTEMİNİN GELİŞTİRİLMESİ**

Hüseyin Koray MISIRLIOĞLU

0000-0002-3473-5721

ONKOLOJİ ANABİLİM DALI

Translasyonel Onkoloji

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Doç. Dr. Gizem ÇALIBAŞI KOÇAL

0000-0002-3201-4752

İZMİR

HAZİRAN 2021

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE
ETİK BEYANI

Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırlayıp sunduğum “*Açık Kaynaklı Klinik Kansere Verilerinin R-Shiny Uygulaması ile Yapay Zeka Tabanlı Web Arayüzü Destek Sisteminin Geliştirilmesi*” başlıklı Yüksek Lisans tezinde elde ettiğim verileri, bilgileri, belgeleri akademik ve etik kurallar çerçevesinde elde ettiğimi, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tezimde yararlandığım eserlere bilimsel kurallara uygun atıfta bulunarak kaynak gösterdiğimi, tezimin özgün olduğunu, tezimin çalışma ve yazımında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Hüseyin Koray MISIRLIOĞLU

İÇİNDEKİLER

İÇİNDEKİLER	i
TABLolar DİZİNİ	v
ŞEKİLLER DİZİNİ	vi
KISALTMALAR.....	viii
TEŞEKKÜR.....	xi
ÖZET.....	1
ABSTRACT	3
1. GİRİŞ VE AMAÇ	5
1.1. PROBLEMİN TANIMI VE ÖNEMİ	5
2. GENEL BİLGİLER.....	7
2.1. KARAR DESTEK SİSTEMLERİ	7
2.1.1. Klinik Karar Destek Sistemleri.....	8
2.2. YAPAY ZEKA	12
2.2.1. Klinik Karar Destek Sistemlerinde Yapay Zeka Uygulamaları	14
2.2.2. Kanser Kliniğinde Yapay Zeka Temelli Klinik Karar Destek Sistemleri	15
2.3. MAKİNE ÖĞRENMESİ	18
2.3.1. Makine Öğrenmesi Çeşitleri	18
2.3.1.1. Etiketlere Göre Makine Öğrenmesi Çeşitleri	19
Denetimli Öğrenme.....	19
Denetimsiz Öğrenme.....	20
Pekiştirmeli Öğrenme	20
2.3.1.2. Özellikler Temelinde Makine Öğrenmesi Çeşitleri.....	20
El Yapımı Özellik Tabanlı Makine Öğrenimi Teknikleri	20
El Yapımı Olmayan Özellik Tabanlı Makine Öğrenimi Teknikleri ..	21
2.3.2. Makine Öğrenmesi Süreci	21
2.3.2.1. Öğrenme Probleminin Tanımı	21
2.3.2.2. Ham Veri Analizi	21
2.3.2.3. Veri Hazırlama	22
Kayıp Değerlerin İşlenmesi	22
Normalizasyon	23

<i>Uç Noktaların Düzeltilmesi</i>	24
2.3.2.4. <i>Algoritma Değerlendirme</i>	24
<i>C4,5 Karar Ağacı Algoritması</i>	24
<i>Naive Bayes Sınıflandırıcı Algoritması</i>	26
<i>Rastgele Orman Algoritması</i>	27
<i>Destek Vektör Makineleri</i>	28
2.3.2.5. <i>Sonuçların Geliştirilmesi</i>	30
<i>Karışıklık Matrisi</i>	32
<i>Alıcı İşlem Karakteristik Eğrisi Analizi</i>	34
2.3.2.6. <i>Sonuçların Sunumu</i>	36
2.3.3. <i>Derin Öğrenme</i>	37
2.4. KANSER	38
2.4.1. <i>Kanser Epidemiyolojisi</i>	38
2.4.2. <i>Kolorektal Kanser</i>	40
2.4.2.1. <i>Kolorektal Kanser Epidemiyolojisi</i>	40
2.4.2.2. <i>Kolorektal Kanser Risk Faktörleri</i>	40
<i>Değiştirilebilir Risk Faktörleri</i>	41
<i>Değiştirilemez Risk Faktörleri</i>	41
<i>Diğer Faktörler</i>	41
2.4.2.3. <i>Kolorektal Kanser Patogenezi</i>	42
<i>Sağ ve Sol Kolorektal Kanseri</i>	43
<i>Konsensüs Moleküler Alt Tipleri</i>	44
2.4.3. <i>NCI GDC Kanser Veri Portalı</i>	45
2.4.4. <i>AJCC TNM Evreleme Sistemi</i>	45
2.4.4.1. <i>T Evresi: Birincil Tümör</i>	46
2.4.4.2. <i>N Evresi: Bölgesel Lenf Düğümü Metastazı</i>	47
2.4.4.3. <i>M Evresi: Uzak Metastaz</i>	48
3. GEREÇ VE YÖNTEM	49
3.1. <i>ARAŞTIRMANIN TİPİ</i>	49
3.2. <i>ARAŞTIRMANIN YERİ VE ZAMANI</i>	49
3.3. <i>ARAŞTIRMANIN EVRENİ VE ÖRNEKLEMİ/ ÇALIŞMA GRUPLARI</i>	49

3.4. ÇALIŞMA MATERYALİ.....	49
3.5. ARAŞTIRMANIN DEĞİŞKENLERİ.....	49
3.6. VERİ TOPLAMA ARAÇLARI	50
3.6.1. TCGAbiolinks	50
3.7. ARAŞTIRMA PLANI VE TAKVİMİ	50
3.8. VERİLERİN DEĞERLENDİRİLMESİ.....	51
3.8.1. İstatistiksel Değerlendirme	51
3.8.1.1. Klinik Verinin Analizi	52
Demografik Verinin Analizi.....	52
Kaplan-Meier Sağkalım Analizi	52
3.8.2. Biyoinformatik Analiz	52
3.8.2.1. Gen Zenginleştirme Analizi	53
3.8.2.2. Makine Öğrenmesi.....	53
Algoritma Değerlendirme.....	54
Sonuçların Geliştirilmesi.....	55
3.8.3. Klinik Karar Destek Aracı: SurvCOCA Shiny Web Arayüzü.....	55
3.9. ARAŞTIRMANIN SINIRLILIKLARI.....	55
3.10. ETİK KURUL.....	56
4. BULGULAR.....	57
4.1. KLİNİK VERİLERİN ANALİZ SONUÇLARI.....	57
4.1.1. Demografik Verilerin Analiz Sonuçları.....	57
4.1.2. Kaplan-Meier Sağkalım Analiz Sonuçları.....	67
4.1.2.1. Cinsiyete Göre Sağkalım	67
4.1.2.2. Beden Kitle İndeksine Göre Sağkalım	68
4.1.2.3. Daha Önce Kansere Teşhisi Alma Durumuna Göre Sağkalım ..	69
4.1.2.4. Birincil Teşhise Göre Sağkalım	70
4.1.2.5. Rezeksiyon veya Biyopsi Bölgesine Göre Sağkalım	71
4.1.2.6. Tümör Evresine Göre Sağkalım.....	72
4.1.2.7. T Evresine Göre Sağkalım	73
4.1.2.8. N Evresine Göre Sağkalım.....	74
4.1.2.9. M Evresine Göre Sağkalım.....	75
4.2. GEN EKSPRESYON VERİLERİNİN ANALİZ SONUÇLARI	76

4.2.1. Ekspresyon Verilerinin Seçimi	76
4.2.1.1. KEGG Veri Tabanı	76
4.2.1.2. MSigDB Cancer Hallmark Gen Seti.....	76
4.3. MAKİNE ÖĞRENMESİ ALGORİTMA SONUÇLARI	77
4.3.1. Klinik ve Genetik Verilerle Oluşturulan Veri Setlerinin Performans Sonuçları	77
4.3.1.1. Klinik Parametrelerin Performans Sonuçları	77
4.3.1.2. Klinik Parametreler ile Glikoliz- Glukoneogenez Yolağındaki Genlerin Ortak Performans Sonuçları.....	77
4.3.1.3. Klinik Parametreler ile Glikan Degredasyon Yolağındaki Genlerin Ortak Performans Sonuçları.....	78
4.3.1.4. Klinik Parametreler ile Pantotenat ve CoA Biyosentezi Yolağındaki Genlerin Ortak Performans Sonuçları	78
4.3.1.5. Klinik Parametreler ile Apoptoz Yolağındaki Genlerin Ortak Performans Sonuçları	79
4.3.1.6. Klinik Parametreler ile mTORC1 Sinyal Yolağındaki Genlerin Ortak Performans Sonuçları.....	79
4.3.1.7. Klinik Parametreler ile Tüm Yolaklardaki Genlerin Ortak Performans Sonuçları	80
4.4. SAĞKALIM TAHMİNLEYEN KLİNİK KARAR DESTEK ARACI: SurvCOCA	80
5. TARTIŞMA	82
6. SONUÇ VE ÖNERİLER.....	87
7. KAYNAKLAR.....	88
8. EKLER.....	108
EK 1: ETİK KURUL RAPORU	108
EK 2: ÖZGEÇMİŞ.....	110

TABLÖLAR DİZİNİ

Sayfa No

Tablo 1. Karışıklık matrisi.....	32
Tablo 2. AJCC kolorektal kanser TNM sınıflandırması	46
Tablo 3. Araştırma planı ve takvimi.....	51
Tablo 4. Cinsiyete göre olgularda yaşam süresi.....	67
Tablo 5. Beden kitle indeksine göre olgularda yaşam süresi	68
Tablo 6. Daha önce kanser teşhisi alma durumlarına göre olgularda yaşam süresi ..	69
Tablo 7. Birincil teşhise göre olgularda yaşam süresi	70
Tablo 8. Rezeksiyon veya biyopsi bölgesine göre olgularda yaşam süresi.....	71
Tablo 9. Tümör evresine göre olgularda yaşam süresi.....	72
Tablo 10. T evresine göre olgularda yaşam süresi	73
Tablo 11. N evresine göre olgularda yaşam süresi.....	74
Tablo 12. M evresine göre olgularda yaşam süresi	75
Tablo 13. Gen zenginleştirme analizi (KEGG)	76
Tablo 14. Gen zenginleştirme analizi (Cancer Hallmark).....	76
Tablo 15. Klinik parametrelerin performans sonuçları	77
Tablo 16. Klinik parametreler ile glikoliz- glukoneogenez yolağındaki genlerin ortak performans sonuçları.....	77
Tablo 17. Klinik parametreler ile glikan degradasyon yolağındaki genlerin ortak performans sonuçları.....	78
Tablo 18. Klinik parametreler ile pantotenat ve CoA biyosentezi yolağındaki genlerin ortak performans sonuçları.....	78
Tablo 19. Klinik parametreler ile apoptoz yolağındaki genlerin ortak performans sonuçları	79
Tablo 20. Klinik parametreler ile mTORC1 sinyal yolağındaki genlerin ortak performans sonuçları.....	79
Tablo 21. Klinik Parametreler ile Tüm Yolaklardaki Genlerin Ortak Performans Sonuçları	80

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1. Makine öğrenmesi algoritma çeşitleri	19
Şekil 2. Karar ağacı algoritmaları şematik gösterim.....	25
Şekil 3. Bayes algoritmaları şematik gösterim	26
Şekil 4. Topluluk algoritmaları şematik gösterim.....	28
Şekil 5. Örnek tabanlı algoritmalar şematik gösterim	29
Şekil 6. Lojistik regresyon eğrisi eşik seçimi	31
Şekil 7. ROC eğrilerinin karşılaştırılması ve AUC.....	36
Şekil 8. Sağ ve Sol Kolon Kanserleri arasındaki farklılıklar	44
Şekil 9. AJCC kolorektal kanser sınıflandırması- Birincil tümör (T evresi)	47
Şekil 10. Kolorektal kanser Japon sınıflaması kanser lenf düğümü haritası	48
Şekil 11. Makine öğrenmesi süreci.....	54
Şekil 12. Olguların cinsiyete göre dağılımları	57
Şekil 13. Olguların ırka göre dağılımları	58
Şekil 14. Olguların beden kitle indeksi durumlarına göre dağılımları.....	58
Şekil 15. Olguların daha önce kanser teşhisi alma durumlarına göre dağılımları	59
Şekil 16. Olguların birincil teşhis durumlarına göre dağılımları	60
Şekil 17. Olguların gruplandırılmış birincil teşhis durumlarına göre dağılımları	60
Şekil 18. Olguların doku veya organ menşei durumlarına göre dağılımları.....	61
Şekil 19. Kolonun bölümleri.....	62
Şekil 20. Olguların rezeksiyon veya biyopsi bölgesi durumlarına göre dağılımları..	62
Şekil 21. Olguların tümör evresine göre dağılımları	63
Şekil 22. Olguların T evresine göre dağılımları.....	64
Şekil 23. Olguların gruplandırılmış T evresine göre dağılımları.....	64
Şekil 24. Olguların N evresine göre dağılımları	65
Şekil 25. Olguların M evresine göre dağılımları	66
Şekil 26. Olguların sağkalım durumlarına göre dağılımları	66
Şekil 27. Cinsiyete göre sağkalım.....	67
Şekil 28. Beden kitle indeksine göre sağkalım	68

Şekil 29. Daha önce kanser teşhisi alma durumuna göre sağkalım	69
Şekil 30. Birincil teşhise göre sağkalım.....	70
Şekil 31. Rezeksiyon veya biyopsi bölgesine göre sağkalım	71
Şekil 32. Tümör evresine göre sağkalım.....	72
Şekil 33. T evresine göre sağkalım.....	73
Şekil 34. N evresine göre sağkalım.....	74
Şekil 35. M evresine göre sağkalım.....	75
Şekil 36. Sağkalım tahminleyen klinik karar destek aracı: SurvCOCA Shiny web arayüzü parametre ayar ekranı	81
Şekil 37. Sağkalım tahminleyen klinik karar destek aracı: SurvCOCA Shiny web arayüzü karar çıktısı ekranı	81

KISALTMALAR

ACC	: Doğruluk, Accuracy
ADH1C	: Alkol Dehidrojenaz-1C (Sınıf I), Gama Polipeptidi, Alcohol Dehydrogenase 1C (Class I), Gamma Polypeptide
AI	: Yapay Zeka, Artificial Intelligence
AJCC	: Amerikan Kanser Komitesi, American Joint Committee on Cancer
AKR1A1	: Aldo-Keto Redüktaz Ailesi-1 Üye-A1, Aldo-Keto Reductase Family1 Member-A1
ANSI	: Amerikan Ulusal Standartlar Enstitüsü, American National Standards Institute
APC	: WNT Sinyal Yolunun APC Düzenleyicisi, APC Regulator Of WNT Signaling Pathway
ASCO	: Amerikan Klinik Onkoloji Topluluğu, American Society of Clinical Oncology
AUC	: ROC eğrisi altında kalan alan, Area under the ROC curve
BCAT1	: Dallı Zincirli Amino Asit Transaminaz-1, Branched Chain Amino Acid Transaminase-1
BKİ	: Beden Kitle İndeksi
BRAF	: B-Raf Proto-Onkogen, Serin / Treonin Kinaz, B-Raf Proto-Oncogene, Serine/Threonine Kinase
CAPN2	: Kalpain-2, Calpain-2
CASP2	: Kaspaz-2, Caspase-2
CCG	: Kanser Genomik Merkezi, Center for Cancer Genomics
CI	: Güven aralığı, Confidence interval
CIMP	: CpG adası metilatör fenotipi, CpG island methylator phenotype
CMS	: Konsensüs moleküler alt tip sınıflandırması, Consensus molecular subtype classification
CNN	: Evrişimli Sinir Ağı, Convolutional Neural Network
CoA	: Koenzim A
DSO	: Dünya Sağlık Örgütü
DVM	: Destek Vektör Makinesi, (Support Vector Machine)

EGFR	: Epidermal büyüme faktörü reseptörü, Epidermal Growth Factor Receptor
ERR	: Hata oranı, Error Rate
ESK	: Elektronik Sağlık Kayıtları
FHIR	: Hızlı Sağlık Hizmetleri Birlikte Çalışabilirlik Kaynakları, Fast Healthcare Interoperability Resources
FNR	: Yanlış negatif oranı, False Negative Rate
FPR	: Yanlış pozitif oranı, False Positive Rate
FS	: F puanı, F Score
GDC	: Ortak Genom Verileri, Genomic Data Commons
GKDS	: Grup Karar Destek Sistemleri
HL7	: Uluslararası Sağlık Seviyesi Yedi, Health Level Seven International
KA	: Karar Ağacı, Decision Tree
KDS	: Karar Destek Sistemi
KEGG	: Kyoto Genler ve Genom Ansiklopedisi, Kyoto Encyclopedia of Genes and Genomes
KKDS	: Klinik Karar Destek Sistemi
KRAS	: KRAS Proto-Onkogen, GTPaz, KRAS Proto-Oncogene, GTPase
LRA	: Lojistik Regresyon Analizi
MAN2B2	: Mannosidaz Alfa Sınıf-2B Üye-2, Mannosidase Alpha Class 2B Member-2
MSI	: Mikrosatellit instabilite, Microsatellite instability
MSigDB	: Moleküler İmza Veritabanı, Molecular Signatures Database
mTORC1	: Rapamisin kompleksi 1'in memeli/mekanik hedefi
N/A	: Uygulanabilir değil, Not Applicable veya mevcut değil, Not Available
NB	: Naive Bayes
NBC	: Naive Bayes Sınıflandırıcı
NCI	: Ulusal Kanser Enstitüsü, National Cancer Institute
NPV	: Negatif tahmin değeri, Negative Predictive Value
PFKM	: Fosfofruktokinaz, Kas, Phosphofructokinase, Muscle
PÖ	: Pekiştirmeli Öğrenme
PPV	: Pozitif tahmin değeri, Positive Predictive Value

RAS	: G-proteini türünde bir proto-onkogen
REC	: Duyarlılık, Sensitivity, Recall, True Positive Rate
RO	: Rastgele Orman, Random Forest
ROC	: Alıcı işlem karakteristiği, Receiver Operating Characteristic
SeD	: Sosyoekonomik Durum
Shiny	: shinyapps.io by RStudio
SPC	: Özgüllük, Specificity, True Negative Rate
SPSS	: Sosyal Bilimler için İstatistik Paketi
TCEA1	:Transkripsiyon Uzama Faktörü-A1, Transcription Elongation FactorA1
TCGA	: Kanser Genom Atlası, The Cancer Genome Atlas
TCGA-COAD	: TCGA Kolon adenokarsinom projesi
TOMM40	: Dış Mitokondriyal Membranın Translokası-40, Translocase Of Outer Mitochondrial Membrane-40
TP53	: Tümör Proteini P53
UPA	: Uygulama Programlama Arayüzü
UV	: Ultraviyole
VEGF	: Vasküler Endotelyal Büyüme Faktörü, Vascular Endothelial Growth Factor
WFO	: “Watson for Oncology”
WKDS	: Web Tabanlı Karar Destek Sistemleri

TEŞEKKÜR

Çalışmalarımız için tüm araştırmacılara uygun ortamı sağlayan Onkoloji Enstitüsü Müdürümüz Sn. Prof. Dr. Nur OLGUN ve Sağlık Bilimleri Enstitüsü Müdürümüz Sn. Prof. Dr. Hülya ELLİDOKUZ'a,

Tez projemin planlanması, yürütülmesi ve yazım aşamasında bilimsel görüşlerini esirgemeyip, eğitim süremde doğruyu öğrenmem için bana değerli vakitlerini ayıran Translasyonel Onkoloji Anabilim Dalı Başkanımız Sn. Prof. Dr. Yasemin BAŞBINAR ve danışman hocam Sn. Doç. Dr. Gizem ÇALIBAŞI KOÇAL'a,

Yürütücüsü olduğu projeden kaynak yaratarak bizlere lisansüstü eğitimimizde burs desteği sağlayan Sn. Doç. Dr. Zerrin IŞIK'a,

Tezimin en başından en sonuna kadar her aşamada her türlü konuda fikrine danıştığım, sorularımı sabırla cevaplayarak bilgisini paylaşan, yardımını ve büyük desteğini hep yanımda hissettiğim sevgili Asım LEBLEBİCİ'ye,

Eğitim sürem boyunca aynı ortamı paylaştığımız, benzer sıkıntılarla karşı karşıya kalıp birlikte öğrendiğimiz ve dayanıştığımız tüm çalışma arkadaşlarıma,

Lisansüstü eğitimime başlamama vesile olan Fatma ÖZDEN'e,

Aldığım her kararda yanımda olan, kardeşim Eray MISIRLIOĞLU, babam Mehmet MISIRLIOĞLU'na, koşulsuz sevgi ve fedakarlığı ile desteğini her zaman yanımda hissettiğim biricik melek annem F. Özden BİLDİRİCİ'ye, sevgiyle bağlılık içerisinde bulunan cânım geniş ailemin tüm üyelerine ve birlikte büyüdüğümüz sevgili dostlarım B.Ö., E.D.Ö., K.Ş., M.Ç., O.G., M.O.K., Y.Ç., ve Y.Ö'a,

En derin duygularla teşekkürlerimi sunarım.

Tezimi kanser hastalığından vefat eden ve her zaman bilim insanı olmamı isteyen rahmetli dedem Emekli Nüfus Md. Mustafa BİLDİRİCİ ile ben bu satırları yazarken kanser hastalığı ile mücadele eden ve mezuniyetimi dört gözle bekleyen sevgili eşi anneannem Ayşe BİLDİRİCİ'ye ithaf ederim.

İzmir, Haziran, 2021

H. Koray MISIRLIOĞLU

AÇIK KAYNAKLI KLİNİK KANSER VERİLERİNİN R-SHINY UYGULAMASI İLE YAPAY ZEKA TABANLI WEB ARAYÜZÜ DESTEK SİSTEMİNİN GELİŞTİRİLMESİ

Yüksek Lisans Tezi

Hüseyin Koray MISIRLIOĞLU

DOKUZ EYLÜL ÜNİVERSİTESİ SAĞLIK BİLİMLERİ ENSTİTÜSÜ

Onkoloji Anabilim Dalı

ÖZET

Kanser hastalarında hastalığın prognozu, sınıflandırılması, tedavi stratejilerinin optimizasyonu ve klinik çalışmaların tasarımı için önemlidir. Bu çalışma, kolon kanseri tanısı alan hastalarda hastaya ait sağkalımın sınıflandırmaya dayalı makine öğrenmesi teknikleri kullanılarak tahminlenebilmesi amacıyla yapıldı.

Çalışmaya NCI'nın açık erişimli GDC portalındaki "TCGA-COAD" projesinden alınan 454 hastaya ait klinik ve genomik verileri dahil edildi. Klinik verilere uygulanan Kaplan-Meier sağkalım analizi sonucunda anlamlı çıkan parametrelerle gen ekspresyon verilerinin analizi sonucunda sağkalımla anlamlı ilişki içerisinde olduğu görülen dokuz farklı gen risk grubu olarak kullanıldı. Analizler R dilinde yazılan kodlarla gerçekleştirildi. R dilinde yazılan kodları geliştirme aracı olarak RStudio kullanıldı. Risk grupları sınıf etiketi olarak kullanılarak tahminler gerçekleştirildi. Seçilen klinik ve genomik parametreler; Naive Bayes, Destek Vektör Makinesi, Rastgele Orman ve C4.5 Karar Ağacı makine öğrenmesi algoritmalarıyla değerlendirildi. 10 kat çapraz doğrulama yöntemiyle test verisetleri oluşturularak model değerlendirmesi yapıldı. Tahminleme modeli Shiny aracılığıyla web arayüzü kullanılarak kamuya açık yapay zeka tabanlı klinik karar desteği sistemine dönüştürüldü.

Çalışmanın sonuçlarına göre; en doğru sağkalım tahmininde bulunan makine öğrenmesi algoritma modelinin Rastgele Orman algoritması olduğu, klinik ve genomik verilerin birlikte kullanılmasının algoritma performansını arttırdığı (%76,2) görüldü.

Onkoloji alanında prognoz ve saękalım sonucuna odaklanan yapay zeka temelli klinik karar destek sistemi modelleri; hasta yařam kalitesini ciddi anlamda etkileyebilecek kemoterapi tedavilerine bařvurulmadan önce tedavinin uygulanmaya deęer olup olmadıęı konusunda hekim, hasta ve hasta yakınları için bir öngörü olması bakımından önemlidir. Prognoz ve saękalım analizi yapabilen buna benzer sistemlerin geliştirilip iyileřtirilmesi hastalıęın yönetimini daha kolay hale getirecek ve hasta için daha doęru kararlar alınmasında yardımcı olacaktır.

Anahtar Sözcükler: klinik karar destek, kolon kanseri, makine öğrenmesi, saękalım analizi, sınıflandırma

Tezin sayfa adedi: 111 sayfa

Danışman: Doç. Dr. Gizem ÇALIBAŐI KOÇAL

DEVELOPMENT OF ARTIFICIAL INTELLIGENCE BASED WEB INTERFACE SUPPORT SYSTEM WITH R-SHINY APPLICATION FROM OPEN SOURCE CLINICAL CANCER DATA

Master Thesis

Huseyin Koray MISIRLIOGLU

DOKUZ EYLUL UNIVERSITY HEALTH SCIENCES INSTITUTE

Department of Oncology

ABSTRACT

Accurate prognosis of cancer patients is important for optimizing treatment strategies, and designing clinical trials. This study was conducted to estimate the survival of the patients diagnosed with colon cancer using machine learning techniques based on classification.

The study included clinical and genomic data of 454 colon cancer patients from the NCI “TCGA-COAD” project on the open access GDC portal. Clinical and genomic parameters that were found to be significantly associated with survival were used as risk groups. Analyzes were carried out with codes written in R language. Estimates were made using risk groups as class labels. Selected clinical and genomic parameters; was evaluated with four different algorithms. Model evaluation was performed by creating test data. The prediction model was transformed into AI-based CDSS (Clinical Decision Support System) using the web interface via Shiny.

According to the results of the study; It was seen that the machine learning algorithm that predicted the most accurate survival was the Random Forest. In addition, it was seen that the use of all clinical and genomic data together increased the algorithm performance (76.2%).

AI-based CDSS models that are focus on survival outcome in the field of oncology are important in terms of providing foresight for physicians and patients about whether it is worth applying the chemotherapy treatment which might seriously affect the patients. Developing and improving such systems will make the

management of the disease easier and help make more accurate decisions for the patient.

Keywords: classification, clinical decision support, colon cancer, machine learning, survival analysis

Page number: 111

Advisor: Assoc. Prof. Gizem CALIBASI KOCAL



1. GİRİŞ VE AMAÇ

1.1. PROBLEMİN TANIMI VE ÖNEMİ

Günümüzdeki teknoloji veri toplamasını ve depolanmasını çok kolay hale getirmiştir. Her alanda giderek artan devasa veri toplulukları büyük veri kavramını ortaya çıkarmıştır. Büyük verinin tek başına var olması hayatı kolaylaştıran uygulamalardaki problemlerin çözümü için yeterli değildir. Bu muazzam veri yığınları analizler yapıp işe yarayan çıkarımlar yapılmadan bir çöplükten ibarettir. Bu yüzden büyük verinin analiz edilmesi, birçok alanda olduğu gibi sağlık alanında da işe yarayan çıkarımların yapılarak fonksiyonel ürün ve teknolojilere dönüşüm için olmazsa olmazdır. Fakat bu noktada karşılaşılan sorun insan zihin gücünün ve zamanının bu karmaşık ve düzensiz devasa veri yığınlarını kategorize edip analizleyerek faydalı çıkarımlar yapmasına yetmeyecek kısıtlılıkta olmasıdır. Bu sebeple bu işlemlerin çeşitli yazılımlara öğretilerek bilgisayarlar tarafından yapılmasını sağlama ihtiyacı doğmuştur. Bu sayede de matematik, istatistik ve programlama disiplinlerinin bir arada kullanılması ile yapay zeka bilimi ortaya çıkmıştır. Yapay zeka tekniklerinin geliştirilmesi ile de büyük veri girdilerinin çıktılarını her seferinde değerlendirerek öğrenip çıkarım yapan ve sonuçlarını giderek iyileştiren “*makine öğrenmesi, machine learning*” kavramı ortaya çıkmıştır. Bilgisayar sistemleri yazılım ve donanımdan meydana gelse de buradaki makine kavramı daha çok tek başına yazılım bileşenlerini ifade etmektedir. Makine öğrenmesi en sade şekilde; bir bilgisayarın bir görevi yapmayı öğrenmesi olarak tanımlanabilir (1). Makine öğrenmesine sahip bir sistemin bir görevi her gerçekleştirmesinde bir öncekinden daha iyi performans sergilemesi gerekmektedir. Böyle bir sistemde öğrenme durumu bol miktarda veri ve algoritmalar ile gerçekleşmektedir. Sınıflandırma; bilinmeyen bir veriyi bilinen ve önceden tanımlanmış kategorilere yerleştirerek kullanılan makine öğrenmesinin temel uygulama alanlarından birisidir (2).

Sağlık bilimleri de makine öğrenmesi teknolojisinin faydalarından yararlanan birçok alandan bir tanesidir. Kanseri hastaların doğru prognozu, hastaların sınıflandırılması, tedavi stratejilerinin optimizasyonu ve klinik çalışmaların tasarımı için önemlidir. Hem klinik özellikler hem de moleküler veriler bu amaçla, örneğin belirli zaman noktalarında sansürlenmiş hastaların hayatta kalmasını tahmin etmek için

kullanılabilir. Genom apında gen ekspresyonu, metilasyon, protein ekspresyonu, kopya sayısı deęiřiklięi ve somatik mutasyon verileri dahil olmak üzere oklu omik verileri, kanser alıřmalarında giderek daha yaygın hale gelmektedir.

Veri madencilięi, büyük ve karmařık veriden anlamlı ve kurallı bilgiler elde edilmesi yani, bilgi keřfidir. Veri madencilięindeki problemlerin özümünde eřitli platformlar kullanılır. Veri madencilięinde kullanılan ücretli ve ücretsiz programlar vardır. Açık kaynak kodlu olması ve sunduęu topluluk desteęi sayesinde R programlama dili arařtırmacılar arasında yaygın olarak kullanılmaktadır (3). Prognoz ve saękalım analizi; kanser hastası olan bireylerin hastalık seyirlerinin nasıl ilerleyeceęini ve nasıl sonuçlanacaęını tahmin edip ön görmek üzere yapılan veri madencilięi esaslı bir veri analiz iřlemidir. Bu analiz sonucunda hastalık seyri iyiye ve kötüye gidecek olan hastalar belirlenir ve bu sayede daha etkin tedavi yöntemlerinin uygulanması konusunda hekim tarafından hızla karar alınması saęlanabilir. R programlama dili sürekli geliştirilen binlerce pakete sahiptir. Bu alıřmada, bu paketlerden biri olan Shiny kullanımı planlanmıřtır (4). Shiny, yazılan R kodlarının bir web arayüzü aracılıęıyla paylařılmasını saęlayan bir R paketidir (5). Bu sayede Shiny ile R programlama dilinde yazılan kodlar, HTML, CSS ya da JavaScript bilgisine gerek duymadan web arayüzlerinden kullanıcıların erişimine sunulabilmekte ve web destekli uygulamalar geliştirilerek uygulama dinamik bir yapıda web üzerinden kullanılabilmektedir (4,6,7).

Bu tez alıřmasında, açık kaynaklı klinik kanser verilerinin, hesaplamalı yaklařımlar sunan yapay zeka tabanlı modeller ile R- Shiny uygulamasının birlikte kullanımı sonucu oluřturulacak bir web arayüzü destek sisteminin hayatta kalma tahmini için kullanımı amaçlanmıřtır. Bu alıřmanın gerek hastalara ait veri setleri üzerinde yapılacak olması, analizin yerel bilgisayarda yapılması yerine erişim izni olan kullanıcılar tarafından web üzerinden erişilerek uzaktan yapılabilmesine imkan saęlaması ve onkoloji alanında lkemizde hasta saękalımı hakkında daha önceden R ile web tabanlı olarak yapılmamıř olması özelliklerinden dolayı özgünlük taşımaktadır.

2. GENEL BİLGİLER

2.1. KARAR DESTEK SİSTEMLERİ

Günümüzde her alanda verilerin fazlalığı durumu, bu devasa veri yığınlarının insan zihni ile tek tek incelenip analiz edilmesini imkansız hale getirmiştir (8). Herhangi bir alanda komplike sorunların çözümü için veri yığınlarının arasındaki ilişkiyi çok büyük veri kümelerini bir arada kullanarak anlamlandırabilen ve bu anlamlı çıktıları bilişim teknolojisini kullanarak karar öneren sistemler geliştirilmiştir (9). Sağlık alanında da insanlardaki muhakeme yeteneğinin bir benzerini sergileme özelliğinde olan kompleks algoritmalara sahip yazılımlar olan klinik karar destek modülleri kullanılmaktadır (10).

Karar destek sistemleri (KDS) kendi başına karar vermekten ziyade karmaşık verileri kullanıp analiz ederek karar verici kullanıcıların daha doğru ve etkin karar vermelerine yardımcı olma amacı ile çeşitli modellerde tasarlanan bilgisayar yazılım sistemlerinin genel adıdır (8,9,11).

KDS'leri verilecek kararın yeterliğinden ziyade etkinliğini geliştirmeye yarar. Bu açıdan daha çok karar vericiye destek sunmaktadırlar. Bu sayede karar verici olan kullanıcılar çeşitli çözümleri deneyerek verileri yeniden inceleme fırsatına sahip olurlar (12).

Karar desteği kavramı ilk olarak 1960'ların ilk yıllarında Carnegie Teknoloji Enstitüsü'nde gerçekleştirilen "örgütsel karar verme" çalışmaları sonucu ortaya çıkmıştır (13).

KDS yazılımlarındaki algoritmaların temelinde "Eğer- Öyleyse" sorgusu yer almaktadır. Bu sorguya göre en iyi olasılık tahminlenmeye çalışılır. Ortaya çıkan değerler tanımlara uygunsa kesin sonuca ulaşılmış olunur (14).

Veri tabanı, verileri analiz eden modelleme algoritmaları, formülleri çalıştıran bileşenler ve kullanıcı sistem arayüzü KDS'leri yazılımlarını oluşturan ana unsurlardır (8,15).

Sunduğu yardım çeşidine göre KDS’ler; iletişim odaklı KDS, belgeye dayalı KDS, veri temelli KDS ve model odaklı KDS şeklinde ayrılabilirler. İletişim odaklı KDS; aynı şey üzerinde çalışan birden fazla kişiyi destekleyerek işbirliği sağlar. Bu KDS’lerine “Google Döküman”’lar örnek olarak verilebilir. Belgeye dayalı KDS’leri; bilgileri işleyerek yapılandırılmalarını sağlarlar. Veri temelli KDS’leri; aynı tür spesifik olgulara yönelik programlanmış KDS’ler olup, özel problem çözme imkanı sağlarlar. Model odaklı KDS’leri; spesifik bir konudaki değişen parametrelerin analizi konusunda son kullanıcılara destek sağlarlar (16). Bunlar dışında sundukları çalışma koşullarına göre grup karar destek sistemleri (GKDS) ve web tabanlı karar destek sistemleri (WKDS) bulunmaktadır. Bu iki KDS’nin çalışma şekilleri birbirine her ne kadar benzese de, GKDS’ni kullanan çalışma ekibi verileri kendi sistemleri üzerinden tek platformda paylaşabilirlerken, WKDS’ni kullanan topluluklar kişisel bilgisayarları ile bağlandıkları network üzerinden çalışma ekipleri, işbirlikçileri ve gelecekteki araştırma- geliştirmeciler ile problem özelindeki çıktılarını ve düşüncelerini paylaşabilmektedirler. Bu açıdan WKDS’leri sunduğu rahat çalışma imkanının yanında açık kaynaklı ve rahat erişilebilir olması sebebiyle bilimsel araştırmalar için daha uygun bir model olarak görünmektedir.

Özetle KDS’lerin karakteristik özellikleri; bireysel ve ekip çalışmalarında karar verme süreçlerine yardımcı olması, değiştirilebilir ve güncellenebilir esneklikte olması, etkileşimli, kolay kullanılabilir olması, modelleme ve analiz becerilerine sahip olması, verilere erişim özelliklerinin olması, hızlı cevap alınabilmesi ve tek başına veya web-tabanlı çalışabilme özelliklerine sahip olması olarak sıralanabilir.

KDS’leri verilen kararı desteklemek için kullanılabilecek objektif verilerin elde edilmesi ve karar verme sürecinin hızlandırılması sayesinde destek aşamasında büyük faydalar sağlamaktadır.

2.1.1. Klinik Karar Destek Sistemleri

Klinik karar destek sistemleri (KKDS) sağlık uzmanları tarafından geçmiş zamanda hastalardaki çıktılar kullanılarak oluşturulan veri yığınlarını şimdiki zamanda incelenen olgudaki hasta girdileri ile karşılaştırıp değerlendirme ve tavsiyeler sunarak sağlık uzmanlarına alacakları kararda en uygunu seçmeleri konusunda destek

sunan sistemlerdir (9). KKDS'ler klinik olarak anlamlıdır ve her bir hastanın teşhis ve tedavi süreçlerinde dolayısı ile sağlığı üzerinde ölçülebilir bir etkiye sahiptir (17).

KKDS'leri; hasta güvenliğini arttırması, klinik yönetim ve idari işlevleri daha kolay hale getirmesi, uygun tedavinin seçilmesi, hataların önlenmesi, hastalığın seyrini tahminlemeye yardımcı olması, hastalığın yönetiminde minimum risk maksimum fayda tedavi seçeneklerini sunması, güncel sağlık veri yığınları ile kişisel sağlık verilerini girdi olarak kullandığı için kişiye özgü değişkenleri analiz ederek bireyselleştirilmiş tıp anlayışıyla doğru hastaya doğru teşhis ve tedavi yönetim seçeneklerini ön plana çıkarabilir olması, sağlık alanlarındaki eğitim- öğrenme dengesini hızlandırması, karar verme sürecini objektif bilgiye dayalı hale getirerek alınacak olan kararın herkes tarafından kabul görmesini sağlaması, karar verme sürecini kolaylaştırıp hızlandırarak zamandan tasarruf sağlar. Tüm bu avantajlarla KKDS'leri hasta bakımı teşhis ve tedavi sürecinde hastaların en iyi şekilde değerlendirilmesi ve hizmet almalarını sağlayarak verimliliği arttırması ve buna bağlı olarak maliyet kalemlerindeki azalmaya olumlu katkı sağlaması sebebi ile önem kazanmaktadır (9,15,18).

Öte yandan bu sistemlerin klinik entegrasyonunda karşılaşılan zorluklar, büyük ve heterojen veri setlerinin entegrasyonu ve yasal uyumluluk dahil olmak üzere çok fazladır (17). Klinik verilerdeki karışıklığa bağlı olarak ortaya çıkan tasarım problemleri, uygulamaların yüzde yüzünün başarılı sonuç vermemesi, parçalı iş akışı modeli ile klinisyenlerin iş akışını bozması, uygunsuz uyarılar ile sağlık uzmanı kullanıcıların güvenliğini zedeleyip kullanmaktan vazgeçmelerine neden olan uyarı yorgunlukları yaratması, diğer tam otomatize sistemlerde de olduğu gibi KKDS kullanıcılarının sistemi hesap makinesi gibi doğruluğunun doğrulanmasına gerek olmayan bir sistem olarak görebilme ihtimallerine bağlı ortaya çıkan aşırı güvenme sonucu mesleki donanımlarını ve zihinsel becerilerini daha az kullanmaya alıştırmaları sorunu, KKDS'lerinin bazılarının çok karmaşık kullanıcı arayüzlerine sahip olması nedeniyle yüksek derecede bilgisayar okuryazarlığı gerektirmesi durumu, sistemin güncel verilerle içeriğinin sürekli güncellenmesi ve teknik bakımının zor olması, düşük veri kalitesinin ve yanlış içeriğin operasyonel etkisi KKDS'lerinin bazı dezavantajları arasında gösterilebilir (18).

Bu engellere rağmen, KKDS'lerin özellikle karmaşık karar verme gerektiren disiplinlerde kullanımı ile ve hem verimliliği hem de hassasiyetinin geliştirilebilme potansiyeline sahip olduğu görülmektedir. Bu gelişimler zaman içerisinde rastgele kontrollü denemeler ile KKDS'lerinin etkinliklerinin kanıtlanması ve düzenleyici kurumlar tarafından onaylanan standartlar aracılığı ile tanınacaktır (17). Örneğin; 1987 yılında kurulan sağlık bakım standartları organizasyonu HL7 (Health Level Seven International); Amerikan Ulusal Standartlar Enstitüsü (ANSI) tarafından akredite edilmiş standartlar geliştiren, kapsamlı bir çerçevede sağlık hizmetlerinin uygulanması, yönetimi, sunulması, değerlendirilmesi için klinik desteği destekleyen elektronik sağlık bilgilerinin değişimi, entegrasyonu, paylaşımı ve alınması ile ilgili standartları sağlamaya adanmış bir amacı gütmeyen bir organizasyondur (19). HL7, KKDS'lerinin gelişimine katkı sağlayacak, "kaynaklar" olarak bilinen veri formatlarını ve öğelerini açıklayan bir standart ve elektronik sağlık kayıtları (ESK)'nın değişimi için bir uygulama programlama arayüzü (UPA) olan "Hızlı Sağlık Hizmetleri Birlikte Çalışabilirlik Kaynakları (Fast Healthcare Interoperability Resources, FHIR)"i oluşturmuştur (20). FHIR; kaliteli verilere daha hızlı, gerçek zamanlı erişim sağlaması, ESK'lardan otomatik veri alımını ve standartlara dayalı UPA'ların kullanımıyla kaliteli verilerin gönderilmesini sağlayarak raporlama yükünü azaltması, birlikte çalışabilirliği teşvik etmesi ve elektronik klinik kalite ölçücü sistemler ile KKDS'leri arasındaki uyumu iyileştirmesi gibi faydaları ile katkılar sunmaktadır (20).

Yeni nesil KKDS'lerinin klinik faydayı en üst düzeye çıkarması için, popülasyon temelli risk sınıflandırmasının ötesine geçerek kişiselleştirilmiş ve kesin risk tahminlerine doğru ilerlemesi gerekmektedir. Doğru tedaviyi doğru hastaya doğru zamanda sunarak tıbbi daha kişiselleştirilmiş bir yaklaşım ile katkıda bulunacaklardır (17). Genetik klinik karar destek sistemi araçları da KKDS'lerinin başarısını arttırma potansiyeli taşımaktadır (21).

KKDS'leri onkoloji, cerrahi, anesteziyoloji ve yoğun bakım alanlarında bu yaklaşımla kullanılmaya başlamışlardır. Bu alanlarda;

Onkolojik cerrahide; cerrahi girişime uygun iskelet metastazı olan hastaların sağkalımlarını tahminleyen PathFx (22),

Cerrahide; hastaya özgü ameliyat sonrası komplikasyon risklerini tahmin eden NSQIP (23), ameliyat sonrası komplikasyonları tahmin eden POSSUM (24), hastayı hastaya özgü ameliyat ile ilgili riskler konusunda bilgilendiren ve NSQIP'nin iyileştirilmiş hali olan DS3 (25),

Anesteziyolojide; kayıt tutma, kalite güvencesi ve araştırma sunan elektronik sağlık kaydı uygulaması AIMS (26),

Yoğun bakımda; yoğun bakım ünitesine kabul edilen hastalar için hastalığın ciddiyetini tahmin eden APACHE (27,28), hastane içi sağkalımı tahmin eden SOFA (29), hastalıkların zamansal sırası açısından hastaların evrimini tanımlayan ACUDES (26) günümüzde kullanılmakta olan KKDS'lerinden bazıları olarak gösterilebilir.

Onkoloji alanındaki KKDS'lerine bir diğer örnek ise; OncoSim web tabanlı mikro simülasyon modeli platformudur. OncoSim; Kanada genelinde kanser kontrolünü hızlandırarak kanıta dayalı politika ve karar almaya yardımcı olmak için 2007 yılında federal hükümet tarafından geliştirilmiş olup bir dizi spesifik kanser modelinden oluşmaktadır. Nüfus düzeyinde politika ve karar verme için geleneksel kaynakları artırmaya yönelik bir araç olarak tasarlanmıştır (30). Klinikten alınan verilerle, uzman görüşlerini ve literatür bilgisini birleştirerek, sigara içme, yetersiz fiziksel aktivite gibi 27 farklı risk faktörüne göre sağlık ve ekonomik sonuçları öngörebilen OncoSim platformu şimdilik meme kanseri, kolorektal kanser, akciğer kanseri ve servikal kanser projelerini içermektedir (31).

Matematiksel bir simülasyon olan OncoSim-Kolorektal projesinde; hastalığın ilerlemesi, Kanada hayatta kalma verileri kullanılarak modellenmiştir (32). İl/ bölge ve ulusal düzeyde kolorektal kanser insidansı, yaşam yılları, kaliteye göre ayarlanmış yaşam yılları ve sağlık hizmetleri maliyetleri gibi sonuçları yansıtan modelin girdileri, Kanada hayati istatistikleri, toplum sağlığı anketleri, kanser kayıtları, tarama programı veri tabanları, idari veri tabanları ve hakemli literatür dahil olmak üzere çok çeşitli kaynaklardan gelmektedir (33,34). Model, kolorektal kanser tarama stratejilerini, değişen tarama yöntemlerini (dışkı testleri, sigmoidoskopi ve kolonoskopi), yaşı, tarama aralığını vb. karşılaştırmak için uygulanmıştır (33–35). Model, sigmoidoskopi taraması denemelerinde ve dışkıda gizli kan testi taramasına yönelik bir meta-analiz

çalışmasında gözlemlenen kolorektal kanser taramasının etkilerini yeniden üretmiştir (34,36). OncoSim-Kolorektal çıktıları ile proje haricinde gözlemlenen verilerin karşılaştırılmasına dair bir çalışmanın sonucuna göre modelin projeksiyonlarının 4 farklı randomize denemenin gözlemlenen güven aralıklarında olduğu gösterilmiştir (34).

2.2. YAPAY ZEKA

Yapay zeka (Artificial Intelligence, AI); “bir makinenin, önceden belirlenmiş bir hedefe ulaşmak için algılanan ortama (belirli girdilere) yanıt olarak bir eyleme karar vermede tipik insan bilişine özgü entelektüel süreçleri bağımsız olarak çoğaltma becerisi” olarak tanımlanabilir (37).

AI, sağlık hizmetinin geleceğini dönüştüren ölçeklerde karar destek sistemi sağlayarak öğrenme kapasitesini artırır. AI'nın klinik uygulamalarda da potansiyeli oldukça fazladır. Mevcut sağlık hizmetlerinde yapılandırılmamış veriler muazzam bir hızla artmaktadır. Yüksek çözünürlüklü tarama ve multiparametrik görüntüleme, gelişmiş genom dizileme, çoklu omik uygulamaları, çok çeşitli tarama biyobelirteçleri ve ayrıntılı hasta kayıtları ile tek bir hastadan onlarca gigabayttan oluşan devasa veri kümelerinin ortaya çıkabilir. Hücresel düzeyde moleküler veriler ile doku yapısına ilişkin diğer biyolojik verilerin normal ve hastalıklı doku işlevi ile nasıl ilişkili olabileceğine dair derin bir anlayış geliştirmek için çok sayıda çok modlu, çok boyutlu veriden oluşan bu kaynaşmış akışlar, AI aracılığıyla entegre edilebilir ve bütünsel olarak yorumlanabilir (37). Günümüzde AI; hastalık teşhisi ve prognozunda, tedavi optimizasyonunda ve sonuç tahmininde, ilaç geliştirmede ve halk sağlığında uygulanmıştır (38). Bunlara ek olarak robotik cerrahide; lazer göz ameliyatı ve saç ekimi gibi basit prosedürlerdeki rutin adımlarda, görüntü analizinde; X ışınlarının, retina taramaları ve diğer görüntülerin incelemesine yardımcı olmak için kullanılması prensibindeki otomatik sistemlerde, genetik analizde; genom tarama verilerinden hızlı içgörüler elde eden sistemlerde, patolojide; biyopsi örneklerini analiz etmede, KKDS'lerde; hastanelerde septik şoku tahmin etmede, sanal hemşirelikte; hekimlerin muayene ziyaretleri arasındaki sürede hastaları kontrol edip otomatik uyarılar almalarının sağlanmasında, tıbbi yönetimde; özel hastane ve sağlık sigortası

şirketlerinin faturalandırma ve sigorta gibi görevlerdeki verimliliklerinin arttırılmasında, akıl ve ruh sağlığında; cep telefonu ve sosyal medya verilerinin incelenerek depresyonun izlenmesi gibi uygulamalarda da kullanılmakta, kullanım alanı, etkinlik ve verimliliklerinin arttırılması için de çalışmalar yapılmaya devam edilmektedir (39).

Pek çok çalışma, hastalığın zamanında ve kesin teşhisinde AI'nın etkinliğine, verimliliğine ve potansiyeline bakmaktadır. Olgu sayısı 1634 olan, kanserli ve sağlıklı akciğer dokusunun görüntüleri üzerinde yapılan bir çalışmada, algoritma sağlıklı vakalar ile iki yaygın akciğer kanseri türünü üç patoloğun ortak değerlendirmesi ile aynı şekilde doğru olarak ayırt etmiştir (40). Başka bir çalışmada % 6'sından fazlası depresyondan etkilenen Amerika Birleşik Devletleri yetişkin nüfusunda AI kullanılarak görüntü ısı haritası örüntü tanınması ile majör depresif bozukluğun tahmin edilmesinin %74 oranında doğru olduğu gösterilmiştir (41).

AI algoritmaları ile bilgisayarlar; iyi huylu hastalığı olan slaytları kanserli olanlardan güvenilir bir şekilde ayırabilir ve şüpheli slaytları insan incelemesine bırakabilir. Bu nedenle, insan uzmanlarından her zaman AI sistemlerine nasıl performans göstereceklerini öğretmeleri, bu sistemlerin tasarlandığı gibi çalıştığından emin olmaları ve istenmeyen sonuçlarla (aşırı teşhis, yetersiz teşhis veya yanlış teşhis gibi) hızlı bir şekilde başa çıkmaları istenecektir. Uzman klinisyenler, yorumlamadaki hatalar ve yanlış mantıksal çıkarımlar için çıktıları sürekli olarak izlemek zorunda kalacaklarından yapay zeka, insan uzmanlığını artıracak ancak onların yerini almayacaktır. Einstein'a göre, “Bilgisayarlar inanılmaz derecede hızlı, doğru ve aptaldır. İnsanlar inanılmaz derecede yavaş, yanlış ve zekidir. Birlikte hayal gücünün ötesinde güçlüdürler” (42).

AI Sistemlerinin hasta verilerinin hepsine yerel ve değişen sağlık politikalarındaki farklılıklar nedeniyle erişememesi ve erişebildiği verilerde de bazı parametrelerde hatalı veya eksik girdiler olması nedeniyle yaşadığı zorluklar; sağlık kurumlarının, kendi özel ve gizlilik kontrollü verilerini paylaşmayı kabul etmeleri ve ortak bir hasta bilgi veri giriş protokolü izlenmesi ile çözülebilir. Böylece AI sistemleri yıllar boyunca takip edilen milyonlarca hastadan öğrenebilir (39). AI algoritmalarının

yeterince eğitilmesi ve performans göstermesi için verilerin ve etiketlerinin mümkün olduğunca tutarlı olması gerekir. AI tabanlı KDS uygulamalarının uygun şekilde eğitilmesini, çalıştırılmasını ve düzenlenmesini sağlamak için klinisyenler, veri bilimciler, bilgisayar araştırmacıları ve mühendisler arasında işbirliği gereklidir.

2.2.1. Klinik Karar Destek Sistemlerinde Yapay Zeka Uygulamaları

Veri sistemlerinin çoğunda çok büyük miktarda veri analitik amaçlarla kaydedilir ve saklanır. Saklanan bu büyük miktardaki veri çoğu zaman optimum karar verme için ek fayda sağlamaz. Hatta bu verileri toplamak, depolamak ve işlemek çok karmaşık ve maliyetli olabilir. Örneğin, yüksek verimli olan mikro-dizileme verilerinin kullanıldığı tümör sınıflandırması, sınıflandırma hatalarının azaltılmasına katkıda bulunmayan çok sayıda gürültülü özelliğin varlığı nedeniyle zordur. Bu tür problemler için genel amaç, sınıflar arasında oldukça farklılaşan sınırlı sayıda gen bulmaktır. Bunun için ise sistemlerin bilişsel becerilere sahip olması gerekmektedir (43).

Tedavi etkinliği ve sonuç tahmini, hastalık yönetimi stratejilerin ve kişiselleştirilmiş bakım planları için önemli klinik uygulama alanlarıdır. Kanserde kişiselleştirilmiş tıbbın hedeflerinden biri de, hastaya özgü tümörlerin genomik verilerinden bireye özgü optimal ilaç tedavilerinin doğru tahmin edilmesidir (44). On yıl önce, kanser sonuçlarını tahmin etmek için yalnızca moleküler ve klinik bilgiler kullanıldı. Genomik, proteomik ve görüntüleme teknolojileri de dahil olmak üzere yüksek verimli teknolojilerin geliştirilmesiyle, yeni tür girdi parametreleri oluşmuş, bu parametreler toplanmış ve tahmin için kullanılmıştır. Histolojik veya patolojik değerlendirmeler de dahil olmak üzere geniş bir örneklem boyutu ve entegre multimodal veri türleri ile bu yöntemler, kansere duyarlılığın, sonuç tahmininin ve prognozun doğruluğunu önemli ölçüde (% 15-% 25) artırabilir (45,46). Kişiselleştirilmiş tıp başarısı, algoritmanın büyük omik verilerini klinik olarak eyleme geçirilebilir tahminlere çevirme yeteneğine bağlıdır. Özellikle kişiye özel kanser tedavilerine olanak tanınması sebebiyle tümörlerin gen ifadelerine göre sınıflandırılması büyük ilgi görmektedir. Gen ifadesine dayalı tümör sınıflandırması ile kanseri tanımlamak için mikrodizi verilerinin kullanılması gerekir (43).

2.2.2. Kanser Kliniğinde Yapay Zeka Temelli Klinik Karar Destek Sistemleri

Kanser hastalarının hayatta kalmasında prognostik değişkenleri değerlendirmek için yaygın olarak kullanılan Cox orantısal risk modeli, sonucun ortak değişkenlerin doğrusal bir kombinasyonu olduğunu varsaydığından, bir kanser hastasının sonucunu doğru bir şekilde tahmin etmek için çok basit olabilir. Cox orantısal risk modeli ile derin öğrenme tabanlı bir sağkalım tahmin algoritmasının sağkalım performanslarını karşılaştıran bir çalışmada algoritmanın daha iyi performans gösterdiğine ulaşılmıştır (47).

Bilişsel bilgisayar sistemleri, sağlık analitiğinde yeniliklere yol açmaktadır. Onkologlar, büyük teknoloji şirketleriyle birlikte yeni süper hesaplama araçlarını tıbbi karar alma sürecine entegre etmekte. Örneğin Amerikan Klinik Onkoloji Topluluğu (American Society of Clinical Oncology ASCO)'nun onkoloji alanındaki AI temelli KKDS'si "CancerLinQ", Amerika Birleşik Devletleri'ndeki sağlık hizmeti enformasyon teknolojisi sistemlerinden kanser olgularına ait verileri toplayıp analiz ederek hasta bakımını iyileştirmek için değerli bilgiler ve araçlar sunar. CancerLinQ platformu, Amerika Birleşik Devletleri'nde kanser ile mücadele için hekimler tarafından yönetilen tek büyük veri analitik platformudur. CancerLinQ veritabanı, kanser tedavisini gerçek dünyadaki tüm değişkenliği ile yansıtarak klinisyenlerin her hastanın deneyimlerinden öğrenmesine olanak tanımaktadır (48). Bu konuda bir diğer örnek olarak IBM'in "Watson Health" isimli yeni girişimiyle, giderek artan sağlık verilerini ve klinik bilgileri kullanarak ileri seviyede KKDS'leri geliştirmeyi hedeflemesi gösterilebilir.

IBM tarafından 2010 yılında doğal dilde sorulan sorulara cevap vermek için geliştirilen AI programı Watson'ın 2011 yılında Jeopardy isminde, yarışmacılara cevaplar ile ipuçlarının sunulup doğru soruları tahmin etmeleri gereken, bir yarışma programında insanlara karşı aldığı zafer bu programın ilk başarılarından olmuştur. Bu başarı Watson'un doğal dil işleme konusundaki olağanüstü becerisini göstermiştir. Oyunu oynamak için kelime oyunları ile dolu karmaşık ipuçlarını ayrıştırması, olası yanıtları bulmak için büyük metin veri tabanlarını araştırması ve en iyi olanı

belirlmesi gereken Watson üstün bir arama motoru olmamasına rağmen yalnızca anahtar kelimelere dayalı olarak belge döndürmekle kalmayarak bir cümledeki “varlıkları” haritalamak ve aralarındaki ilişkileri anlamak için yüzlerce algoritma kullandı. Bu büyük başarıdan sonra 2011 yılında IBM, Watson’ı sağlık hizmetleri sektörüne entegre etme çabasına başladı. Tıbbi literatür ve uzman kişilerin yönetimiyle desteklenen vaka odaklı öğrenme sayesinde gerçek hasta geçmişlerini takip edip ortaya çıkan anlamlı bilgileri alınacak klinik kararlarla ilişkilendirmeyi hedefleyen şirket o zamandan beri, bu hedefe yönelik tıp için yapay zeka destekli yeni araçlar geliştirmeyi amaçlayan ortaklıklar hakkında yaklaşık 50 duyuru yaptı. Bu ortaklıkların çoğu henüz ürüne dönüşmemiş olsa da, bu projeler içinde başarıya ulaşmış olanlar; 2012 Mart ayında duyurulan “Memorial Sloan Kettering Kanser Merkezi” ile birlikte yürütülen “Watson for Oncology (WFO)” ismindeki “Kanser için klinik karar destek sistemi” projesi, 2014 Eylül’ünde duyurulan “Mayo Clinic” ile birlikte yürütülen “Watson for Clinical Trial Matching” ismindeki “Klinik araştırma eşleştirme aracı” projesi, 2015 Nisan’ında duyurulan “Medtronic” ile birlikte yürütülen “Sugar.IQ app” ismindeki “Kişiselleştirilmiş diyabet yönetimi için tüketici uygulaması” projesi, 2015 Mayıs’ında duyurulan “Kuzey Carolina Üniversitesi” ile birlikte yürütülen “Watson for Genomics” ismindeki “Kanser için genomik analiz aracı” projesi ve 2016 Ekim ayında başlayan “Quest Diagnostics” ile birlikte yürütülen “Watson for Genomics from Quest Diagnostics” ismindeki “Kanser için genomik analiz aracı” projeleridir.

IBM’in tüm bu projeleri içerisinde en çok duyurulan ve ses getiren projesi, WFO, Watson’un "bilişsel" yeteneklerini kullanıp büyük verileri hastalar için kişiselleştirilmiş kanser tedavilerine dönüştürmeyi amaçlayarak onkoloji alanına odaklanmış durumda (39,49). Son yıllarda yapılan bazı çalışmalar WFO’nun kanser tedavisi önerilerini onkologların önerileriyle karşılaştırmıştır. Tayland’da 211 vaka üzerinde yapılan çalışmada kolorektal kanser için %89, akciğer kanseri için %91, meme kanseri için %76 ve mide kanseri için %78 olmak üzere %83 genel uyum oranına ulaşılmıştır (50). Hindistan’da 638 meme kanseri vakası ile gerçekleştirilen retrospektif gözlemsel çalışmada uyum oranı %73, Güney Kore’de 656 kolon kanseri vakası ile gerçekleştirilen çalışmada ise genel uyum oranı %49 olarak açıklanmıştır (51,52). IBM’in “Watson for Genomics” projesinin etkinliği hakkında başka bir çalışmada Watson AI algoritması çalışmaya katılan kanser hastalarının %32’sinde

insan incelemesiyle belirlenememiş potansiyel olarak önemli mutasyonları tespit etti (53). Tüm bu çalışmalar sonucunda WFO, uzman tıbbi onkolog hekimlerin sınırlı sayıda olduğu veya olmadığı merkezler için ve tıp öğrencileri, asistan hekimler gibi henüz yeterli tecrübeye sahip olmayan klinisyenler için kanser tedavisine karar vermede literatür bilgisini de içeren iyi bir yardımcı referans aracı olarak görülmektedir. Ayrıca kemoterapi için seçenekler arayan uzman onkologlar için de AI hesaplama metodolojisi sunan potansiyel bir KKDS destek aracı olarak görülebilir.

Bilişsel hesaplama maharetine sahip AI teknolojisinin, sağlık uzmanlarının hastalarla etkileşime girme ve klinik verileri yorumlama becerisini kazanamayacağı için insan uzmanların yerini alamayacağı açık olsa da büyük, sürekli olarak tutulan ve giderek artacak olan veri kümeleri nedeniyle kanser tedavisi ile ilgili birçok görevin gerçek zamanlı bir analiz gerektirdiği bir noktaya geldik. Günlük olarak üretilen muazzam miktardaki yeni klinik verinin kapsamlı analizini yapma gibi görevleri gerçekleştirmek de insanlar için pek uygun olmadığından bilişsel hesaplama yaklaşımına sahip bu teknoloji, hastaların önemli bir kısmının tedavi ve bakımını iyileştirebilecek potansiyeldedir. Bu sebeple genel perspektiften bakıldığında Watson ve benzeri diğer tüm AI temelli KKDS araçları yakın gelecekte insan hekimlerin yerini alacak robot bir hekim gibi hizmet vermekten çok, hekimlerin çoğunun ellerinin altında bulunmasından memnun olacağı, tıp alanında uzman bir kütüphaneci robot olarak hizmet vereceğe benzemektedir.

Yapay zeka temelli sağkalım tahmini, tahmin doğruluğunu artırabilir ve klinisyenlere hem sağkalım için daha etkin tedavi seçeneklerine karar vermede hem de gereksiz tedavilerden kaçınmada yol gösterebilir. Bu çalışmada da benzer şekilde AI teknolojisini kullanarak kanser kliniğinde “Evre 4” kolon kanseri hastaları için sağkalım tahmini yapılarak hastanın yaşam kalitesini çok büyük ölçüde etkileyebilecek yüksek yan etki potansiyeline sahip kemoterapi tedavisinin uygulanıp uygulanmaması konusunda artı eksi analizi yapılarak böylesi hassas bir sürecin en doğru kararlarla yürütülmesi adına hekime ve hastaya bilgi verici yardımcı bir karar destek sisteminin ortaya koyulması hedeflenmiştir.

2.3. MAKİNE ÖĞRENMESİ

Makine öğrenmesi; yapay zekanın bir alt kümesi gibi düşünülebilir. Makine öğrenmesi eylemleri belirlemek için istatistiksel araçlar kullanarak verileri analiz eden algoritmaların geliştirilmesini ve konuşlandırılmasını içerir (54). Makine öğrenmesi algoritmaları dinamiklidir. Yani daha fazla veri kullanıldıkça iyileşme, öğrenme eğilimindedir. Çoğu makine öğrenimi algoritması bir dizi değişkeni etiketleyerek yine bir dizi sonuç değişkeni ile eşleyen matematiksel modellerden oluşur. Algoritmalar, açık özellikleri analiz ederek etiketleri tahmin edebilmek için "eğitim" olarak bilinen bir süreçte optimize edilir (54).

Makine öğrenmesi; klinikte; hastalık tahmini ve teşhisi, tedavi etkinliği ve sonuç tahmininde, translasyonel tıpta; ilaç keşfi ve yeniden kullanımında, *in silico* klinik denemelerde, halk sağlığında; salgın hastalık tahmininde yüksek potansiyele ve etkilere sahiptir (38). Makine öğrenmesi; segmentasyon ve füzyon biyopsisi gibi teşhis ve izlem için kullanılan tanısal görüntüleme işlemlerinde, robotik cerrahi ve cerrahi becerilerin değerlendirilmesi gibi tedavi müdahalelerinde, otomatik teşhis için kullanılan dijital patoloji işlemlerinde, genomikte sınıflandırıcılarda ve birleştirilmiş veri akışlarını analiz etme gibi işlemlerde de kullanılabilir (37).

ESK'lar; sağlık hizmetleri bilgilerini belgelemek ve paylaşmak için etkili araçlardır. Veri kümeleri için özel olarak tasarlanmış makine öğrenmesi tabanlı modellemeyi entegre etmek, potansiyel komplikasyonların tespitini kolaylaştırma, prognostiklerin doğruluğunu iyileştirme, verimli ve yüksek hızda teşhis sağlayarak kişiselleştirilmiş bir düzeyde sonuçları iyileştirme ve sağlık hizmeti kaynak kullanımını iyileştirerek sağlık bakımı maliyetlerini daha da düşürme potansiyeline sahiptir (55).

2.3.1. Makine Öğrenmesi Çeşitleri

Makine öğrenmesi teknikleri, genel olarak etiket türüne ve özelliğin türüne göre sınıflandırılabilir. Etiketlere dayanarak makine öğrenmesi; denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olarak üçe ayrılabilir. Özellikler

temelinde ise; el yapımı veya el yapımı olmayan özellik tabanlı teknikler olarak sınıflandırılabilir (56).



Şekil 1. Makine öğrenmesi algoritma çeşitleri

2.3.1.1. Etiketlere Göre Makine Öğrenmesi Çeşitleri

Denetimli Öğrenme

Denetimli öğrenme yöntemleri, karmaşık ve çok faktörlü hastalık sınıflandırması için doğrusal olmayan ilişkileri yakalamada etkili araçlardır. Denetimli öğrenme, uzmanlar tarafından etiketlenmiş açık veri setlerine dayanır. Tasarlanmış özellik algoritmaları, tahmin edilen etiketler ile “kesin referans” olarak adlandırılan bilinen etiketler arasındaki farkın bir ölçümü olan tahmin hatasını en aza indirecek şekilde eğitilir. NB (Naive Bayes) sınıflandırması, doğrusal ve lojistik regresyon, DVM (Destek Vektör Makineleri) ve RO (Rastgele Ormanlar, Random Forest) bu tür algoritmaların örnekleri arasında gösterilebilir (54,57,58).

Denetimli öğrenme yönteminin denendiği Abedi ve arkadaşları tarafından gerçekleştirilmiş, 260 hastadan oluşan bir kohort çalışmasında, ilgili modelin akut

serebral iskeşmiyi, eğitilmiş acil tıbbi müdehale ekiplerine göre daha iyi teşhis edebildiğı gösterilmiştir (59).

Denetimsiz Öğrenme

Denetimsiz öğrenmede algoritma, karşılık gelen etiketler olmadan, tek başına eğitim verilerinin özellikleri temelinde örnekleri farklı sınıflara en uygun şekilde ayırır. Bu tür algoritmaların örnekleri, sağlanan verilerdeki benzer örneklerin gruplarını keşfetmek için kullanılan k-ortalama kümeleme, temel bileşen analizi ve otomatik kodlayıcılardır (54,60,61).

Pekiştirmeli Öğrenme

Türkçe bazı kaynaklarda takviyeli öğrenme olarak da geçen PÖ (Pekiştirmeli Öğrenme, Reinforcement Learning), tipik olarak sıralı işleyen bir algoritma ailesini içerir. Her adımda, "aracı" olarak adlandırılan bir takviye algoritması, geçmiş ve şimdiki özelliklere dayalı olarak gelecekteki bir adımda özellikler üzerinde hareket eder ve bunları tahmin eder ve tahmine dayalı olarak bir ödöl veya ceza atanır. Algoritmalar genellikle gelecekteki tahmini ödölün toplamı olarak beklenen getiriyi en üst düzeye çıkarmak için her aşamada gerçekleştireceğı eylemi seçmeye yönelik bir politika öğrenir (62).

2.3.1.2. Özellikler Temelinde Makine Öğrenmesi Çeşitleri

El Yapımı Özellik Tabanlı Makine Öğrenimi Teknikleri

El yapımı özellik tabanlı makine öğrenimi teknikleri, veri setinde önceden belirlenmiş sınırsız sayıda açık özelliğın çıkarılmasını içerir. Özellikler, uzmanların genellikle teşhis veya karar süreçlerinde aradıkları ilgili bilgilere dayanmaktadır. Bu ayırt edici özellikler genellikle deneyime ve uzman bilgisine dayanır ve nitel muhakemeye tabidir.

El yapımı özellik tabanlı teknikler; kendi bilişsel süreçlerini takip edebilme ile karakterizedir. Yeniden üretilebilme ve test edilebilebilme sayesinde ilgili işleme adımlarının tam olarak anlaşılmasını sağlama avantajına sahiptir. Bu özellik,

düzenleyici onaylarını ve makine öğrenmesinin klinik uygulamaya çevrilmesini kolaylaştırır.

El Yapımı Olmayan Özellik Tabanlı Makine Öğrenimi Teknikleri

El yapımı olmayan özellik tabanlı teknikler, öğrenmenin bir parçası olarak ham verilerin işlenmesini içerir. Bu yöntemlerde, algoritma öğrenir ve ardından tahmin hatasını veya diğer sınıflandırma performansı ölçüsünü optimize etmek için veri alanından açık bir etiketleme yapmadan kendi özelliklerini çıkaracak şekilde adapte olur. Bu yöntemler verilerle ölçeklenir; yani, eğitim için her zamankinden daha büyük veri setleri kullanıldıkça, performansları artar ve sonuçta ortaya çıkan özellikler insanlar tarafından yorumlanabilir olmayabilir. Bu tür tekniklerin çağdaş bir örneği, yapay sinir ağlarına dayanan ve üstün problem çözme yetenekleri sunan “Derin Öğrenme” (Deep Learning)’dir (61,63,64).

2.3.2. Makine Öğrenmesi Süreci

Makine öğrenmesi ile bir problemin çözümü için sistematik bir süreç izlenmelidir. Bu çalışmada süreç; öğrenme probleminin tanımlanması, veri analizi, veri hazırlama, algoritma değerlendirme, sonuçların geliştirilmesi ve sonuçların sunumu şeklinde izlenmiştir.

2.3.2.1. Öğrenme Probleminin Tanımı

Öğrenme problemin tanımlanması sürecin ilk adımıdır. Çalışmanın gereksinimi belirlendikten sonra bu gereksinime yönelik makine öğrenmesi modelinin hedefinin belirlenmesi gerekmektedir. Bu çalışmada görev; kolorektal kanseri tanımlı hastaların sağkalımlarının tahmin edilmesi için hastaların sınıflandırılması, performans; doğru sınıflandırılan hastaların yüzdesi, deneyim; sağkalım veya ölümle sonuçlanan olgulara ait eşdeğer alınan kayıt ve gözlemler (65).

2.3.2.2. Ham Veri Analizi

Makine öğrenmesi için ikinci aşama ham veri analizidir. Bu aşama için önce ham verinin toplanması gerekir. Toplanacak veri yapılandırılmış veya

yapılandırılmamış olabilir. Veri tabanlarından alınan veriler genelde yapılandırılmış verilerdir. Bu çalışmada kullanılan veri de yapılandırılmış veridir. Tanımlanan öğrenme problemine uygun olarak alınan yapılandırılmış ham verinin önce analiz edilmesi gerekir. Bu hazırlık süreci oldukça zaman alıcı bir aşamadır. Bu aşama ulaşılan veriyi anlama işlemi için verideki değerlerin nümerik ve kategorik olarak belirlenip ayrılması, ayrılan değerler üzerinde maksimum, minimum, açıklık, ortalama, mod, medyan veya kartil hesaplamalarının yapılarak hesaplama sonuçlarının daha iyi anlaşılması için grafikleştirilerek görsele dönüştürülmesi işlemlerini içerebilir. Veri ön işleme aşaması sonucunda verinin nasıl hazırlanacağı ve hangi ön işleme adımlarından geçeceği belirlenmiş olur (66).

2.3.2.3. Veri Hazırlama

Algoritma değerlendirme sürecine geçilmeden analizi yapılan ham verinin ön işleme alınarak hazırlanması gerekir. Bu süreç verideki gürültülerden kurtulmak için verinin temizlenmesi, bütünleştirilmesi, dönüştürülmesi, verinin indirgenmesi, ayrıklaştırılması adımlarından geçmesi ile tamamlanır. Bu adımlar kullanılan veri setine göre değişkenlik gösterebilir (67).

Kayıp Değerlerin İşlenmesi

Alınan ham verideki bazı değişkenlerde kayıp değerler olabilmektedir. Kayıp değerler verinin alındığı veri bankasına yüklenirken oluşan hatalar veya ölçüm sonucu gibi insan kaynaklı hatalardan kaynaklanabileceği gibi ölçülmemiş değerlerden kaynaklı da olabilmektedir. Kayıp değerler algoritmada sınıflandırma performansına olumsuz etki edebileceği gibi analiz sonucundaki kesinliği de azaltabilmektedir. Eksik veri kaynaklı boşluklar hesaplamayı zorlaştırarak veri dağılımını yanlış yönlendirebilmektedir. Bu yüzden kayıp değerlerin işlenmesi ve tamamlanması gerekmektedir. Veri hazırlama aşamasının bir parçası olan kayıp değerlerin işlenmesi için bazı işlemler uygulanabilir;

- Aynı örnekleme ait kayıp değer çok fazla ise örneklemin tüm satırı silinebilir,

- Kayıp değer ile ilgili tahmin önemsiz ise kayıp değerlerin tümü “N/A” (Uygulanabilir değil, Not Applicable veya mevcut değil, Not Available) sabiti ile gösterilebilir.
- Kayıp değerlerin fazla olduğu parametredeki değerlerin ortalaması sınıf değerlerine göre alınıp bu ortalama parametredeki kayıp değerler yerine kullanılabilir.
- Kategorik veri parametrelerindeki kayıp değerlere ilgili parametredeki mod değeri atanabilir.
- Kayıp değerler sınıflandırmaya dayalı makine öğrenmesi algoritmalarıyla da tahmin edilebilirler veya kayıp değerlerin tahmini için regresyon yöntemi uygulanabilir (68–75).

Normalizasyon

Verideki her nümerik değer parametresi aynı aralıklarda olmayabilir. Çalışmadaki insanların yaş değerleri 18-90 aralığında yer alabilirken, aynı çalışmada bir ilçede yaşayan insan nüfusu değeri 600-950.000 gibi bir aralıkta yer alabilir. Böyle durumlarda büyük değer aralığına sahip parametrenin hedef değışkeni etkilememesi için veriye normalizasyon uygulanması gerekebilir. Normalizasyon işlemi için formül 2.1’deki “Min-max” denklemi, formül 2.2’deki “Z-score” denklemi yada formül 2.3’deki “Ondalık ölçekleme” denklemi kullanılabilir (67).

$$v' = \frac{v - \min_A}{\max_A - \min_A} \times (\text{yeni_max}_A - \text{yeni_min}_A) + \text{yeni_min}_A \quad (2.1)$$

$$v' = \frac{v - \bar{A}}{\sigma_A} \quad (2.2)$$

$$v' = \frac{v}{10^j} \quad j: \text{Max}(|v'|) < 1 \text{ koşulunu sağlayan en küçük tam sayı} \quad (2.3)$$

Uç Noktaların Düzeltilmesi

Bir veri grubundaki değerlerden bariz şekilde ayrı duran değerlere uç değerler yada uç noktalar denir. Uç noktalar veri grubunun merkezi eğilim ölçülerini ciddi şekilde etkileyebilir ve sonuca olumsuz etki edebilirler. Veri girişinde oluşan hatalar, deneyde gerçekleşen olan dışı gözlem sonuçları, değerlerin değişken kombinasyonu şeklinde yazılmaları uç noktaların oluşumuna sebep olabilir. Uç noktaların ilgili veri grubundan temizlenmesi veriyi daha sağlıklı analiz ederek işleme sokabilmek adına oldukça önemlidir (76–78).

2.3.2.4. Algoritma Değerlendirme

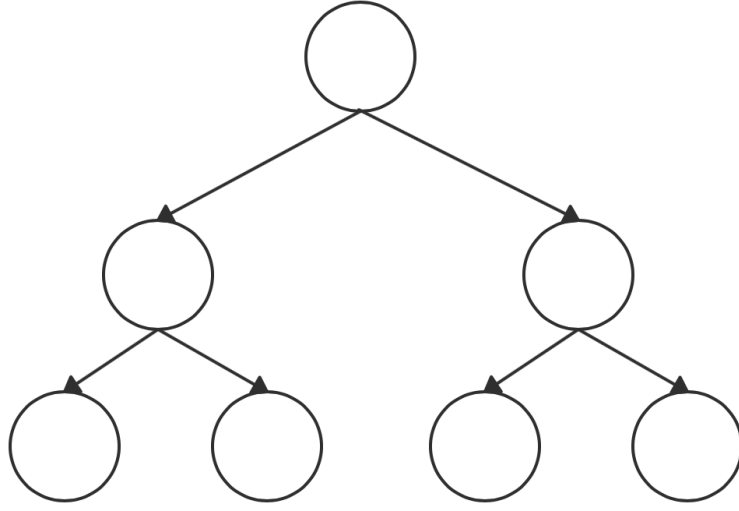
Analiz edilen ve hazırlanan ham veri artık çeşitli makine öğrenmesi algoritma modelleri ile işlenmeye hazırdır. Bu aşamada kullanılacak algoritmanın eğitilip değerlendirilmesi gerekmektedir. Algoritma değerlendirme sürecinde veri üzerinde farklı modeller kullanılarak veri içinde var olan örüntü tespit edilir. Kullanılan algoritma modelleri ile amaca yönelik görevler yönlendirilip yönlendirilen görevlerin öğrenme problemine yönelik olarak algoritma modellerini üreten makine öğrenme algoritmaları tarafından çözülmesi beklenir.

Algoritma değerlendirme aşamasındaki amaç; öğrenme problemine yönelik test veri setinde en doğru çıktıyı sağlayabilecek algoritma modelinin seçilmesidir. Öğrenme algoritması seçerken aynı öğrenme algoritması farklı girdi kombinasyonları ile denenebilir yada farklı öğrenme algoritmaları aynı girdilerle kullanılabilir (79–81).

Bu çalışmada kullanılan “Karar Ağacı (Decision Tree C4,5)” algoritması, “Naive Bayes Sınıflandırıcı” algoritması, “Rastgele Orman (Random Forest)” algoritması ve “Destek Vektör Makineleri (Support Vector Machines)” algoritması alt başlıklarda kısaca açıklanmıştır.

C4,5 Karar Ağacı Algoritması

Karar ağacı algoritmaları; verideki niteliklerin gerçek değerini temel alarak alınan kararların bir modelini oluşturur. Belirli bir kayıt için bir tahmin kararı verilene kadar kararlar ağaç yapılarında çatallanırlar.



Şekil 2. Karar ağacı algoritmaları şematik gösterim

Karar ağaçları, sınıflandırma ve regresyon sorunları için veriler üzerinde eğitilir. Karar ağaçları genellikle hızlı ve doğrudur ve makine öğreniminde favori yöntemlerdendir. En popüler karar ağacı algoritmaları;

- Sınıflandırma ve Regresyon Ağacı (CART, Classification and Regression Tree),
- Yinelemeli Dikotomiser 3 (ID3, Iterative Dichotomiser 3),
- C4,5 ve C5,0,
- Ki-kare Otomatik Etkileşim Algılama (CHAID, Chi-squared Automatic Interaction Detection),
- Decision Stump,
- M5 ve
- Koşullu Karar Ağaçları şeklindedir.

Bu çalışmada C4,5 karar ağacı algoritması kullanılmıştır. C4,5 karar ağacı algoritmasında kazanç oranı söz konusudur. Kazanç oranı bilgiye dayalı bir ölçümdür ve C4,5 karar ağacı algoritmasının kullandığı varsayılan bölme kriteridir. Algoritma tarafından kazanılan bilgiye bölme bilgisi uygulanarak normalizasyon gerçekleştirilir (67,82).

$$BölmeBilgisi_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right) \quad (2.4)$$

Formül 2.4'deki v değer sayısını, A ise eğitim verisindeki niteliği göstermektedir. Eğitim verisinin v parçaya bölündüğünde oluşan bilgiye **BölmeBilgisi_A** denilebilir. Kazanç oranını en yüksek değerde sağlayan nitelik algoritma tarafından bölme işleminin gerçekleştirileceği nitelik olarak belirlenektedir.

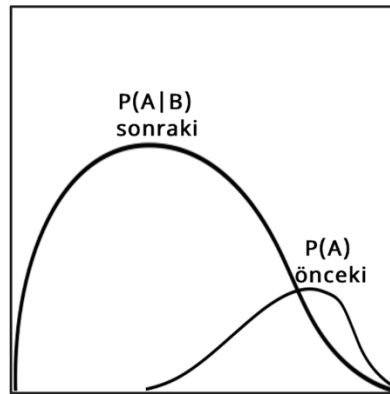
$$KazançOranı(A) = \frac{Kazanç(A)}{BölmeBilgisi_A} \quad (2.5)$$

C4,5 karar ağacı algoritmasının avantajı hem kategorik hem de nümerik veriler ile çalışabilmesidir (83).

Naive Bayes Sınıflandırıcı Algoritması

Bayesci yöntemler, sınıflandırma ve regresyon gibi problemler için Bayes'in Teoremini açıkça uygulayan yöntemlerdir. En popüler Bayes algoritmaları;

- Naive Bayes Sınıflandırıcı
- Gaussian Naive Bayes
- Çok Terimli (Multinomial) Naive Bayes
- Ortalamalı Tek Bağımlı Tahminciler (AODE, Averaged One-Dependence Estimators)
- Bayes İnanç Ağı (BBN, Bayesian Belief Network)
- Bayes Ağı (BN, Bayesian Network) şeklindedir.



Şekil 3. Bayes algoritmaları şematik gösterim

Naive Bayes Sınıflandırıcı (NBC, Naive Bayes Classifier) algoritması diğer Bayesci yöntemler gibi Bayes Teoremi'nin (Formül 2.6) uygulanması esasına dayanır. Sınıflandırıcı uygulanması sonucu girdi örneğinin hedeflenen niteliğin sınıf değerinde olup olmadığında dair olasılıklar hesaplanır. $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m)$ kümesini örnek kümesi olarak, $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n)$ kümesini de sınıf kümesi olarak varsayarsak Naive Bayes sınıflandırıcı algoritmasının hedefi $P(\mathbf{A}|\mathbf{B})$ “sonraki” olasılık değerinin mümkün olan en büyük değerine ulaşmaktır. Bu sayede maksimum sonsal hipoteze ulaşılmış olunur (67).

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{A}) \times P(\mathbf{A})}{P(\mathbf{B})} \quad (2.6)$$

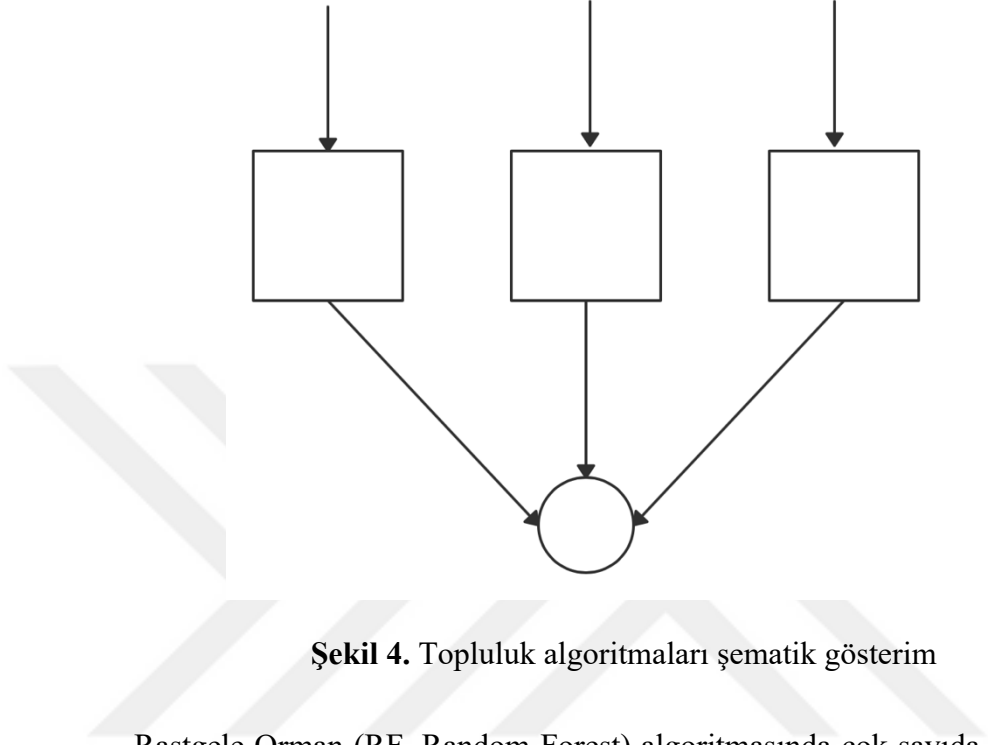
$P(\mathbf{B})$ “marjinalleştirme” olasılık değeri tüm sınıflarda aynı olduğu için bir değişken gibi değerlendirilmez. Bu durumda $P(\mathbf{B}|\mathbf{A}) \times P(\mathbf{A})$ sırasıyla “anlık olasılık” ve “önceki” olasılık değerlerinin çarpımının maksimum olması $P(\mathbf{A}|\mathbf{B})$ “sonraki” olasılık değerini de maksimum yapacaktır. Bu sayede her seferinde hem (tecrübe ile bilgiye dönüşmüş olan) $P(\mathbf{A})$ “önceki” olasılık değeri hem de $P(\mathbf{A}|\mathbf{B})$ “sonraki” olasılık değerleri artırılarak önceki deneyimlerden kazanımla makine öğrenimi gerçekleşmiş olur.

Rastgele Orman Algoritması

Topluluk yöntemleri bağımsız olarak eğitilmiş ve tahminleri genel tahmini yapmak için birleştirilen çok sayıda daha zayıf modelden oluşan modellerdir. Hangi tür zayıf öğrencilerin bir araya getirileceği ve bunların hangi kombinasyonda ve nasıl birleştirileceği konusunda çok çaba harcansa da topluluk algoritmaları oldukça güçlü algoritmalar. Bu sebeple sıklıkla tercih edilirler. En popüler topluluk algoritmaları;

- Artırma (Boosting),
- Çuvallama (Bootstrapped Aggregation),
- AdaBoost,
- Karıştırma/ Ağırlıklı Ortalama,
- Yığınlama/ Yığılmış Genelleme (Stacked Generalization),
- Gradyan Artırma Makineleri (GBM, Gradient Boosting Machines),

- Gradyan Artırılmış Regresyon Ağaçları (GBRT, Gradient Boosted Regression Trees),
- Rastgele Orman (RF, Random Forest) şeklindedir.



Şekil 4. Topluluk algoritmaları şematik gösterim

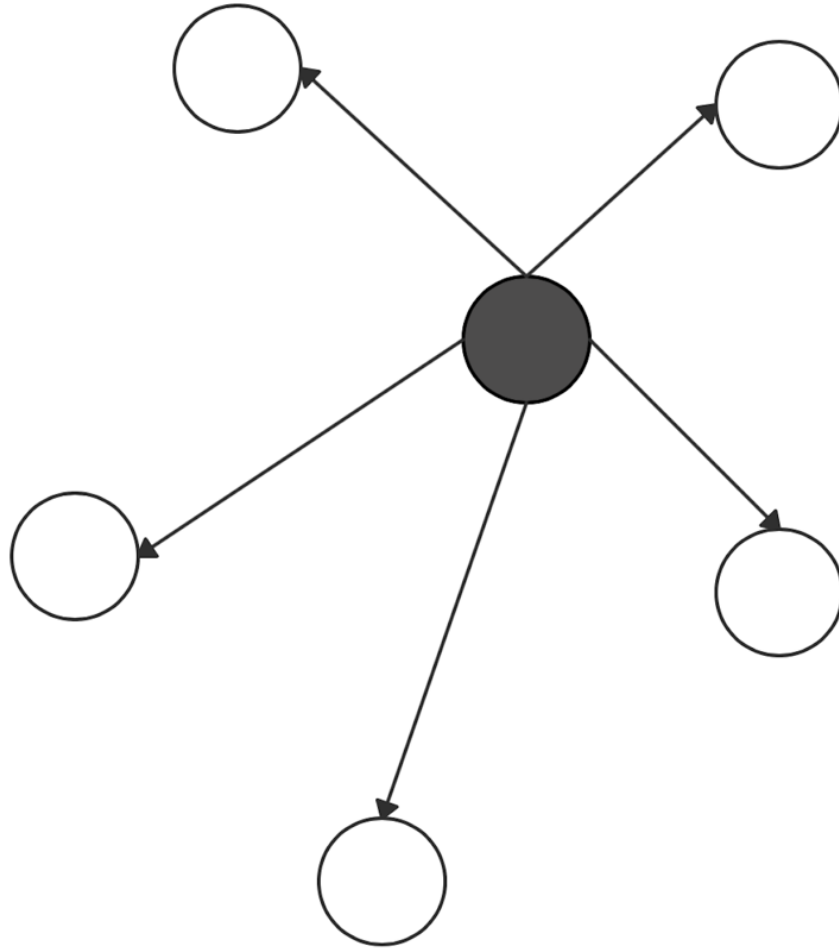
Rastgele Orman (RF, Random Forest) algoritmasında çok sayıda karar ağacı üretilir. Amaç sınıflandırma esnasında sınıflandırma değerinin yükseltilmesidir. Birer birer oluşturulan karar ağaçlarının bir araya gelmesiyle karar ormanı oluşmuş olur. Şekil 4. deki karelerin her biri Şekil 2. deki karar araçlarını temsil etmektedir. Karar ağaçlarına göre anlaması ve yorumlaması daha kompleks algoritmalardır. Daha kesin sonuçlar verebilirler. Karar ağacı algoritmalarına göre aşırı uyum gösterme riski daha azdır. Sınıflandırmada karar ormanındaki ağaçlar içerisinde oyu en yüksek olan seçilir. Regresyon durumunda ise ağaçların çıktı ortalamaları alınabilir. Kararlı algoritmalar olmalarının yanında hem sınıflandırma hem de regresyon için kullanılabilir olmaları büyük bir artıdır. Sınıflandırma açısından avantajlı olsalar da regresyon açısından dezavantajlı sayılabilirler.

Destek Vektör Makineleri

Örnek tabanlı öğrenme modeli, model için önemli veya gerekli görülen eğitim verilerinin örnekleriyle ilgili bir karar problemidir. Bu tür yöntemlerde örnek bir veri

tabanı oluşturulur. Benzerlik ölçüsü kullanılarak en iyi eşleşmeyi bulmak ve tahmin yapmak için yeni veriler oluşturulan örnek veri tabanı ile karşılaştırılır. Bu nedenle örnek tabanlı yöntemler “Kazananların tümünü alan yöntemler” veya “Bellek tabanlı öğrenme yöntemleri” olarak da adlandırılır. Depolanmış örnekler ve örnekler arasında kullanılan benzerlik ölçülerinin temsiline odaklanır. En popüler örnek tabanlı öğrenme algoritmaları;

- k-En Yakın Komşu (kNN, k-Nearest Neighbor),
- Vektör Nicemesini Öğrenme (LVQ, Learning Vector Quantization),
- Kendi Kendini Düzenleyen Harita (SOM, Self-Organizing Map),
- Yerel Ağırlıklı Öğrenme (LWL, Locally Weighted Learning),
- Destek Vektör Makineleri (SVM, Support Vector Machines) şeklindedir.



Şekil 5. Örnek tabanlı algoritmalar şematik gösterim

2.3.2.5. Sonuçların Geliştirilmesi

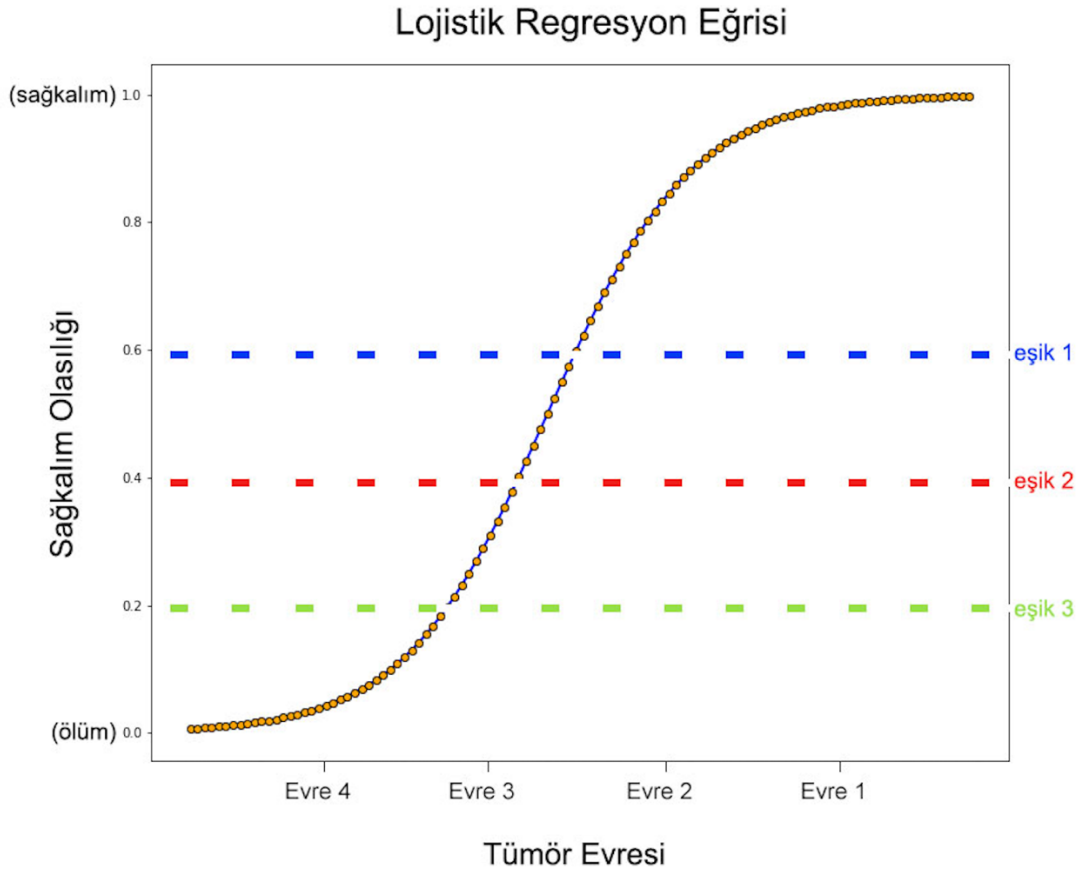
Makine öğrenmesi için kullanılan algoritmaların performans sonuçlarının değerlendirilmesi gerekir.

Makine öğrenmesi için kullanılan algoritmalar aldıkları girdileri lojistik regresyon eğrilerinde ikili karşılaştırmaya sokarak sınıflandırır (Şekil 6.). Bu analizde lojistik regresyon eğrisinin üzerindeki her noktanın y eksenindeki değeri ilgili bağımlı değişkenin olasılığını ($P(x)$) olasılık fonksiyonunun çıktısı) göstermektedir. Aynı grafikte eğrinin üzerindeki her noktanın x eksenindeki değeri ise ilgili çalışmaya ait bağımsız değişkene ait değeri göstermektedir. Şekil 6.'daki lojistik regresyon eğrisinde bu tez çalışmasına ait bir bağımsız değişken olan *tümör evresi* ile bağımlı değişken olan *sağkalım olasılığı* örnek olarak seçilmiştir.

Öncelikle sınıflandırma için bir eşik değeri gerekir. Bu eşik değeri ilk denemede rastgele seçilir. Örneğin Şekil 6.'daki *eşik 1* değerinin rastgele seçilen ilk eşik olduğunu varsayalım. Bu durumda Evre 1 ve Evre 2 hastalar için eğri üzerindeki y değeri eşik 1'in üzerinde kaldığından ($P(x) > 0,6$) algoritma bağımlı değişkenin sonucunu sağkalım olarak sınıflandırırken, Evre 3 ve Evre 4 hastalar için eğri üzerindeki y değeri eşik 1'in altında kaldığından algoritma bağımlı değişkenin sonucunu ölüm olarak sınıflandıracaktır. Bu sınıflandırmalar algoritmanın yaptığı ilk tahminler olarak düşünülebilir. Kullanılan algoritma modeli gerçekleştirdiği sınıflandırma ile hastalara ait gerçek sağkalım sonuçlarını doğru tahmin etmiş olabileceği gibi yanlış tahmin etmiş de olabilir. İşte bu yüzden sınıflandırma sonuçları ve gerçek sonuçlar karışıklık matrisine yerleştirilerek bir tablo oluşturulur (Tablo 1.). Alt başlıkta daha detaylı bahsedilen Karışıklık matrisi (Confusion matrix), performans sonuçlarının değerlendirilmesi için kullanılan performans değerlendirme ölçülerinden biridir. Karışıklık matrisi tamamlandıktan sonra sağkalım olasılık eşiği 0,6 olduğu durumda lojistik regresyonu değerlendirmek için Formül 2.14'e göre Duyarlılık (REC) ve Formül 2.15'e göre Özgüllük (SPC) değerleri hesaplanır ve hesaplanan sonuçlar kullanılan algoritma modeli tarafından saklanır.

Algoritma, eşik değeri değiştirerek yeni sınıflandırmalar ve ardından yeni karışıklık matrisleri oluşturup tablo sonuçlarına göre REC ve SPC değerlerini

hesaplayarak yukarıda detaylıca bahsedilen işlemleri tekrarlar. Bu sefer oluşan yeni REC ve SPC değerlerini bir önceki eşik değeri kullanımı ile oluşan REC ve SPC sonuçlarıyla karşılaştırarak REC ve SPC değerlerinin iyileşip iyileşmediğini kontrol eder. İyileşme olduysa eşik değerini değişiklik yaptığı yönde, olmadıysa aksi yönde değiştirerek en iyi REC ve SPC değerlerine erişesiye kadar tüm bu işlemleri tekrarlar (84).



Şekil 6. Lojistik regresyon eğrisi eşik seçimi

En iyi eşik değerinin hangisi olduğuna karar verme süreci için sadece en iyi REC ve SPC değerlerinin bulunması tek başına yeterli olmamaktadır. Aynı zamanda yine bir diğer alt başlıkta anlatılan ROC (Alıcı işlem karakteristiği, Receiver Operating Characteristic) eğrisinin oluşturularak AUC (Bu eğri altında kalan alan, Area Under the Curve)'nin hesaplanması gerekmektedir. Bu hesaplama ile ilgili bilgiye *Alıcı İşlem Karakteristik Eğrisi Analizi* başlığında yer verilmiştir.

Karışıklık Matrisi

Karışıklık matrisi sınıflandırıcıları karşılaştırarak hangisinin daha etkili olduğunu bulmaya yarayan yöntemlerden biridir. Karışıklık matrisi kontenjans tablosu, olabilirlik çizelgesi yada hata matrisi olarak da isimlendirilir (85). Algoritmanın tahmin ettiği çıktılar ile gerçek çıktılar karşılaştırılarak sınıflandırma sonuçları ikili olarak değerlendirilir. Hem gerçek çıktının pozitif hem de tahmin çıktısının pozitif olduğu durumlar doğru pozitif (dp), gerçek çıktının pozitif tahmin çıktısının negatif olduğu durumlar yanlış negatif (yn), gerçek çıktının negatif tahmin çıktısının pozitif olduğu durumlar yanlış pozitif (yp) ve hem gerçek çıktının negatif hem de tahmin çıktısının negatif olduğu durumlar doğru negatif (dn) olarak değerlendirilir (Tablo 1.) (86–88).

Tablo 1. Karışıklık matrisi

		Gerçek		
		Pozitif	Negatif	Toplam
Tahmin	Pozitif	Doğru Pozitif (dp)	Yanlış Pozitif (yp)	$tPoz$
	Negatif	Yanlış Negatif (yn)	Doğru Negatif (dn)	$tNeg$
	Toplam	poz	neg	m

Tablo 1.'e göre;

$$m = dp + yp + yn + dn \quad (2.7)$$

$$tPoz = dp + yp \quad (2.8)$$

$$tNeg = yn + dn \quad (2.9)$$

$$poz = dp + yn \quad (2.10)$$

$$neg = yp + dn \quad (2.11)$$

olmaktadır. Sınıflandırmanın doğruluğu (*accuracy*) ve hata oranı (*error rate*) Formül 2.12 ve Formül 2.13'e göre hesaplanmaktadır.

$$\text{Doğruluk (ACC)} = \frac{dp+dn}{m} = \frac{dp+dn}{dp+yp+yn+dn} \quad (2.12)$$

$$\text{Hata oranı (ERR)} = 1 - \text{ACC} \quad (2.13)$$

Pozitif sınıf etiketleri sınıflandırıcı tarafından tahmin edilirken bu tahminin etkinliği **duyarlılık** (*sensitivity, recall* yada *true positive rate*) olarak adlandırılmaktadır. Toplam pozitif örnek içerisindeki doğru sınıflandırılan pozitif sınıf etiketine sahip örnek sayısı duyarlılık sonucunu verir (Formül 2.14).

$$\text{Duyarlılık (REC)} = \frac{dp}{\text{poz}} = \frac{dp}{dp+yn} \quad (2.14)$$

Negatif sınıf etiketleri sınıflandırıcı tarafından tahmin edilirken bu tahminin etkinliği ise **özgüllük** (*specificity, true negative rate*) olarak adlandırılmaktadır. Toplam negatif örnek içerisindeki doğru sınıflandırılan negatif sınıf etiketine sahip örnek sayısı özgüllük sonucunu verir (Formül 2.15).

$$\text{Özgüllük (SPC)} = \frac{dn}{\text{neg}} = \frac{dn}{yp+dn} \quad (2.15)$$

Toplam pozitif tahmin edilen örnek içerisindeki doğru sınıflandırılan pozitif sınıf etiketine sahip örnek sayısı **pozitif tahmin değeri** (*positive predictive value*) (Formül 2.16), Toplam negatif tahmin edilen örnek içerisindeki doğru sınıflandırılan negatif sınıf etiketine sahip örnek sayısı **negatif tahmin değeri** (*negative predictive value*) verir (Formül 2.17).

$$\text{Pozitif tahmin değeri (PPV)} = \frac{dp}{tPoz} = \frac{dp}{dp+yp} \quad (2.16)$$

$$\text{Negatif tahmin değeri (NPV)} = \frac{dn}{tNeg} = \frac{dn}{yn+dn} \quad (2.17)$$

Duyarlılık ve pozitif tahmin değerlerinin harmonik ortalaması iki ölçünün birden değerlendirilmesini sağlar ve **F puanı** (*F score*) olarak adlandırılır (Formül 2.18).

$$\text{F Puanı (FS)} = \frac{2 \times \text{PPV} \times \text{REC}}{\text{PPV} + \text{REC}} = \frac{2 \times \frac{dp}{dp+yp} \times \frac{dp}{dp+yn}}{\frac{dp}{dp+yp} + \frac{dp}{dp+yn}} \quad (2.18)$$

Tüm negatif örnekler içerisinde aslında negatif olup yanlışlıkla pozitif olarak sınıflandırılan sınıf etiketine sahip örneklerin sayısı **yanlış pozitif oranını** (*false positive rate*) verir. (Formül 2.19). Yanlış pozitif oranı özgüllüğün değildir.

$$\text{Yanlış pozitif oranı (FPR)} = 1 - \text{SPC} = \frac{yp}{neg} = \frac{yp}{yp+dn} \quad (2.19)$$

Tüm pozitif örnekler içerisinde aslında pozitif olup yanlışlıkla negatif olarak sınıflandırılan sınıf etiketine sahip örneklerin sayısı ise, **yanlış negatif oranını** (*false negative rate*) verir. (Formül 2.20). Yanlış negatif oranı duyarlılığın değildir.

$$\text{Yanlış negatif oranı (FNR)} = 1 - \text{REC} = \frac{yn}{poz} = \frac{yn}{dp+yn} \quad (2.20)$$

Alıcı İşlem Karakteristik Eğrisi Analizi

ROC (Alıcı işlem karakteristiği, Receiver Operating Characteristic) eğri analizinin temel amacı ilgili algoritma modelinin lojistik regresyon sonucuna göre yaptığı sınıflandırma sonuçlarına dair karmaşık bilgilerin tümünü basitçe özetlemektir. ROC grafiğinde y eksenini doğru pozitif oranı yani REC değerini gösterirken, x eksenini ise yanlış pozitif oranı yani FPR değerini göstermektedir.

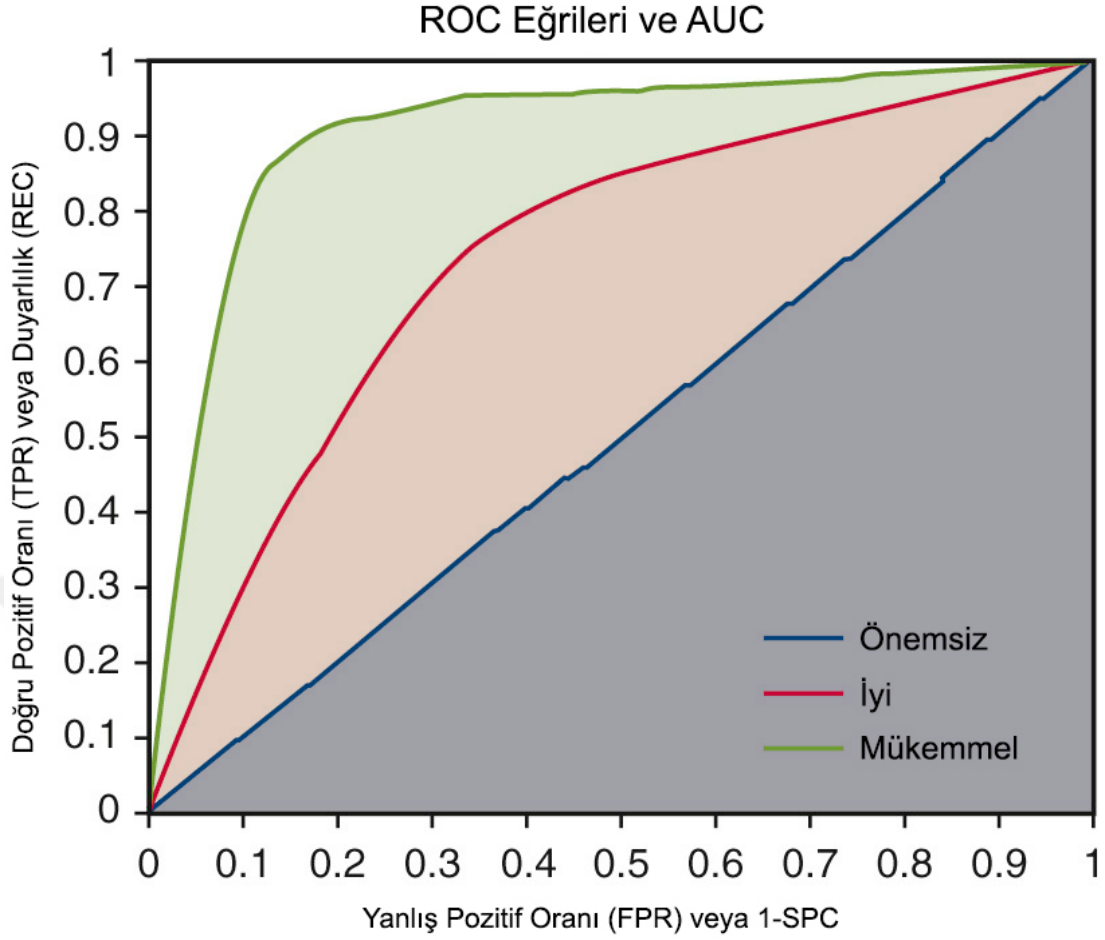
Lojistik regresyon eğrisinde sınıflandırma için kullanılan eşik değerlerin her biri için hesaplanan REC ve FPR sonuçları ROC grafiğinde minimum 0, maksimum 1 olan x ve y eksenlerinde ilgili alanda nokta bilgisi oluşturur. ROC grafiğinde eğimi 45 derece olan $y=x$ doğrusu üzerindeki diğer her nokta gibi (0,0) noktası ile (1,1) noktası istenmeyen durumların olduğu noktalar.

- (1,1) noktası; hem REC hem FPR değerinin maksimum olduğu mutlak pozitif sınıflandırmaları yani rastgele tahmini gösterirken,
- (0,0) noktası ise; pozitif sınıflandırmanın hiçbir zaman elde edilemediği yani yine rastgele tahminin yapıldığına işaret etmektedir.
- (0,1) noktası; REC değerinin maksimum olduğu, FPR değerinin ise minimum olduğu mükemmel sınıflandırmayı işaret ederken,
- (1,0) noktası ise; REC değerinin minimum olduğu, FPR değerinin ise maksimum olduğu olabilecek en kötü sınıflandırmayı işaret eder.

Kısaca ROC grafiđi oluřurken ortaya ıkacak tm noktaların (0,1) noktasına mmkn olduđunca yakın olması arzu edilir.

Lojistik regresyon eđrisinde sınıflandırma iin kullanılan eřik deđerlerin her biri iin hesaplanan REC ve FPR sonuları ROC grafiđine bir nokta bilgisi olarak yerleřtirilip bu noktalar birleřtirildiđinde ortaya bir eđri ıkar (ROC eđrisi). Bu eđrilerin altında kalan alana ise AUC (Area Under the Curve) denmektedir (řekil 7.). AUC'nin temel amacı ise makine đrenimi sınıflandırma problemlerinde kullanılan algoritma tahminlerinin sonularını deđerlendirip ortaya algoritma performans sonucu ıkarmaktır. AUC alanının byklđne gre algoritmanın tahminlerinin dođruluđu anlařılır. Alan bydđđ zaman mkemmel tahmine, kldđđ zaman kt yada nemsiz tahmine yaklařılır.

AUC sonularına gre đrenme problemi iin kullanılan farklı algoritma modelleri karřılařtırılıp hangi makine đrenimi algoritma modelinin problemin zm iin daha etkin performans gsterdiđi bulunmuř olur (89,90).



Şekil 7. ROC eğrilerinin karşılaştırılması ve AUC

2.3.2.6. Sonuçların Sunumu

Eğitilen makine öğrenmesi algoritmaları test edilip performansları değerlendirildikten sonra öğrenme probleminin çözümü için kullanılmaya hazırdır. Daha önceki adımların herhangi birinde oluşan bir sorunun çözümü için her seferinde birer adım geriye dönülüp bir önceki adım kontrol edilir. Sorun kontrol edilen adımda değilse kontrol edilen adımdan öncesine dönülerek diğer adımlar teker teker kontrol edilir. Sorunun bulunduğu adımdaki problem giderilerek makine öğrenimi sürecine devam edilir (Şekil 11.).

Makine öğrenmesi probleminin çözümü tamamlandığında öğrenen makine algoritması kullanılmaya hazırdır. Bu aşamada model ilgili çalışmaya göre uygulamaya geçirilip sunulmalıdır. Bu çalışmada kolon kanserinde hastalara ait klinik ve genetik verilerle sağkalım tahmini yapabilen modelin son kullanıcılarla

buluşturulabilmesi için Shiny aracı kullanılarak ilgili klinik karar destek aracının web arayüzüne aktarılarak kullanıma sunulması hedeflenmiştir.

2.3.3. Derin Öğrenme

Derin öğrenme de; makine öğrenmesinin bir alt kümesi gibi düşünülebilir. Yapay zeka teknolojisinin en özelleşmiş halidir. Derin öğrenme modelleri hem denetimli hem de denetimsiz öğrenme ortamlarına uygulanabilir. Derin öğrenme, özellikle sayısallaştırılmış görüntülere ve örüntü tanımaya uygulandığında verimli olmaktadır. Bu uygulamaya en güzel örnek sıklıkla kullanılan bir yapay sinir ağı olan CNN (Evrişimli Sinir Ağı, Convolutional Neural Network)'dır (91). CNN'lere girdi olarak ham bir görüntü verilir ve temelde yatan özellikleri bir eğitim süreci ile öğrenirler. CNN, her biri önceki katmandan girdi alan, onu işleyen ve sonraki katmana çıktı veren birden çok katmandan oluşur. Her katmanın çıktısı, kendinden bir sonra gelen katmanın girdisi olacak şekilde işlenir. İlk katmanın girdisi ilgilenilen veri kümesi iken, son katmanın çıktısı bir sınıflandırma etiketi veya veri kümesinin diğer değerlendirilebilir özelliğidir. Tıpta ve hesaplamalı biyolojide son 5 yılda tıbbi görüntü analizi, intraoperatif görüntüleme ve genomik için farklı derin öğrenme yaklaşımları önerilmiş ve uygulanmıştır (92–97).

Derin öğrenme tekniklerinin test edilmesi makine öğrenmesi tekniklerine göre daha zordur. Tamamen bağımsız ve düzenleyici amaçlar için gizli doğrulama veri setlerine uygulama gerektirir. Bununla birlikte, derin öğrenme, bilgisayarın yeni özellikleri 'görmesini' sağlama veya insan oksipital lobunun kapasitesini aşacak kadar ince olan dijitalleştirilmiş kodlanmış bilgilerdeki zayıf sinyalleri tanımasını sağlama potansiyeline sahiptir. Kritik olarak aşağı akış sonuçlarına bağlı olabilir ve bu sonuçları tahmin edebilir. Bu özellik, tahmin edilen sonuçlara göre özelleştirilmiş hassas bakım yollarının sağlanmasını sağlayarak sağlık hizmeti sunumunu olumlu bir şekilde geliştirebilir.

Gürültülü veriler ve deneysel sınırlamalar modellerin klinik kullanımını azaltsa da, derin öğrenme yöntemleri, katmanlı otomatik kodlama analizleri yoluyla verilerin boyutluluğunu azaltarak bu sınırlamaları giderebilir. Örneğin; DeepBTS olarak isimlendirilen derin öğrenme modeli ile özellikle evre-1 hastalarda küçük hücreli

olmayan akciğer kanseri prognozlarının geleneksel modellerden daha iyi tespit edilebilir (98). Başka bir çalışmada bazal hücreli karsinomu saptamak ve malign lezyonları iyi huylu lezyonlardan ayırmak için cildin 308 histopatoloji bölgesinden 1.400'den fazla görüntünün analizi sonucu derin öğrenme algoritmasının uzmanlara kıyasla $>90\%$ tanısal doğruluk sağlayabildiği gösterilmiştir (99). Ve yine başka bir çalışmaya göre de yoğun ve yoğun olmayan meme dokusunun belirlenmesi için yürütülen meme mamogramlarının 41.000'den fazla dijital tarama ile incelenmesi sonucunda, 10.763 derin öğrenme değerlendirmesinin 94% 'ü uzman radyologlar tarafından doğru kabul edilmiştir (100).

2.4. KANSER

2.4.1. Kanser Epidemiyolojisi

Kanser, dünyada yıllık 8.8 milyon ile iskemik kalp hastalığı ve diğer kalp hastalıkları ile beraber en çok ölüme sebep olan nedenlerdendir (101). Türkiye'de ise yaklaşık 20% ile iskemik kalp hastalığı ve diğer kalp hastalıklarından sonra ikinci ölüm nedenidir (102). Kanser coğrafi dağılım, cinsiyet, yaş, sosyoekonomik durum, mesleki özellikler gibi epidemiyolojik özellikler olduğu bilinmektedir (103). Kanser sıklığı bölgelere, ülkelere hatta aynı ülkenin farklı yerleşim yerlerine göre değişmektedir. Bunun nedeninin kalıtım ya da çevresel risk etmenleri olabileceği düşünülmektedir. Kanser sıklığı ve türleri cinsiyet ile ilişkili olarak da farklılık göstermektedir. Türkiye'de her 100.000 kişide en sık görülen ilk 10 kanser erkeklerde; trakea-bronş-akciğer (53.8), prostat (31.3), kolorektal (21.8), mesane (19.7), mide (14.2), non-hodgkin lenfoma (7.2), larinks (6.4), böbrek (6.3), tiroid (6.3), pankreas (5.9) iken kadınlarda; meme (41.6), tiroid (24.2), kolorektal (13.2), uterus korpusu (10.1), trakea-bronş-akciğer (8.2), mide (6.4), over (5.9), non-hodgkin lenfoma (5.1), uterus serviksi (4.2), beyin-sinir sistemi (3.9) şeklindedir (104). Genel olarak kanser riski yaşla birlikte artmaktadır. Kanser türlerinde de yaşa bağlı görülme sıklıkları değişmektedir. Sosyoekonomik durumu (SeD) düşük olan ülkelerde kanser sıklık oranı, kanser vakalarının kayıtları, değişik kanserlerin görülme sıklığı, kanser patolojileri, klinik kanser karakterlerinde artış olduğu görülmüştür. Bu ülkelerdeki SED düşük bireylerde kanserin temel risklerinin daha fazla görülmesinin; bu kişilerin sağlık hizmetlerine daha az ulaşan grup olmaları, yetersiz ve dengesiz beslenme,

üreme sağlığı, enfeksiyonlar, mesleksel karsinojenlere maruziyet, alkol ve sigara kullanımı ile ilişkili olduğu düşünülmektedir. Kanseri sıklığı ve türlerinin farklılık gösterdiği bir diğer parametre ise meslektir. Kanseri oluşumunda mesleksel etmenlerin etkisi ilk olarak 1775 yılında Pott tarafından, baca temizleyiciliği ve skrotum kanseri ilişkisinden yola çıkarak ortaya atılmıştır. Daha sonra yapıştırıcılar içindeki çözücüde bulunan, benzen adlı kimyasal maddenin lösemi riskini, aşırı güneşte kalmanın (UV radyasyonu) açık havada çalışanlarda cilt kanseri riskini, hepatit B-C ile karaciğer kanseri ilişkisinin sağlık çalışanları üzerindeki riski, asbestin ise endüstri çalışanlarında mezotelyoma ve akciğer kanseri riskini arttırdığı anlaşılmıştır. Kanseri oluşumunda fiziksel karsinojen, UV ve iyonizan radyasyon, kimyasal karsinojenler, asbest, sigara dumanı bileşenleri, aflatoksin, arsenik, biyolojik karsinojenler, virüs, bakteri, parazit gibi çevresel faktörlerin yanı sıra yaşlanma ve genetik faktörlerde etkilidir (105). Araştırmacıların yaptıkları çalışmalara göre; risk faktörlerinin %90-95 kadarını çevresel faktörler oluştururken, %5-10 kadarını ise genetik faktörler oluşturmaktadır. Çevresel faktörlerin ise %30-35'ini diyet, %25-30'unu tütün kullanımı, %15-20'sini enfeksiyonlar, %10-20'sini obezite, %4-6'sini alkol kullanımı, ve %10-15'ini diğer durumlar oluşturmaktadır. Kanserle ilgili ölümlerin 1/3'ü 5 risk faktörüne bağlıdır. Bunlar; BKİ (Beden Kitle İndeksi)'nin yüksek olması, meyve ve sebze alımının düşük olması, fiziksel aktivite eksikliği, tütün kullanımı, ve alkol kullanımıdır (106). Obezite küresel boyutta önemli bir sağlık sorunudur. DSO; obeziteyi "Sağlığı bozacak ölçüde vucutta aşırı yağ birikmesi" olarak tanımlamıştır. Yine DSO'ye göre dünyada obezite 1975'li yıllardan beri 3 katına ulaşmıştır. Obezite ve kanserin doğrudan ilişkisi 2014- 2015 yıllarında gösterilmiştir (107). Sigara ise dünyada tek başına en önemli kanser nedenidir. Dünyada kansere bağlı ölümlerin %22'sinden sorumludur. Her yıl yaklaşık 6 milyon insanın ölümüne neden olmaktadır. Tütün kullanımı tüm kanserlerin %15'lik kısmından sorumludur (DSO, 2018). Yine Hepatit B, C enfeksiyonlarının karaciğer kanserine, HPV'nin serviks kanserine, *H.Pylori*'nin mide kanserine, Epstein Barr virüsünün Burkitt Lenfoma'ya ve nazofarengeal karsinoma, HIV'nin ise Kaposi sarkomu ve non-hodgkin lenfomaya neden olduğu gösterilmiştir. Hiç doğum yapmamış kadınların meme, endometrium ve over kanseri için risk etmeni olmaları, bunun yanında erken cinsel ilişki ve çok sayıda cinsel eşliliğin de serviks kanseri açısından risk etmeni olması, üreme sağlığı

etmenlerinin kanser ile ilişkisini göstermektedir. Tüm bu epidemiyolojik bilgiler ışığında; çevresel etmenler kanser oluşumunda genetik etmenlerin önüne geçmektedir. Bu durum genetik faktörlerin de değerlendirme dışında tutulması gerektiği anlamına gelmemektedir.

2.4.2. Kolorektal Kanser

2.4.2.1. Kolorektal Kanser Epidemiyolojisi

20- 30 sene öncesinde nadiren teşhis edilen kolorektal kanser günümüzde yılda yaklaşık 900.000 ölümle dünyanın dördüncü en ölümcül kanseridir (108). Kolorektal kanser, dünya çapında her yıl teşhis edilen tüm kanserlerin ve kansere bağlı ölümlerin yaklaşık %10'unu oluşturmaktadır (109). Kadınlarda en sık görülen ikinci ve erkeklerde en sık görülen üçüncü kanserdir. Kadınlarda insidans ve ölüm oranı erkeklere göre yaklaşık %25 daha düşüktür. Bu oranlar gelişmiş ülkelerde görülme sıklığının artmasının yanında coğrafi olarak da değişiklik gösterir (110). Gelişmekte olan ülkelerde devam eden ilerlemeyle, dünya çapında kolorektal kanser insidansının 2035 yılında 2- 5 milyon yeni vakaya yükseleceği öngörülmektedir (111). Azalan gidişat yalnızca çok gelişmiş ülkelerde görülmektedir. Bunun nedeni bu ülkelerin popülasyonlarının yaşam tarzı ve diyet değişikliklerinin yanında daha da önemli olarak ülke çapında tarama programlarına ve kolonoskopi alımının artmasına bağlanmıştır (112). Bu durumun tersine, kolorektal kanser semptomları ile sağlık kuruluşlarına başvuran 50 yaşın altındaki hastalarda özellikle rektal kanser ve sol taraflı kolon kanserinde endişe verici bir artış gözlenmiştir (113,114). Genetik, yaşam tarzı, obezite ve çevresel faktörlerin rolü olsa da, bu artışın nedenleri tam olarak anlaşılamamıştır.

2.4.2.2. Kolorektal Kanser Risk Faktörleri

Kolorektal kanserin gelişiminde hem kalıtsal hem de çevresel risk faktörleri rol oynar. Yapılan epidemiyolojik çalışmalarda, erkek cinsiyeti ve artan yaş, hastalık insidansı ile ilişkili bulunmuştur (115,116). Kolorektal kanser için risk faktörlerini en temelde değiştirilebilir, değiştirilemez ve diğer faktörler olarak üçe ayırabiliriz;

Değiştirilebilir Risk Faktörleri

Sigara kullanımı, alkol kullanımı, işlenmiş et tüketimi, kırmızı et tüketimi, yetersiz sebze/meyve tüketimi, fazla vücut yağ oranı ve obezite kolorektal kanser riskini arttıran değiştirilebilir risk faktörleridir (117–120).

Fiziksel aktivite, tam tahıllı gıdaların tüketimi, lifli diyet, günlük süt ürünleri tüketimi, balık tüketimi, badem, ceviz, fındık gibi kuruyemişlerin tüketimi, D ve C vitamini başta olmak üzere yeterli vitamin alımı ve kalsiyum takviyesi kolorektal kanser riskini azaltan değiştirilebilir risk faktörleridir.

Değiştirilemez Risk Faktörleri

Kalıtsal kolorektal kanser sendromları ve pozitif aile öyküsü değiştirilemez kalıtsal risk faktörleriyken, etnik köken, erkek cinsiyetinde olmak, tip 2 diyabet, inflamatuvar bağırsak hastalığı değiştirilemez diğer risk faktörleridir (121).

Diğer Faktörler

Aspirin veya steroid yapıda olmayan anti-inflamatuvar ilaçların kullanımı, menopozal hormon tedavisi, statin içerikli ilaçların kullanımı kolorektal kanser riskini azaltan diğer risk faktörleridir.

Kolorektal kanserli tüm hastaların yaklaşık %10-20'sinde pozitif aile öyküsü yer almaktadır. Pozitif aile öyküsünde kolorektal kanserden etkilenen akraba sayısı ve derecesi, bu kişilerin kolorektal kanser tanısı aldıkları yaş gibi risk faktöründe değişiklik gösteren etmenler söz konusudur (122,123). İkiz kardeşler üzerinde ve farklı aileler üzerinde yapılan çalışmalara göre kolorektal kanserin kalıtsallığının %12 ile %35 arasında değiştiği düşünülmektedir (124,125). Kolorektal kanser riski ile ilgili genom çapındaki birçok çalışma yaygın tek nükleotid polimorfizmlerini içeren kansere yatkınlık genlerini başarıyla tanımlamış olsa da, kolorektal kanser kalıtımına neden olan faktörlerin çoğu hala belirsizdir ve daha fazla çalışma gerektirmektedir (126). Teşhisin konulması, kanseri önlemek için önemli bir gözetim imkanı sağlamasının yanında risk altındaki akrabalar için de uygun sürveyans tavsiyeleri sağlaması açısından oldukça önemlidir. Uzun süredir devam eden iltihaplı bağırsak

hastalığı olan ve daha önce kolorektal kanser veya adenom öyküsü olan hastalar kolorektal kanser için yüksek risk altında olduklarından hassas gözetim altında tutulmalıdır.

Kalıtsal kolorektal kanserler; polipozis olmayan “Lynch sendromu” ve polipozis olmayan ailesel kolorektal kanserler ile polipozis sendromlar olarak iki alt gruba ayrılabilirler. Polipozis sendromlar, artan polip sayısı ve polip tipi ile hekim tarafından kolayca tanınabilirken Lynch sendromu az sayıda adenoma neden olması ve adenomların morfolojik yapılarının sporadik lezyonlara benzemesi sebebiyle sıklıkla gözden kaçabilmektedir. Bu nedenle Lynch sendromuna sahip hastalardaki tümörlerin moleküler analizleri bu sendromun teşhisini kolaylaştırmak için kullanılmaktadır. Yapılan moleküler analizler sonucunda bu sendroma DNA yanlış eşleşme tamir mekanizmasındaki hataların neden olduğu anlaşılmıştır. Ayrıca immünohistokimya analizleri bu sendrom nedeniyle oluşan tümörlerin yanlış eşleşme tamir proteinlerinin de eksik olduğunu göstermiştir. Hızlı çalışan adenom-karsinom yolağı nedeniyle Lynch sendromuna sahip bireylerin 20’li yaşlarından itibaren her yıl veya maksimum iki yılda bir sık sık kolonoskopi yaptırmaları önerilmektedir (127,128). Ayrıca bu bireyler ince bağırsak, mide, yumurtalık, renal pelvis, üreter ve hepatobiliyer sistem kanserleri gibi endometriyal kanserler ve diğer maligniteler için de yüksek risk altındadır.

Kolon mikrobiyotası araştırmaları *Fusobacterium nucleatum* ve *Bacteroides fragilis* gibi spesifik bakteri türleri enfeksiyonlarının da kolorektal kanser riskini artırabileceğini göstermiştir (129–131).

2.4.2.3. Kolorektal Kanser Patogenezi

Kolorektal kanserlerin çoğu poliplerden kaynaklanmaktadır. Bu süreç; anormal bir kript hücresinin kısaca polip olarak da adlandırılan neoplastik öncü bir lezyona, bu lezyonun da yaklaşık 10-15 yıllık bir süre içinde kolorektal kansere dönüşmesi şeklinde ilerler. Kolorektal kanserlerin çoğunun menşe hücresinin bir kök hücre veya kök hücre benzeri bir hücre olduğu varsayılmaktadır (132,133). Bu kanser/kanser benzeri kök hücreler, tümör baskılayıcı genleri inaktifleştirip onkogenleri aktive eden genetik ve epigenetik değişikliklerin birikimi sonucu ortaya çıkmaktadır. Tümörün

oluşumu ve ilerlemesi için gerekli olan kanser kök hücresi kolon kript hücre tabakasının altında bulunur (132,133). Bu kanser kök hücrelerinin büyümesini kontrol eden düzenleyici mekanizmaların araştırılması, olası terapötik ajanlar ve önleyici tedavi için umut verici bir araştırma alanıdır (134,135).

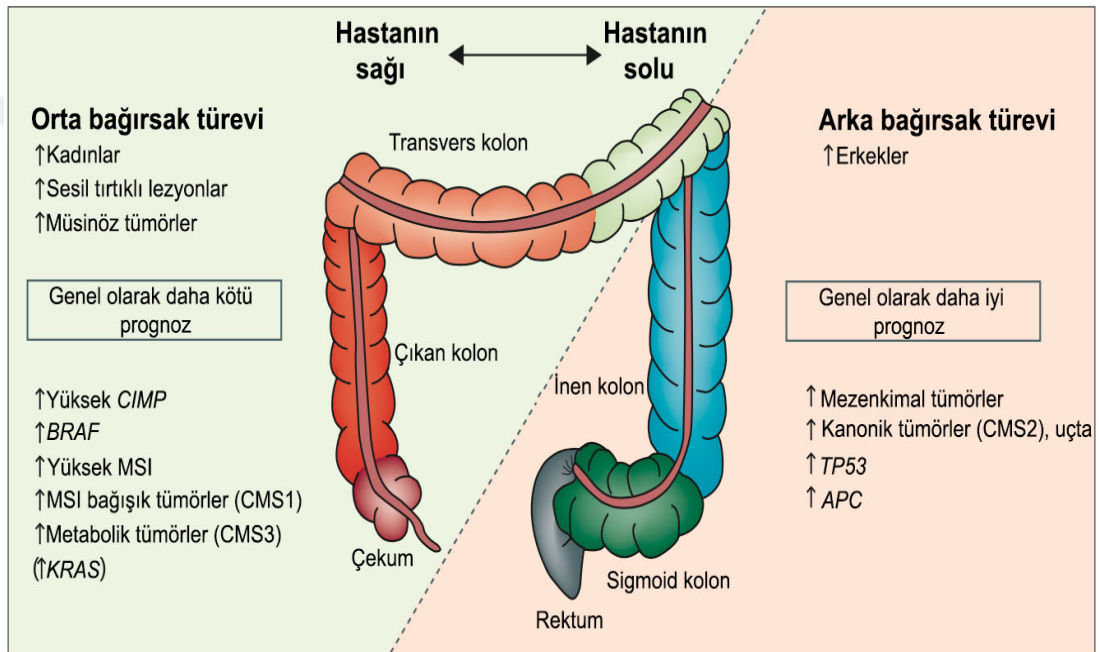
Kolorektal kanserlerin %70-90'ına yol açan kromozomal instabilite dizisi olarak da adlandırılan adenom-karsinom yolağı ile kolorektal kanserlerin %10-20'sine yol açan tırtıklı neoplazi yolağı farklı iki ana öncü lezyon yolağıdır. Bu yolaklar sıralı çok sayıda farklı genetik ve epigenetik olayları içerir (136). Kromozomal instabilite APC geninin mutasyonu ile başlar ve RAS aktivasyonu veya TP53 geninin fonksiyon kaybı ile devam eder. Tırtıklı neoplazi yolağı ise mikrosatellit stabil ve instabil kanserlere yol açan CpG adası metilasyonu ile karakterize RAS mutasyonları, RAF mutasyonları ve epigenetik instabilite ile ilişkilidir.

Sağ ve Sol Kolorektal Kanserleri

Sağ (proksimal) kolorektal kanserlerinin moleküler özellikleri sol (distal) kolorektal kanserlerden farklıdır. Moleküler farklılıkların yanı sıra sol ve sağ kolorektal kanserler arasında embriyolojik ve anatomik farklılıklar da vardır. Bu farklılıklar özellikle metastatik durum için anahtar bir role sahiptir ve anti-EGFR ilaçlara karşı oluşacak yanıtın öngörülebilmesini sağlaması konusunda giderek daha fazla kabul görmektedir (137,138).

Sağ kolorektal kanserler orta bağırsak (midgut) kökenlidir ve kolonun; çekum, çıkan kolon ve transvers kolon bölümlerinde görülür. Sol kolorektal kanserler ise arka bağırsak (hindgut) kökenli olup kolonun inen kolon ve sigmoid kolon bölümlerinde ve rektumda görülür. Sağ kolorektal kanserler daha sıklıkla kadın bireylerde görülürken, sol kolorektal kanserler daha sıklıkla erkek bireylerde görülmektedir. Sağ kolorektal kanserlerde müsinöz tümörler daha fazla görülür ve sesil tırtıklı lezyonlara (sessile serrated lesions) fazlaca rastlanır. Sağ kolorektal kanserlerde yüksek CIMP'e (CpG adası metilatör fenotipi) ve yüksek MSI'ye (mikrosatellit instabilite) daha sıklıkla rastlanır. Sağ kolorektal kanserlerde *BRAF* ve *KRAS* genlerinin ekspresyon seviyeleri yüksekken, Sol kolorektal kanserlerde *TP53* ve *APC* genlerinin ekspresyon seviyeleri yüksektir. Genel olarak sağ kolorektal kanserlerinin prognozu daha kötü seyrederken,

sol kolorektal kanserlerin prognozu daha iyi seyreder. Sağ kolorektal kanserlerin prognozunun kötü seyrindeki bu durum kanserlerin tüm evreleri için geçerli değildir. Bu kötü seyir anti-EGFR ve anti-VEGF tedavilerine yanıtla ilişkili olarak metastatik durumlarda daha sık görülür. Sağ kolorektal tümörlerin çoğunda MSI yüksek olduğundan bu tümörler, kemoterapiye nispeten dirençlidirler. Kötü prognozda bu durumun etkisi de söz konusudur. Günümüzde artan tedavi seçenekleri sayesinde bu tümörler için immünoterapi imkanı mevcut olduğundan bu tümörlerde alınan sonuçlar gelişmekte ve değişmektedir (Şekil 2.) (108,139–144).



Şekil 8. Sağ ve Sol Kolon Kanserleri arasındaki farklılıklar

Konsensüs Moleküler Alt Tipleri

Kolorektal kanserler 2014 yılında gen ekspresyonu temelinde dört moleküler alt tipte sınıflandırıldı (konsensüs moleküler alt tipleri, consensus molecular subtypes, CMS) (145). Her CMS için ayrı ve benzersiz gen ve yollar bulunmaktadır. CMS1 mikrosatellit instabilite bağışık tümörler, CMS2 kanonik tümörler, CMS3 metabolik tümörler ve CMS4 mezenkimal tümörleri ifade eden alt tiplerdir. Sağ kolorektal kanserler daha çok CMS1 ve CMS3 tümörlerdir (146). Tümörlerin sağ veya sol kolonda olması ve mutasyon durumu (RAS veya RAF) sistematik tedavilerin

seçilmesinde yardımcı faktörler olsa da CMS sınıflandırması klinik çalışmalarda prognostik ve prediktif belirteç olarak araştırılmaktadır.

2.4.3. NCI GDC Kanseri Veri Portalı

Amerikan NCI (Ulusal Kanser Enstitüsü, National Cancer Institute) tarafından CCG (Kanser Genomik Merkezi, Center for Cancer Genomics) çatısı altında TCGA (Kanser Genom Atlası, The Cancer Genome Atlas) ismiyle 2006 yılında genom dizileyip kanserden sorumlu genetik mutasyonları kataloglamak amacıyla başlatılan proje giderek kapsamlı hale gelerek 20.000’den fazla birincil kanserin moleküler karakterizasyonunu sağladı. Başlatıldığından bu yana proje kapsamında 2,5 petabayttan fazla genomik, epigenomik, transkriptomik ve proteomik veri üretildi. Üretilen veriler araştırma topluluğundaki herkesin kullanabilmesi amacıyla kamuya açık durumdadır.

NCI, TCGA ve diğer programlar ve çalışmalardan gelen genomik verileri de kapsayacak birleşik bir veri sistemi olan GDC veri portalını (Ortak Genom Verileri, Genomic Data Commons) 2016 yılında kullanıma sundu. GDC veri portalı farklı kanser türleri ile ilgili farklı veri tiplerini barındıran kamuya açık büyük veri ortamıdır (147–149).

2.4.4. AJCC TNM Evreleme Sistemi

Kanser sürveyansı ve bakım iyileştirilmesi amacıyla multidisipliner kuruluşlarla işbirliği içerisinde kanserin sınıflandırılması ve yönetimi için kanıta dayalı sistemlerin geliştirilmesi, tanıtımı ve sürdürülmesi konularında faaliyet gösteren ve dünya çapında tanınırlığı olan (150) AJCC (American Joint Committee on Cancer, Amerikan Ortak Kanser Komitesi)’nin kanser evreleme kılavuzu 6. ve 7. Baskılarında yer alan “TNM Evre Gruplama Sistemi”ne göre TNM evreleme sistemi aşağıdaki gibidir (151–154);

Tablo 2. AJCC kolorektal kanser TNM sınıflandırması

Evre Sınıflandırması			
Evre	T	N	M
0	Tis	N0	M0
I	T1-T2	N0	M0
IIA	T3	N0	M0
IIB	T4	N0	M0
IIIA	T1-T2	N1	M0
IIIB	T3-T4	N1	M0
IIIC	Herhangi	N2	M0
IV	Herhangi	Herhangi	M1

2.4.4.1. T Evresi: Birincil Tümör

TX: Birincil tümörün varlığının değerlendirilememesi durumu,

T0: Birincil tümöre rastlanmaması durumu,

Tis: *in situ* karsinom tanım başlangıcı; tümörün intraepitelyal muskularis mukozasına veya lamina propria kadar gelişmiş olma durumu,

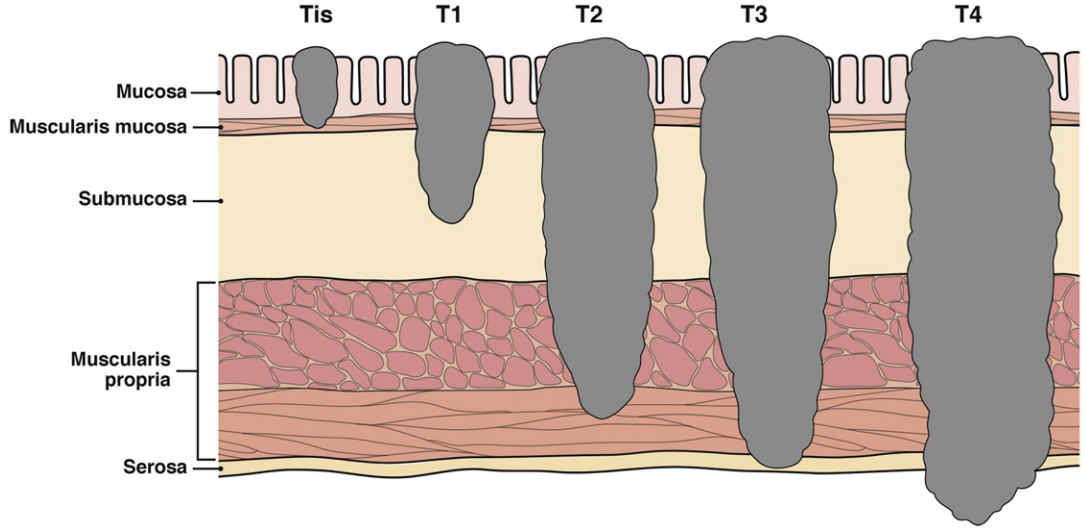
T1: Tümörün muskularis mukozasını geçerek submukozayı istila etmiş olma durumu,

T2: Tümörün muskularis propriaya kadar gelişmiş olma durumu,

T3: Tümörün muskularis propriadan seröz altı membrana kadar veya peritonealize olmayan perikolik/perirektal dokulara yayılmış olma durumu,

T4a: Tümörün seröz altı membrandan geçip visseral peritonu delmiş ancak diğer yakın dokulara veya organlara henüz ulaşmamış olma durumu,

T4b: Tümörün seröz altı membrandan geçip visseral peritonu delerek yakındaki diğer dokulara veya organlara yapışmış, istila etmiş olma durumu (151,155).



Şekil 9. AJCC kolorektal kanser sınıflandırması- Birincil tümör (T evresi)

2.4.4.2. N Evresi: Bölgesel Lenf Düğümü Metastazı

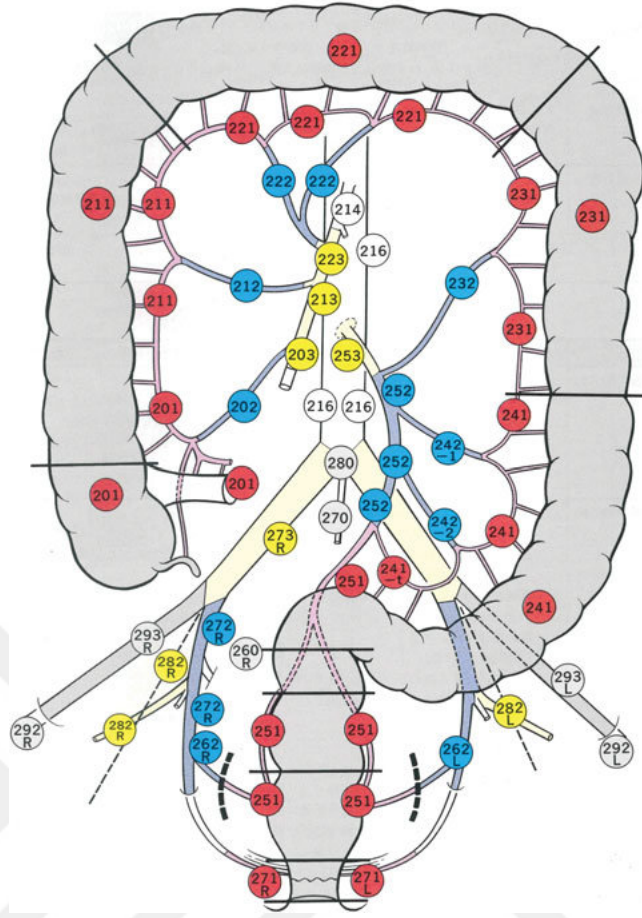
NX: Bölgesel (tümörün yakınında) lenf düğümü metastazı varlığının değerlendirilememesi durumu,

N0: Bölgesel lenf düğümü metastazının olmaması durumu,

N1: Bir, iki veya üç bölgesel lenf düğümünde metastaz varlığı durumu,

N2: Dört veya daha fazla bölgesel lenf düğümünde metastaz varlığı durumu

(151,156).



Şekil 10. Kolorektal kanser Japon sınıflaması kanser lenf düğümü haritası

2.4.4.3. *M* Evresi: Uzak Metastaz

MX: Uzak metastazın varlığının değerlendirilememesi durumu,

M0: Uzak metastazının olmaması durumu,

M1: Uzak metastazının olması durumu (151).

3. GEREÇ VE YÖNTEM

3.1. ARAŞTIRMANIN TİPİ

Bu çalışma tanımlayıcı bir araştırmadır.

3.2. ARAŞTIRMANIN YERİ VE ZAMANI

Çalışma girişimsel olmayan etik kurul onayı ve tez önerisi alındıktan sonra Dokuz Eylül Üniversitesi Onkoloji Enstitüsü Translasyonel Onkoloji Anabilim Dalı'nda yasal yüksek lisans eğitim süresi içinde yapılmıştır (2018- 2021 yılları arasında).

3.3. ARAŞTIRMANIN EVRENİ VE ÖRNEKLEMİ/ ÇALIŞMA GRUPLARI

Kamuya açık klinik veri portallarından retrospektif olarak tarama gerçekleştirileceği için ilgili tüm veriye ulaşılması planlanmaktadır. Çalışmada kontrol grubu bulunmamaktadır.

3.4. ÇALIŞMA MATERYALİ

Çalışma materyali olarak GDC veri portalından erişilen TCGA-COAD projesi kapsamındaki kolorektal kanser klinik ve genomik gen ekspresyon verileri kullanılmıştır.

3.5. ARAŞTIRMANIN DEĞİŞKENLERİ

Klinik veriler kapsamında değerlendirilen cinsiyet, ırk, tümör evresi, T evresi, N evresi, M evresi, daha önce kanser teşhisi almış veya almamış olma durumu, doku veya organ menşei, rezeksiyon veya biyopsi bölgesi, beden kitle endeksi, birincil teşhis ile genomik veriler kapsamında değerlendirilen glikoliz/glukoneogenez yolağı, glikan bozunması yolağı, pantotenat ve CoA biyosentez yolağı, apoptoz yolağı, mTORC1 sinyal yolağı ve bu yolaklarda rol alan ADH1C geni, AKR1A geni, BCAT1 geni, CAPN2 geni, CASP2 geni, MAN2B2 geni, PFKM geni, TCEA1 geni, TOMM40 geni araştırmanın bağımsız değişkenleridir.

Hastanın sağkalım tahmini ise araştırmanın bağımlı değişkenidir.

3.6. VERİ TOPLAMA ARAÇLARI

Çalışmada veri toplama aracı olarak TCGAbiolinks kullanılmıştır. TCGAbiolinks paketi R ortamında GDC veri portalı verilerine erişimi sağlar, analiz ve görselleştirmeler ile ilgili fonksiyonlar sağlamaktadır. Tcgabiolinks paketi kullanılarak kolon adenokarsinom (TCGA-COAD) projesinden klinik ve gen ekspresyon verileri indirilmiştir.

3.6.1. TCGAbiolinks

TCGAbiolinks, kullanıcıların GDC'de depolanan kanser verilerini sorgulamasına, indirmesine ve bütünleşik analizlerini gerçekleştirmesine olanak tanıyan bir R/ Bioconductor paketidir. TCGAbiolinks R-sürümünün tüm özelliklerini içeren, ek olarak kullanıcılara analiz adımlarında da kolaylıklar sağlayan ve kullanıcıların yerel bilgisayarlarında çalışabilen bir grafik kullanıcı arabirimine (TCGAbiolinksGUI) de sahiptir (157–162).

3.7. ARAŞTIRMA PLANI VE TAKVİMİ

Araştırma literatür taraması ile başladı. Öğrenme problemi belirlendikten sonra araştırma için gerekli olan klinik ve genomik veriler NCI- GDC veri portalından toplanıp tanımlandı. Toplanan verilerde ön işlem basamakları uygulanarak veriler gürültüden kurtarıldı, temizlendi. Daha sonra gen zenginleştirme analizi ile eğitim kümesi tanımlandı ve ardından algoritma seçimi yapılarak eğitime başlandı. Eğitim gerçekleştikten sonra test veri seti ile model değerlendirmesi yapılarak öğrenmenin gerçekleşip gerçekleşmediği kontrol edildi. Öğrenmenin gerçekleştiği anlaşıncı sınıflandırıcılar atandı ve web arayüzü oluşturuldu (163) (Tablo 3., Şekil 11.).

Tablo 3. Araştırma planı ve takvimi

İP No	İş Paketi Adı/Tanımı	Kim(ler) Tarafından Yapılacağı	ZAMAN				
			Eylül-Ekim 2020	Kasım-Aralık 2020	Ocak-Şubat 2021	Mart-Mayıs 2021	Haziran-Temmuz 2021
1	Literatür araştırması ve ön hazırlığın yapılması (Gerekli izinlerin alınması)	Hüseyin Koray MISIRLIOĞLU	X				
2	Veri tespiti ve verilerin kaydı	Hüseyin Koray MISIRLIOĞLU Asım LEBLEBİCİ Prof.Dr.Hülya ELLİDOKUZ		X			
3	Verilerin analizi ve değerlendirilmesi	Hüseyin Koray MISIRLIOĞLU Prof.Dr.Hülya ELLİDOKUZ Doç. Dr. Gizem ÇALIBAŞI KOÇAL			X	X	
4	Gen zenginleştirme analizi ve anlamlandırma	Hüseyin Koray MISIRLIOĞLU Asım LEBLEBİCİ				X	
5	Çalışmanın yazılması, çıktıların ulusal ve uluslararası dergi ve kongrelerde paylaşılması	Hüseyin Koray MISIRLIOĞLU Doç. Dr. Gizem ÇALIBAŞI KOÇAL				X	X

3.8. VERİLERİN DEĞERLENDİRİLMESİ

3.8.1. İstatistiksel Değerlendirme

SPSS; IBM firmasına ait istatistiksel analiz gerçekleştirmeyi sağlayan bir bilgisayar yazılımıdır (164).

TCGA-COAD projesindeki 454 hastaya ait klinik verinin istatistiksel analizi SPSS 24.0 (SPSS, Chicago, IL, USA) programı ile gerçekleştirildi. Bağımsız iki grubun karşılaştırılmasında non parametrik Mann- Whitney U testi, 2’den fazla grupların Kruskal Wallis testi, kategorik verilerin karşılaştırılmasında Fisher exact testi uygulandı. Faktörlerin mortalite ile yaşam sürelerine etkilerini incelemek için Kaplan-Meier- Log Rank (Mantel-Cox) analizi kullanıldı. Univariante analizden anlamlı gelen parametreler çok değişkenli (multivariate) Cox Regresyon Analizi ile değerlendirildi. Kantitatif veriler tablolarda ortalama $\pm$ standart sapma ve medyan (minimum- maksimum) değerler şeklinde, kategorik veriler sayı ve yüzdelerle (%)

ifade edilmiştir. Veriler %95 güven aralığında incelenip, p değeri 0,05 ‘den küçük olarak anlamlı kabul edilmiştir.

3.8.1.1. Klinik Verinin Analizi

Demografik Verinin Analizi

Demografik verideki cinsiyet, ırk, tümör evresi, T evresi, N evresi, M evresi, daha önce kanser teşhisi almış veya almamış olma durumu, doku veya organ menşei, rezeksiyon veya biyopsi bölgesi, beden kitle endeksi, birincil teşhis parametrelerine ait değişkenlerin frekans, kümülatif frekans ve yüzdeleri SPSS programı ile analiz edilerek verinin nasıl dağıldığı analiz edildi. Çalışmada kullanılan demografik verilerin analizine yönelik grafikler “*Microsoft 365 Excel for Mac 16.49 versiyonu*” kullanılarak çizilmiştir.

Kaplan-Meier Sağkalım Analizi

Verideki cinsiyet, ırk, tümör evresi, T evresi, N evresi, M evresi, daha önce kanser teşhisi almış veya almamış olma durumu, doku veya organ menşei, rezeksiyon veya biyopsi bölgesi, beden kitle endeksi, birincil teşhis parametrelerinin sağkalım ile ilişkisi Kaplan-Meier- Log Rank (Mantel-Cox) sağkalım analizi ile değerlendirildi. Anlamlı olan parametreler (%95 güven aralığı, $p < 0,05$) makine öğrenmesi algoritması için seçildi. Çalışmada kullanılan Kaplan-Meier Sağkalım Analizi’ne yönelik sağkalım fonksiyonu grafikleri RStudio ve SPSS kullanılarak çizilmiştir.

3.8.2. Biyoinformatik Analiz

TCGA-COAD projesindeki 454 hastaya ait gen ekspresyon verisi TCGAbiolinks ile indirildikten sonra indirilen verilere normalizasyon ve filtrasyon uygulandı. Gen ekspresyon verilerinde, primer solid tümör alt tipi seçilerek birden fazla veri içeren örnekler elendi. Sağ ve sol kolon bilgisi baz alınarak gruplar arasında gen ekspresyon değişimlerini gösterebilmek için gen ekspresyon analizi yapıldı. Bu analizde katlanma oranı (fold change) hesaplandı. p-değerlerini hesaplamak için student t-test analizi uygulandı. "FDR- False positive rate" yöntemi kullanılarak p-

değerleri düzeltildi. Her ikili grup için istatistiksel olarak anlamlı gen listeleri, katlanma oranı mutlak değer > 1.0 ve FDR < 0.05 olan genler alınarak oluşturuldu.

Ekspresyonu anlamlı olarak değişen genler için gen zenginleştirme/ yolak analizi yapıldı ve makine öğrenmesi algoritmaları uygulanarak değerlendirildi.

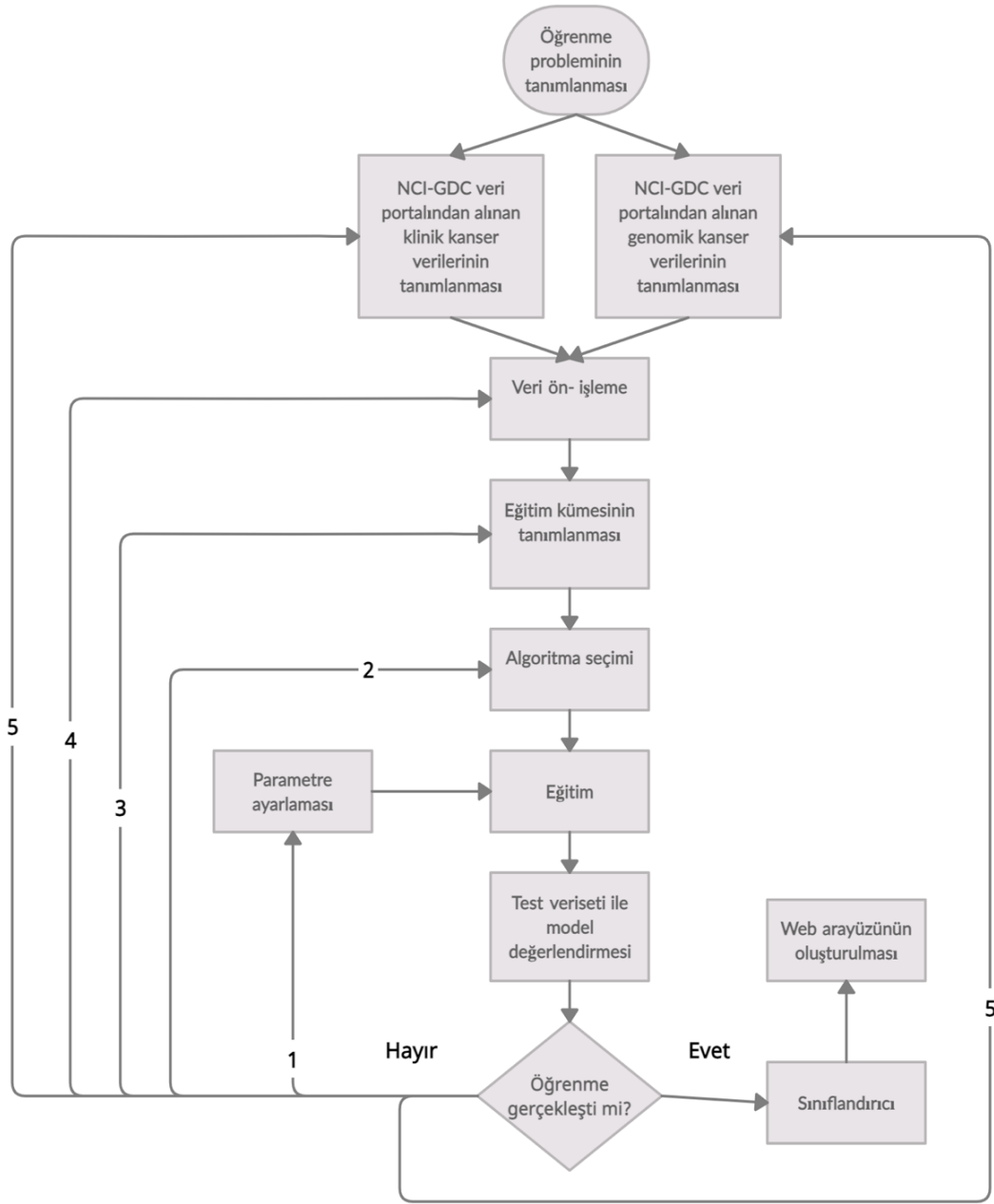
3.8.2.1. Gen Zenginleştirme Analizi

Enrichr, gen listelerinin kolektif işlevleri için çeşitli görselleştirme özetleri sağlayan, açık kaynak web tabanlı sezgisel bir zenginleştirme analizi aracıdır (165,166).

EnrichR paketi kullanılarak MSigDB (Moleküler İmza Veritabanı, Molecular Signatures Database)'deki "Cancer Hallmark" gen seti ile KEGG (Kyoto Genler ve Genom Ansiklopedisi, Kyoto Encyclopedia of Genes and Genomes) veri tabanında gen zenginleştirme analizleri yapıldı. İstatistiksel olarak anlamlı olan yollar/ genler makine öğrenmesinde öz nitelik olarak seçildi.

3.8.2.2. Makine Öğrenmesi

Seçilen yollardaki genler Bölüm 2.3.2'de yer alan makine öğrenmesi sürecindeki adımlar izlenerek değerlendirildi (Şekil 11.).



Şekil 11. Makine öğrenmesi süreci

Algoritma Değerlendirme

C4,5 Karar Ağacı, Naive Bayes, Rastgele Orman ve Destek Vektör Makineleri makine öğrenme algoritmalarının eğitimleri ayrı ayrı gerçekleştirildi. Eğitim sonucu algoritmaların öğrenimlerinin etkinliğini ve algoritmalar tarafından yapılan tahminlerin ne kadar doğru olduğunu anlamak için algoritma performans değerlendirme analizleri gerçekleştirildi.

Karmaşıklık tabloları 10 kat çapraz doğrulama yöntemi ile oluşturulan test setlerinden oluşturuldu. Test setlerinde elde edilen değerlerin ortalaması kullanıldı.

Sonuçların Geliştirilmesi

Tüm biyoinformatik analizler R programlama dili ile gerçekleştirilmiştir. R hem istatistik analizlerini mümkün kılan, hem de makine öğrenim teknikleri ve analizlerinin uygulanabildiği ücretsiz bir ortamdır. Kullanıcılar oluşturdukları kaynak kodları yayınlatabilmektedirler. Bu durum Genel Kamu Lisansı'na tabidir. Kodların yayınlanması diğer kullanıcıların da daha önce geliştirilen kodlardan ücretsiz yararlanabilmelerini sağlamaktadır. Bu sayede kodlar kullanıcılar tarafından devamlı geliştirilebilir niteliktedir. Bu çalışma kapsamında da R dilinde yazılmış çok sayıda kod örneği incelenmiştir. R programlama dilinin yazım ve derleme ortamı R konsol ve RStudio'dur. Bu çalışmada kod yazım ve derleme aracı olarak RStudio kullanılmıştır. RStudio sağladığı esnek bütünleşik geliştirme platformu sayesinde kod derleme sürecinde kolaylıklar sağlamaktadır.

3.8.3. Klinik Karar Destek Aracı: SurvCOCA Shiny Web Arayüzü

Elde edilen klinik ve transkriptomik verilerin işlenmesi ile oluşturulan makine öğrenimi tahmin algoritmaları R-Shiny paketi kullanılarak karar destek web arayüzüne dönüştürüldü (Şekil 30., Şekil 31.). Shiny, yazılan R kodlarının bir web arayüzü aracılığıyla paylaşılmasını sağlayan bir R paketidir (4,7,167). Shiny ile web ortamına aktarımı sağlanan klinik karar destek sistemi shinyapps.io aracılığı ile yayınlandı (<https://cancerapp.shinyapps.io/shiny/>) (168). TCGA-COAD verilerinden öğrenerek sağkalım tahminleyen klinik karar destek aracına “SurvCOCA” ismi verildi. İsim için literatür araştırması yapılarak daha önce herhangi bir çalışmada veya KKDS’de bu ismin kullanılmadığına emin olunmuştur.

Analiz sonuçlarının tümü Bölüm 4.’te yer almaktadır.

3.9. ARAŞTIRMANIN SINIRLILIKLARI

Çalışmanın sınırlılığı kamuya açık veri portallarındaki bilgilerle kısıtlı olmasıdır.

3.10. ETİK KURUL

Dokuz Eylül Üniversitesi Girişimsel Olmayan Etik Kurulu’nun 23.11.2020 tarih ve 2020/28-29 karar numaralı dosyası ile “Açık Kaynaklı Klinik Kanseri Verilerinin R-Shiny Uygulaması ile Yapay Zeka Tabanlı Web Arayüzü Destek Sisteminin Geliştirilmesi” isimli yüksek lisans araştırma tezi için etik kurul onayı alınmıştır.



4. BULGULAR

4.1. KLİNİK VERİLERİN ANALİZ SONUÇLARI

Hastalara ait demografik verilerin analiz sonuçları ile Kaplan-Meier sağkalım analiz sonuçları aşağıdaki gibidir.

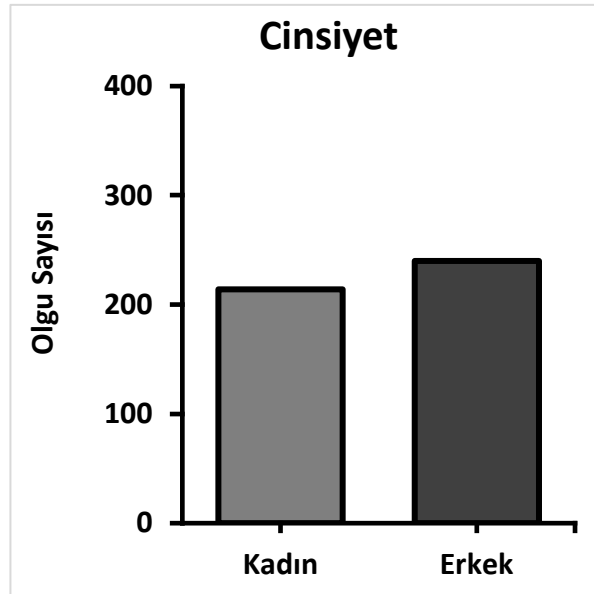
4.1.1. Demografik Verilerin Analiz Sonuçları

Çalışmaya Ulusal Kanser Enstitüsü Kanser Genom Atlas Programı'nın GDC Veri Portalında bulunan TCGA verilerinden COAD (Kolon adenokarsinom) isimli proje dahil edildi. Projede yer alan hastalar karışık alt tipli adenokarsinom, nöroendokrin farklılaşmalı adenokarsinom, adenokarsinom - NOS, adenoskuamöz karsinom, karsinom - NOS, müsinöz adenokarsinom ve papiller adenokarsinom - NOS tanıları almış 454 hastadan oluşmaktadır. 454 hasta içerisinde cinsiyet, ırk, BKİ, daha önce kanser teşhisi alma durumu, birincil teşhis, doku veya organ menşei, rezeksiyon veya biyopsi bölgesi, tümör evresi, T evresi, N evresi, M evresi ve sağkalım parametreleri incelenerek aşağıdaki verilere ulaşıldı.

454 olgunun;

%47,14'ü (n:214) kadın,

%52,86'sı (n:240) erkek bireylerden oluşmaktadır.



Şekil 12. Olguların cinsiyete göre dağılımları

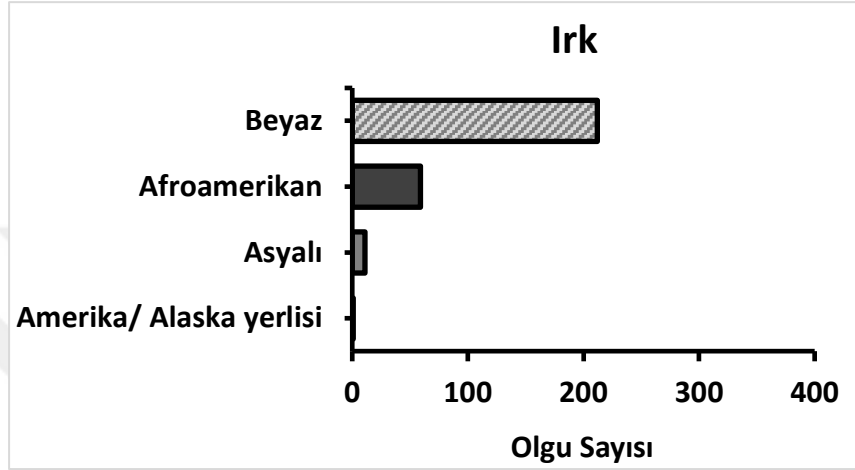
454 olgunun içerisinde ırkı belli olan 283 olgunun;

%74,91'i (n:212) Beyaz,

%20,85'i (n:59) Afroamerikan,

%3,89'u (n:11) Asyalı,

%0,35'i (n:1) Amerika yerlisi veya Alaska yerlisi bireylerden oluşmaktadır.

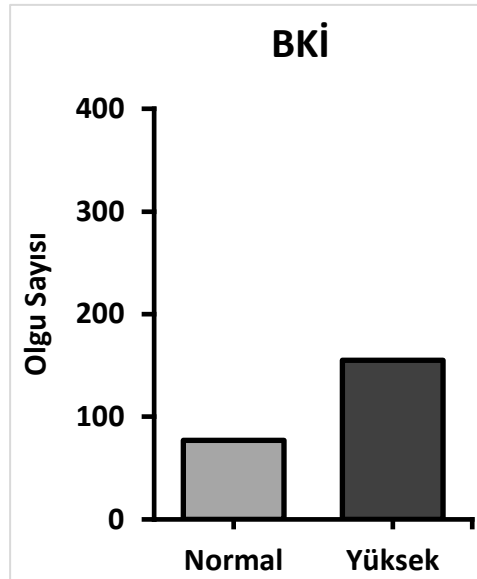


Şekil 13. Olguların ırka göre dağılımları

454 olgunun içerisinde BKİ'i belli olan 232 olgunun;

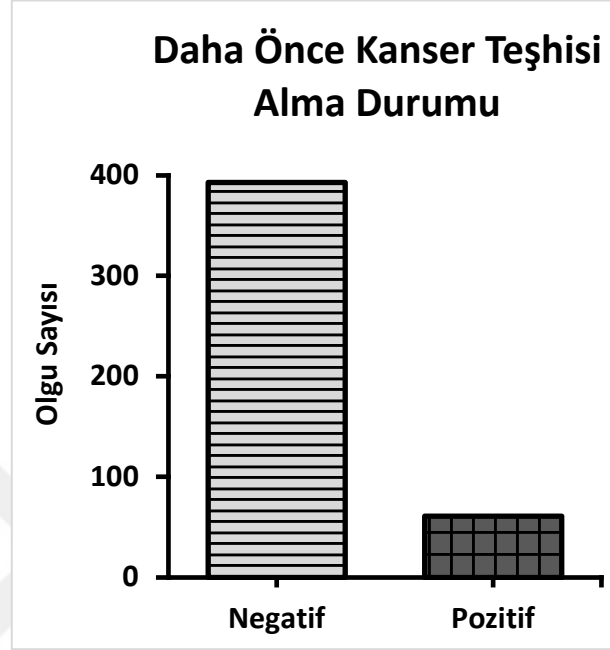
%33,19'u (n:77) normal ve

%66,81'i (n:155) yüksek BKİ'ne sahip bireylerden oluşmaktadır.



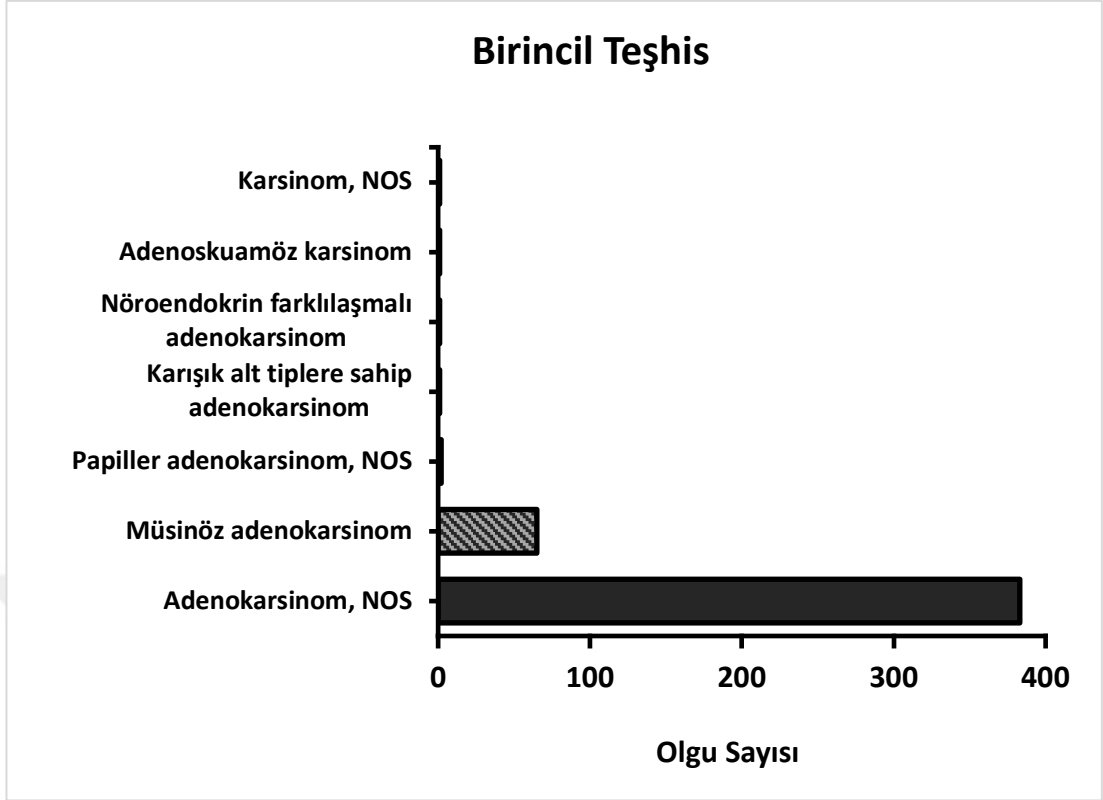
Şekil 14. Olguların beden kitle indeksi durumlarına göre dağılımları

454 olgunun daha önce kanser teşhisi alma durumuna göre dağılımları;
%86,56 (n:393) daha önce kanser teşhisi almamış (negatif) ve
%13,44 (n:61) daha önce kanser teşhisi almış (pozitif) şeklindedir.



Şekil 15. Olguların daha önce kanser teşhisi alma durumlarına göre dağılımları

454 olgunun birincil teşhis durumuna göre dağılımları;
%84,36 (n:383) Adenokarsinom- NOS,
%14,32 (n:65) Müsinöz adenokarsinom,
%0,44 (n:2) Papiller adenokarsinom- NOS,
%0,22 (n:1) Karışık alt tiplere sahip adenokarsinom,
%0,22 (n:1) Nöroendokrin farklılaşmalı adenokarsinom,
%0,22 (n:1) Adenoskuamöz karsinom ve
%0,22 (n:1) Karsinom- NOS şeklindedir.

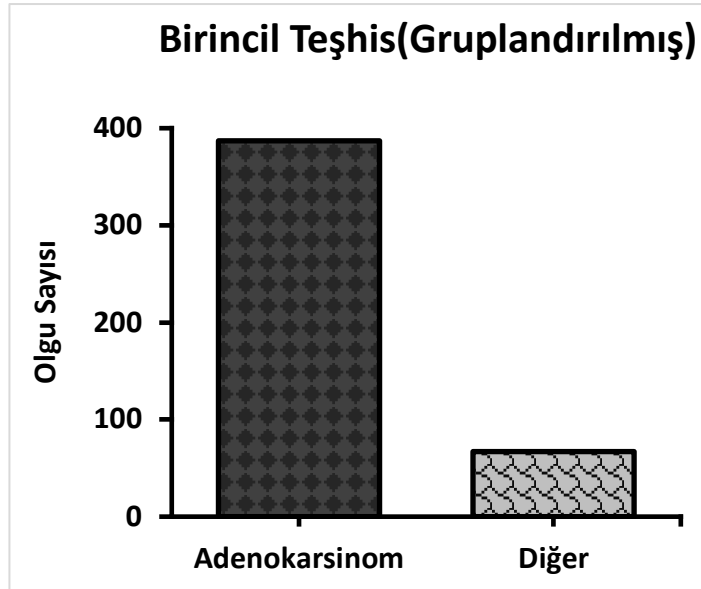


Şekil 16. Olguların birincil teşhis durumlarına göre dağılımları

454 olgunun birincil teşhis durumuna göre dağılımları;

%85,24 (n:387) Adenokarsinom ve

%14,76 (n:67) diğerleri şeklindedir.



Şekil 17. Olguların gruplandırılmış birincil teşhis durumlarına göre dağılımları

454 olgunun doku veya organ menşesinde göre dağılımları;

%24,67 (n:112) “Sigmoid kolon”,

%22,69 (n:103) “Kolon”,

%19,82 (n:90) “Çıkan kolon”,

%19,16 (n:87) “Çekum”,

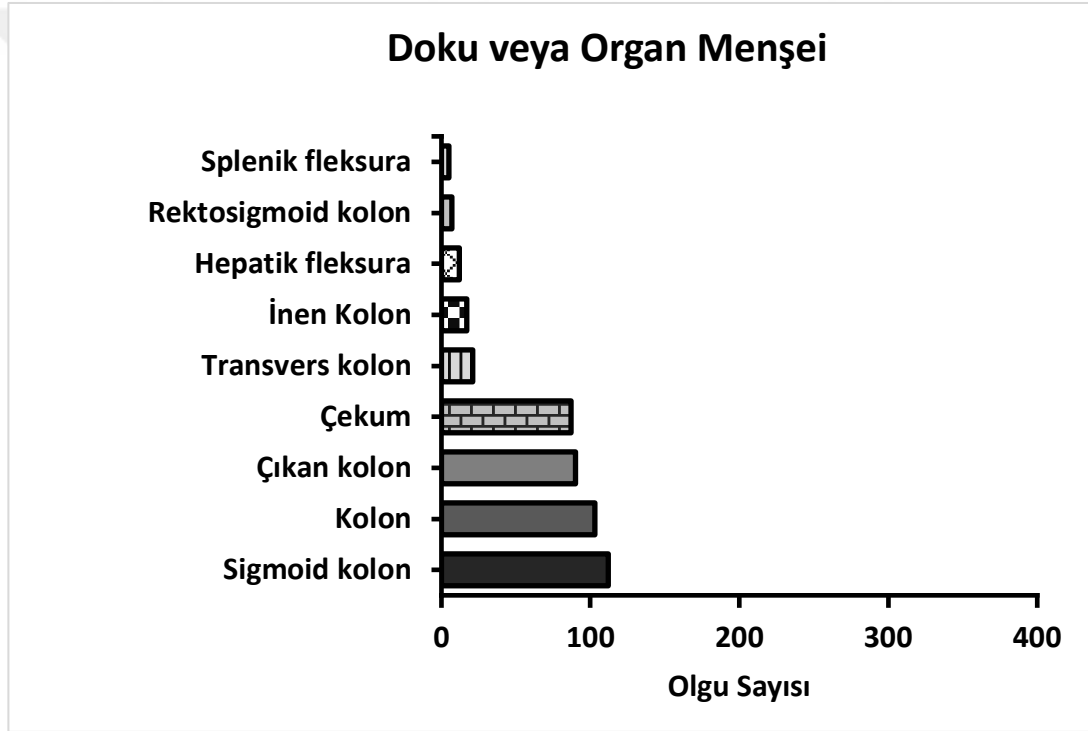
%4,63 (n:21) “Transvers kolon”,

%3,74 (n:17) “İnen kolon”,

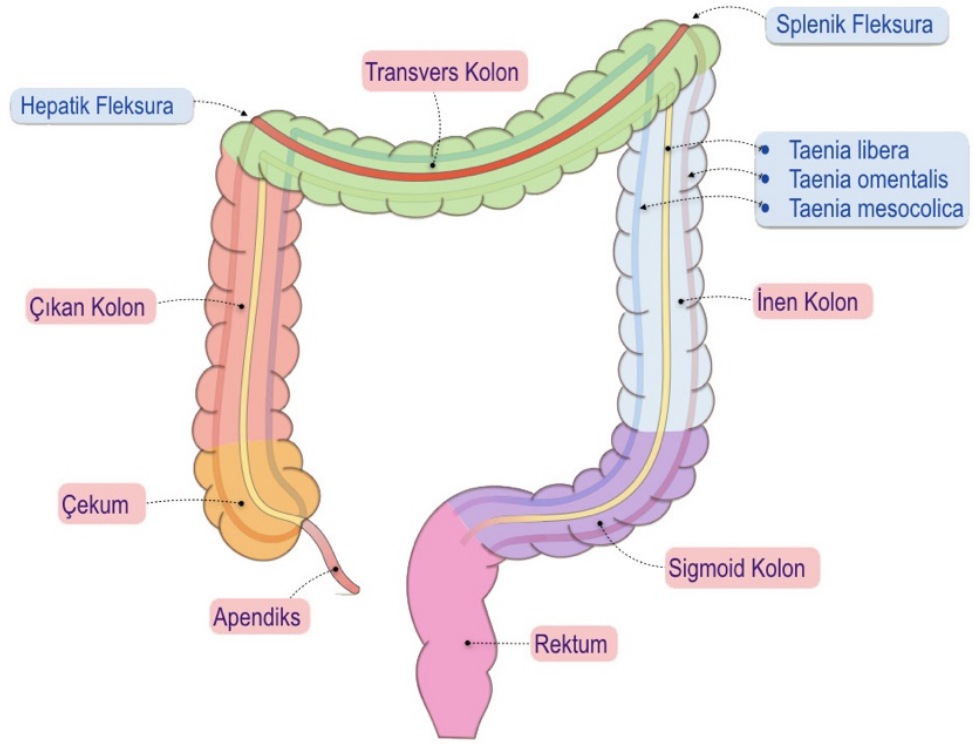
%2,64 (n:12) “Hepatik fleksura”,

%1,54 (n:7) “Rektosigmoid kolon” ve

%1,10 (n:5) “Splenic fleksura” şeklindedir (Şekil 16, Şekil 17) (169–171).

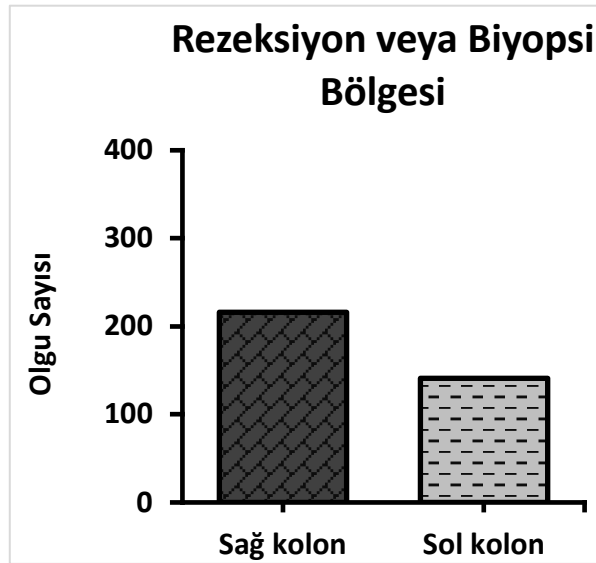


Şekil 18. Olguların doku veya organ menşei durumlarına göre dağılımları



Şekil 19. Kolonun bölümleri

454 olgunun içerisinde rezeksiyon veya biyopsi bölgesi belli olan 357 olgunun;
 %60,50'si (n:216) sağ kolon kanseri ve
 %39,50'si (n:141) sol kolon kanseridir.



Şekil 20. Olguların rezeksiyon veya biyopsi bölgesi durumlarına göre dağılımları

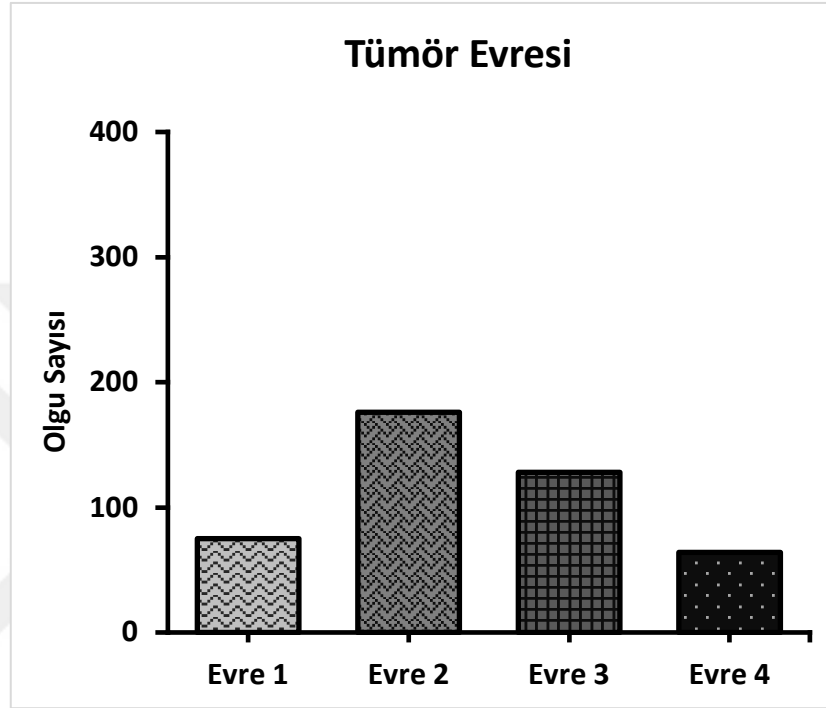
454 olgunun içerisinde tümör evresi belli olan 443 olgunun;

%16,93'ü (n:75) “Evre 1”,

%39,73'ü (n:176) “Evre 2”,

%28,89'u (n:128) “Evre 3” ve

%14,45'i (n:64) “Evre 4” şeklindedir.



Şekil 21. Olguların tümör evresine göre dağılımları

454 olgunun içerisinde T evresi belli olan 453 olgunun;

%2,43'ü (n:11) T1 evresi,

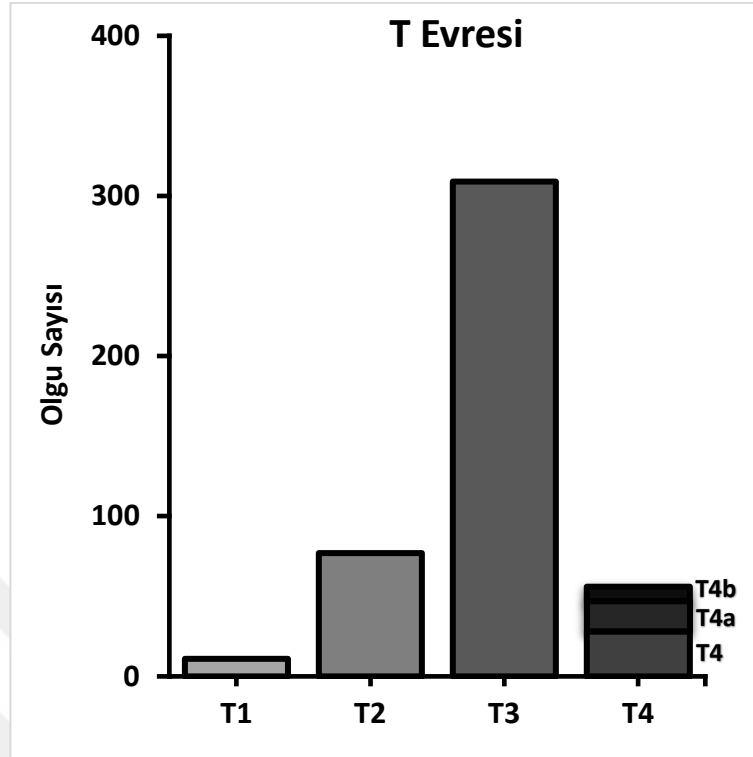
%17,00'ı (n:77) T2 evresi,

%68,21'i (n:309) T3 evresi,

%6,18'i (n:28) T4 evresi,

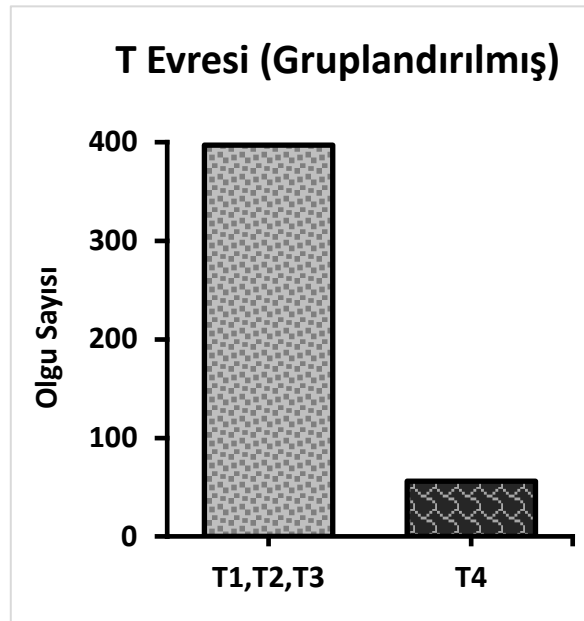
%4,19'u (n:19) T4a evresi ve

%1,99'u (n:9) T4b evresi şeklindedir.



Şekil 22. Olguların T evresine göre dağılımları

454 olgunun içerisinde T evresi belli olan gruplandırılmış 453 olgunun; %87,64'ü (n:397) T1,T2 ve T3 evreleri ve %12,36'sı (n:56) T4 evresi şeklindedir.



Şekil 23. Olguların gruplandırılmış T evresine göre dağılımları

454 olgunun N evresi durumuna göre dağılımları;

%58,81 (n:267) N0 evresi,

%16,08 (n:73) N1 evresi,

%3,30 (n:15) N1a evresi,

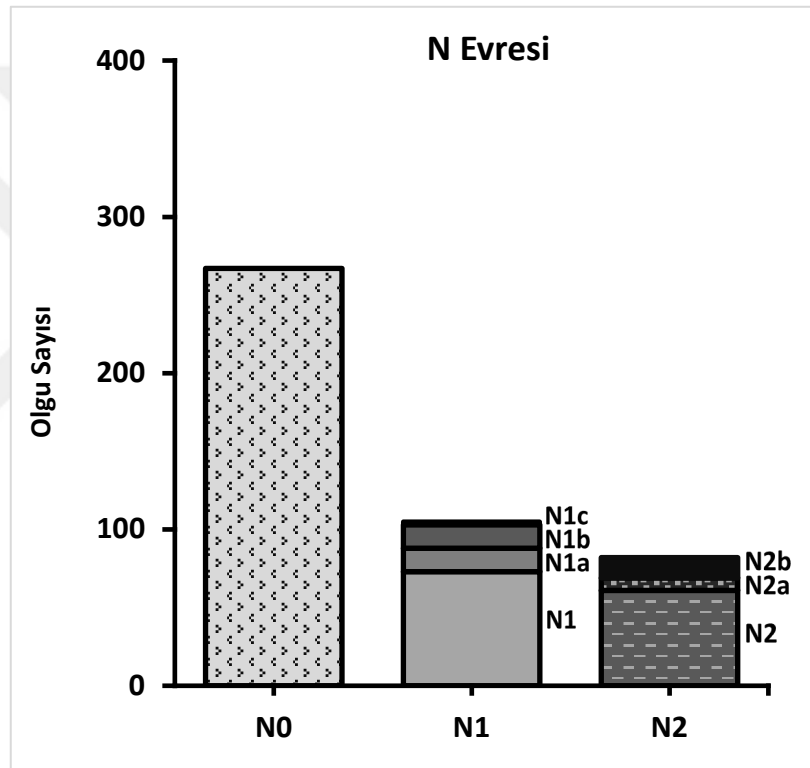
%3,30 (n:15) N1b evresi,

%0,44 (n:2) N1c evresi,

%13,44 (n:61) N2 evresi,

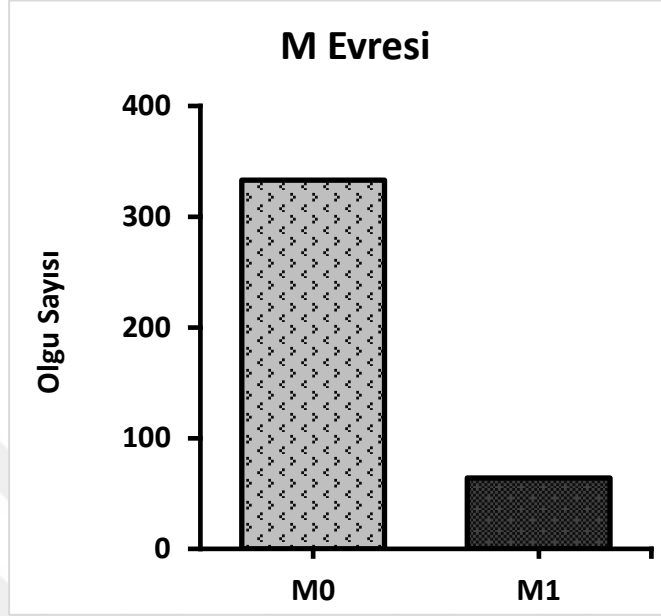
%1,76 (n:8) N2a evresi ve

%2,86 (n:13) N2b evresi şeklindedir.



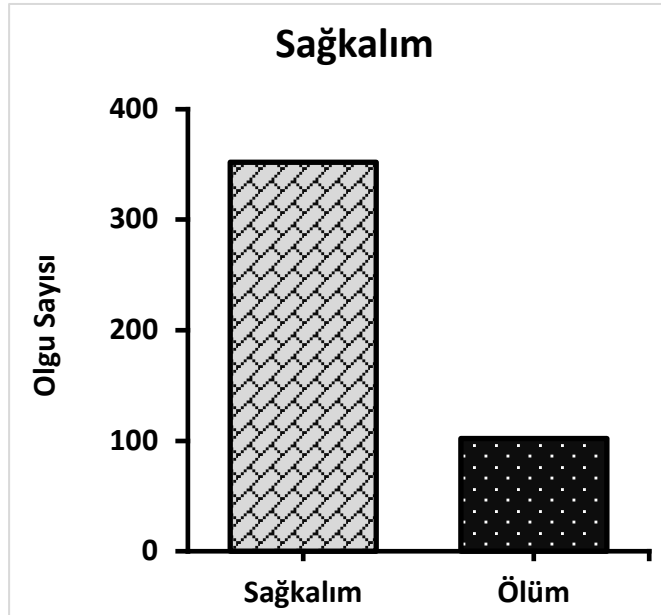
Şekil 24. Olguların N evresine göre dağılımları

454 olgunun içerisinde M evresi belli olan 397 olgunun;
%83,88'i (n:333) M0 evresi ve
%16,12'si (n:64) M1 evresi şeklindedir.



Şekil 25. Olguların M evresine göre dağılımları

454 olgunun sağkalım durumları;
%77,53 (n:352) sağkalım ve
%22,47 (n:102) ölüm ile sonuçlanmıştır.

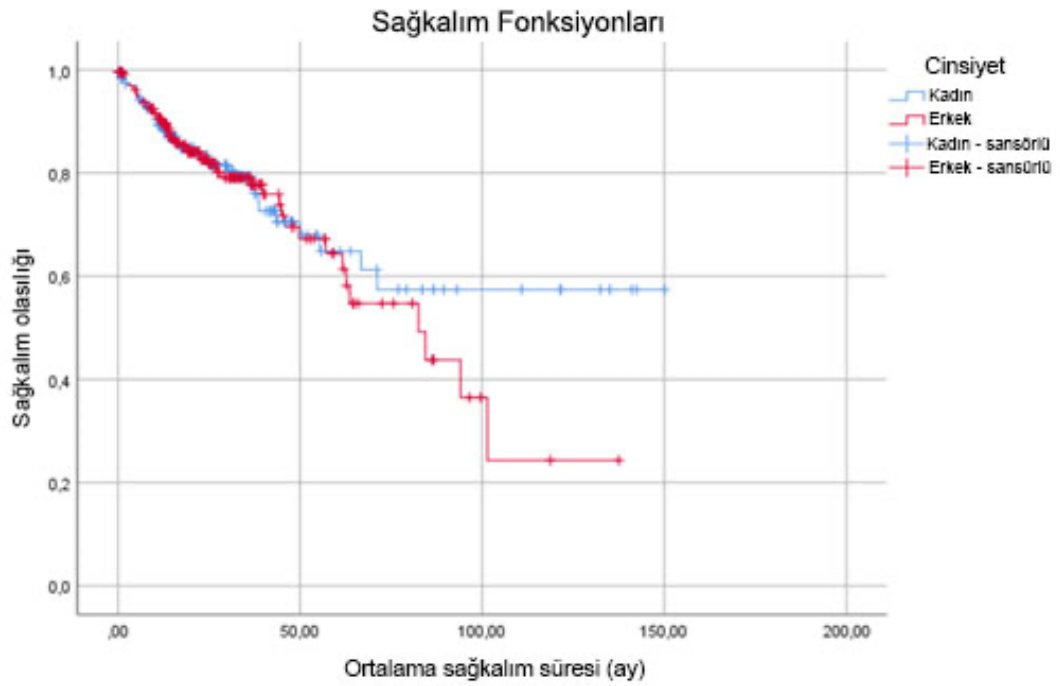


Şekil 26. Olguların sağkalım durumlarına göre dağılımları

4.1.2. Kaplan-Meier Sağkalım Analiz Sonuçları

Olgularda cinsiyet, BKİ, daha önce kanser teşhisi alma durumu, birincil teşhis, rezeksiyon veya biyopsi bölgesi, tümör evresi, T evresi, N evresi ve M evresi değişkenleri için yapılan Kaplan- Meier sağkalım (172) analizlerine göre; “sağkalım olasılığı- ortalama sağkalım süresi (ay)” grafikleri ve grafiklere ait tablolar aşağıdaki gibidir.

4.1.2.1. Cinsiyete Göre Sağkalım



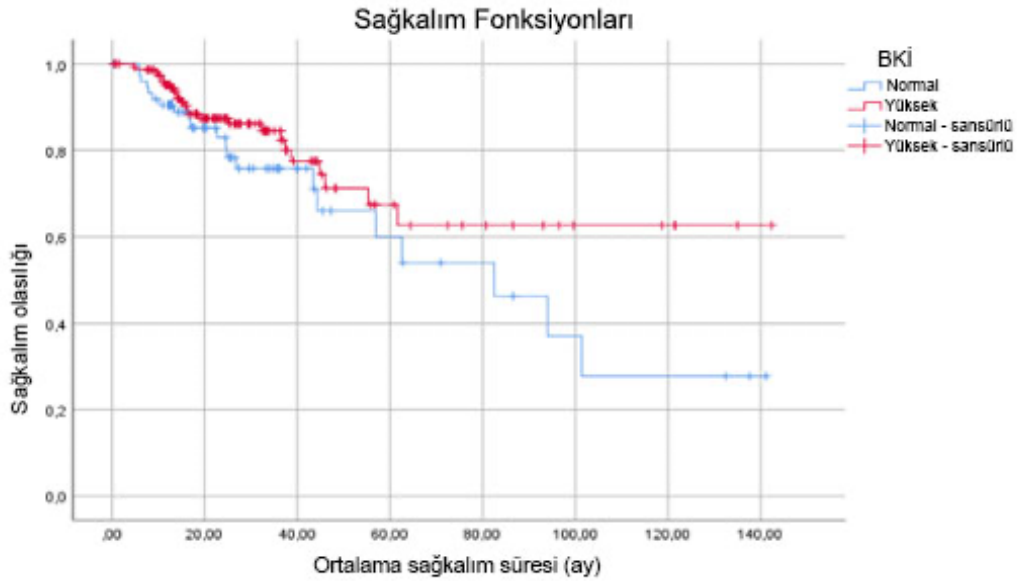
Şekil 27. Cinsiyete göre sağkalım

Tablo 4. Cinsiyete göre olgularda yaşam süresi

	Cinsiyete Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Kadın	100,50	7,29	86,21-114,79
Erkek	77,49	6,75	64,27-90,72
Log Rank test: 0,334, SD: 1, p: 0,563			

Sağkalımda cinsiyetler arasında anlamlı fark yoktur (Log Rank test: 0,334, SD: 1, p: 0,563). Erkeklerde tahmini ortalama sağkalım ($\pm$ SH) 77,49 ay $\pm$ 6,75, %95 CI (confidence interval, güven aralığı) 64,27- 90,72. Kadınlarda tahmini ortalama sağkalım ($\pm$ SH) 100,50 ay $\pm$ 7,29, %95 CI 86,21- 114,79'dur.

4.1.2.2. Beden Kitle İndeksine Göre Sağkalım



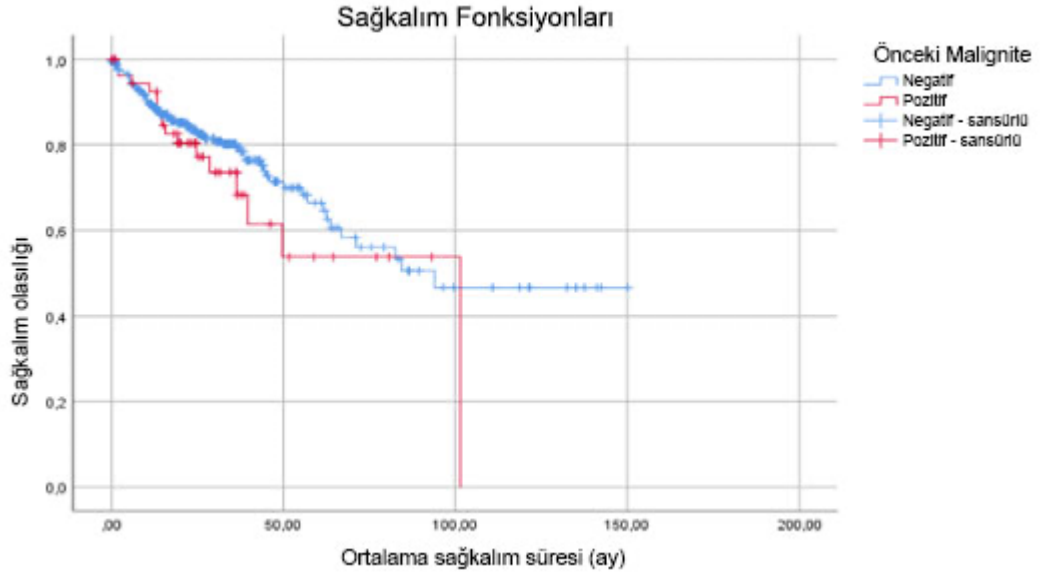
Şekil 28. Beden kitle indeksine göre sağkalım

Tablo 5. Beden kitle indeksine göre olgularda yaşam süresi

	Beden Kitle İndeksine Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Normal	79,00	9,39	60,60-97,41
Yüksek	102,17	7,79	86,91-117,43
<i>Log Rank test: 2,555, SD:1, p:0,110</i>			

Beden kitle indeksine göre (normal veya yüksek) olgularda sağkalım açısından anlamlı bir farka rastlanmamıştır (Log Rank test: 2,555, SD: 1, p: 0,110). Beden kitle indeksi normal olan olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 79,00 ay $\pm$ 9,39, %95 CI 60,60- 97,41. Beden kitle indeksi yüksek olan olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 102,17 ay $\pm$ 7,79, %95 CI 86,91- 117,43'dür.

4.1.2.3. Daha Önce Kansere Teşhisi Alma Durumuna Göre Sağkalım



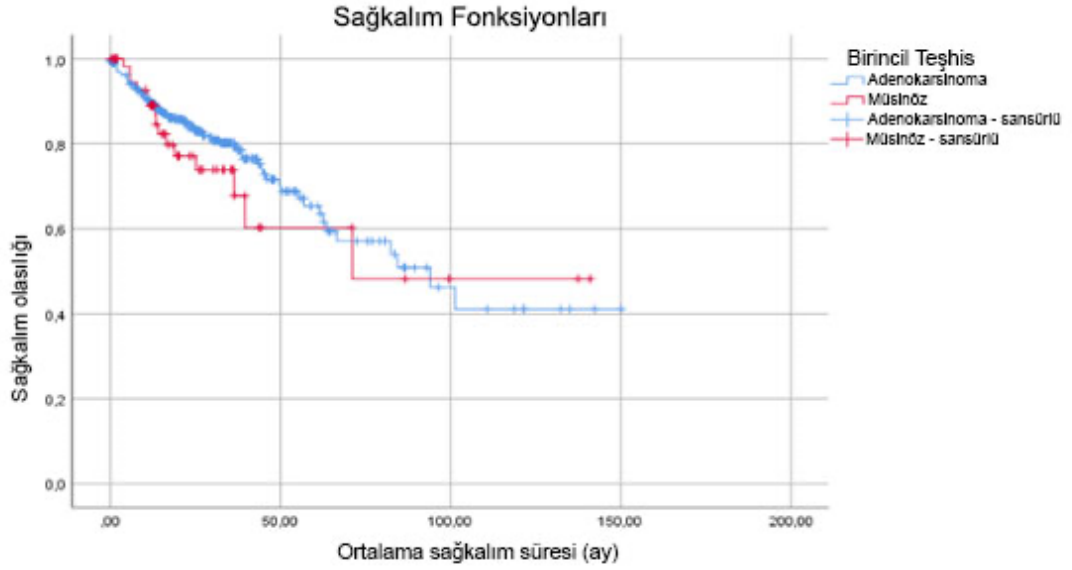
Şekil 29. Daha önce kanser teşhisi alma durumuna göre sağkalım

Tablo 6. Daha önce kanser teşhisi alma durumlarına göre olgularda yaşam süresi

	Daha Önce Kansere Teşhisi Alma Durumlarına Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Negatif	93,99	5,95	82,32-105,65
Pozitif	67,02	7,75	51,84-82,20
<i>Log Rank test: 1,164, SD:1, p:0,281</i>			

Daha önce kanser tanısı almış hastalar (pozitif) ile daha önce kanser tanısı almamış olan hastalar (negatif) arasında sağkalım açısından anlamlı farka rastlanmamıştır (Log Rank test: 1,164, SD: 1, p: 0,281). Negatif olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 93,99 ay $\pm$ 5,95, %95 CI 82,32- 105,65. Pozitif olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 67,02 ay $\pm$ 7,75, %95 CI 51,84- 82,20'dir.

4.1.2.4. Birincil Teşhise Göre Sağkalım



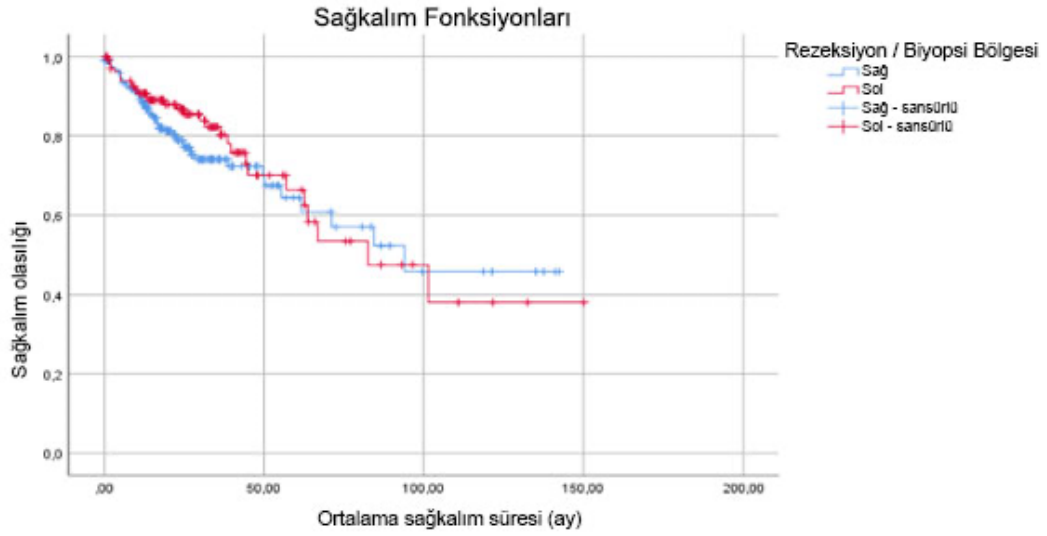
Şekil 30. Birincil teşhise göre sağkalım

Tablo 7. Birincil teşhise göre olgularda yaşam süresi

	Birincil Teşhise Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Adenokarsinom	91,29	6,19	79,16-103,43
Müsinöz adenokarsinom	85,28	12,64	60,51-110,04
<i>Log Rank test: 0,622, SD:1, p:0,430</i>			

Birincil teşhise göre (adenokarsinom ve müsinöz adenokarsinom) olgularda sağkalım açısından anlamlı bir farka rastlanmamıştır (Log Rank test: 0,622, SD: 1, p: 0,430). Birincil teşhisi adenokarsinom olan olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 91,29 ay $\pm$ 6,19, %95 CI 79,16- 103,43. Birincil teşhisi müsinöz adenokarsinom olan olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 85,28 ay $\pm$ 12,64, %95 CI 60,51- 110,04'dır.

4.1.2.5. Rezeksiyon veya Biyopsi Bölgesine Göre Sağkalım



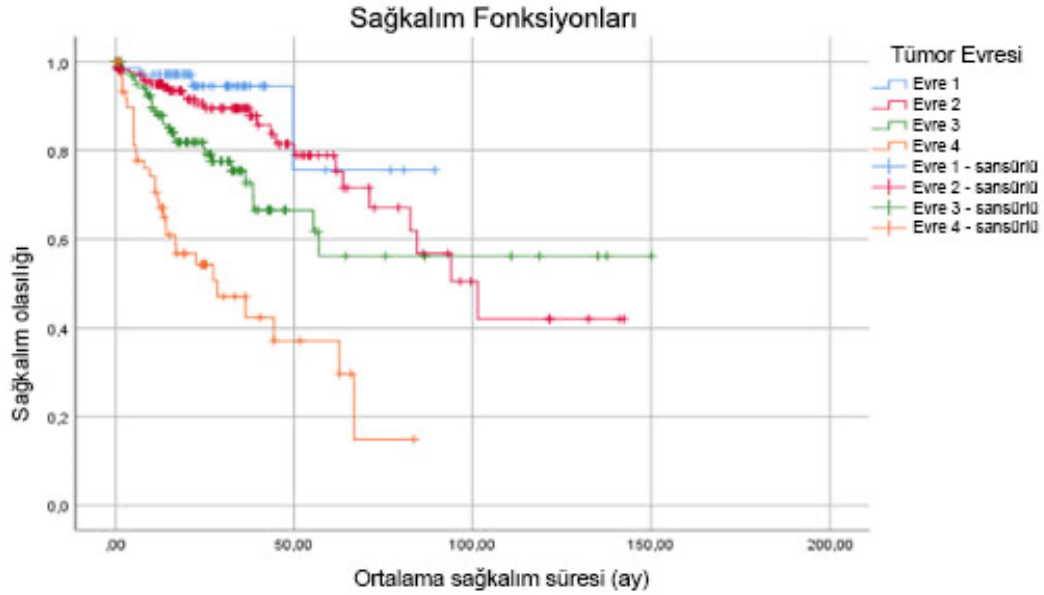
Şekil 31. Rezeksiyon veya biyopsi bölgesine göre sağkalım

Tablo 8. Rezeksiyon veya biyopsi bölgesine göre olgularda yaşam süresi

	Rezeksiyon veya Biyopsi Bölgesine Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Sağ	88,58	7,01	74,85-102,32
Sol	89,55	9,03	71,85-107,26
Log Rank test: 0,659, SD:1, p:0,417			

Sağkalım açısından rezeksiyon veya biyopsi bölgeleri (sağ veya sol) arasında anlamlı bir farka rastlanmamıştır (Log Rank test: 0,659, SD: 1, p: 0,417). Sağ taraftan alınan rezeksiyon veya biyopsilere ait olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 88,58 ay $\pm$ 7,01, %95 CI 74,85- 102,32. Sol taraftan alınan rezeksiyon veya biyopsilere ait olgularda tahmin edilen ortalama sağkalım ($\pm$ SH) 89,55 ay $\pm$ 9,03, %95 CI 71,85- 107,26'dır.

4.1.2.6. Tümör Evresine Göre Sağkalım



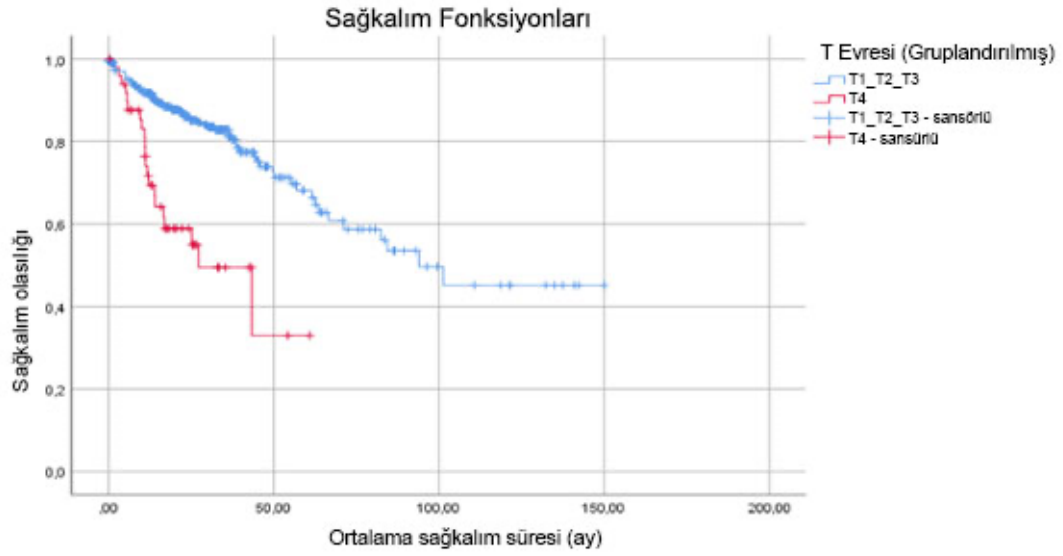
Şekil 32. Tümör evresine göre sağkalım

Tablo 9. Tümör evresine göre olgularda yaşam süresi

	Tümör Evresine Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
Evre 1	77,68	7,05	63,85-91,50
Evre 2	96,61	7,49	81,93-111,28
Evre 3	97,09	9,33	78,81-115,38
Evre 4	37,15	4,87	27,60-46,69
<i>Log Rank test: 57,061, SD:3, p:0,000</i>			

Sağkalım açısından tümör evre 1, evre 2, evre 3 ve evre 4 olgular arasında anlamlı fark gözlenmiştir (Log Rank test: 57,061, SD: 3, p:0,00). Evre 1 için tahmin edilen ortalama sağkalım ($\pm$ SH) 77,68 ay $\pm$ 7,05, %95 CI 63,85- 91,50. Evre 2 için tahmin edilen ortalama sağkalım ($\pm$ SH) 96,61 ay $\pm$ 7,49, %95 CI 81,93- 111,28. Evre 3 için tahmin edilen ortalama sağkalım ($\pm$ SH) 97,09 ay $\pm$ 9,33, %95 CI 78,81- 115,38. Evre 4 için tahmin edilen ortalama sağkalım ($\pm$ SH) 37,15 ay $\pm$ 4,87, %95 CI 27,60-46,69'dur.

4.1.2.7. T Evresine Göre Sağkalım



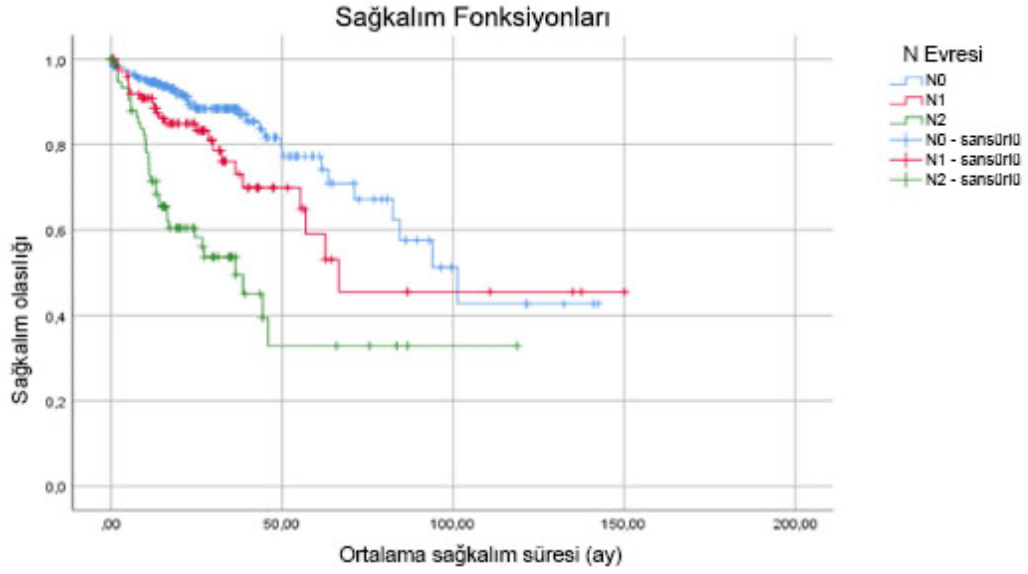
Şekil 33. T evresine göre sağkalım

Tablo 10. T evresine göre olgularda yaşam süresi

	T Evresine Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
T1_T2_T3	95,50	5,86	84,03-106,98
T4	33,96	4,24	25,64-42,27
Log Rank test: 26,465, SD:1, p:0,000			

Bir arada (ortak kümede) değerlendirilen T1_T2_T3 evreleri ile bunlardan ayrı değerlendirilen T4 evresi arasında sağkalımda anlamlı fark gözlenmiştir (Log Rank test:26,465, SD: 1, p:0,000). T1_T2_T3 için tahmin edilen ortalama sağkalım ($\pm$ SH) 95,50 ay $\pm$ 5,86, %95 CI 84,03- 106,98. T4 için tahmin edilen ortalama sağkalım ($\pm$ SH) 33,96 ay $\pm$ 4,24, %95 CI 25,64- 42,27’ dür.

4.1.2.8. N Evresine Göre Sağkalım



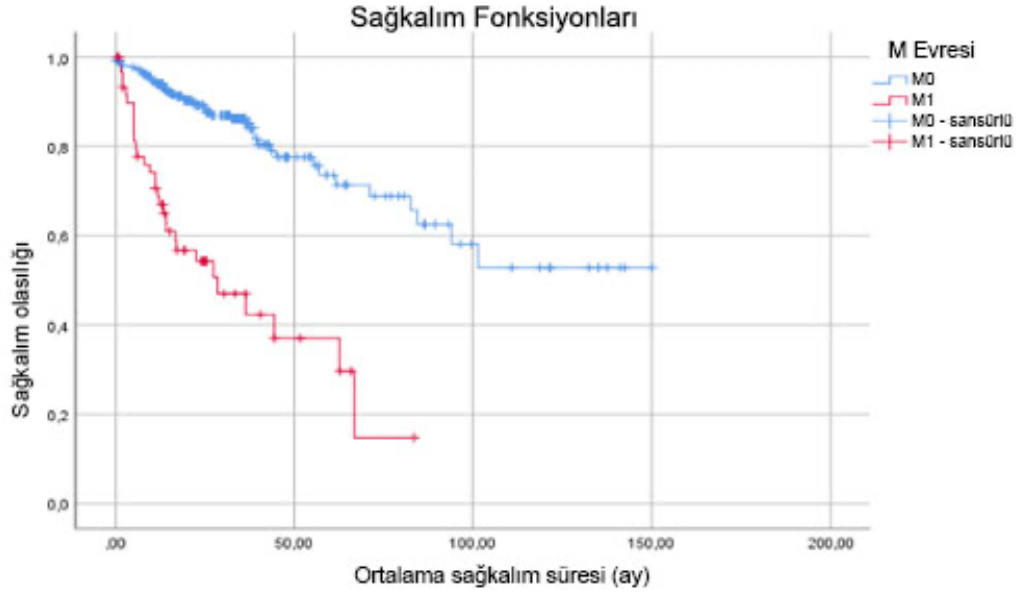
Şekil 34. N evresine göre sağkalım

Tablo 11. N evresine göre olgularda yaşam süresi

	Lenf Nodu Evresine Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
N0	96,58	7,04	82,78-110,38
N1	89,37	10,81	68,18-110,55
N2	53,28	8,06	37,49-69,06
<i>Log Rank test: 40,541, SD:2, p:0,000</i>			

Ayrı ayrı değerlendirilen N0, N1 ve N3 evreleri arasında sağkalımda anlamlı fark gözlenmiştir (Log Rank test: 40,541, SD: 2, p:0,000). N0 için tahmin edilen ortalama sağkalım ($\pm$ SH) 96,58 ay $\pm$ 7,04, %95 CI 82,78- 110,38. N1 için tahmin edilen ortalama sağkalım ($\pm$ SH) 89,37 ay $\pm$ 10,81, %95 CI 68,18- 110,55. N2 için tahmin edilen ortalama sağkalım ($\pm$ SH) 53,28 ay $\pm$ 8,06, %95 CI 37,49- 69,06'dir.

4.1.2.9. M Evresine Göre Sağkalım



Şekil 35. M evresine göre sağkalım

Tablo 12. M evresine göre olgularda yaşam süresi

	Uzak Metastaza Göre Olgularda Yaşam Süresi (Ay)		
	Ortalama	Standart Hata	%95 Güven Aralığı
M0	104,96	6,30	92,62-117,30
M1	37,15	4,87	27,60-46,69
Log Rank test: 54,291, SD:1, p:0,000			

Uzak metastazın olmaması (M0 evresi) ve uzak metastazın olması (M1 evresi) durumlarına göre sağkalımda anlamlı fark gözlenmiştir (Log Rank test:54,291, SD: 1, p:0,00). M0 için tahmin edilen ortalama sağkalım ($\pm$ SH) 104,96 ay $\pm$ 6,30, %95 CI 92,62- 117,30. M1 için tahmin edilen ortalama sağkalım ($\pm$ SH) 37,15 ay $\pm$ 4,87, %95 CI 27,60- 46,69'dur.

4.2. GEN EKSPRESYON VERİLERİNİN ANALİZ SONUÇLARI

4.2.1. Ekspresyon Verilerinin Seçimi

Ekspresyon verilerinin seçimi için KEGG ile MSig veri tabanlarında yolak ve genler taranarak gen zenginleştirme analizleri gerçekleştirildi.

4.2.1.1. KEGG Veri Tabanı

Tablo 13. Gen zenginleştirme analizi (KEGG)

KEGG	P değeri	Yolak	Eşleşen	Genler
1	0,001	Glikoliz/Glukoneogenez	3/68	ADH1C; AKR1A1; PFKM
2	0,044	Diğer glikan bozunması	1/18	MAN2B2
3	0,046	Pantotenat ve CoA biyosentezi	1/19	BCAT1
4	0,049	Apoptoz	2/143	CAPN2; CASP2

KEGG veri tabanındaki glikoliz/ glukoneogenez yolağında aktif rol oynayan 68 gen içerisinde üç tanesi (ADH1C; AKR1A1; PFKM) ekspresyon verileri ile anlamlı eşleşme gösterirken (p:0,001), diğer glikan bozunması isimli yolakta bulunan 18 gen içerisinde biri (MAN2B2) (p:0,044), pantotenat ve CoA biyosentezi yolağında 19 genden biri (BCAT1) (p:0,046) ve apoptoz yolağında 143 genden iki tanesi (CAPN2; CASP2) ekspresyon verileri ile anlamlı eşleşme (0,049) gösterdi.

4.2.1.2. MSigDB Cancer Hallmark Gen Seti

Tablo 14. Gen zenginleştirme analizi (Cancer Hallmark)

Hallmark	P Değeri	Yolak	Eşleşen	Genler
1	0,014	mTORC1 Sinyali	3/200	TOMM40; TCEA1; BCAT1

MSig veri tabanı “Cancer Hallmark Gen Seti”ndeki mTORC1 sinyal yolağında bulunan 200 gen içerisinde üç tanesi (TOMM40; TCEA1; BCAT1) ekspresyon verileri ile anlamlı eşleşme gösterirdi (p:0,014).

4.3. MAKİNE ÖĞRENMESİ ALGORİTMA SONUÇLARI

4.3.1. Klinik ve Genetik Verilerle Oluşturulan Veri Setlerinin Performans Sonuçları

İstatistiksel anlamlılık gösteren klinik ve genetik parametreler üzerinde uygulanan karar ağacı, destek vektör makineleri, rastgele orman ve Naive Bayes algoritmaları model değerlendirmesine sokularak performans sonuçları hesaplandı. Karşılaşılan performans sonuçları aşağıdaki gibidir.

4.3.1.1. Klinik Parametrelerin Performans Sonuçları

Tablo 15. Klinik parametrelerin performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%96,6	%28,8	%83,5	%69,7	%82,3	%67,8
DVM	%93,0	%16,3	%80,5	%38,2	%76,7	%54,6
RO	%83,9	%45,0	%85,0	%42,9	%75,7	%72,2
Naive Bayes	%88,3	%41,3	%84,8	%48,5	%78,3	%73,7

Klinik parametrelerin genomik veriler olmadan girdi olarak alındığı durumda makine öğrenim algoritma modellerinin performans değerlendirme sonuçları Tablo 15.'de gösterilmiştir.

4.3.1.2. Klinik Parametreler ile Glikoliz- Glukoneogenez Yolağındaki Genlerin Ortak Performans Sonuçları

Tablo 16. Klinik parametreler ile glikoliz- glukoneogenez yolağındaki genlerin ortak performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%92,6	%28,8	%82,9	%51,1	%79,1	%66,2
DVM	%93,0	%17,5	%80,8	%40,0	%77,0	%55,2
RO	%91,6	%31,3	%83,2	%50,0	%78,8	%73,8
Naive Bayes	%88,6	%36,3	%83,8	%46,0	%77,5	%69,9

4.3.1.3. Klinik Parametreler ile Glikan Degredasyon Yolağındaki Genlerin Ortak Performans Sonuçları

Tablo 17. Klinik parametreler ile glikan degradasyon yolağındaki genlerin ortak performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%95,6	%28,8	%83,3	%63,9	%81,5	%67,5
DVM	%93,0	%16,3	%80,5	%38,2	%76,7	%54,6
RO	%93,6	%38,8	%85,1	%62,0	%82,0	%74,1
Naive Bayes	%86,6	%43,8	%85,1	%46,7	%77,5	%73,2

4.3.1.4. Klinik Parametreler ile Pantotenat ve CoA Biyosentezi Yolağındaki Genlerin Ortak Performans Sonuçları

Tablo 18. Klinik parametreler ile pantotenat ve CoA biyosentezi yolağındaki genlerin ortak performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%97,0	%28,8	%83,5	%71,9	%82,5	%67,9
DVM	%93,0	%15,0	%80,3	%36,4	%76,5	%54,0
RO	%90,6	%36,3	%84,1	%50,9	%79,1	%74,2
Naive Bayes	%41,3	%80,0	%88,5	%26,8	%49,5	%71,7

4.3.1.5. Klinik Parametreler ile Apoptoz Yolağındaki Genlerin Ortak Performans Sonuçları

Tablo 19. Klinik parametreler ile apoptoz yolağındaki genlerin ortak performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%93,6	%33,8	%84,0	%58,7	%81,0	%71,2
DVM	%93,0	%16,3	%80,5	%38,2	%76,7	%54,6
RO	%90,3	%32,5	%83,3	%47,3	%78,0	%75,9
Naive Bayes	%63,1	%68,8	%88,3	%33,3	%64,3	%72,4

4.3.1.6. Klinik Parametreler ile mTORC1 Sinyal Yolağındaki Genlerin Ortak Performans Sonuçları

Tablo 20. Klinik parametreler ile mTORC1 sinyal yolağındaki genlerin ortak performans sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%96,0	%31,3	%83,9	%67,6	%82,3	%66,7
DVM	%93,3	%16,3	%80,6	%39,4	%77,0	%54,8
RO	%94,0	%32,5	%83,8	%59,1	%81,0	%74,1
Naive Bayes	%83,9	%48,8	%85,9	%44,8	%76,5	%71,8

4.3.1.7. Klinik Parametreler ile Tüm Yolaklardaki Genlerin Ortak Performans Sonuçları

Tablo 21. Klinik Parametreler ile Tüm Yolaklardaki Genlerin Ortak Performans Sonuçları

Değerlendirme Algoritmalar	Duyarlılık (REC)	Özgüllük (SPC)	PPV	NPV	Doğruluk (ACC)	AUC
KA	%85,6	%43,8	%85,0	%44,9	%76,7	%68,1
DVM	%94,3	%13,8	%80,3	%39,3	%77,2	%54,0
RO	%95,0	%26,3	%82,7	%58,3	%80,4	%76,2
Naive Bayes	%74,5	%61,3	%87,7	%39,2	%71,7	%71,1

**4.4. SAĞKALIM TAHMİNLEYEN KLİNİK KARAR DESTEK ARACI:
SurvCOCA**

TCGA-COAD verilerinden öğrenen sağkalım tahmin aracı Shiny ile web ortamına aktarılmıştır. URL bilgisi “<https://cancerapp.shinyapps.io/shiny/>” olan açık erişime sahip bu KKDS’ye (SurvCOCA) ait hesaplama öncesi parametre ayar ekranı ile hesaplama sonrası karar çıktısı ekranları Şekil 30. Ve Şekil 31. de verilmiştir.

Prediction Tool for TCGA-COAD Cancer Data

AI Based DSS.

Koray Misirlioglu & Asim Leblebici & Prof.Dr.Yasemin Basbinar

A Web Tool based Decision Tree & SVM Algorithms & Random Forest & Naive Bayes

Clinical Values:

Tumor Stage

☒ stage 1

☐ stage 2

☐ stage 3

☐ stage 4

AJCC Pathologic T

☒ T1_T2_T3

☐ T4

AJCC Pathologic N

☒ N0

☐ N1

☐ N2

AJCC Pathologic M

☒ M0

☐ M1

OS by Month

0 29 151 90,000

Expression Values:

ADH1C

0 6,769 90,000

BCAT1

0 6,769 90,000

AKR1A1

0 6,769 90,000

CAPN2

0 6,769 90,000

CASP2

0 6,769 90,000

Expression Values:

MAN2B2

0 23,280 90,000

PFKM

0 11,949 90,000

TCEA1

0 22,956 90,000

TOMM40

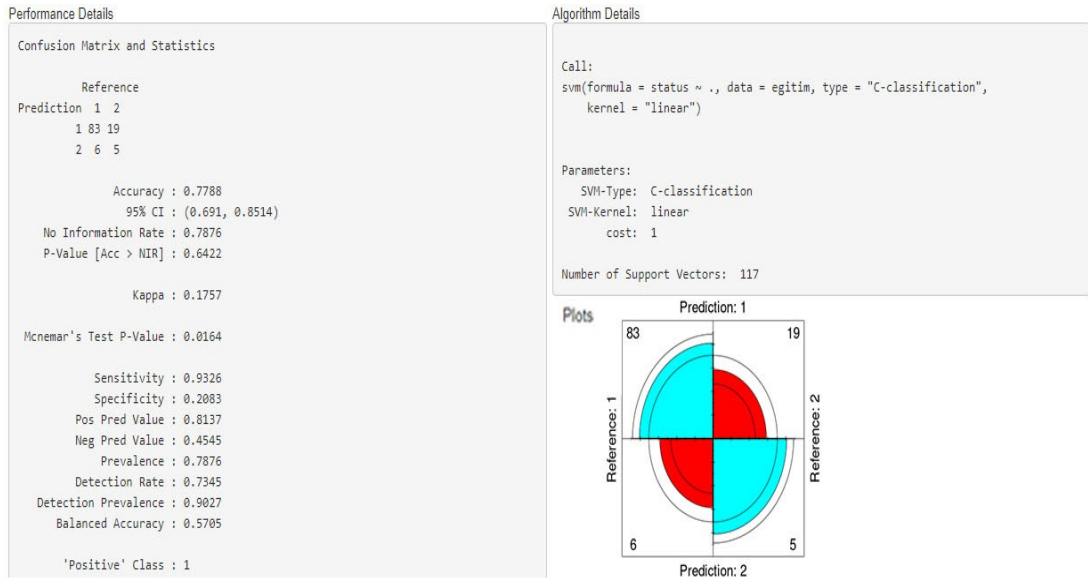
0 36,229 90,000

Machine Learning: method

Random Forest

Calculate!

Şekil 36. Sağkalım tahminleyen klinik karar destek aracı: SurvCOCA Shiny web arayüzü parametre ayar ekranı



Şekil 37. Sağkalım tahminleyen klinik karar destek aracı: SurvCOCA Shiny web arayüzü karar çıktısı ekranı

5. TARTIŞMA

Bu çalışma makine öğrenimi teknikleri kullanılarak yapay zeka destekli klinik karar destek sisteminin kolon kanseri hastalarında sağkalım tahmininde bulunması amacıyla gerçekleştirildi.

Hastalara uygulanan kemoterapi tedavisinde kullanılan ilaçlar, hastaların yaşam kalitelerini uygulama süresi boyunca yan etki ve kontrendikasyonlara bağlı olarak olumsuz etkileyebilmektedir. Kullanılan ilaca ve hastanın kişiye özgü metabolizmasının bu ilaçlara gösterdiği reaksiyonlara bağlı olarak oluşan; mide bulantısı, kusma, ishal, saç dökülmesi, iştah kaybı, yorgunluk, ateş, ağız kuruluğu, ağız yaraları, ağrı, kabızlık, ciltte oluşan morluklar, kanama, akciğer dokusunda hasar, kalp sorunları, kısırlık, böbrek sorunları, sinir hasarı, tat ve koku duyusunun bozulması, laktoz intoleransı, kilo kaybı, boğazda şişme ve yutkunma güçlüğü, çok fazla kilo alma gibi birçok durum hastaları ciddi anlamda olumsuz etkileyebilmektedir. Bu olumsuz etkiler hastaların yaşam kalitesini oldukça aşağı çekmektedir. Bu noktada sağkalım tahminleme ilerlemiş veya ölümcül kanserli (Evre 4) hastalar için daha da önemli görünmektedir. Özellikle hekimleri tarafından bilgilendirilen ileri yaştaki Evre 4 hastalar veya uzak metastaza sahip (M1) “Evre 4” hastalar böyle ciddi yan etkilere maruz kalmaktansa kemoterapi tedavisini almamayı tercih edebilmektedir. Bu doğrultuda tedavinin uygulanıp uygulanmaması, uygulanırsa oluşacak getiri ve götürüleri, uygulanmazsa oluşacak artı ve eksilerin değerlendirilip en uygun kararın alınabilmesi açısından sağkalım analizi ve sağkalımın tahminlenmesi büyük önem taşımaktadır.

Ölümcül düzeydeki kanser hastalarında hekimlerin sağkalım süresi tahminlerinin gerçekleşen sağkalım süreleri ile karşılaştırılmasına ilişkin bir çalışma hekimlerin fazlasıyla iyimser davranıp gerçek sağkalım sürelerinden daha fazla sürelerde tahminlerde bulunduklarını açığa çıkarmıştır (173). Bu sonuca göre sağkalımın klinik tahminini içeren doğruluk payı yüksek prognostik modellere ihtiyaç vardır. Sağkalım çalışmalarının tasarımı ve uygulanmasındaki metodolojik gelişmeler, prognostik belirsizliği azaltabilir ve nihayetinde ölümcül hastalar ve aileleri için daha iyi bakım sağlayabilir (174).

Yapılan çalışmada gerçekleştirilen Kaplan-Meier Sağkalım Analizi sonuçlarına göre (Bölüm 4.1.2.); kadın hastalar ile erkek hastalar arasında, normal beden kitle indeksine sahip hastalar ile yüksek beden kitle endeksine sahip hastalar arasında, daha önce kanser teşhisi almamış hastalar ile daha önce kanser teşhisi almış hastalar arasında, birincil teşhisi adenokarsinom olan hastalar ile birincil teşhisi müsinöz adenokarsinom olan hastalar arasında, rezeksiyon/ biyopsi bölgesi sağ kolon olan hastalar ile rezeksiyon/ biyopsi bölgesi sol kolon olan hastalar arasında sağkalım açısından anlamlı bir ilişkiye rastlanmazken tümör evresine göre (evre 1, evre 2, evre 3, evre 4), T evresine göre (T1_T2_T3 ile T4 arasında), N evresine göre (N0, N1, N2) ve M evresine göre (M0, M1) sağkalım açısından anlamlı farklara rastlanmıştır.

Kanser sinyal yolları, kanser gelişimini ve ilaç direncini düzenleyen sinyal zincirlerini tanımlayan önemli ve yorumlanabilir kavramlardır. Kişiye özgü sinyal yolları ile hasta sağkalımı arasındaki ilişkilerin araştırılması önemlidir. Çoklu omik verilerin klinik faktörlere entegre edilmesi sinyal yollarının kanser hastalarının hayatta kalması üzerindeki etkilerinin anlaşılmasını kolaylaştırabilir. Kişiye özgü sinyal yollarının etkilerinin ve sağkalım ile ilişkisinin anlaşılması hastanın sağkalım süresi ve sonucu ile ilgili en yüksek korelasyonlu sinyal yollarını hedefleyen yeni terapötik ilaçları belirlenmesi için oldukça önemlidir (175). Bu kapsamda yapılan gen ekspresyon verilerinin analizleri ile gen zenginleştirme analizlerine göre (Bölüm 4.2.1.); glikoliz/ glukoneogenez yolağında aktif rol oynayan ADH1C; AKR1A1; PFKM genleri, glikan bozunma yolağındaki MAN2B2 geni, pantotenat ve CoA biyosentezi yolağındaki BCAT1 geni, apoptoz yolağındaki CAPN2; CASP2 genleri ve mTORC1 sinyal yolağındaki TOMM40; TCEA1; BCAT1 genlerinin ekspresyon verileri ile anlamlı eşleşme gösterdiği görülmüştür.

Sağkalım ile anlamlı ilişki içerisinde olduğu belirlenen klinik ve genomik verilere ait parametreler sağkalım tahmini gerçekleştirmesi amacıyla dört farklı makine öğrenmesi algoritmasının eğitiminde kullanılmıştır. Kullanılan algoritmalar C4,5 Karar Ağacı, Destek Vektör Makineleri, Rastgele Orman ve Naive Bayes sınıflandırıcı algoritma modelleridir. Algoritma modellerindeki eğitim süreçlerinin ne denli başarılı geçtiğini dolayısı ile modellerin öğrenme ve doğru tahminde bulunabilme becerilerinin hangi düzeyde olduğunun belirlenebilmesi için performans

değerlendirme analizleri gerçekleştirilmiştir. Makine öğrenmesi algoritma modellerinde yapılan performans değerlendirme sonuçlarına göre; sadece klinik parametreler ile öğrenme gerçekleştiğinde en etkin tahminde bulunan algoritma modelinin Naive Bayes sınıflandırıcı algoritması (%73,7) olduğu, klinik parametreler ile glikoliz-glukoneogenez yolağındaki genlerin birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%73,8) olduğu, klinik parametreler ile glikan degradasyonu yolağındaki genlerin birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%74,1) olduğu, klinik parametreler ile pantotenat ve CoA biyosentezi yolağındaki genlerin birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%74,2) olduğu, klinik parametreler ile apoptoz yolağındaki genlerin birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%75,9) olduğu, klinik parametreler ile mTORC1 sinyal yolağındaki genlerin birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%74,1) olduğu ve klinik parametreler ile tüm yolaklardaki genlerin hep birlikte girdi olarak kullanılarak öğrenmenin gerçekleşmesi durumunda en etkin tahminde bulunan algoritma modelinin Rastgele Orman algoritması (%76,2) olduğu görülmüştür.

Tüm sonuçlar birlikte değerlendirildiğinde en doğru sağkalım tahmininde bulunan modelin Rastgele Orman algoritması olduğu görülmektedir. Rastgele Orman algoritma modeline en yakın performansı ise Naive Bayes algoritma modelinin gösterdiği görülmektedir. En düşük performansı ise Destek Vektör Makineleri algoritması göstermiştir. C4,5 Karar Ağacı algoritması ise performans etkinliğinde üçüncü sıradadır. (Tablo 15-21.).

Literatürde daha önce yapılan çalışmalara göre hem kategorik hem de nümerik değerlerin girdi olarak kullanıldığı veri setlerindeki çalışmalarda elde edilen performans sonuçlarının yalnızca kategorik değerlerin girdi olarak kullanıldığı veri setlerindeki çalışmalarda elde edilen performans sonuçlarından daha başarılı olduğu

belirtilmiştir (3). Bu sebeple hedef değışkene etki eden faktörlerin sayısal değerlerinin de makine öğrenimi sürecine dahil edilmesi öğrenim etkinliğini dolayısı ile performans sonuçlarını arttıracaktır.

Bu tez çalışmasını literatürdeki diğer çalışmalardan ayıran özelliklerden birisi; C4,5 Karar Ağacı, Destek Vektör Makineleri, Rastgele Orman ve Naive Bayes makine öğrenimi algoritmalarının hepsinin yer aldığı ve kullanıcılara istedikleri algoritmaları seçme opsiyonunun sunulduğu kolon kanseri hastalarında sağkalım tahminleme araçlarının, bir web portalında herkesin açık erişimine sunulmuş olmasıdır. Bu erişimi mümkün kılan Shiny; R kodlarının arka planda gerçek zamanlı olarak çalıştığı bir sistemde çalışan kodlara yine eş zamanlı erişimi web aracılığı ile sağlayan bir R paketidir.

Literatürde teşhise yönelik klinik karar destek sistemlerinin prognoz veya sağkalıma yönelik klinik karar destek sistemlerine göre çok daha fazla sayıda olduğu görülmüştür. Bu durum klinik karar destek sistemlerinde prognoz veya sağkalım tahmininin uzman klinisyenler ve araştırmacılar tarafından fazla önemli görülmemesi sebebi ile ortaya çıkmış olabilmekle birlikte, prognoz veya sağkalım tahminine yönelik benzer çalışmaların az sayıda var olması sebebiyle çalışmayı değerli kılmaktadır.

Çalışma kolon kanseri kliniğinde ileri evre hastalara ilişkin prognoz ve sağkalım öngörülerini ile tedavi etkileri arasındaki kritik dengenin kurulmasının sağlanması bakımından önemlidir (176).

Çalışmanın gelişime açık yönleri algoritma performans sonuçları başta olmak üzere, sağkalıma yönelik tahmininin sağkalım ihtimali yüzdesi ile sunuluyor olması yerine sağkalım veya ölüm şeklinde ikili sonuç ile sunuluyor olması ve web ortamındaki kullanıcı arayüzüdür.

Bu çalışmada elde edilen sonuçların geliştirilmesi için;

- Daha kalabalık veri grupları ile çalışmaya ihtiyaç vardır.
- Sağkalım tahmin doğruluğunu arttırabilmek için kullanılabilecek başka risk faktörlerinin de (egzersiz yapma durumu, alkol kullanımı, aile kanser öyküsü vb.) değerlendirilmeye katılması gerekmektedir.
- Tez çalışması kapsamında olmayan diğer makine öğrenimi algoritma modellerinin de (C5 Karar Ağacı, ID3 Karar Ağacı, Yapay Sinir Ağları, k-En Yakın Komşu Algoritması vb.) değerlendirilerek daha yüksek performanslı sınıflandırıcılar elde edilmesi sağlanabilir.
- Shiny ile web ortamına aktarılan klinik karar destek sisteminin sağkalıma yönelik tahminini sağkalım ihtimali yüzdesi şeklinde sunması daha faydalı olacaktır.
- Web ortamındaki kullanıcı arayüzünün başka araçlar kullanılarak daha basit, daha anlaşılır, daha ergonomik, daha kullanıcı dostu ve daha estetik bir görünüme kavuşturulması oluşturulan klinik karar destek sisteminin kullanımının daha çok kişi tarafından tercih edilmesini sağlaması ve buna benzer uygulamaların başkaları tarafından da yapılmasının teşvik edilip yaygınlaştırılmasının sağlanması açısından önemli görünmektedir.

6. SONUÇ VE ÖNERİLER

Biyoinformatik biliminin onkoloji alanında bir uygulaması olan özgün tez çalışmasında kolon kanseri hastalarına ait klinik ve genomik veri setleri üzerinde tahmine dayalı makine öğrenmesi algoritma modelleri kullanılarak veriden öğrenip karar veren klinik karar destek sistemi üretilmiştir. Üretilen klinik karar destek sistemi ile değişiklik gösteren koşullar altında hastalığa özgü sağkalım ve prognozun belirlenmesine yönelik bir araç geliştirilmeye çalışılarak bilime katkıda bulunmak amaçlanmıştır.

Klinik parametreler ve yolaklardaki genlerin tümünün girdi olarak kullanılmasıyla gerçekleşen öğrenme durumunda diğer girdi kombinasyonları ile gerçekleştirilen analiz sonuçlarına görece daha yüksek algoritma performansına ulaşılması (%76,2), makine öğrenmesi etkinliği için daha çok verinin daha iyi sonuçlar verdiğini göstermektedir.

Bu sebeple bilimsel literatürde de bilindiği üzere makine öğrenimi modellerinde daha kalabalık veri grupları ile çalışılmasına ihtiyaç vardır (177,178). Bunun için ülkemizde ve dünyada sağlık kuruluşlarında ayrı ayrı toplanan sağlık verilerinin araştırma amacıyla bir platform üzerinde toplanması ve bilim insanlarının erişimine açık olması önem arz etmektedir. Dünyada başka ülkelerde örnekleri olan buna benzer uygulamaların ülkemizde de teşvik edilmesi gerekmektedir (30).

Bu ve benzeri çalışmalar ile belirli kanser türlerine özgü olduğu bilinen biyobelirteçlerin klinik karar destek sistemlerinde daha da sıklıkla kullanılması önerilmektedir.

Bu çalışmanın sonuçlarının ve etkinliğinin geliştirilmesi için çalışmada ele alınmayan modeller ile devam ettirilmesi mühim olmakla birlikte bu tezin sonuçları benzer çalışmalar yapmayı planlayan araştırmacılar ve gelecekte yapılacak diğer çalışmalara katkı sağlaması açısından önemlidir.

7. KAYNAKLAR

1. Simon HA. The Sciences of the Artificial. Cambridge, Mass; 1996.
2. Harrington. Machine Learning in Action | Guide books [Internet]. 2012 [a.yer 23 Mart 2021]. Eriřim adresi: <https://dl.acm.org/doi/abs/10.5555/2361796>
3. Kartal E, Balaban E. M.E. Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri Ve Mes Kardiyolojik Risk Değerlendirmesine İliřkin Bir Uygulama. Çağlayan Kitapevi; 2016.
4. RStudio, Shiny [Internet]. [a.yer 23 Mart 2021]. Eriřim adresi: <https://shiny.rstudio.com/>
5. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, vd. shiny: Web Application Framework for R [Internet]. 2021 [a.yer 23 Mart 2021]. Eriřim adresi: <https://CRAN.R-project.org/package=shiny>
6. Kuhn M. topepo/caret [Internet]. 2021 [a.yer 23 Mart 2021]. Eriřim adresi: <https://github.com/topepo/caret>
7. team shinyapps io. shinyapps.io user guide [Internet]. [a.yer 23 Mart 2021]. Eriřim adresi: <https://docs.rstudio.com/shinyapps.io/>
8. Karar Destek Sistemleri [Internet]. [a.yer 13 Nisan 2021]. Eriřim adresi: https://acikders.ankara.edu.tr/pluginfile.php/3574/mod_resource/content/0/13ncuHaftaKararDestekSistemleri.pdf
9. Mendi B. Sağlık Biliřimi ve Güncel Uygulamalar [Internet]. 1. bs. C. 1. Nobel Tıp; 2016 [a.yer 02 Nisan 2021]. 254 s. Eriřim adresi: <https://nobeltip.com/product/168/saglik-bilisimi-ve-guncel-uygulamalar>
10. Persidis A, Persidis A. Medical Expert Systems: An Overview. J Manag Med. 01 Ocak 1991;5(3):27-34.

11. Karar Destek Sistemi [İnternet]. [a.yer 13 Nisan 2021]. Erişim adresi: <https://e-saglik.gov.tr/TR,7080/karar-destek-sistemi.html>
12. Davis GB, Olson MH. Management Information Systems. 2nd edition. New York: McGraw-Hill Inc.,US; 1985. 708 s.
13. Keen PGW. Decision Support Systems: An Organizational Perspective. Reading, Mass: Addison-Wesley; 1978. 264 s.
14. Umar H. Capabilities of computerized clinical decision support systems: the implications for the practicing dental professional. J Contemp Dent Pract. 15 Şubat 2002;3(1):27-42.
15. Şeker ŞE. Karar Destek Sistemleri (KDS, Decision Support Systems, DSS) [İnternet]. Bilgisayar Kavramları. 2014 [a.yer 03 Nisan 2021]. Erişim adresi: <http://bilgisayarkavramlari.com/2014/05/12/karar-destek-sistemleri-kds-decision-support-systems-dss/>
16. Gachet A. Building Model-Driven Decision Support Systems with Dicodess. Zürich: Vdf Hochschulverlag AG; 2004.
17. Belard A, Buchman T, Forsberg J, Potter BK, Dente CJ, Kirk A, vd. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. J Clin Monit Comput. Nisan 2017;31(2):261-71.
18. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. Npj Digit Med. 06 Şubat 2020;3(1):1-10.
19. Health Level Seven International - Homepage | HL7 International [İnternet]. [a.yer 23 Nisan 2021]. Erişim adresi: <https://www.hl7.org/>
20. Fast Healthcare Interoperability Resources® (FHIR) | eCQI Resource Center [İnternet]. [a.yer 23 Nisan 2021]. Erişim adresi: <https://ecqi.healthit.gov/fhir>

21. Sebastian A, Carroll JC, Oldfield LE, Mighton C, Shickh S, Uleryk E, vd. Effect of genetics clinical decision support tools on health-care providers' decision making: a mixed-methods systematic review. *Genet Med*. Nisan 2021;23(4):593-602.
22. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: an application of a bayesian belief network. *PloS One*. 2011;6(5):e19956.
23. New ACS NSQIP Surgical Risk Calculator [Internet]. American College of Surgeons. [a.yer 19 Nisan 2021]. Eriřim adresi: <http://www.facs.org/media/press-releases/2013/risk-calculator0813>
24. Scott S, Lund JN, Gold S, Elliott R, Vater M, Chakrabarty MP, vd. An evaluation of POSSUM and P-POSSUM scoring in predicting post-operative mortality in a level 1 critical care setting. *BMC Anesthesiol*. 18 Kasım 2014;14(1):104.
25. Norton WE, Hosokawa PW, Henderson WG, Volckmann ET, Pell J, Tomeh MG, vd. Acceptability of the decision support for safer surgery tool. *Am J Surg*. Haziran 2015;209(6):977-84.
26. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. *Yearb Med Inform*. 15 Ağustos 2014;9:14-20.
27. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. Mayıs 2006;34(5):1297-310.
28. Ledoux D, Finfer S, McKinley S. Impact of operator expertise on collection of the APACHE II score and on the derived risk of death and standardized mortality ratio. *Anaesth Intensive Care*. Ekim 2005;33(5):585-90.
29. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, vd. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems

of the European Society of Intensive Care Medicine. Intensive Care Med. Temmuz 1996;22(7):707-10.

30. Gauvreau CL, Fitzgerald NR, Memon S, Flanagan WM, Nadeau C, Asakawa K, vd. The OncoSim Model: Development and Use for Better Decision-Making in Canadian Cancer Control. Curr Oncol. Aralık 2017;24(6):401-6.
31. OncoSim fact sheets [Internet]. Canadian Partnership Against Cancer. [a.yer 20 Nisan 2021]. Erişim adresi: <https://www.partnershipagaincancer.ca/tools/oncosim/fact-sheets/>
32. Boswell-Purdy J, Flanagan WM, Roberge H, Le Petit C, White KJ, Berthelot J-M. Population health impact of cancer in Canada, 2001. Chronic Dis Can. 2007;28(1-2):42-55.
33. Coldman AJ, Phillips N, Brisson J, Flanagan W, Wolfson M, Nadeau C, vd. Using the Cancer Risk Management Model to evaluate colorectal cancer screening options for Canada. Curr Oncol. Nisan 2015;22(2):e41-50.
34. Coldman A, Pader J, Gauvreau C, Memon S, Fitzgerald N, Flanagan W, vd. Simulating results from trials of sigmoidoscopy screening using the OncoSim microsimulation model. J Cancer Policy. 01 Mart 2018;15:52-8.
35. Coldman A, Flanagan W, Nadeau C, Wolfson M, Fitzgerald N, Memon S, vd. Projected effect of fecal immunochemical test threshold for colorectal cancer screening on outcomes and costs for Canada using the OncoSim microsimulation model. J Cancer Policy. 01 Eylül 2017;13:38-46.
36. Heresbach D, Manfredi S, D'halluin PN, Bretagne J-F, Branger B. Review in depth and meta-analysis of controlled trials on colorectal cancer screening by faecal occult blood test. Eur J Gastroenterol Hepatol. Nisan 2006;18(4):427-33.
37. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nat Rev Urol. Temmuz 2019;16(7):391-403.

38. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. *Am J Med.* 01 Temmuz 2019;132(7):795-801.
39. Strickland E. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectr.* Nisan 2019;56(4):24-31.
40. Razavian N, Tsirigos A. Pathologists meet their match in tumour-spotting algorithm. *Nature.* 17 Eylül 2018;561(7724):436-7.
41. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Res Neuroimaging.* 30 Haziran 2017;264:1-9.
42. Albert Einstein: Computers are incredibly fast, accurate, and stupid. Human beings are incredibly slow, inaccurate, and brilliant. Together they are powerful beyond imagination. [İnternet]. [a.yer 18 Nisan 2021]. Erişim adresi: <https://www.quotes.net/quote/38086>
43. Ghaddar B, Naoum-Sawaya J. High dimensional data classification and feature selection using support vector machines. *Eur J Oper Res.* 16 Mart 2018;265(3):993-1004.
44. Haverty PM, Lin E, Tan J, Yu Y, Lam B, Lianoglou S, vd. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature.* 19 Mayıs 2016;533(7603):333-7.
45. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2015;13:8-17.
46. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* 11 Şubat 2007;2:59-77.

47. Kim DW, Lee S, Kwon S, Nam W, Cha I-H, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. 06 Mayıs 2019;9(1):6994.
48. Home | ASCO CancerLinQ [İnternet]. [a.yer 06 Mayıs 2021]. Erişim adresi: <https://www.cancerlinq.org/>
49. IBM Watson Collaboration Aims to Improve Oncology Decision Support Tools [İnternet]. Cancer Network. [a.yer 19 Nisan 2021]. Erişim adresi: <https://www.cancernetwork.com/view/ibm-watson-collaboration-aims-improve-oncology-decision-support-tools>
50. Suwanvecho S, Suwanrusme H, Sangtian M, Norden AD, Urman A, Hicks A, vd. Concordance assessment of a cognitive computing system in Thailand. *J Clin Oncol*. 20 Mayıs 2017;35(15_suppl):6589-6589.
51. Somashekhar SP, Sepúlveda M-J, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C, vd. Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann Oncol*. 01 Şubat 2018;29(2):418-23.
52. Lee W-S, Ahn SM, Chung J-W, Kim KO, Kwon KA, Kim Y, vd. Assessing Concordance With Watson for Oncology, a Cognitive Computing Decision Support System for Colon Cancer Treatment in Korea. *JCO Clin Cancer Inform*. 03 Nisan 2018;(2):1-8.
53. Patel NM, Michelini VV, Snell JM, Balu S, Hoyle AP, Parker JS, vd. Enhancing Next-Generation Sequencing-Guided Cancer Care Through Cognitive Computing. *The Oncologist*. Şubat 2018;23(2):179-85.
54. Pattern Classification, 2nd Edition | Wiley [İnternet]. Wiley.com. [a.yer 18 Nisan 2021]. Erişim adresi: <https://www.wiley.com/en-us/Pattern+Classification%2C+2nd+Edition-p-9780471056690>

55. Rivers EP, McIntyre L, Morro DC, Rivers KK. Early and innovative interventions for severe sepsis and septic shock: taking advantage of a window of opportunity. CMAJ Can Med Assoc J J Assoc Medicale Can. 25 Ekim 2005;173(9):1054-65.
56. Types of Machine Learning Algorithms [Internet]. 7wData. 2020 [a.yer 20 Mayıs 2021]. Erişim adresi: <https://www.7wdata.be/visualization/types-of-machine-learning-algorithms-2/>
57. Encyclopedia of Statistical Sciences, 16 Volume Set, 2nd Edition | Wiley [Internet]. Wiley.com. [a.yer 18 Nisan 2021]. Erişim adresi: <https://www.wiley.com/en-us/Encyclopedia+of+Statistical+Sciences%2C+16+Volume+Set%2C+2nd+Edition-p-9780471150442>
58. Bishop C. Pattern Recognition and Machine Learning [Internet]. New York: Springer-Verlag; 2006 [a.yer 18 Nisan 2021]. (Information Science and Statistics). Erişim adresi: <https://www.springer.com/gp/book/9780387310732>
59. Abedi V, Goyal N, Tsivgoulis G, Hosseinichimeh N, Hontecillas R, Bassaganya-Riera J, vd. Novel Screening Tool for Stroke Using Artificial Neural Network. Stroke. Haziran 2017;48(6):1678-81.
60. Everitt B, Howell DC. Encyclopedia of statistics in behavioral science. Hoboken, N.J.: John Wiley & Sons; 2005.
61. Ian G, Yoshua B, Aaron C. Deep Learning [Internet]. MIT Press; 2016 [a.yer 18 Nisan 2021]. Erişim adresi: <https://mitpress.mit.edu/books/deep-learning>
62. Russell SJ. Artificial intelligence : a modern approach. Prentice Hall; 2010.
63. Bishop CM. Neural Networks for Pattern Recognition. USA: Oxford University Press, Inc.; 1995. 482 s.
64. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. Mayıs 2015;521(7553):436-44.

65. Mitchell TM. Machine Learning. 1st edition. New York: McGraw-Hill Education; 1997. 432 s.
66. UCI Machine Learning Repository [İnternet]. [a.yer 26 Mayıs 2021]. Erişim adresi: <http://archive.ics.uci.edu/ml/index.php>
67. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Haryana, India; Burlington, MA; 2011.
68. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. Appl Artif Intell. 01 Mayıs 2003;17(5-6):519-33.
69. Carpenter JR, Smuk M. Missing data: A statistical framework for practice. Biom J [İnternet]. [a.yer 26 Mayıs 2021];n/a(n/a). Erişim adresi: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000196>
70. Durrant GB. Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review. :37.
71. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge ; New York; 2006.
72. Rahman MM, Davis DN. Machine Learning-Based Missing Value Imputation Method for Clinical Datasets. İçinde: Yang G-C, Ao S, Gelman L, editörler. IAENG Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering 2012 [İnternet]. Dordrecht: Springer Netherlands; 2013 [a.yer 26 Mayıs 2021]. s. 245-57. (Lecture Notes in Electrical Engineering). Erişim adresi: https://doi.org/10.1007/978-94-007-6190-2_19
73. R.s S, R N. Evaluation of Three Simple Imputation Methods for Enhancing Preprocessing of Data with Missing Values. Int J Comput Appl. 31 Mayıs 2011;21(10):14-9.
74. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating [İnternet]. New York: Springer-Verlag; 2009 [a.yer 26

Mayıs 2021]. (Statistics for Biology and Health). Erişim adresi: <https://www.springer.com/gp/book/9780387772431>

75. Suthar B, Patel H, Goswami A. A Survey: Classification of Imputation Methods in Data Mining.
76. Cheng JG. OUTLIER MANAGEMENT IN INTELLIGENT DATA ANALYSIS. :163.
77. Grubbs FE. Procedures for Detecting Outlying Observations in Samples. Technometrics. Şubat 1969;11(1):1-21.
78. Anderson RE, Tatham RL, Black WC, Hair JF. Multivariate Data Analysis. 5th edition. Upper Saddle River, N.J: Prentice Hall College Div; 1998. 730 s.
79. Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data [Internet]. Cambridge: Cambridge University Press; 2012 [a.yer 27 Mayıs 2021]. Erişim adresi: <http://ebooks.cambridge.org/ref/id/CBO9780511973000>
80. Rai P. Model Selection and Feature Selection. . 2011;103.
81. Vellido A, Martín-Guerrero JD, Lisboa PJG. Making machine learning models interpretable. İçinde: In Proc European Symposium on Artificial Neural Networks, Computational In^{TEL}ligence and Machine Learning. 2012.
82. Quinlan JR. Improved Use of Continuous Attributes in C4.5. J Artif Intell Res. 01 Mart 1996;4:77-90.
83. Kohavi R, Quinlan JR. Handbook of data mining and knowledge discovery. Klösgen W, Zytkow JM, editörler. USA: Oxford University Press, Inc.; 2002. 1025 s.
84. Bircan H. Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama. :24.
85. Japkowicz N, Shah M. Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press; 2011. 423 s.

86. Nizam H, Akın SS. Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. :6.
87. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 01 Kasım 2003;56(11):1129-35.
88. Flach PA, Lachiche N. Naive Bayesian Classification of Structured Data. Mach Learn. 01 Aralık 2004;57(3):233-69.
89. Ferraris VA. Commentary: Should we rely on receiver operating characteristic curves? From submarines to medical tests, the answer is a definite maybe! J Thorac Cardiovasc Surg. 01 Haziran 2019;157(6):2354-5.
90. Tomak L, Bek Y. İşlem Karakteristik Eğrisi Analizi ve Eğri Altında Kalan Alanların Karşılaştırılması. J Exp Clin Med. 21 Ocak 2011;27(2).
91. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, vd. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. Aralık 1989;1(4):541-51.
92. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, vd. A survey on deep learning in medical image analysis. Med Image Anal. Aralık 2017;42:60-88.
93. Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 21 Haziran 2017;19:221-48.
94. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. Eylül 2017;10(3):257-73.
95. Alam IS, Steinberg I, Vermesh O, van den Berg NS, Rosenthal EL, van Dam GM, vd. Emerging Intraoperative Imaging Modalities to Improve Surgical Precision. Mol Imaging Biol. Ekim 2018;20(5):705-15.

96. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol.* 01 Temmuz 2016;12(7):878.
97. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, vd. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci.* 27 Mart 2018;115(13):E2970-9.
98. Lee B, Chun SH, Hong JH, Woo IS, Kim S, Jeong JW, vd. DeepBTS: Prediction of Recurrence-free Survival of Non-small Cell Lung Cancer Using a Time-binned Deep Neural Network. *Sci Rep.* 06 Şubat 2020;10(1):1952.
99. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. İçinde: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, editörler. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013.* Berlin, Heidelberg: Springer; 2013. s. 403-10. (Lecture Notes in Computer Science).
100. Lehman CD, Yala A, Schuster T, Dontchos B, Bahl M, Swanson K, vd. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology.* Ocak 2019;290(1):52-8.
101. J F, I S, M E, R D, S E, C M, vd. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 v1.0 [Internet]. [a.yer 21 Mart 2021]. Erişim adresi: <https://publications.iarc.fr/Databases/Iarc-Cancerbases/GLOBOCAN-2012-Estimated-Cancer-Incidence-Mortality-And-Prevalence-Worldwide-In-2012-V1.0-2012>
102. Forouzanfar MH, Afshin A, Alexander LT, Anderson HR, Bhutta ZA, Biryukov S, vd. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet.* 08 Ekim 2016;388(10053):1659-724.

103. Plummer M, Martel C de, Vignat J, Ferlay J, Bray F, Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob Health*. 01 Eylül 2016;4(9):e609-16.
104. BW Steward, Wild C. WHO | World Cancer Report 2014 [İnternet]. WHO. World Health Organization; [a.yer 23 Mart 2021]. Erişim adresi: https://www.who.int/cancer/publications/WRC_2014/en/
105. The global initiative for cancer registry development [İnternet]. The Global Initiative for Cancer Registry Development. [a.yer 23 Mart 2021]. Erişim adresi: <https://gicr.iarc.fr/about-the-gicr/>
106. Anand P, Kunnumakara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, vd. Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Pharm Res*. 01 Eylül 2008;25(9):2097-116.
107. Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M, vd. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol*. 01 Ocak 2015;16(1):36-46.
108. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet Lond Engl*. 19 Ekim 2019;394(10207):1467-80.
109. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. Kasım 2018;68(6):394-424.
110. Cancer today [İnternet]. [a.yer 24 Mayıs 2021]. Erişim adresi: <http://gco.iarc.fr/today/home>
111. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 01 Nisan 2017;66(4):683-91.

112. Ouakrim DA, Pizot C, Boniol M, Malvezzi M, Boniol M, Negri E, vd. Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. *BMJ*. 06 Ekim 2015;351:h4970.
113. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, vd. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst*. 01 Ağustos 2017;109(8).
114. Kasi PM, Shahjehan F, Cochuyt JJ, Li Z, Colibaseanu DT, Merchea A. Rising Proportion of Young Individuals With Rectal and Colon Cancer. *Clin Colorectal Cancer*. Mart 2019;18(1):e87-95.
115. Bailey CE, Hu C-Y, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, vd. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg*. Ocak 2015;150(1):17-22.
116. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, vd. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. Temmuz 2018;68(4):250-81.
117. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. *JAMA*. 17 Aralık 2008;300(23):2765-78.
118. Cai S, Li Y, Ding Y, Chen K, Jin M. Alcohol drinking and the risk of colorectal cancer death: a meta-analysis. *Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP*. Kasım 2014;23(6):532-9.
119. Kyrgiou M, Kalliala I, Markozannes G, Gunter MJ, Paraskevaidis E, Gabra H, vd. Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ*. 28 Şubat 2017;356:j477.

120. Chan DSM, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, vd. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PloS One*. 2011;6(6):e20456.
121. Krämer HU, Schöttker B, Raum E, Brenner H. Type 2 diabetes mellitus and colorectal cancer: meta-analysis on sex-specific differences. *Eur J Cancer Oxf Engl* 1990. Haziran 2012;48(9):1269-82.
122. Henrikson NB, Webber EM, Goddard KA, Scrol A, Piper M, Williams MS, vd. Family history and the natural history of colorectal cancer: systematic review. *Genet Med Off J Am Coll Med Genet*. Eylül 2015;17(9):702-12.
123. Schoen RE, Razzak A, Yu KJ, Berndt SI, Firl K, Riley TL, vd. Incidence and mortality of colorectal cancer in individuals with a family history of colorectal cancer. *Gastroenterology*. Kasım 2015;149(6):1438-1445.e1.
124. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*. 10 Mayıs 2002;99(2):260-6.
125. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, vd. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 13 Temmuz 2000;343(2):78-85.
126. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, vd. Estimating the heritability of colorectal cancer. *Hum Mol Genet*. 15 Temmuz 2014;23(14):3898-905.
127. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW, vd. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol*. Şubat 2015;110(2):223-62; quiz 263.

128. Vasen HFA, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, vd. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut*. Haziran 2013;62(6):812-23.
129. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, vd. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun*. 30 Ekim 2015;6(1):8727.
130. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, vd. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. *Gastroenterology*. Ağustos 2018;155(2):383-390.e8.
131. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, vd. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 14 Ağustos 2013;14(2):207-15.
132. Medema JP. Cancer stem cells: the challenges ahead. *Nat Cell Biol*. Nisan 2013;15(4):338-44.
133. Nassar D, Blanpain C. Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annu Rev Pathol*. 23 Mayıs 2016;11:47-76.
134. de Sousa e Melo F, Kurtova AV, Harnoss JM, Kljavin N, Hoeck JD, Hung J, vd. A distinct role for Lgr5⁺ stem cells in primary and metastatic colon cancer. *Nature*. 29 Mart 2017;543(7647):676-80.
135. Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, vd. Visualization and targeting of LGR5⁺ human colon cancer stem cells. *Nature*. 11 Mayıs 2017;545(7653):187-92.

136. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, vd. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. Temmuz 2012;487(7407):330-7.
137. Loree JM, Pereira AAL, Lam M, Willauer AN, Raghav K, Dasari A, vd. Classifying Colorectal Cancer by Tumor Location Rather than Sidedness Highlights a Continuum in Mutation Profiles and Consensus Molecular Subtypes. *Clin Cancer Res Off J Am Assoc Cancer Res*. 01 Mart 2018;24(5):1062-72.
138. Lenz H-J, Ou F-S, Venook AP, Hochster HS, Niedzwiecki D, Goldberg RM, vd. Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). *J Clin Oncol Off J Am Soc Clin Oncol*. 01 Ağustos 2019;37(22):1876-85.
139. Baran B, Mert Ozupek N, Yerli Tetik N, Acar E, Bekcioglu O, Baskin Y. Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterol Res*. Ağustos 2018;11(4):264-73.
140. Lee GH, Malietzis G, Askari A, Bernardo D, Al-Hassi HO, Clark SK. Is right-sided colon cancer different to left-sided colorectal cancer? – A systematic review. *Eur J Surg Oncol EJSO*. 01 Mart 2015;41(3):300-8.
141. Salem ME, Weinberg BA, Xiu J, El-Deiry WS, Hwang JJ, Gatalica Z, vd. Comparative molecular analyses of left-sided colon, right-sided colon, and rectal cancers. *Oncotarget*. 21 Eylül 2017;8(49):86356-68.
142. Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H, vd. Comparison of 17,641 Patients With Right- and Left-Sided Colon Cancer: Differences in Epidemiology, Perioperative Course, Histology, and Survival. *Dis Colon Rectum*. Ocak 2010;53(1):57-64.
143. Warschkow R, Sulz MC, Marti L, Tarantino I, Schmied BM, Cerny T, vd. Better survival in right-sided versus left-sided stage I - III colon cancer patients. *BMC Cancer*. 28 Temmuz 2016;16(1):554.

144. Lal N, White BS, Goussous G, Pickles O, Mason MJ, Beggs AD, vd. KRAS Mutation and Consensus Molecular Subtypes 2 and 3 Are Independently Associated with Reduced Immune Infiltration and Reactivity in Colorectal Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 01 Ocak 2018;24(1):224-33.
145. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Sonesson C, vd. The consensus molecular subtypes of colorectal cancer. Nat Med. Kasım 2015;21(11):1350-6.
146. Lee MS, Menter DG, Kopetz S. Right Versus Left Colon Cancer Biology: Integrating the Consensus Molecular Subtypes. J Natl Compr Cancer Netw JNCCN. Mart 2017;15(3):411-9.
147. The Cancer Genome Atlas Program - National Cancer Institute [Internet]. 2018 [a.yer 25 Mayıs 2021]. Erişim adresi: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
148. The Cancer Genome Atlas - Timeline and Milestones - National Cancer Institute [Internet]. 2018 [a.yer 26 Mayıs 2021]. Erişim adresi: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/history/timeline>
149. GDC [Internet]. [a.yer 25 Mayıs 2021]. Erişim adresi: <https://portal.gdc.cancer.gov/>
150. AJCC - What is the AJCC? [Internet]. [a.yer 20 Mayıs 2021]. Erişim adresi: <https://cancerstaging.org/About/what-is-the-ajcc/Pages/whatisajcc.aspx>
151. AJCC6thEdCancerStagingManualPart1.pdf [Internet]. [a.yer 19 Mayıs 2021]. Erişim adresi: <https://cancerstaging.org/references-tools/deskreferences/Documents/AJCC6thEdCancerStagingManualPart1.pdf>
152. AJCC6thEdCancerStagingManualPart2.pdf [Internet]. [a.yer 19 Mayıs 2021]. Erişim adresi: <https://cancerstaging.org/references-tools/deskreferences/Documents/AJCC6thEdCancerStagingManualPart2.pdf>

153. Edge SB, American Joint Committee on Cancer, editörler. AJCC cancer staging manual. 7th ed. New York: Springer; 2010. 648 s.
154. Colorectal Cancer Stages | Rectal Cancer Staging | Colon Cancer Staging [Internet]. [a.yer 19 Mayıs 2021]. Erişim adresi: <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/staged.html>
155. Shaukat A, Kaltenbach T, Dominitz JA, Robertson DJ, Anderson JC, Cruise M, vd. Endoscopic Recognition and Management Strategies for Malignant Colorectal Polyps: Recommendations of the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 01 Kasım 2020;159(5):1916-1934.e2.
156. Colon JS for C of the, Rectum. Japanese Classification of Colorectal, Appendiceal, and Anal Carcinoma: the 3d English Edition [Secondary Publication]. *J Anus Rectum Colon*. 2019;3(4):175-95.
157. Colaprico A, Silva TC, Olsen C, Garofano L, Garolini D, Cava C, vd. TCGAbiolinks: TCGAbiolinks: An R/Bioconductor package for integrative analysis with GDC data [Internet]. Bioconductor version: Release (3.12); 2021 [a.yer 18 Mayıs 2021]. Erişim adresi: <https://bioconductor.org/packages/TCGAbiolinks/>
158. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, vd. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 05 Mayıs 2016;44(8):e71.
159. TCGAbiolinks.pdf [Internet]. [a.yer 23 Mayıs 2021]. Erişim adresi: <https://bioconductor.org/packages/release/bioc/manuals/TCGAbiolinks/man/TCGAbiolinks.pdf>
160. TCGAbiolinks | NCI Genomic Data Commons [Internet]. [a.yer 18 Mayıs 2021]. Erişim adresi: <https://gdc.cancer.gov/content/tcgabiolinks>

161. Silva TC, Colaprico A, Olsen C, D'Angelo F, Bontempi G, Ceccarelli M, vd. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Research. 28 Aralık 2016;5:1542.
162. Silva TC, Colaprico A, Olsen C, D'Angelo F, Bontempi G, Ceccarelli M, vd. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Research. 29 Haziran 2016;5:1542.
163. Kotsiantis S. Supervised Machine Learning: A Review of Classification Techniques. Inform Slov. 01 Ocak 2007;31:249-68.
164. SPSS Statistics - Genel Bakış [İnternet]. 2020 [a.yer 25 Mayıs 2021]. Erişim adresi: <https://www.ibm.com/tr-tr/products/spss-statistics>
165. Enrichr [İnternet]. Erişim adresi: maayanlab.cloud/Enrichr/
166. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, vd. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 15 Nisan 2013;14(1):128.
167. team shinyapps io. Chapter 2 Getting Started | shinyapps.io user guide [İnternet]. [a.yer 23 Mart 2021]. Erişim adresi: <https://docs.rstudio.com/shinyapps.io/getting-started.html#installation>
168. Shiny [İnternet]. [a.yer 31 Mayıs 2021]. Erişim adresi: <https://cancerapp.shinyapps.io/shiny/>
169. Yeo CJ. Shackelford's Surgery of the Alimentary Tract, E-Book. Elsevier Health Sciences; 2017. 2599 s.
170. Brandt CP, Malangoni MA, Priebe PP, Temes TT. Shackelford's Surgery of the Alimentary Tract, 5th Edition. Ann Surg. Ağustos 2002;236(2):261.
171. Bullard Dunn KM, Rothenberger DA. Colon, Rectum, and Anus. İçinde: Brunicaudi FC, Andersen DK, Billiar TR, Dunn DL, Hunter JG, Matthews JB, vd., editörler. Schwartz's Principles of Surgery [İnternet]. 10. bs New York, NY:

McGraw-Hill Education; 2015 [a.yer 20 Mayıs 2021]. Erişim adresi: accessmedicine.mhmedical.com/content.aspx?aid=1117749699

172. Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. Eylül 2010;143(3):331-6.
173. Glare P, Virik K, Jones M, Hudson M, Eychmuller S, Simes J, vd. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ*. 24 Temmuz 2003;327(7408):195.
174. Viganó A, Bruera E, Jhangri GS, Newman SC, Fields AL, Suarez-Almazor ME. Clinical Survival Predictors in Patients With Advanced Cancer. *Arch Intern Med*. 27 Mart 2000;160(6):861-8.
175. Feng J, Zhang H, Li F. Investigating the relevance of major signaling pathways in cancer survival using a biologically meaningful deep learning model. *BMC Bioinformatics*. 05 Şubat 2021;22(1):47.
176. Cooper K, Tappenden P, Cantrell A, Ennis K. A systematic review of meta-analyses assessing the validity of tumour response endpoints as surrogates for progression-free or overall survival in cancer. *Br J Cancer*. Kasım 2020;123(11):1686-96.
177. Alpaydin E. *Machine Learning: The New AI*. MIT Press; 2016. 225 s.
178. Domingos P. A few useful things to know about machine learning. *Commun ACM*. Ekim 2012;55(10):78-87.

8. EKLER

EK 1: ETİK KURUL RAPORU

DOKUZ EYLÜL ÜNİVERSİTESİ GİRİŞİMSSEL OLMAYAN ARAŞTIRMALAR ETİK KURUL KARARI

Sayın Prof.Dr. Hülya Ellidokuz

Araştırmanıza ilişkin Kurulumuz kararı aşağıda sunulmuştur.

Bilgilerinizi ve gereğini rica ederiz.

ETİK KOMİSYONUN ADI	DOKUZ EYLÜL ÜNİVERSİTESİ GİRİŞİMSSEL OLMAYAN ARAŞTIRMALAR ETİK KURULU
AÇIK ADRES	Dokuz Eylül Üniversitesi Tıp Fakültesi Dekanlığı 2. Kat İnciraltı-İZMİR
TELEFON	0 232 412 22 54-0 232 412 22 58
FAKS	0 232 412 22 43
E-POSTA	etikkurul@deu.edu.tr

BAŞVURU BİLGİLERİ	DOSYA NO:	5733-GOA
	ARAŞTIRMA	UZMANLIK TEZİ ☐ MÜNFERİT ARAŞTIRMA ☐ ÖÇM ☐ YÜKSEKLİSANS ☒ DOKTORA ☐
	ARAŞTIRMANIN AÇIK ADI	Açık Kaynaklı Klinik Kanser Verilerinin R-Shiny Uygulaması ile Yapay Zeka Tabanlı Web Arayüzü Destek Sisteminin Geliştirilmesi
	ARAŞTIRMA PROTOKOL KODU	
	SORUMLU ARAŞTIRMACI ÜNVANI/ADI/SOYADI ve UZMANLIK ALANI	Prof.Dr. Hülya Ellidokuz Onkoloji Enstitüsü
	ARAŞTIRMAYA KATILAN MERKEZLER	TEK MERKEZ ☐ ÇOK MERKEZLİ ☒

DEĞERLENDİRİLEN BELGELER	Belge Adı	Tarihi	Versiyon Numarası	Dili
	Araştırmacı dilekçesi	16.11.2020		Türkçe ☒ İngilizce ☐ Diğer ☐

KARAR BİLGİLERİ	Karar No:2020/28-29	Tarih:23.11.2020				
	Prof.Dr. Hülya Ellidokuz'un sorumlusu olduğu "Açık Kaynaklı Klinik Kansere Verilerinin R-Shiny Uygulaması ile Yapay Zeka Tabanlı Web Arayüzü Destek Sisteminin Geliştirilmesi" isimli klinik araştırmaya ait 16.11.2020 tarihli sorumlu araştırmacı dilekçesine ilişkin; -Sorumlu araştırmacı görevinin Dr. Öğr.Üyesi Gizem Çalıbaşı Koç'a devredilmesi ile ilgili belgeler incelenerek bilgi edinilmiş ve uygun bulunmuştur.					
ETİK KURUL BİLGİLERİ						
ÇALIŞMA ESASI	Dokuz Eylül Üniversitesi Girişimsel Olmayan Araştırmalar Etik Kurulu İşleyiş Yönergesi İy Klinik Uygulamaları Kılavuzu					
ETİK KURUL ÜYELERİ						
Unvanı/Adı/Soyadı	Uzmanlık Alanı	Kurumu	Cinsiyet	Araştırma ile ilişkili mi?		İmza
Prof.Dr.Sadık Kıvanç METİN (Başkan)	Kalp ve Damar Cerrahisi	DEU Tıp Fakültesi Kalp Damar Cerrahisi Anabilim Dalı	Erkek	E ☐	H ☒	
Prof.Dr. Sermin ÖZKAL (Başkan Yardımcısı)	Tıbbi Patoloji	DEU Tıp Fakültesi Tıbbi Patoloji A.D	Kadın	E ☐	H ☒	
Prof.Dr.Serkan YENER	Endokrinoloji	DEU Tıp Fakültesi İç Hastalıkları Anabilim Dalı	Erkek	E ☐	H ☒	Katılmadı
Prof.Dr.Pınar TUNCEL	Tıbbi Biyokimya	DEU Tıp Fakültesi Tıbbi Biyokimya Anabilim Dalı	Kadın	E ☐	H ☒	
Doç.Dr.Seher Özyürek	Muskuloskeletal Fizyoterapi - Fizik Tedavi ve Rehabilitasyon	DEU Fizik Tedavi ve Rehabilitasyon Yüksek Okulu	Kadın	E ☐	H ☒	
Doç.Dr.Nil Hocaoglu AKSAY	Tıbbi Farmakoloji	DEU Tıp Fakültesi Tıbbi Farmakoloji Anabilim Dalı	Kadın	E ☐	H ☒	Katılmadı
Doç.Dr.Murat BEKTAŞ	Çocuk Sağlığı ve Hastalıkları Hemşireliği	DEU Hemşirelik Fakültesi Çocuk Sağlığı ve Hastalıkları Hemşireliği	Erkek	E ☐	H ☒	
Doç.Dr.Tufan ÇANKAYA	Tıbbi Genetik	Tıbbi Genetik Anabilim Dalı	Erkek	E ☐	H ☒	
Doç.Dr.Ayfer DAYI	Davranış Fizyolojisi	DEU Tıp Fakültesi Fizyoloji Anabilim Dalı	Kadın	E ☐	H ☒	
Doç.Dr.Korcan DEMİR	Pediyatrik Endokrinoloji	DEU Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı	Erkek	E ☐	H ☒	
Doç.Dr.Mahmut Cem ERGON	Tıbbi Mikrobiyoloji	DEU Tıp Fakültesi Tıbbi Mikrobiyoloji Anabilim Dalı	Erkek	E ☐	H ☒	
Öğr.Gör.Dr.Kıvanç YÜKSEL	Biyoistatistik ve Tıbbi Bilişim	Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Bilişim A.D	Erkek	E ☐	H ☒	Katılmadı
Av.Esra FIRTINA	Avukat	DEU Rektörlüğü Hukuk Müşavirliği	Kadın	E ☐	H ☒	Katılmadı
Mehmet Erhan ÖZKUL	Sağlık mensubu olmayan üye	D.E.U Tıp Fakültesi İdari Mali İşler	Erkek	E ☐	H ☒	

EK 2: ÖZGEÇMİŞ



HÜSEYİN KORAY MISIRLIOĞLU

Kişisel Bilgiler

İletişim Bilgileri

İletişim Adresi

Telefon

E-posta

İnternet Sayfası

Öğrenim Bilgileri

08 Ağustos 2018 - Şu Anda (2 yıl 10 ay)
Yüksek Lisans, Tezli Program, DOKUZ EYLÜL ÜNİVERSİTESİ, TÜRKİYE
SAĞLIK BİLİMLERİ ENSTİTÜSÜ, TRANSLASYONEL ONKOLOJİ (YL) (TEZLİ)
Diploma Numarası: -
Ağırlıklı Genel Not Ortalaması: 3.92 / 4.0

01 Eylül 2010 - 24 Ekim 2016 (6 yıl 2 ay)
Lisans, Anadal/Normal Öğretim, GEBZE TEKNİK ÜNİVERSİTESİ, TÜRKİYE
TEMEL BİLİMLER FAKÜLTESİ, MOLEKÜLER BİYOLOJİ VE GENETİK PR.
Diploma Numarası: TEM-LS-79-27
Ağırlıklı Genel Not Ortalaması: 2.98 / 4.0

Yabancı Dil Bilgileri

İNGİLİZCE (Okuma: İyi, Yazma: İyi, Konuşma: İyi)

Bilimsel Teknolojik Faaliyet Alanları

Bilimsel Teknolojik Faaliyet Alanı Bilgileri

Temel Bilimler -- Yaşam Bilimleri -- Moleküler Biyoloji ve Genetik -- Hayvan
Moleküler Genetiği

Temel Bilimler -- Yaşam Bilimleri -- Moleküler Biyoloji ve Genetik -- Genomiks

Temel Bilimler -- Yaşam Bilimleri -- Moleküler Biyoloji ve Genetik -- Kanser
Moleküler Biyolojisi

Sağlık Bilimleri -- Tıp -- Temel Tıp Bilimleri -- Tıbbi Biyoloji

Temel Bilimler -- Yaşam Bilimleri -- Biyoinformatik -- Biyoenformasyon

Ar-Ge Yetkinlik

Makaleler

A. PAKDEMİRLİ, F. TOKSOZ, A. KARADAG, H. K. MİSİRLİOĞLU, Y. BASBİNAR, H. ELLİDOKUZ & O. ACIKGOZ, Role of mesenchymal stem cell-derived soluble factors and folic acid in wound healing, TURKISH JOURNAL OF MEDICAL SCIENCES, 2019, 1300-0144, 49, 3, 914-921.

TÜBİTAK Burs ve Destekleri

Proje Bilgileri

318S276, Kanserde İmmünoterapi Etkinliğinin İyileştirilmesine Yönelik Makine Öğrenmesine Dayalı Yeni Gen Keşfi ve İlaç Yeniden Konumlandırma Platformunun Geliştirilmesi, 1003 - Öncelikli Alanlar (2. Aşama), Burslu, Yürürlükte, ARDEB, SBAG - Sağlık Bilimleri Araştırma Destek Grubu, Projeye Katılma/Ayrılma Tarihleri: 01.03.2020 - 01.01.2023, Proje Başlangıç/Bitiş Tarihleri: 01.01.2020 - 01.01.2023.

Panelistlik/İzleyicilik/Raportörlük Sayısı

Hakemlik/Panelistlik/Dış Danışmanlık Sayısı	ARDEB/BİDEB 0	TEYDEB 0	Toplam 0
İzleyicilik/Danışmanlık Sayısı	ARDEB/BİDEB 0	TEYDEB 0	Toplam 0
Raportörlük Sayısı	ARDEB/BİDEB 0	TEYDEB 0	Toplam 0