Tıp Bilişimi Derneği

16. Tıp Bilişimi Kongresi

22-23 MAYIS 2025, ANKARA

Ana Tema: Sağlıkta Multiomiks

BİLDİRİLER KİTABI

16th Medical Informatics Congress

Main Theme: Multiomics in Healthcare

PROCEEDINGS

Editörler:
Kemal TURHAN, Ahmet YARDIMCI, Adil ALPKOÇAK, Abdullah Utku ŞENOL, Aslı Suner KARAKÜLAH
Kemal Hakan GÜLKESEN, Uğur BİLGE, Oğuz DİCLE, Filiz İŞLEYEN, Muammer ALBAYRAK
Mehmet Erşan KALAYCI, Sinem ÖZDEMİR, Ayşe Banu PAK, Işık ÇAKMAK

https://turkmia.net/kongre2025

TIP BİLİŞİMİ DERNEĞİ YAYINLARI

Sertifika No: 7163
Yayın Tarihi: 07/07/2025 Yayın No:3

16. Tıp Bilişimi Kongresi Bildiriler Kitabı

Proceedings of 16th Turkish Congress of Medical Informatics Association

Editörler / *Editors*

Kemal Turhan, Ahmet Yardımcı, Adil Alpkoçak, Abdullah Utku Şenol, Aslı Suner Karakülah, Kemal Hakan Gülkesen, Uğur Bilge, Oğuz Dicle, Filiz İşleyen, Muammer Albayrak, Mehmet Erşan Kalaycı, Sinem Özdemir, Ayşe Banu Pak, Işık Çakmak

Dijital olarak basılmış ve çevrimiçi olarak dağıtılmıştır.
268 sayfadır.

Ücretsizdir.

© 2025 Her hakkı saklıdır.

Bu kitaptaki makalelerdeki fikirler ilgili yazar(lar)a aittir. İlgili yazar(lar)ın yazılı izin alınmadan, makalelerin kısmen veya tamamı, herhangi bir platformda yeniden yayınlanamaz.

ISBN 978-605-72564-2-3

978-605-72564-2-3

16. TIP BİLİŞİMİ KONGRESİ

16th Turkish Congress of Medical Informatics Association

22-23 Mayıs 2025, Ankara, Türkiye

Hacettepe Üniversitesi
Sıhhiye Kültür Merkezi

BİLDİRİLER KİTABI

Proceedings

Editörler *Editors*

Kemal Turhan, Ahmet Yardımcı, Adil Alpkoçak, Abdullah Utku Şenol, Aslı Suner Karakülah, Kemal Hakan Gülkesen, Uğur Bilge, Oğuz Dicle, Filiz İşleyen, Muammer Albayrak, Mehmet Erşan Kalaycı, Sinem Özdemir, Ayşe Banu Pak, Işık Çakmak

Organizasyon *Organized by*

Tıp Bilişimi Derneği

Destekleyen Kuruluşlar *In Cooperation with*

EFMI
EUROPEAN FEDERATION for MEDICAL INFORMATICS

Ana Sponsor *Main Sponsored by*

ILYOME

Sponsorlar *Sponsored by*

AKGÜN

oteo

mergentech

KONGRE BAŞKANI Congress Chair

Kemal TURHAN, İzmir Ekonomi Üniversitesi

DÜZENLEME KURULU Editorial Board

Kemal TURHAN, İzmir Ekonomi Üniversitesi

Ahmet Yardımcı, Akdeniz Üniversitesi

Kemal Hakan Gülkesen, Akdeniz Üniversitesi

Adil Alpkoçak, İzmir Bakırçay Üniversitesi

Abdullah Utku Şenol, Akdeniz Üniversitesi

Aslı Süner Karakülah, Ege Üniversitesi

Oğuz Dicle, Dokuz Eylül Üniversitesi

Tunca Doğan, Hacettepe Üniversitesi

Ali Güner, Karadeniz Teknik Üniversitesi

Burçin Kurt, Karadeniz Teknik Üniversitesi

Mehmet Kıvrak, Recep Tayyip Erdoğan

Filiz İşleyen, Sağlık Bakanlığı

Muammer Albayrak, Karadeniz Teknik Üniversitesi

KONGRE SEKRETERİ Congress Secretary

Filiz İşleyen, Sağlık Bakanlığı

Muammer Albayrak, Karadeniz Teknik Üniversitesi

Mehmet Erşan Kalaycı, Karadeniz Teknik Üniversitesi

BİLİMSEL KURUL Scientific Committee

Ahmet Yardımcı, Akdeniz Üniversitesi (Başkan)

Kemal Turhan, İzmir Ekonomi Üniversitesi

Adil Alpkoçak, İzmir Bakırçay Üniversitesi

Afsun Ezel Türkmen Esatoğlu, Ankara Üniversitesi

Ahmet Ademoğlu, Boğaziçi Üniversitesi

Ali Yılmaz, Kırıkkale Üniversitesi

Alper Selver, Dokuz Eylül Üniversitesi

Aslı Suner Karakülah, Ege Üniversitesi

Azim Çelik,GE Healthcare

Barış Emre Dayanç, İzmir Ekonomi Üniversitesi

Burçin Kurt, Karadeniz Teknik Üniversitesi

Cavit Boz, Karadeniz Teknik Üniversitesi

Cemil Çolak, İnönü Üniversitesi

Cengizhan Öztürk ,Boğaziçi Üniversitesi

Demet Gündüz, Akdeniz Üniversitesi

Ebru Akçapınar Sezer, Hacettepe Üniversitesi

Emel Timuçin, Acıbadem Mehmet Ali Aydınlar Üniversitesi

Engin Yenilmez, Karadeniz Teknik Üniversitesi

Engin Yenilmez, Karadeniz Teknik Üniversitesi

Esin Öztürk Işık, Boğaziçi Üniversitesi

Fatih Gürcan, Karadeniz Teknik Üniversitesi

Filiz İşleyen, Sağlık Bakanlığı

Gökhan Ertaş,Yeditepe Üniversitesi

Gökhan Karakülah, Dokuz Eylül Üniversitesi

Hakan Abacıoğlu, İzmir Ekonomi Üniversitesi

Hanefi Özbek , İzmir Bakırçay Üniversitesi

Hasan Yiğit, Sağlık Bilimleri Üniversitesi

Kaan Kurtel, İzmir Ekonomi Üniversitesi

Kemal Hakan Gülkesen, Akdeniz Üniversitesi

Kılıçhan Gürleyik,University of Pittsburgh

Leman Tomak, Ondokuz Mayıs Üniversitesi

Mehmet Kıvrak, Recep Tayyip Erdoğan Üniversitesi

Mehmet Serdar Güzel, Ankara Üniversitesi

Mehmet Ünlütürk, Yaşar Üniversitesi

Menderes Tarcan, Osmangazi Üniversitesi

Mine Duman, Karadeniz Teknik Üniversitesi

Muammer Albayrak, Karadeniz Teknik Üniversitesi

Murat Canpolat,Akdeniz Üniversitesi

Murat Osman Ünalır, Ege Üniversitesi

Mustafa Mahir Ülgü, Sağlık Bakanlığı

Mustafa Necmi İlhan, Gazi Üniversitesi

Mustafa Özmen, Hacettepe Üniversitesi

Nese Zayim, Akdeniz Üniversitesi

Neşe Zaim, Akdeniz Üniversitesi

Nurhan Gürsel Özmen, Karadeniz Teknik Üniversitesi

Oğuz Dicle, Dokuz Eylül Üniversitesi

YEREL DÜZENLEME KOMİTESİ
Local Organizing Committee

Sinem Özdemir, Karadeniz Teknik Üniversitesi

Ayşe Banu Pak, Trabzon Üniversitesi

Işık Çakmak, Karadeniz Teknik Üniversitesi

Orhan Er, İzmir Bakırçay Üniversitesi

Osman Uğur Sezerman, Acıbadem Mehmet Ali Aydınlar Üniversitesi

Öznur Özkasap, Koç Üniversitesi

Pembe Keskinoğlu, Dokuz Eylül Üniversitesi

Pınar Özbay, Boğaziçi Üniversitesi

Selen Bozkurt, Emory Üniversitesi

Semih Utku, Dokuz Eylül Üniversitesi

Sibel Kul, Karadeniz Teknik Üniversitesi

Suat Özdemir, Hacettepe Üniversitesi

Şükrü Özen,Akdeniz Üniversitesi

Timur Köse, Ege Üniversitesi

Tolga Berber, Karadeniz Teknik Üniversitesi

Tunca Doğan, Hacettepe Üniversitesi

Tuncay Aydoğan,Isparta Uygulamalı Bilimler Üniversitesi

Uğur Bilge, Akdeniz Üniversitesi

Uras Tos, Yaşar Üniversitesi

Utku Şenol, Akdeniz Üniversitesi

Yaşar Güneri Şahin, İzmir Ekonomi Üniversitesi

Yeşim Aydın Son, Ortadoğu Teknik Üniversitesi

Yılmaz Kemal Yüce, Alanya Alaaddin Keykubat Üniversitesi

Yiğit Ali Üncü, Akdeniz Üniversitesi

Zeliha Aydın Kasap, Karadeniz Teknik Üniversites

GRAFİK TASARIM
Graphic Design

Belgin İpek, Karadeniz Teknik Üniversitesi

İLETİŞİM ADRESİ E-mail Address
kongre2025@turkmia.net

KONGRE WEB Congress Web
https://turkmia.net/kongre2025/

İçindekiler *Contents*	Sayfa No *Page #*
Önsöz *Foreword*	viii
Erken Dönem Parkinson Teşhisinde Beyin Bölgelerinin BağlantısallığınınAraştırılması için Biyobelirteç Önerisi Neslihan BAKİ ve Nurhan GÜRSEL ÖZMEN	*1*
Crohn ve Ülseratif Kolit Hastalıklarında Doku ve KanTranskriptomları ile Diferansiyel Ekspresyon ve Yolak İncelemesi Sinem ÖZDEMİR, Işık ÇAKMAK, Kemal TURHAN	*9*
Kalp Yetmezliği Hastalarında Böbrek Fonksiyonu Sınıflamasına İlişkin Üç Yönlü ROC Analizi Selçuk ORHAN, Burçin KURT	*10*
A Comparative Study of BiLSTM, BiLSTM +CRF, and SpaCy Models for De identification of Turkish Electronic Health Records Fatih DİLMAÇand Adil ALPKOÇAK	*15*
Fall Risk Prediction from Clinical Patient Notes with Machine Learning Mehmet YİĞİT, Hilal YİĞİT, Kemal TURHAN	*24*
Meme Kanseri Alt Tiplerinde Gen Ekspresyonu ve İlaç Duyarlılığı İlişkisinin Multiomik Analizi Adem SONUVAR, Esra TOKUR SONUVAR, Uğur BİLGE	*30*
Effect of FICZ-Induced AhR Activation on Lipid Profiles in Cancer Cells using Lipidomics Sema DURUOĞLU, Cemil Can EYLEM, Emirhan NEMUTLU, Cansu ÖZDEMIR SAKA, İdil YET	*45*
Investigating Synthetic Lethal Interactions Using Machine Learning Approaches in Cancer Datasets Mehmet Tevfik KARA, İdil YET	*51*
Prediction of Metastasis Risk Using Machine Learning Based on BRAF Mutation and miRNA Expression in Thyroid Cancer Mehmet Erşan KALAYCI, Kemal TURHAN	*55*
Predicting Type 1 Diabetes from Gut Microbiome Data: A Machine Learning Approach Combining Alpha/Beta Diversity and PICRUSt2-Inferred Metabolic Pathways Açelya DALGIÇ, İdil YET	*56*
Retrosentez İçin Makine Öğrenimi Kullanımı Güven ŞAHİN, Mehmet Erşan KALAYCI, Kemal TURHAN	*61*
FonokardiyogramVerileri ile Üfürüm Teşhisi: Derin Öğrenme Tabanlı Bir Yaklaşım Yasin GÖRMEZ, Fatma Hilal YAĞIN, Abdulvahap PINAR	*65*
Development of a Mobile Calculator for Mortality Prediction in Elderly Individuals Using the Multidimensional Prognostic Index İhsan DOĞAN, Pembe KESKİNOĞLU	*70*
Sağlıkta Biyoinformatik Dönüşümüne Yerli Bir Yanıt: Ilyome Deneyimi ILYOME	*72*
Bazı Schiff Bazları ve Amin Türevlerinin in Siliko Değerlendirilmesi: Moleküler Kenetlenme ve ADME Analizi Işık ÇAKMAK , Sinem ÖZDEMİR , Nurhan GÜMRÜKÇÜOĞLU	*74*

Model Context Protocol is Necessary for AI Agents in the Healthcare Domain Aslı BOYRAZ	*76*
Classification of Acute Myeloid Leukemia (AML) Gene Expression Dataset and Identification of Feature Importances Using Machine Learning Methods Hüseyin KUTLU, Cemil ÇOLAK	*80*
Önceden Eğitilmiş Yapay Zeka Modelleri ile Miyokard Enfarktüsünün Tahmini ve Sınıflandırılması Cemil ÇOLAK, Abdulvahap PINAR, Ahmet Kadir ARSLAN, Fatma Hilal YAĞIN	*93*
Biostatapps: Bir Kullanıcı Dostu Veri Bilimi ve Yapay Zekâ Temelli Yazılım Platformu Ahmet Kadir ARSLAN, Cemil ÇOLAK, Hasan UCUZAL, Emek GÜNDOĞAN, Şeyma YAŞAR, Fatma Hilal YAĞIN, Zeynep KÜÇÜKAKÇALI, İpek BALIKÇI ÇİÇEK, Harika G.G. BAĞ	*99*
Key Challenges in Developing Explainable Ensemble Learning Models for Classification of Atherosclerotic Heart Diseases Ahmet Kadir ARSLAN, Hasan UCUZAL, Cemil ÇOLAK	*103*
Atherosclerotic Cardiovascular Disease Prediction and Categorization Using Pre-trained Artificial Intelligence Models Furkan ŞAKİROĞLU, Cemil ÇOLAK, Mehmet Cengiz ÇOLAK	*105*
Sağlık Hizmetlerinde Dijital Onam ve Hukuki Çerçevesi Bilgenur KARA, Filiz İŞLEYEN	*109*
Kimyasal Bileşiklerin Antimikrobiyal Özelliğinin Makine Öğrenmesi Yöntemleriyle Belirlenmesi Sinem ÖZDEMİR, Işık ÇAKMAK, Nurhan GÜMRÜKÇÜOĞLU	*118*
Tıp Bilişimi Araştırmalarında Tematik Eğilimler ve Yapay Zekanın Etkisi: 2018-2024 Dönemi Bibliyometrik Analizi Adem SONUVAR, Ahmet YARDIMCI, Kemal Hakan GÜLKESEN, Esra TOKUR SONUVAR	*120*
Bibliometric Analysis of Machine Learning Applications in Psoriasis Treatment: Health Informatics and Economics Perspectives Mehtap ÇAKMAK BARSBAY	*125*
Akciğer Kanserinin Sınıflandırılmasında Makine Öğrenmesi Modellerinin Karşılaştırmalı Performans Analizi Funda İPEKTEN, Abdulvahap PINAR, Elif Çelik GÜRBULAK	*126*
State Detection of Brain Tumors and Patient Life Expectancy Prediction with Explainable Artificial Intelligence-Based Multimodal System Ahmet Caner TAT, Emirhan YILDIZ, Salih KIZILIŞIK	*136*
The Impact of Large Language Models on Health and Medical Informatics Ruken ERDEM DEMİR, Mehmet Erşan KALAYCI, Sinem ÖZDEMİR, Kemal TURHAN	*148*
The Role of Radiomics in Dentistry in the Age of Digital Transformation and Artificial Intelligence Ali RASAT, Selmi TUNÇ	*155*
Kişiselleştirilmiş Tıp Uygulamaları için Multi-Omik Verilerin Bütünleştirilmesi: NoSQL Yaklaşımı Muammer ALBAYRAK	*167*
Blockchain Technology in Healthcare Data Zeliha AYDIN KASAP, Mehmet Erşan KALAYCI	*179*
Transformers-based Artificial Intelligence Models in the Prediction and Classification	*182*

of Heart Valve Diseases Hasan UCUZAL, Cemil ÇOLAK, Mehmet Cengiz ÇOLAK	
Artificial Intelligence and Machine Learning in Eating Disorders Öykü Peren TÜRK, Nural ERZURUM ALİM	*186*
Identification of Genes Affecting Type 2 Diabetes Mellitus Based on GEO Database Ayşe Banu PAK	*196*
Integration of Artificial Intelligence into Medical Education Curriculum Selçuk AKTURAN	*198*
Survival Analysis with Artificial Intelligence Approach in Gastric Cancer Ali GÜNER, Burçin KURT	*206*
Metabolomik Profilden Akciğer Kanser Türü Tahmini Duygu KORKMAZ YALÇIN, Yasin GÖRMEZ, Aslı BOZ, Neşe ZAYİM	*207*
Evaluation of Some Agricultural Pesticides Against Halyomorpha halys by Molecular Docking and Identification of Potential Candidates Ayşe Banu PAK	*209*
Bibliometric Analysis of Histopathology Studies Utilizing AlexNet in Web of Science Serdar YANIK	*211*
Kardiyovasküler Hastalık Tahmini: Keşifsel Veri Analizi ve Modelleme Yaklaşımı Mehmet KIVRAK, Handan DUMAN, Hakan DUMAN	*222*
Dijital Okur Yazarlık Eğitimi İçin Öğretim Üyelerinin Hazır Bulunuşluk Durumu Kemal Hakan GÜLKESEN, Utku ŞENOL, Volkan DOĞRU, Yeşim ŞENOL, Ayşe Gülbin ARICI	*232*
Fonksiyonel Manyetik Rezonans Görüntüleme ile Ruhsal Hastalıkların Yapay Zeka Kullanılarak Tahmin Edilmesi: Bibliyometrik Bir Analiz Cansu GÜNDOĞAN TOKER	*236*
Fair and Efficient Transmission of Multi-Omics Data via NOMA: A Hybrid Machine Learning Approach using Random Forest and KNN Seda KIRTAY, Veysel Gokhan BOCEKCI, Kazim YILDIZ, Murat KOCAK	*242*
Exploring the trends of Multiomics Research: A Comprehensive Bibliometric Analysis with Interactive Visualization Tools (BiblioMaps) Murat KOÇAK, Seda KIRTAY, Zafer AKÇALI	*246*
GPT-4'ün Hasta Notlarından Klinik Bulguları Belirleme Yetisinin Değerlendirilmesi: NBME Veri Seti ile Karşılaştırmalı Bir Çalışma Esra TOKUR SONUVAR, Adem SONUVAR, Nevruz İLHANLI, Ahmet YARDIMCI	*261*

ÖNSÖZ Foreword

Sayın Tıp Bilişimi Araştırmacıları,

Tıp Bilişimi Derneği olarak, hızla gelişen bilim ve teknolojinin sağlık alanındaki yansımalarını ülkemizle buluşturmak ve bu alanın akademik altyapısını daha da güçlendirmek amacıyla çıktığımız yolda, bir bilimsel buluşmayı daha başarıyla tamamlamanın gururunu yaşıyoruz. Bu yıl, Hacettepe Üniversitesinde düzenlediğimiz 16. Tıp Bilişimi Kongresi, "Sağlıkta Multi-Omics" ana temasıyla disiplinlerarası iş birliklerini destekleyen, bilgi üretimini teşvik eden ve geleceğin sağlık sistemine yön verecek teknolojileri ele alan bir platform olmuştur.

Multi-Omics veri yaklaşımları, yapay zeka ve büyük veri analitiğiyle birleştiğinde, yalnızca araştırma dünyasını değil, klinik karar süreçlerinden tedavi planlamasına kadar birçok alanda sağlık hizmetlerini dönüştürme gücüne sahiptir. Kongremiz, bu dönüşümün tartışıldığı, farklı bakış açılarıyla zenginleşen ve sağlık bilişimi alanında yeni araştırma ağlarının kurulduğu bir ortam yaratmıştır.

Tıp Bilişimi Kongresi'nde sunulan bilimsel bildirilerin yer aldığı bu kitabın, siz değerli okuyucularımıza ilham vermesini ve yeni projelere kapı aralamasını temenni ediyoruz. Bu sürece katkı sunan tüm araştırmacılara, kurumlara, destekçilere ve organizasyon ekibine içten teşekkürlerimizi sunarız.

Dear Medical Informatics Researchers,

As the Medical Informatics Association, we are proud to conclude another successful scientific gathering in our ongoing mission to integrate the rapid advancements of science and technology into the healthcare field and to strengthen the academic foundations of medical informatics in our country. The 16th Medical Informatics Congress, organized at Hacettepe University with the main theme of "Multi-Omics in Healthcare", served as a platform to encourage interdisciplinary collaboration, promote scientific discoveries and examine technologies that shape the future of healthcare.

When combined with artificial intelligence and big data analytics, Multi-Omics approaches have the transformative power to influence not only the research landscape but also clinical decision-making and personalized medicine. Our congress has offered a vibrant space for discussing this transformation and has paved the way for new academic and professional networks.

We hope that this proceedings book, which contains the scientific contributions presented during the 16th Congress, will inspire you, our esteemed readers, and open the door to new ideas and collaborative projects. We extend our heartfelt thanks to all researchers, institutions, sponsors, and the organizing committee who contributed to this meaningful endeavor.

Prof. Dr. Kemal TURHAN
Kongre Başkanı *Congress Chair*

Erken Dönem Parkinson Teşhisinde Beyin Bölgelerinin BağlantısallığınınAraştırılması için Biyobelirteç Önerisi

Neslihan BAKİ[1] ve Nurhan GÜRSEL ÖZMEN[2]

[1] Karadeniz Teknik Üniversitesi, Biyoistatistik ve Tıp Bilişimi, Trabzon
[2]Karadeniz Teknik Üniversitesi, Makine Mühendiliği Bölümü, Trabzon
neslihanbaki520@gmail.com,gnurhan@ktu.edu.tr

Özet. Parkinson hastalığı (PH), motor işlev bozukluklarıyla karakterize ilerleyici nörodejeneratif bir hastalıktır. Hastalığın erken tanısı tedavi seçimini etkilediğinden oldukça önem arz etmektedir. Bu çalışmada, EEG verisi ile PH'nin erken evresinde düşük frekans bantlarındaki bağlantısallık değişimlerinin beyin bölgelerine göre değerlendirilmesi amaçlanmıştır. Çalışma, açık kaynaklı veri seti üzerinden yaş ve cinsiyete göre eşleştirilmiş 14 PH'li ve 14 sağlıklı kontrolden (SK) elde edilen, 31 kanallı EEG verisi ile gerçekleştirilmiştir. Çalışmada, güç spektral yoğunluğu (PSD) ve EEG bantları için Ağırlıklı Karşılıklı Sembolik Bilgi (wSMI) öznitelikleri çıkarılmış ve sınıflandırma performansları CatBoost algoritması ile değerlendirilmiştir. Çalışmada bölgesel bağlantısallık ilişkisini görmek amacıyla frontal bölge, santral bölge ve oksipital bölge elektrotlarına yönelik ayrı değerlendirme yapılmıştır. Değerlendirme sonunda tüm 31 kanal verisi kullanılması durumunda ve frontal bölge kanalları kullanılması durumunda yüksek başarı elde edilmiştir. 18 kanallı santral bölgede 12 öznitelik kullanıldığında %95.00 doğruluk sağlanmıştır. Oksipital bölgede wSMI Teta, wSMI Delta ve wSMI (Teta)/wSMI (Delta) oranı kullanılmasına rağmen, sınıflandırma performansı düşük çıkmıştır. wSMI'nin aynı bölgede belirgin biçimde düşük olmasının, PH'nin erken evresinde bağlantısal bütünlüğün korunduğunu, buna rağmen sistemik bir değişimin başladığı çıkarılabilir. Son olarak, önerilen wSMI (Delta) özniteliği tek başına santral bölgede %93.33 doğrulukla, erken dönem PH için anlamlı bir biyobelirteç olarak önerilebilir.

Anahtar Sözcükler: Parkinson Hastalığı, EEG Sinyal İşleme, CatBoost, wSMI, Biyobelirteç

Biomarker Proposal for Investigating Connectivity of Brain Regions in Early Parkinson's Diagnose

Neslihan BAKİ[1] and Nurhan GÜRSEL ÖZMEN[2]

[1]Karadeniz Technical University, Biostatistics and Medical Informatics, Trabzon
[2]Karadeniz Technical University, Mechanical Engineering, Trabzon
neslihanbaki520@gmail.com,gnurhan@ktu.edu.tr

Abstract. Parkinson's disease (PD) is a progressive neurodegenerative disease characterized by motor dysfunction. Early diagnosis of the disease is of great importance as it affects treatment selection. In this study, it is aimed to evaluate the connectivity changes in low frequency bands according to brain regions in the early stage of PD using EEG data. The study was conducted with 31-channel EEG data obtained from 14 PD patients and 14 healthy controls (HC) matched according to age and gender from an open-source dataset. In the study, power spectral density (PSD) and Weighted Mutual Symbolic Information (wSMI) features were extracted for EEG bands, and the classification performances were evaluated with the CatBoost algorithm.In the study, separate evaluations were made for the frontal region, central region and occipital region electrodes in order to see the regional connectivity relationship. At the end of the evaluation, high success was achieved when all 31 channel data was used and when the frontal region channels were used. When 12 features were used in the central region with 18 channels, 95.00% accuracy was achieved. Findings regarding the occipital region showed that although wSMI Theta, wSMI Delta and wSMI (Theta)/wSMI (Delta) ratio were used in this region, the classification performance was low. The fact that wSMI was significantly lower in the same region may suggest that connectivity integrity is preserved in the early stage of PD, yet a systemic change has begun. Finally, the proposed wSMI (Delta) attribute alone can be proposed as a significant biomarker for early PH with an accuracy of 93.33% in the central region.

Keywords: Parkinson's Disease, EEG Signal Processing, CatBoost, wSMI, Biomarker

1 Giriş

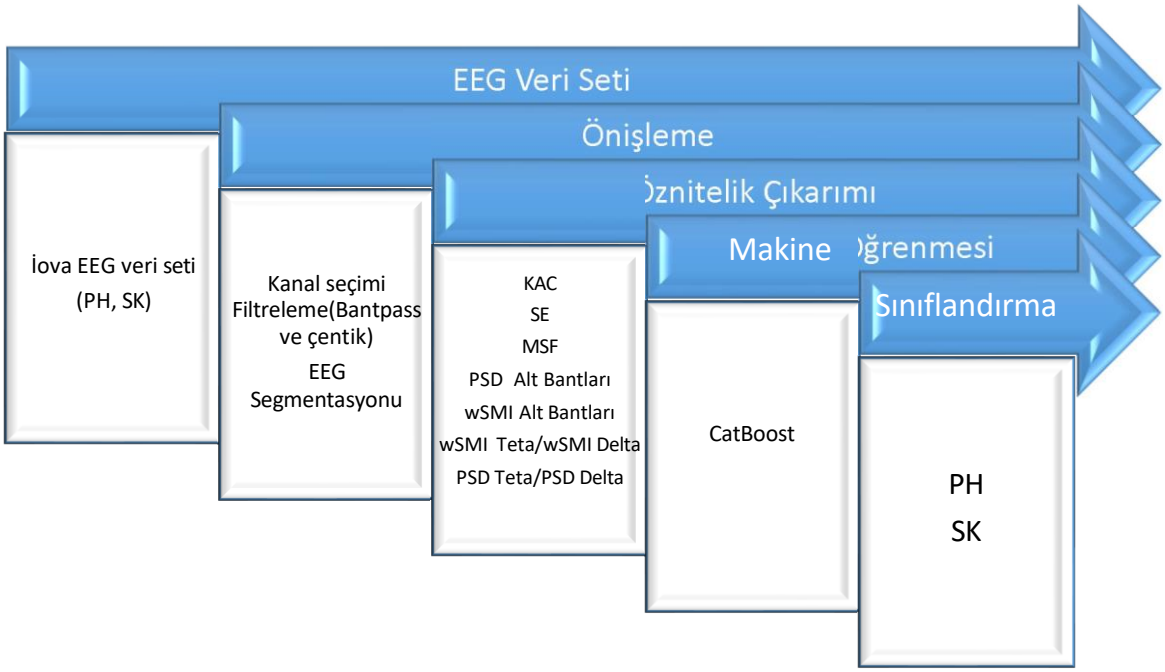
Parkinson hastalığı (PH), dünya genelinde nüfusun yaklaşık %1'ini etkileyen, ilerleyici ve yaygın görülen nörodejeneratif bir hastalıktır. Genetik ve çevresel nedenlerin etkileşimi sonucu ortaya çıkan ve çok sayıda hücresel olayı etkileyen bir durumdur [1]. Orta beyin bölgesindeki substantia nigra pars kompaktadaki dopaminerjik nöron kaybına bağlı olarak gelişerek, tremor, rijidite ve bradikinezi gibi motor belirtilerle seyreden bir hastalıktır [2]. Hastalığın erken döneminden itibaren motor bozukluklarla birlikte bu hastalarda, mekan algısı ve oryantasyonu gibi vizyospasyal yeti kayıpları da görülmektedir [3].
Ancak bu semptomlar, nörodejeneratif sürecin ileri evrelerinde ortaya çıkar ve tanıda gecikmelere neden olur. Bu nedenle, erken tanı için objektif, nörofizyolojik temelli yeni objektif biyobelirteçlere ihtiyaç vardır. Erken teşhis, tedavi seçimini etkilemektedir.
Son yıllarda, manyetoensefalografi (MEG) ve elektroensefalografi (EEG) temelli araştırmalar, PH teşhisinde öncü bir rol oynamıştır. EEG, yüksek zaman çözünürlüğü, non-invaziv yapısı ve sinaptik düzeydeki işlevsel değişiklikleri yansıtabilme kapasitesi sayesinde, PH'nin tanısında umut vadeden bir teknik olarak öne çıkmaktadır. Beyin aktivitesinin değerlendirilmesinde EEG kayıtlarının kullanımı oldukça yerindedir; zira PH, esasen subkortikal düzeydeki bozulmalarla ilişkili bir hastalıktır [6]. PH'de EEG kullanımı üzerine yürütülen çok sayıda çalışma bulunmaktadır [7, 8]. Bu çalışmaların önemli bir kısmı, spektral analizlere dayanmakta olup, delta ve teta genliklerinde artış gözlenmiştir [9]. Ancak, klasik spektral öznitelikler, beyin bölgeleri arasındaki bilgi alışverişini yansıtmada sınırlı kalabilmektedir. Bu noktada, beyin bağlantısallığının analizi devreye girmektedir. Beyin bölgeleri arasındaki bağlantıların incelenmesi, beyin aktivitesine dair değerli bilgiler sunmakta ve EEG tabanlı sınıflandırmalarda önemli katkılar sağlamaktadır. Bağlantısallık, genel olarak üç ana başlık altında sınıflandırılmaktadır: yapısal bağlantı, fonksiyonel bağlantı ve etkili bağlantı. Yapısal bağlantı, anatomik yollarla belirlenen fiziksel bağlantıları ifade ederken; fonksiyonel bağlantı, farklı beyin bölgeleri arasındaki istatistiksel bağımlılıkları veya korelasyonları ortaya koyar. Etkili bağlantı ise bu bölgeler arasındaki nedensel ve yönsel ilişkileri değerlendirmeye odaklanır. EEG literatüründe bu tür bağlantıları ölçmek üzere çeşitli metrikler tanımlanmıştır [10]. Bu çalışmada, bilgi kuramına dayalı fonksiyonel bağlantı ölçümleri ön plana çıkarılmıştır.
Bilgi kuramı temelli metrikler, özellikle karşılıklı bilgi tabanlı yöntemler, beyin bölgeleri arasındaki senkronizasyonu daha ayrıntılı bir biçimde ölçme potansiyeline sahiptir [11]. Ağırlıklı Karşılıklı Sembolik Bilgi (wSMI) yöntemi, sinyaller arasındaki ortak bilgi içeriğini simgesel temsiller üzerinden analiz eder. wSMI, PH, bilinç durumları, epilepsi ve alzheimer hastalığı gibi çeşitli klinik durumların ayırt edilmesinde başarıyla kullanılmıştır [12]. Ancak, PH özelinde wSMI'nin bantlar arası karşılaştırmalı analizine dair sistematik bir değerlendirme literatürde henüz yer almamaktadır. Bu çalışmada önerilen wSMI (Delta) söz konusu boşluğu doldurmayı amaçlamakta ve literatürde ilk kez PH test edilmektedir.
EEG verileri üzerinden PH'nin erken teşhisine yönelik bir araştırma sunulmuştur. 14 PH'li ve 14 SK grubundan oluşan 28 kişiye ait dinlenme durumu EEG veri seti üzerinden yeni öznitelikler çıkarılarak, PH ve SK'nın ayırt edilmesi amaçlanmıştır. Bu amaçla CatBoost makine öğrenmesi algoritması kullanılmıştır. Ayrıca bölgesel elektrot verileri üzerinden, hastalığa ait bağlantısallık ilişkisi araştırılmıştır.

2 Yöntem

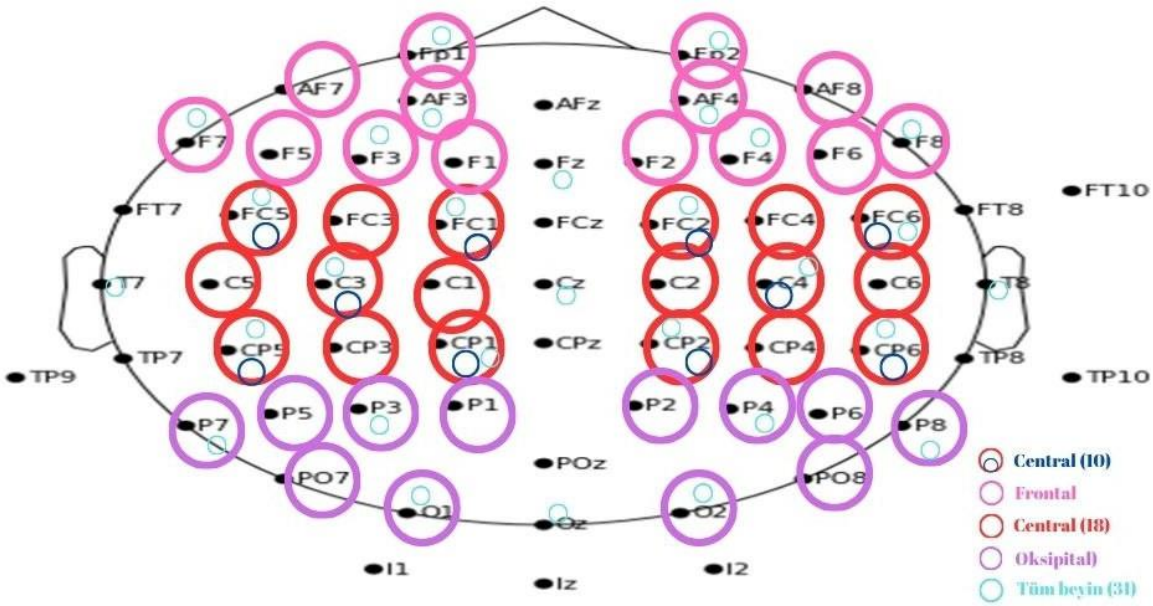
2.1 Veri Seti Tanıtımı ve Önişleme

Bu çalışmada, ABD Iova Üniversitesi tarafından kaydedilip, internet erişimine açılan EEG Iova veri seti kullanılmıştır [13]. Veri seti 14 PH ve 14 kontrol grubuna ait dinlenme durumu EEG kayıtlarından oluşmaktadır. Veri, uluslararası 10-20 sistemine göre kafaya yerleştirilen 64 kanallı elektrotlar ile toplanmıştır. Verideki SK grubu, PH ile yaş ve cinsiyet açısından demografik olarak eşleştirilmiş ve herhangi bir nöropsikiyatrik bozukluk semptomu veya öyküsü olmayan fiziksel ve zihinsel olarak sağlıklı kontrol grubundan oluşmaktadır. Tüm katılımcılara ait yazılı bilgilendirilmiş onay mevcuttur. PH'lilere, ilaçlı durumlarında nöropsikolojik ve anket değerlendirilmesi yapılmış olup, Birleşik Parkinson Hastalığı Derecelendirme Ölçeği (UPDRS) de uygulanmıştır. Dinlenme durumu EEG verileri, 64 kanallı sinterlenmiş Ag/AgCl elektrotlarından 0,1–100 Hz aralığında, 500 Hz örnekleme hızıyla Brain Vision sisteminde, 2 dakika gözler açık olarak kaydedilmiştir. Pz kanalı referans kanal olarak seçilmiştir. Analizler Frontal (14), Oksipital (14), Santral(18), Santral (10) ve tüm beyin (31 kanal) EEG verileri üzerinden yapılmıştır. Kullanılan EEG kanalları, 31 kanal: Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, O1, Oz, O2, CP5, P3, P7, PO3, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz; Santral10 kanal: FC5, FC1, FC2, FC6, C3, C4, CP5, CP1, CP2, CP6; Santral 18 kanal FC1, FC2, FC3, FC4, FC5, FC6, C1, C2, C3, C4, C5, C6, CP1, CP2, CP3, CP4, CP5, CP6; Frontal 14 kanal,Fp1, F3, F7, F4, F8, Fp2, AF7, F1, F5, F3, F6, AF8, AF4, F2; Oksipital 14 kanal, P3, P7, O1, O2,

P4, P8, P1, P5, PO7, PO3, PO4, PO8, P6, P2‘den oluşmaktadır. Burada 40 saniyelik epoklama yapılmış, 0.1-120 Hz bant geçiren filtre ve 60 ve 100 Hz’de çentik filtre uygulanmıştır. EEG veri analizi adımları Şekil 1’de yer almaktadır. Bölgesel seçilen ve tüm beyinin analiz yapılan kanal konumları Şekil 2’de verilmiştir.



Şekil 1. EEG veri analiz aşamaları



Şekil 2. İşlem için seçilen elektrot konumları

2.2. Öznitelik Çıkarımı

Bu çalışmada EEG verisinin teta, alfa, beta, gama bantları üzerinden Güç Spektral Yoğunluğu (PSD) hesaplanmış ve öznitelik olarak kullanılmıştır. Ayrıca Spektral Entropi (SE), Kolmogorov Algoritmik Karmaşıklık (KAC), Medyan Spektral Frekans (MSF), wSMI Delta, wSMI Teta, wSMI Alfa, wSMI Beta, wSMI Gama özniteliklerinden oluşan 12 öznitelik kullanılmıştır.

2.2.1. Ağırlıklı Sembolik Karşılıklı Bilgi (wSMI)

Ağırlıklı Sembolik Karşılıklı Bilgi, iki sinyal arasında doğrusal olmayan bir bilgi paylaşımı endeksinin tahminidir [12]. Tau zamanla ayrılabilen sinyalin k örnekleri dikkate alınmaktadır. Bu k örneklerin, büyüklükleri aralarında bulunan sıra ilişkisine bağlı olarak, bir dizi sembol tanımlamaktadır. Bu tanımlanan semboller, sinyallerin zamansal değişimini temsil etmektedir. Her bir sembol çiftinin ortak olasılığını tahmin ederek, dönüştürülmüş her bir EEG sinyal çifti için tahmin edilmektedir. Bu ifade denklem (1)'de verilmiştir.

$$wSMI(X,Y) = \frac{1}{\log k!} \sum_{x \in X} \sum_{y \in Y} w(x,y)\, p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (1)$$

Burada k, sembolik dönüşüm için kullanılan vektörün boyutu, x ve y sırasıyla X ve Y sinyallerinde bulunan tüm semboller, w(x,y) ağırlık matrisi ve p(x, y) X sinyalinde x simgesinin ve Y sinyalinde y simgesinin birlikte ortaya çıkma olasılığının ortak olasılığıdır. Son olarak, p(x) ve p(y), her sinyaldeki bu simgelerin olasılıklarıdır.

Analizlerde her bir elektrotun medyan değeri hesaplanmış, wSMI parametreleri, wSMI Delta, wSMI Teta, öznitelikleri hesaplanırken sırasıyla, tau = 64, tau = 32 ve k = 3 olarak seçilmiştir. Bu ayrıştırılmış EEG sinyallerinin alt bantlarının PSD'leri üzerinden wSMI'ye ait wSMI Delta, wSMI Teta öznitelikleri hesaplanmıştır.

2.3 Sınıflandırma

2.3.1. Kategorik Arttırma Algoritması (CatBoost)

Kategorik Arttırma Algoritması, Yandex tarafından geliştirilmiş olup, heterojen veri setleriyle etkili bir şekilde çalışabilen Gradient Boosting ve Karar Ağacı algoritmalarından türetilen yeni bir topluluk öğrenme yöntemidir [14]. CatBoost, aynı zamanda temel tahminciler olarak dengeli unutkan ağaçlar kullanarak aşırı öğrenme problemlerinin önüne geçmektedir. Genel olarak, diğer topluluk öğrenme algoritmalarına kıyasla daha yüksek sınıflandırma başarısı göstermektedir.

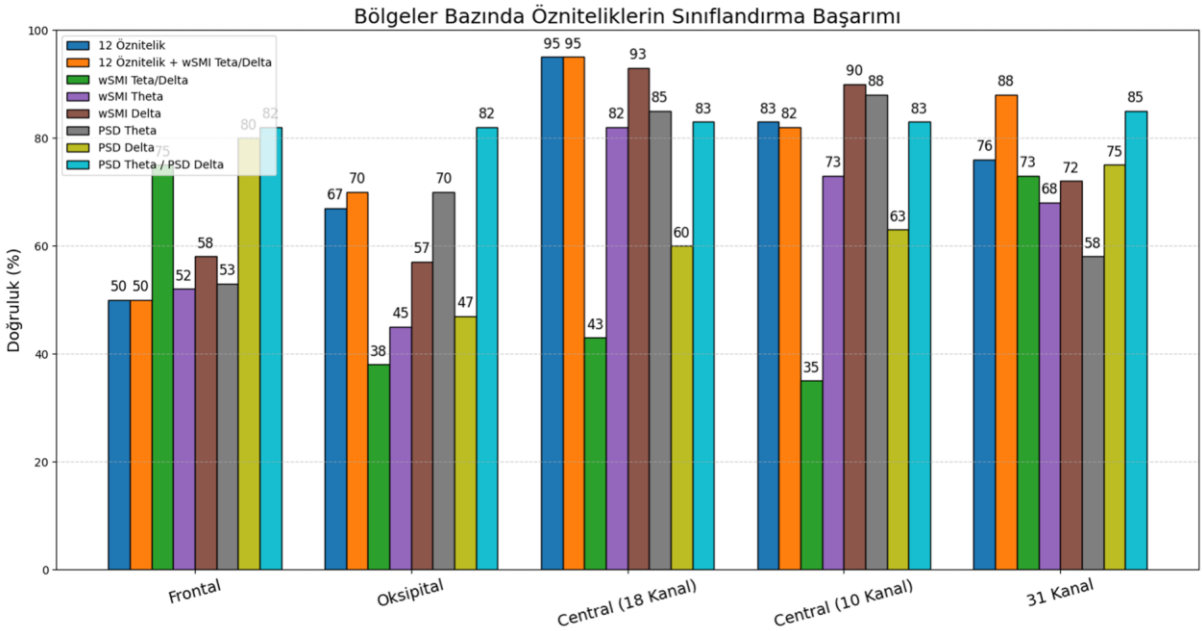
3 Bulgular

Çalışmanın sonuçları tüm elektrotlar (31 kanal) ve bölgesel seçilen elektrotların performansları şeklinde beş farklı durum için verilmiştir. wSMI öznitelikleri tek başına kullanımı ve diğer on iki öznitelik ile birlikte kullanımının sınıflandırma performansına etkileri araştırılmıştır. wSMI Teta/ wSMI Delta Oranı, 12 Öznitelik ve wSMI Teta / wSMI Delta, 12 Öznitelik, wSMI Teta, wSMI Delta, PSD Teta, PSD Delta, PSD Teta / PSD Delta oranının CatBoost algoritması kullanılarak gerçekleştirilen 10 katlı çapraz doğrulama analizlerinde elde edilen sınıflandırma doğrulukları, bölgesel dağılımlarla birlikte Tablo 1'de sunulmuştur. Ayrıca Şekil 3'de de bölge bazlı doğruluk sonuçları yer almaktadır. Bu çalışmada, PH'nin EEG verileri üzerinden sınıflandırılmasında farklı öznitelik setlerinin ve beyin bölgelerinin ayırt edici gücü değerlendirilmiştir.

Tablo 1. Bölgelere göre CatBoost Algoritması ile 10 katlı çapraz doğrulama sonuçları (doğruluk)

Öznitelikler	Frontal	Oksipital	Santral(18)	Santral(10)	Tüm beyin(31)
wSMI (Teta)/ wSMI (Delta)	**75.00 (+/- 30)**	38.33 (+/- 23)	43.33 (+/- 29)	35.00 (+/- 26)	**73.00 (+/- 32)**
12 Öznitelik ve wSMI (Teta) / wSMI (Delta)	50.00 (+/- 21)	70.00 (+/- 31)	**95.00 (+/- 15)**	**81.67 (+/- 24)**	**88.00 (+/- 18)**
12 Öznitelik	50.00 (+/- 21)	66.67 (+/- 29)	**95.00 (+/- 15)**	83.33 (+/- 21)	76.00 (+/- 15)

wSMI Teta	51.67 (+/- 33)	45.00 (+/- 21)	**81.67 (+/- 18)**	**73.33 (+/- 29)**	68.33 (+/- 32)
wSMI Delta	58.33 (+/- 32)	56.67 (+/- 24)	**93.33 (+/- 13)**	**90.00 (+/- 15)**	**71.67 (+/- 21)**
PSD Teta	53.33 (+/- 30)	**70.00 (+/- 22)**	**85.00 (+/- 18)**	**88.33 (+/- 18)**	58.33 (+/- 29)
PSD Delta	**80.00 (+/- 20)**	46.67 (+/- 33)	60.00 (+/- 35)	63.33 (+/- 26)	**75.00 (+/- 17)**
PSD Teta / PSD Delta	**81.67 (+/- 18)**	**81.67 (+/- 24)**	**83.33 (+/- 22)**	**83.33 (+/- 22)**	**85.00 (+/- 18)**



Şekil 3. Bölge bazında Catboost ile 10 katlı çapraz doğrulama sonuçları

Iova veri setinde CatBoost algoritması ve 10 katlı çapraz doğrulama sonuçlarına baktığımızda; **wSMI Delta/Teta oranı**, özellikle frontal bölgede %75.00 (±30) doğrulukla yüksek bir ayırt edicilik göstermiştir. Bu oran, tüm beyin genelinde de benzer şekilde %73.00 (±32) doğruluk sağlamış; ancak oksipital (%38.33 ±23), santral 18 (%43.33 ±29) ve santral10 (%35.00 ±26) bölgelerinde sınıflandırma başarısı düşük kalmıştır. **12 öznitelik ve wSMI Teta/Delta oranının birleştirilmesiyle** oluşturulan öznitelik seti, en yüksek doğruluk değerine %95.00 (±15) ile santral (18 kanal) bölgesinde ulaşmıştır. Aynı öznitelik seti, santral (10 kanal) bölgesinde %81.67 (±24) ve tüm beyin genelinde %88.00 (±18) doğruluk sağlamıştır. Bu kombinasyon, oksipital bölgede %70.00 (±31) ve frontal bölgede %50.00 (±21) doğruluk üretmiştir. **Yalnızca 12 öznitelik** kullanıldığında da benzer sonuçlar gözlenmiştir. Santral (18) bölgesinde %95.00 (±15), santral (10) bölgesinde %83.33 (±21) ve tüm beyinde %76.00 (±15) doğruluk elde edilmiştir. Bu bulgu, öznitelik setinin wSMI oranı eklense de eklenmese de merkezi alanlarda yüksek sınıflandırma başarısı sunduğunu göstermektedir. **wSMI Teta** özniteliği tek başına kullanıldığında, özellikle santral (18 kanal) bölgesinde %81.67 (±18), santral (10 kanal) bölgesinde %73.33 (±29) ve tüm beyinde %68.33 (±32) doğruluk üretmiştir. Oksipital bölgedeki doğruluğu ise %45.00 (±21) ile sınırlı kalmıştır. **wSMI Delta** ise benzer şekilde en yüksek doğruluğa Santral bölgelerde ulaşmıştır (santral 18: %93.33 ±13; santral 10: %90.00 ±15). Tüm beyin genelindeki doğruluğu %71.67 (±21) iken, frontal ve oksipital bölgelerdeki sınıflandırma performansı daha düşüktür. **Geleneksel frekans temelli öznitelikler** olan **PSD Teta** ve **PSD Delta**, özellikle oksipital ve frontal bölgelerde dikkat çekici doğruluklar sağlamıştır. PSD Teta, oksipital bölgede %70.00 (±22), santral (10) bölgede ise %88.33 (±18) doğruluk sağlamıştır. PSD Delta ise frontal bölgede %80.00 (±20), tüm beyinde ise %75.00 (±17) doğrulukla sınıflandırmada etkili olmuştur. Genel olarak, bulgular PH'nin EEG tabanlı sınıflandırılmasında hem öznitelik türlerinin hem de elektrot bölgelerinin belirleyici rol oynadığını göstermektedir. Özellikle santral bölgeden elde edilen veriler, hem bilgi kuramsal hem de frekans temelli özniteliklerle yüksek sınıflandırma başarısı sağlamıştır. Bununla birlikte, oksipital bölgedeki wSMI temelli bağlantısallık özniteliklerinin sınıflandırma başarısının düşük olması, bu bölgedeki bağlantısallığın erken evrelerde daha az bozulmuş olabileceğine işaret etmektedir.

4 Tartışma ve Sonuçlar

Bu çalışmada, PH'de kortikal bilgi alışverişinin frekansa özgü dinamiklerini değerlendirmek amacıyla, bilgi kuramına dayalı bir öznitelik olan wSMI Delta önerilmiştir. EEG sinyalleri arasındaki bilgi paylaşımını ölçmeyi amaçlayan wSMI yöntemi; (a) sinyallerin entropilerini hızlı ve güvenilir bir şekilde tahmin etmesi, (b) doğrusal olmayan bağlantıları tespit edebilmesi ve (c) ortak kaynaklardan kaynaklanan sahte korelasyonlara duyarsız olması özellikleriyle öne çıkmaktadır. wSMI, iki EEG sinyali arasındaki rastgele olmayan ortak dalgalanmaların (bilgi paylaşımı) derecesini değerlendirir ve bilinç durumuna bağlı olarak arttığı bilinmektedir [11]. Literatürde wSMI çoğunlukla tekil frekans bantları üzerinden incelenmişken [11,15], bu çalışmada frekans bantları arasındaki bilgi paylaşımı dengesine odaklanan oransal bir yaklaşım benimsenmiştir. PH bu tür frekanslar arası karşılaştırmaya dayalı wSMI analizi literatürde ilk kez yapılmıştır. Önerilen wSMI özniteliği, özellikle hastalığın erken evresinde gözlenen nörofizyolojik değişimlerle uyumludur. Erken evre PH'de delta ve teta bantlarında relatif güç artışı bildirilmiş olmakla birlikte, entropi temelli analizlerde bu bantlarda bilgi çeşitliliğinin azaldığı (yani daha düşük entropi değerleri) görülmüştür. Teta bandı, özellikle prefrontal korteksin yürütücü işlevler ve bilişsel kontrol süreçleriyle ilişkili bağlantılarında kritik bir role sahiptir [16] ve yapılan çalışmalarda PH'de teta bant dinamiklerinin bilişsel gerilemeyle birlikte kademeli olarak azaldığı ortaya konmuştur [17,18]. Bu bağlamda wSMI oran özniteliğindeki düşüklük, PH'de hem düşük frekanslı kortikal senkronizasyonun göreceli korunmasına hem de bilişsel işlevlerle ilişkili daha yüksek frekanslı teta iletişiminin bozulmasına işaret edebilir. Oran bazlı bu wSMI özniteliği, bölgesel olarak tek başına kullanıldığında sınıflandırma doğruluğunda anlamlı sonuçlar elde edilmiş, klasik özniteliklerle birlikte kullanıldığında ise sınıflandırma performansını anlamlı düzeyde artırmıştır. Çalışmada Santral (sensorimotor) bölge, en yüksek sınıflandırma başarısının elde edildiği bölge olmuştur (%95 doğruluk). Bu bulgu, PH'de motor korteksle ilişkili alanlarda görülen erken evre sinirsel bozulmalarla tutarlıdır [19]. Santral bölgede hem klasik öznitelikler hem de wSMI Delta ve PSD Teta gibi frekans/bağlantısallık temelli ölçütler güçlü ayırt ediciliğe sahiptir. Bu durum, motor sistemin hem spektral gücünde hem de bilgi transfer yapısında farklılıklar oluştuğunu ortaya koymaktadır. Frontal bölgede ise en yüksek doğruluk PSD Teta/Delta oranı ile (%81) elde edilmiştir. Bu durum, PH'de frontal korteksin bilişsel yavaşlamaya bağlı olarak değişen genliklerin dağılımlarını yansıttığına işaret etmektedir. wSMI oranı bu bölgede tek başına %75 doğruluk sağlamış ancak klasik öznitelikler birleştiğinde katkısı sınırlı kalmıştır. Bu sonuç, bozulmaların çoğunlukla frontal lobla ilişkili yürütücü işlevlerde saptandığı bilgisiyle tutarlıdır [20]. wSMI Deltanın, hastalığın erken evresinde özgün bir biyobelirteç olarak sınıflandırma başarısını artırabileceği ve ilerleyen evrelerde hastalık progresyonunu takip etmek için potansiyel bir araç olabilir. Bu çalışmada anlamlı değerler bulunmuş olup, çalışmanın genellenebilir hale gelmesi amacıyla, önerilen wSMI öznitelikleri, farklı veri setleri ile farklı durumlar için değerlendirilebilir ve topoğrafik analizlerle desteklenecek şekilde sürdürülebilir.

Kaynaklar

1. Valls-Solé, Josep, and Francesc Valldeoriola. "Neurophysiological correlate of clinical signs in Parkinson's disease." *Clinical neurophysiology* 113.6 (2002): 792-805.
2. Postuma, Ronald B., et al. "MDS clinical diagnostic criteria for Parkinson's disease." Movement disorders 30.12 (2015): 1591-1601.
3. Mahlknecht, Philipp, et al. "Meta-analysis of dorsolateral nigral hyperintensity on magnetic resonance imaging as a marker for Parkinson's disease." Movement Disorders 32.4 (2017): 619-623.
4. Parkinson's UK Homepage, http://www.parkinsons.org.uk/about-parkinsons/what-is-parkinsons.aspx, last accessed 2025/04/28.
5. Braak, Heiko, et al. "Staging of the intracerebral inclusion body pathology associated with idiopathic Parkinson's disease (preclinical and clinical stages)." Journal of neurology 249 (2002): iii1-iii5.
6. Pappalettera, Chiara, et al. "Analysis of complexity in the EEG activity of Parkinson's disease patients by means of approximate entropy." GeroScience 44.3 (2022): 1599-1607.
7. Stoffers, D., et al. "Slowing of oscillatory brain activity is a stable characteristic of Parkinson's disease without dementia." Brain 130.7 (2007): 1847-1860.
8. Serizawa, Kan, et al. "Comparison of quantitative EEGs between Parkinson disease and age-adjusted normal controls." Journal of Clinical Neurophysiology 25.6 (2008): 361-366.
9. Babiloni, Claudio, et al. "Cortical sources of resting state electroencephalographic rhythms in Parkinson's disease related dementia and Alzheimer's disease." Clinical Neurophysiology 122.12 (2011): 2355-2364.
10. Chiarion, Giovanni, et al. "Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends." *Bioengineering* 10.3 (2023): 372.
11. King, Jean-Rémi, et al. "Information sharing in the brain indexes consciousness in noncommunicative patients." Current biology 23.19 (2013): 1914-1919.
12. Sitt, Jacobo Diego, et al. "Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state." Brain 137.8 (2014): 2258-2270.

13. Anjum, Md Fahim, et al. "Linear predictive coding distinguishes spectral EEG features of Parkinson's disease." Parkinsonism & related disorders 79 (2020): 79-85.
14. Prokhorenkova, Liudmila, et al. "CatBoost: unbiased boosting with categorical features." Advances in neural information processing systems 31 (2018).
15. Melloni, Margherita, et al. "Cortical dynamics and subcortical signatures of motor-language coupling in Parkinson's disease." Scientific reports 5.1 (2015): 11899.
16. Cavanagh, James F., et al. "Diminished EEG habituation to novel events effectively classifies Parkinson's patients." Clinical Neurophysiology 129.2 (2018): 409-418.
17. Schmiedt, Christina, et al. "Theta and alpha oscillations reflect differences in memory strategy and visual discrimination performance in patients with Parkinson's disease." Neuroscience letters 388.3 (2005): 138-143.
18. Güntekin, Bahar, et al. "Abnormalities in auditory and visual cognitive processes are differentiated with theta responses in patients with Parkinson's disease with and without dementia." International Journal of Psychophysiology 153 (2020): 65-79.
19. Wu, Tao, et al. "Regional homogeneity changes in patients with Parkinson's disease." Human brain mapping 30.5 (2009): 1502-1510.
20. Caviness, John N., et al. "Defining mild cognitive impairment in Parkinson's disease." Movement disorders: official journal of the Movement Disorder Society 22.9 (2007): 1272-1277.

Crohn ve Ülseratif Kolit Hastalıklarında Doku ve Kan Transkriptomları ile Diferansiyel Ekspresyon ve Yolak İncelemesi

Sinem ÖZDEMİR[1][0000-0002-6332-0891], Işık ÇAKMAK[1][0000-0003-4758-4077], Kemal TURHAN[2][0000-0001-7871-3025]

[1] Karadeniz Teknik Üniversitesi,Tıp Fakültesi, Temel Tıp, Biyoistatistik ve Tıp Bilişimi,Trabzon,TÜRKİYE
[2] İzmir Ekonomi Üniversitesi Tıp Bilişimi ve Biyoistatistik, İzmir, TÜRKİYE

Abstract. İnflamatuar bağırsak hastalığı (İBH), dünya genelinde artış gösteren kronik ve tekrarlayan inflamatuar bir grup hastalığı ifade eder; en yaygın alt türleri Crohn hastalığı (CH) ve ülseratif kolittir (ÜK). Bu çalışmada, NCBI-GEO'dan indirilen GSE57945 (RNA-seq) ve GSE112057 (mikroarray) veri setleri GEO2R aracıyla analiz edilerek $|\log_2 FC| > 1$ ve adj. $p < 0{,}05$ eşikleriyle diferansiyel ekspresyon tespit edilmiştir. Hem CH hem de ÜK grupları için kontrol gruplarına kıyasla "yüksek" ve "düşük" ekspresyona sahip gen listeleri ayrı ayrı oluşturulmuş; iki veri setinde ortak profilli genler, g:Profiler'ın Python API'si ile GO:BP, KEGG ve Reactome veri tabanlarında zenginleştirme analizine tabi tutulmuştur. CH'de aşırı-eksprese genler, bağışıklık yanıtı, nötrofil aktivasyonu, bakteriyel savunma, fagositoz ve sitokin sinyal yollarıyla ilişkilendirilirken; ÜK'de kompleman sistemi, antijen tanıma, nötrofil aktivasyonu ve sitokin aracılı sinyalizasyonun merkezi rol oynadığı belirlenmiştir. Her iki alt türde, aşırı-eksprese genlerle zenginleştirilmiş 12 ortak biyolojik yolak saptanmış; düşük-ekspresyonlu genlerde ise ortak bir yol bulunamamıştır. Ayrıca, her iki hastalıkta da disakkarit ve polifenol metabolizmasındaki enzimatik süreçlerin baskılandığı gözlenmiştir. Bulgular, İBH'nin moleküler patogenezine dair hem özgün hem de ortak inflamatuvar mekanizmaları ortaya koyarak, hastalık yönetimi ve tedavi stratejileri geliştirilmesine yönelik değerli içgörüler sunmaktadır.

Keywords: Crohn Hastalığı, GEO2R, g:Profiler, İnflamatuar Bağırsak Hastalığı, Ülseratif Kolit

Kalp Yetmezliği Hastalarında Böbrek Fonksiyonu Sınıflamasına İlişkin Üç Yönlü ROC Analizi

Selçuk ORHAN[1 [0000-0003-0528-5255]], Burçin KURT[2 [0000-0001-5781-2382]]

1 Ordu İl Sağlık Müdürlüğü, Ordu, Türkiye
2 Karadeniz Teknik Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Trabzon, Türkiye

selcukorhan52@gmail.com
burcinnkurt@gmail.com

Özet: Bu çalışmanın amacı, üç sınıflı ya da sınıfa indirgenebilen hastalıklar için tanı koyma sürecinde Üç Yönlü ROC Analizi (Three-way ROC Analysis) yöntemi kullanılarak tanı performansının değerlendirilmesi ve biyobelirteçler için uygun kesim noktalarının belirlenmesidir. Üç sınıflı tanı testlerinin sınıfları ayırt etme gücü, ROC yüzeyi ve bu yüzeyin altında kalan hacim (HVUS: Hypervolume Under Surface) ile değerlendirilmektedir. Bununla birlikte, analiz sonucunda üç sınıfı birbirinden en iyi ayıran kesim noktaları da belirlenebilmektedir.
Çalışmada kullanılan açık erişimli veri seti(PhysioNet, Hospitalized patients with heart failure: integrating electronic healthcare records and external outcome data), Çin'in Sichuan Eyaleti'nde yer alan Zigong Dördüncü Halk Hastanesi'nde 2016–2019 yılları arasında kalp yetmezliği nedeniyle hastaneye yatırılan 2.008 hastaya ait elektronik sağlık kayıtları ve takip verilerini içeren retrospektif bir kohort çalışmasından elde edilmiştir.
İstatistiksel analizler için Python programlama dili ve HVUS hesaplaması için açık kaynaklı kütüphaneler kullanılmıştır. Hastaların böbrek fonksiyonlarını değerlendirmede yaygın olarak kullanılan glomerüler filtrasyon hızı (GFR) referans yöntem olarak tercih edilmiş ve ROC analizi çerçevesinde değerlendirilecek yeni parametre olarak ise kan cystatin-C düzeyi (ortalama:1.83, std:0.94, min:0.23, max:10.37) seçilmiştir. Klinik referanslara göre GFR değeri esas alınarak beş kategoriye ayrılan böbrek sağlığı durumu, bu çalışmada benzer özellikteki kategorilerin birleşimiyle; Evre 1 (Sağlıklı), Evre 2 (Orta derecede bozukluk), Evre 3 (İleri derecede bozukluk) olmak üzere üç kategoride değerlendirilmiştir. Üç yönlü ROC analizine göre HVUS 0.7792 olarak hesaplanmış, varyansı 0.0001 olarak elde edilmiştir. Cystatin-C düzeyi için en iyi kesim noktaları için sırasıyla; Evre 1-Evre 2'yi ayırmada"1.52" ve Evre 2-Evre 3'ü ayrırmada "2.52" olarak belirlenmiş, bu değerlere göre testin doğruluğu %74 olarak saptanmıştır.
Elde edilen bulgular, sayısal sonuçlu yeni parametrelerin üç sınıflı hastalık evrelemesinde kullanılabileceğini ve cystatin-C düzeylerinin GFR ile tutarlı şekilde tanı koymada yardımcı olabileceğini göstermektedir.

Three-Way ROC Analysis of Renal Function Classification in Heart Failure Patients

Abstact: The aim of this study is to evaluate the diagnostic performance and to determine appropriate cut-off points for biomarkers by using Three-way ROC Analysis method in the diagnostic process for diseases with three classes or diseases that can be reduced to classes. The power of three-class diagnostic tests to discriminate between classes is evaluated by the ROC surface and the volume under this surface (HVUS: Hypervolume Under Surface). However, the analysis can also identify the cut-off points that best distinguish the three classes from each other.
The open access dataset (PhysioNet, Hospitalized patients with heart failure: integrating electronic healthcare records and external outcome data) used in the study was obtained from a retrospective cohort study of electronic healthcare records and follow-up data of 2,008 patients hospitalized for heart failure between 2016 and 2019 at Zigong Fourth People's Hospital in Sichuan Province, China.

Python programming language was used for statistical analyses and open-source libraries were used for HVUS calculation. Glomerular filtration rate (GFR), which is widely used to assess renal function of patients, was preferred as the reference method and blood cystatin-C level (mean:1.83, std:0.94, min:0.23, max:10.37) was chosen as the new parameter to be evaluated within the framework of ROC analysis. Renal health status, which is divided into five categories based on GFR value according to clinical references, was evaluated in three categories as Stage 1 (Healthy), Stage 2 (Moderately impaired), Stage 3 (Severely impaired) by combining categories with similar characteristics in this study. According to the three-way ROC analysis, HVUS was calculated as 0.7792 with a variance of 0.0001. The best cut-off points for cystatin-C level were determined as "1.52" and "2.52" for differentiating Stage 1-Stage 2 and Stage 2-Stage 3, respectively, and the accuracy of the test was determined as 74% according to these values.
The findings suggest that new parameters with quantitative results can be used in three-class disease staging and that cystatin-C levels can be helpful in making a diagnosis consistent with GFR.

Anahtar Kelimeler: Böbrek fonksiyon sınıflaması, Üç Yönlü Roc Analizi, cystatin-C düzeyi, GFR

1. Giriş

Modern tıpta doğru tanı koyma süreci, yalnızca doğru hastalığı saptamayı değil, aynı zamanda hastalığın evresini de belirlemeyi gerektirmektedir. Özellikle kronik hastalıklarda, örneğin böbrek fonksiyon bozukluklarında, hastalığın evrelenmesi tedavi süreci açısından büyük önem taşımaktadır. Geleneksel olarak iki sınıflı(hastalık/sağlıklı) tanı sistemlerinin yeterli olduğu düşünülse de, birçok klinik durumda hastalık süreci çok daha karmaşık bir seyir izlemekte ve birden fazla evreden oluşmaktadır[1]. Bu nedenle, çok sınıflı(multi-class) tanı yaklaşımlarının istatistiksel olarak modellenmesi, sağlık alanında karar destek sistemleri açısından önemli bir ihtiyaç hâline gelmiştir.

Bu bağlamda, üç sınıflı ROC(Receiver Operating Characteristic) analizleri, çok evreli hastalıkların tanısında doğruluk değerlendirmesi ve sınıflar arası ayırt ediciliği ölçmek için etkili bir yöntem olarak öne çıkmaktadır. Bu analizlerde ROC eğrisinin bir uzantısı olan ROC yüzeyi ve bu yüzeyin altında kalan hacim(Volume Under Surface - VUS), tanı testlerinin üçlü sınıflandırmalardaki performansını istatistiksel olarak değerlendirmek için kullanılmaktadır. ROC yüzeylerinin temel avantajı, sayısal sonuç veren tanı testleriyle birden fazla hastalık evresi arasındaki sınırları belirlemeye imkân tanımasıdır[2]. Elde edilen bulgular, çok sınıflı ROC analizinin yalnızca performans ölçümünde değil, aynı zamanda yeni tanı değişkenlerinin sınıflandırma potansiyelini ortaya koymada da etkili bir araç olabileceğini göstermektedir[3]. Bu çalışma ile de cystatin-C gibi alternatif biyobelirteçlerin klinik karar süreçlerinde kullanılabilirliğini destekleyen istatistiksel bir bakış açısı kazanılabileceği düşünülmektedir.

Amaç, hasta kayıtları üzerinden böbrek fonksiyonunu değerlendiren glomerüler filtrasyon oranı(GFR) aracılığıyla cystatin-C gibi bir biyobelirtecin üç evreli hastalık sınıflamasında tanı performansını değerlendirmektir. Çalışmada, literatürde GFR'ye göre evrelenen böbrek sağlığı durumu için cystatin-C düzeylerinin sınıflar arası ayırt ediciliği, üç yönlü ROC analizi ve HVUS(Hypervolume Under Surface) ölçütü kullanılarak karşılaştırılmış; buna ek olarak en uygun kesim noktaları istatistiksel olarak belirlenmiştir[4].

2. Materyal ve Metod

Bu çalışmada, veri seti olarak açık kaynak platformu olan https://physionet.org adresinde yer alan ve Çin'in Sichuan eyaletindeki Zigong Dördüncü Halk Hastanesi'nde 2016–2019 yılları arasında kalp yetmezliği nedeniyle hastaneye yatırılan 2008 hastanın retrospektif sağlık kayıtları kullanılmıştır[5]. Veri setinde yer alan bilgilerden demografik nitelikte olanlar hemşireler tarafından manuel olarak girilmiş, laboratuar bulguları ise her hasta için hastane kabulünün ilk gününde elde edilmiş kayıtları içermektedir. Veriler hastanenin elektronik sağlık kayıt sisteminden SQL (Structured Query Language) sorguları ile dışa aktarılmış olup, veri seti için harici bir veri kaynağıyla eşleştirme uygulaması bulunmamaktadır.

Böbrek fonksiyonlarının değerlendirilmesinde çeşitli uluslararası kuruluşlar tarafından belirlenen farklı ölçütler kullanılmakla beraber; bu çalışmada, yaygın kullanımı sebebiyle "KDIGO 2012 Kronik Böbrek Hastalığının Değerlendirilmesi ve Yönetimi için Klinik Uygulama Kılavuzu" çerçevesinde glomerüler filtrasyon oranının beş aşamada sınıflandırıldığı evreleme sistemi tercih edilmiştir[6]. Bu kılavuz değerleri referans yöntem oluşturmak amacıyla kullanılmış olup, mevcut durumu üç evrede sınıflayabilmek amacıyla yeniden bir sınıflandırma yapılmıştır(**Tablo1**). Üç sınıflı yeni evreleme sistemi oluşturulurken, aynı rehberde yer alan ve böbrek yetmezliğinin prognoz tahmininde idrar albumin düzeyinin de dahil edilerek sunulduğu sınıflandırmaya göre benzer özellikteki grupların olmasından yola çıkılmıştır.

Tablo 1: Böbrek Fonksiyon Sınıflaması için Kullanılan Evreleme Kategorileri

GFR Kategorileri(ml/min/1.73 m^2)	KDIGO Evreleme	Yeni Oluşturulan Evreleme
≥90	G1 (Normal)	Evre 1(Sağlıklı)
60-89	G2 (Hafif azalma)	
45-59	G3a(Ilımlı azalma)	Evre 2(Orta derecede bozukluk)
30-44	G3b(Ilımlı-ağır azalma)	
15-29	G4 (Ağır azalma)	Evre 3(İleri derecede bozukluk)
<15	G5 (Böbrek Yetmezliği)	

Çalışma çerçevesinde kesim noktası belirlenecek ve kesin olmayan yeni parametre olarak ise düşük molekül ağırlıklı proteinaz inhibitörü olan cystatin-C düzeyi tercih edilmiştir. Cystatin-C, GFR hesaplanması için birkaç parametre ile birlikte de kullanılabilen bir biyobelirteç olup, daha erken dönemde bilgi sağlaması ve kas kütlesinden bağımsız olması özelliklerine sahiptir.

Çalışmada HVUS hesaplandıktan sonra ortak değişken belirleyebilmek amacıyla sayısal değişken olarak 'BMI' ve nominal değişkenler olarak 'gender', 'NYHA.cardiac.function.classification', 'Killip.grade', 'myocardial.infarction','congestive.heart.failure', 'peripheral.vascular.disease', 'cerebrovascular.disease', 'dementia', 'Chronic.obstructive.pulmonary.disease', 'connective.tissue.disease','peptic.ulcer.disease','diabetes', 'moderate.to.severe.chronic.kidney.disease', 'hemiplegia', 'leukemia', 'malignant.lymphoma', 'solid.tumor', 'liver.disease', 'AIDS' istatistiksel analiz çerçevesinde değerlendirmeye tabi tutulmuştur.

İstatistiksel Yöntemler

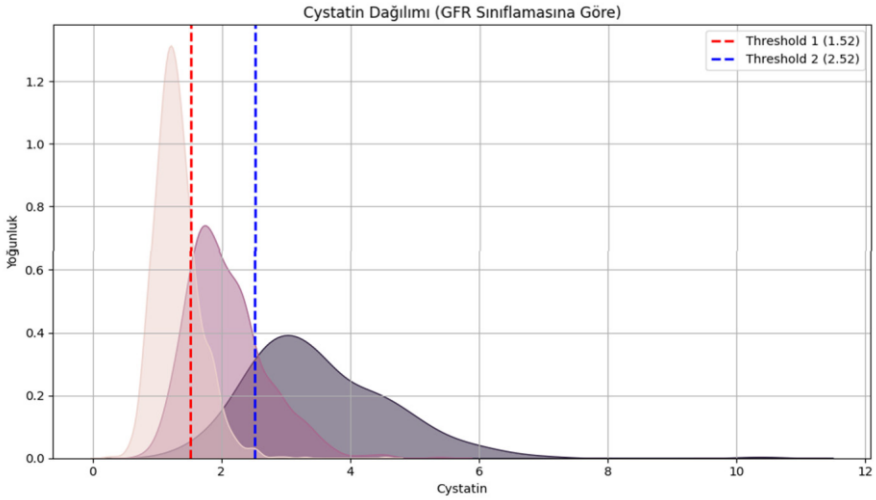
Veri setini daha iyi tanıyabilmek amacıyla öncelikle değişkenler için tanımlayıcı istatistikler verilmiş olup, KNN imputer metoduyla tespit edilen eksik veriler tamamlanmıştır. Referans kılavuza göre oluşturulan üç sınıfta, cystatin-C düzeyi ile ilişkili olabileceği düşünülen ortak değişkenlerin sınıflararası karşılaştırılması Kruskal-Wallis ve Ki-kare ile yapılmıştır. Üç sınıfa göre tanı testinin performansını değerlendirmek ve cystatin-C'nin üç sınıfı ayırmada en iyi kesim noktalarını belirlemek için "Üç yönlü ROC analizi" kullanılmıştır. Ortak değişkenlerin etkisi için "doğrusal regresyon analizi" yöntemi kullanılmıştır. Düzeltilmiş cystatin-C değerleri doğrusal regresyon analizi kullanılarak elde edilmiştir. Çalışmada açık kaynak Python programlama dili vasıtasıyla istatistikler için Statsmodels, üç yönlü Roc analizi için ise HVUS kütüphaneleri kullanılmıştır[7].

3. Bulgular

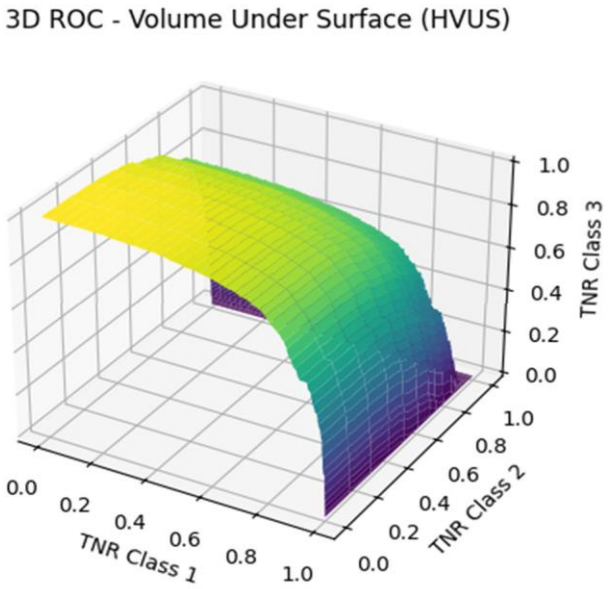
Toplam 2008 kişinin sonuçları değerlendirilmeye alınmıştır. Cystatin-C sütununda eksik değer sayısı(oranı) 41(2.04%), 'glomerular.filtration.rate' sütununda eksik değer sayısı(oranı) 63(3.14%) olup, KNN imputer metoduyla eksik veriler tamamlanmıştır. KDIGO sınıflamasının düzenlenmiş üç evreli haliye Evre 1(Sağlıklı): 1112 kişi, Evre 2(Orta Bozukluk):643 kişi, Evre 3(İleri Bozukluk):253 kişi olacak şekilde dağıldığı görülmüştür. Cystatin-C değerlerinin gruplara göre ortalama±standart sapma(min-max) değerleri sırasıyla 1,31±0.36(0,23-4,55), 2,08±0.61(0,80-5,39), 3,53±1.14(1,25-10,37) bulunmuştur. Cystatin-C'nin referans yöntem olan GFR'ye göre böbrek fonksiyonlarını değerlendirmedeki başarısı üç yönlü ROC analizi ile incelenmiştir. Yüzey altında kalan hacim(HVUS-hypervolume under surface değeri:0.7792, var:0,0001), en iyi kesim değerleri(th1:1.52,

th2:2.52), kesim değerlerine göre Evre 1 için doğruluk %77.3, Evre 2 için doğruluk %65.7 Evre 3 için doğruluk %84.9 olarak hesaplanmıştır(**Şekil 1**).

Şekil 1: Cystatin-C'nin 3 Sınıflı Evrelemeye Göre Dağılım Grafiği ve Kesim Noktaları



Şekil 2. Üç Boyutlu Hypervolume Under Surface Eksenleri ve Yüzey Gösterimi



Ortak değişken düzeltmesine dahil etme kararı için BMI sayısal değişkeni Kruskal-Wallis testiyle, diğer değişkenler ise Ki-kare testiyle karşılaştırılmıştır. Analiz sonucuna göre 'gender', 'Killip.grade', 'myocardial.infarction', 'peripheral.vascular.disease', 'diabetes' anlamlı değişkenleri($p<0,05$) ile, 3 evreli yeni sınıflandırmanın her bir sınıfı için regresyon modeli oluşturulmuştur. Bu regresyon modelleri içerisinde sadece 'gender' değişkeni anlamlı olarak bulunduğundan ortak değişken düzeltmesine dahil edilmiştir. Bu yeni modele göre ise, minimal bir artış ile yüzey altında kalan hacim (HVUS değeri:0.7817, var:0,0001), en iyi kesim noktaları (th1:-0.34, th2:0.61), kesim noktalarına göre doğruluk değerleri(Evre 1 için doğruluk %75.7, Evre 2 için doğruluk %64.6 Evre 3 için doğruluk %86.9) yeniden hesaplanmıştır.

4. Tartışma ve Sonuç

Yapılan bu çalışma ile üç sınıflı bir hastalık durumu ve sürekli sayısal sonuçlu bir tanı testi olması durumunda, tanı testinin performansının üç yönlü ROC analizi ile değerlendirilebileceği gösterilmiştir. Buna göre, hastalık durumu üç sınıflı olduğunda, iki yönlü ROC analizi yerine üç yönlü ROC analizinin kullanılması testin performansının doğru değerlendirilmesi açısından çok daha uygun olacağı düşünülmektedir.
İkiden fazla hastalık sınıfı varlığında, herhangi iki sınıfın verileri birleştirilip tek sınıf olarak değerlendirilebileceği gibi, sınıflardan birini dışarıda bırakarak ikili karşılaştırmalar da yapılabilmektedir. Bu tür değerlendirmelerde sınıflardan biri göz ardı edildiğinden tümel bakış açısı kaybedilebilmektedir. Aynı anda tüm sınıflarla testin performansını değerlendirebilen üç yönlü ROC analizi tercihinin daha doğru olduğu, sonraki aşamada istenirse ikili karşılaştırmaların iki boyutlu ROC analizi ile yapılabildiği güncel literatürde ifade edilmektedir. Ayrıca yöntem, üç sınıflı hastalık durumunu birbirinden ayırmak için kullanılan en iyi kesim noktalarının belirlenmesi konusunda da veri sınıflarının tamamı için tek analizle çözüm sunmaktadır. Her kesim noktasını belirlemede bir grubun dışarıda kalması da bilgi kaybına neden olabilir. Bu nedenle, üç yönlü ROC analizi araştırmacıya aynı anda iki kesim noktasının belirleme imkanı yaratmaktadır.
Üzerinde çalışma yapılan hastalık ve grup özelliklerine VUS değerleri ve kesim noktaları değişebilmektedir. Bahsi geçen durumlarda iyileştirme sağlayabilmek için birden fazla parametrenin kombine olarak kullanımı, ortak değişken düzeltmesi gibi uygulamalar literatürde yer almaktadır[8, 9, 10] Bu çalışmada elde edilen VUS değeri için; testin performans bakımından değerlendirmesi neticesinde, literatürde yer alan simülasyon çalışmalarına göre orta düzeyin üst sınırına yakın bir performans gösterdiği(HVUS:07792) ve ileri çalışmalarda iyileştirme yöntemleri kullanımıyla arttırılabileceği düşünülmektedir. Bu yöntemlerden biri olan ortak değişken düzeltmesinin üç yönlü ROC analizine uyarlanabileceği ve etkilerinin tespiti de denenmiş olup, çalışmamızda bu etki minimal olarak tespit edilebilmiştir. Bu anlamda çalışmanın kısıtlılıkları mevcut olmakla birlikte literatürde yer alan iyileştirme yöntemlerinin dahil edileceği yeni araştırma makalelerine başlangıç teşkil etmesi, hasta kliniğindeki gerçek verilerden elde edilen nispeten büyük bir veri setine ulaşabilmiş olması bakımından katkı sağlayabileceği düşünülmektedir.
Bu çalışmada geliştirilen model prototip özellikte olup; kullanılan parametre benzeri hastaya ait laboratuar değerleri vasıtasıyla sonucun nasıl hesaplandığına dair açıklama ile birlikte hastanenin dijital hasta sistemi içerisinde söz konusu modelin kullanımının, ilgili organ ya da sistem sağlığına ilişkin hızlı bir değerlendirmeye olanak sağlayabileceği düşünülmektedir.

5. Kaynaklar

1. Coolen-Maturi, Tahani. "Three-Group ROC Predictive Analysis for Ordinal Outcomes." Communications in Statistics - Theory and Methods, vol. 46, no. 19, 2017, pp. 9476–9493. https://doi.org/10.1080/03610926.2016.1212074
2. Nardi, Elena. ROC Curves and the Generalization to Multiple Classes. PhD dissertation, Alma Mater Studiorum - Università di Bologna, 2018. Dottorato di Ricerca in Scienze Statistiche, XXXI Ciclo.
3. Dirikoc, A., H. Baser, B. Polat, et al. "Can Biochemical Markers and Ultrasonographical Diameters Be Used to Predict Histopathological Diagnosis in Patients with Primary Hyperparathyroidism?" Indian Journal of Surgery, vol. 86, suppl. 1, 2024, pp. 136–144. https://doi.org/10.1007/s12262-022-03586-y
4. Skaltsa, Konstantina, Lluis Jover, Damian Fuster, and Josep Lluís Carrasco. "Optimum Threshold Estimation Based on Cost Function in a Multistate Diagnostic Setting." Statistics in Medicine, vol. 31, no. 11–12, May 2012, pp. 1098–1109. https://doi.org/10.1002/sim.4369
5. Zhang, Zhongheng, et al. "Hospitalized Patients with Heart Failure: Integrating Electronic Healthcare Records and External Outcome Data." Scientific Data, vol. 8, no. 1, 2021, Article no. 46. Nature Publishing Group, https://doi.org/10.1038/s41597-021-00835-9
6. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl. 2013;3:1-150.
7. Liu, Shun, et al. "An Efficient and User-Friendly Software Tool for Ordered Multi-Class Receiver Operating Characteristic Analysis Based on Python." SoftwareX, vol. 19, 2022, p. 101175. https://doi.org/10.1016/j.softx.2022.101175
8. Fong, Youyi et al. “Combining biyobelirteçs linearly and nonlinearly for classification using the area under the ROC curve.” Statistics in medicine vol. 35,21 (2016): 3792-809. doi:10.1002/sim.6956
9. Đức, Khánh Trần. “Improving Estimation of the Volume under the ROC Surface When Data Are Missing Not at Random.” ArXiv (Cornell University), Cornell University, 2019.
10. Wang, Binhuan. “Variable selection in ROC regression.” Computational and mathematical methods in medicine vol. 2013 (2013): 436493. doi:10.1155/2013/436493
11. Li, Jing, Xiaofeng Zhou, and Jason P. Fine. "A Regression Approach to ROC Surface, with Applications to Alzheimer’s Disease." Science China Mathematics, vol. 55, no. 8, 2012, pp. 1583–1595. https://doi.org/10.1007/s11425-012-4462-3

A Comparative Study of BiLSTM, BiLSTM +CRF, and SpaCy Models for De identification of Turkish Electronic Health Records

Fatih DİLMAÇ[1][0000-0001-5508-8861] and Adil ALPKOÇAK[2][0000-0001-7695-196X]

[1] Department of Computer Engineering, Dokuz Eylül University, İzmir, Turkey
[2] Department of Computer Engineering, Bakırçay University, İzmir, Turkey

Abstract. De-identification of Electronic Health Records (EHRs) is essential to protect patient privacy while enabling the secondary use of clinical data. This study compares traditional deep learning architecture and a fine-tuned natural language processing model for de-identifying Turkish Electronic Health Records (TEHRs). We evaluated four models: BiLSTM, BiLSTM+CRF, BASE SPACY, and FINE-TUNED SPACY model adapted for Turkish named entity recognition. We developed and manually annotated a dataset of 1,000 Turkish clinical documents, each containing sensitive entities. This dataset represents one of the most significant contributions to our study. Our findings show that while BiLSTM+CRF improves entity boundary detection compared to vanilla BiLSTM, the FINTE-TUNED SPACY model delivers competitive performance with minimal domain-specific fine-tuning. The results highlight the potential of FINE-TUNED NLP models for de-identification tasks in low-resource language settings such as Turkish.

Keywords: Turkish electronic health records, de-identification, deep learning, named entity recognition, natural language processing.

1 Introduction

Electronic Health Records (EHRs) contain critical information for clinical research and healthcare improvements but also pose privacy risks due to the inclusion of personally identifiable information (PII). Ensuring the de-identification of EHRs is therefore a necessary step, especially under regulatory frameworks such as Turkey's Personal Data Protection Law (KVKK)[1].

Previous studies have explored various techniques for de-identification, ranging from rule-based systems[2], [3] to advanced deep learning approaches [4], [5], [6], [7], [8]. However, limited work has focused on Turkish clinical texts. Given the morphological richness of Turkish and the scarcity of domain-specific annotated dataset, developing effective models remains challenging.

This study focuses on comparing three different approaches: a Bidirectional Long Short-Term Memory (BiLSTM) model, a BiLSTM combined with a Conditional Random Field (BiLSTM+CRF) layer, and a fine-tuned SPACY model tailored for Turkish named entity recognition. Our goal is to benchmark their effectiveness in accurately identifying sensitive entities in TEHRs.

2 Methods and Materials

The primary objective of this study is to develop a deep learning model for identifying Turkish Electronic Health Records (TEHRs). To achieve this, the study follows two main steps:
Dataset Preparation: The first step involves obtaining and processing the dataset. This includes cleaning, annotating, and preparing the data for model training.

This study utilized a comprehensive dataset comprising anamnesis records, discharge summaries (epicrisis report), surgical reports, and nursing observation notes of all patients hospitalized at Ankara University Hospital in 2010. The initial dataset comprised 641,078 entries. During preprocessing, records containing missing values, non-informative fields, or placeholder characters (e.g., "x", "öykü", etc.) were systematically filtered out. This data curation process resulted in a final corpus of 1,000 records, which we named the Turkish Electronic Health Records (TEHRs).

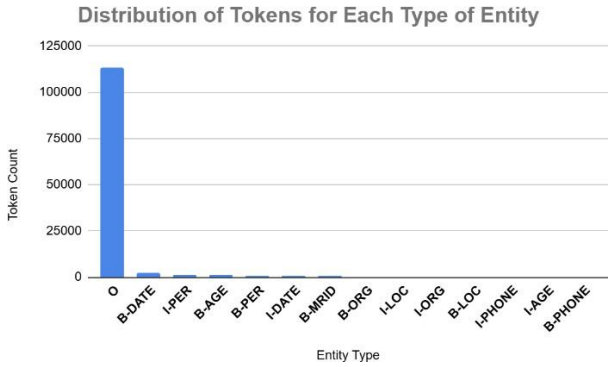


Figure 1 Token Distribution

The study of de-identifying TEHRs involves several key steps. First, we acquired and cleaned the raw dataset, then selected 1000 candidate subsets containing personal information. A subset of the data was manually annotated by three domain experts to identify sensitive entities such as personal names, dates, and similar items, using the standard Inside–Outside–Beginning (IOB) tagging scheme. While entity types such as B-DATE, I-PER, and B-AGE appeared frequently, others—including I-AGE, I-PHONE, and B-PHONE—were significantly underrepresented. The detailed distribution of entity labels is illustrated in Figure 1.

The resulting dataset exhibited a notable class imbalance. This imbalance was found to adversely affect the model's performance, particularly in recognizing rare entity types. To mitigate this issue, we initially employed a weighted loss function to adjust the contribution of each class during training. Additionally, we experimented with two commonly used resampling techniques:

– **Oversampling**: Rare classes were artificially increased by duplicating corresponding samples within the training set.
– **Undersampling**: The number of samples from overrepresented classes was reduced to achieve a more balanced class distribution.

Despite these interventions, neither oversampling nor undersampling yielded substantial gains in model accuracy. In several instances, these approaches led to overfitting or impaired generalization. Consequently, the weighted loss strategy was retained as the most stable and effective solution for addressing class imbalance in our experiments.

Model Development: The second step involves proposing and implementing a novel model for de-identification.

We preprocessed the dataset through tokenization and embedding. We then used the processed data to train and evaluate multiple models, including BiLSTM, BiLSTM+CRF, and SpaCy. To assess their effectiveness, we compared the model's performance using macro average precision, recall, and F1-score metrics. The results highlight the strengths of each model, with BiLSTM+CRF showing strong performance, paving the way for improved de-identification of Turkish EHRs.

We conducted our experiments on a manually annotated dataset comprising 1,000 TEHRs. Utilizing the BIO tagging scheme, we annotated sensitive entities such as personal names, dates, locations, and identification numbers.

2.1 Model Architectures

We evaluate BiLSTM, BiLSTM+CRF and SpaCy. BiLSTM is a sequence labeling model using bidirectional LSTM layers. An enhanced version where a CRF layer is added on top of the BiLSTM output to better model entity boundaries. SPACY is a natural language processing library and adapted for NER tasks. SPACY model was fine-tuned, while BiLSTM architectures were implemented in TensorFlow. Each model was trained on TEHRs using the BIO tagging scheme for sequence labeling.

2.2 Training Details

We trained the BiLSTM and BiLSTM+CRF models from scratch, employing a weighted cross-entropy loss function to address class imbalance. To optimize training, we utilized the Adam optimizer and implemented early stopping. For the SpaCy model,

we evaluated its performance using the native pipeline with minimal fine-tuning. We assessed all models using precision, recall, and F1-score metrics.

3 Results and Discussion

We assessed each model's performance based on macro-averaged precision, recall, and F1-score, as presented in Table 1.

Table 1 Performance Comparison

Model	Precision (%)	Recall (%)	F1-Score (%)
BiLSTM	50.00	47.00	48.00
BiLSTM+CRF	87.00	76.00	80.00
Base-SpaCy	45.00	41.00	43.00
Fine-Tuned SPACY	71.00	52.00	55.00

The BiLSTM+CRF model achieved the highest F1-score at 80.0%, demonstrating the advantage of structured prediction for de-identification tasks. The FINE-TUNED SpaCy model, despite minimal domain adaptation, achieved a competitive F1-score of 55.00%, suggesting that BiLSTM +CRF can be an efficient alternative when annotated data is limited.

The results emphasize the importance of advanced models like BiLSTM +CRF in achieving high accuracy and comprehensive de-identification of sensitive information. Meanwhile, SpaCy -based model, while efficient and scalable, fall short in recall, suggesting limitations in their ability to capture complex dependencies and contextual nuances. This underscores the need for more sophisticated architectures, such as transformers and hybrid models, to address the challenges of de-identifying sensitive information in Turkish Electronic Health Records (TEHRs).

Figure 1 illustrates the comparative performance of various models regarding precision, recall, and F1-score for identifying sensitive information in Turkish Electronic Health Records (TEHRs). Among the models, BiLSTM +CRF achieves the highest scores across all metrics, with an F1 score of 80%, demonstrating its superior ability to generalize and accurately identify sensitive entities. The BiLSTM +CRF model also performs well, particularly when enhanced with additional domain-specific data, showing significant improvements in all metrics. In contrast, BASE-SPACY and FINE-TUNED-SPACY models exhibit lower performance, especially in recall, indicating limitations in handling complex dependencies. The results highlight the effectiveness of hybrid approaches like BiLSTM +CRF in achieving robust and accurate de-identification. In contrast, simpler models like SPACY struggle to match their performance.

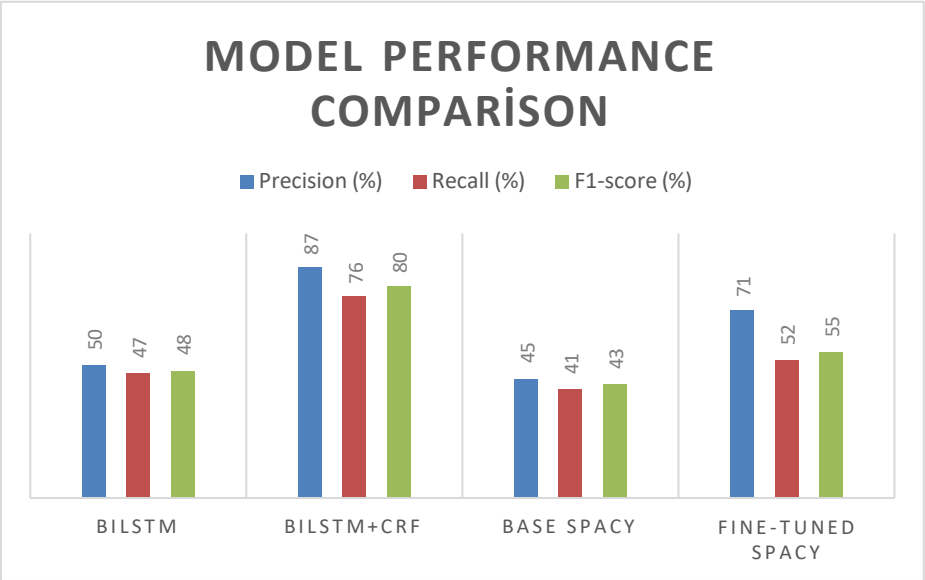


Figure 2 Performance Comparison of Different Models

The BiLSTM model, with an F1-score of 48.00%, demonstrates moderate performance but falls short compared to models enhanced with Conditional Random Fields (CRF). Adding a CRF layer in the BiLSTM +CRF model significantly improves the F1-score to 80.00%, highlighting the importance of CRF layers in better sequence modeling and capturing label dependencies in Named Entity Recognition (NER) tasks.

FINE-TUNED models offer a clear advantage in NER tasks, with notable differences between SpaCy. The BASE-SPACY model underperforms with an F1 Score of 43.00% due to limitations in feature extraction and training objectives. The FINE-TUNED-SPACY model shows improvement, achieving an F1 Score of 55.00%, indicating that fine-tuned on data corpora positively impacts performance. However, these results remain insufficient compared to BiLSTM +CRF architecture.

3.1 Comparative Analysis of Applied Model Performance

Precision Comparison of Models Across Entities

Table 2 presents the precision scores for five different models—BİLSTM, BİLSTM+CRF, BASE SPACY, FINE-TUNED SPACY —across various named entity categories. Precision measures the proportion of correctly identified entities among all entities classified as a given category, providing insight into each model's accuracy in avoiding false positives.

While the BiLSTM model struggles with several categories (e.g., 0.00 precision for B-ORG, I-ORG, B-LOC, B-PHONE, and I-PHONE), the BiLSTM+CRF improves substantially, achieving precision scores above 0.90 in most categories, such as I-ORG (1.00) and B-PHONE (1.00). This suggests that integrating a CRF layer enhances the BİLSTM 's ability to model sequential dependencies, particularly for less frequent entities.
The BASE SPACY model exhibits inconsistent performance, with high precision in some categories, like B-DATE (0.91) and B-MRID (0.93), but fails for others (e.g., 0.00 for I-ORG, B-LOC, B-PHONE, and I-PHONE). FINE-TUNED SPACY improves over the base model in certain tags (e.g., I-ORG: 0.89 and I-LOC: 1.00), leveraging FINE-TUNED embeddings to capture contextual information better.

Table 2 Precision Comparison of Models Across Entities

TAG	BiLSTM	BiLSTM+CRF	BASE SPACY	FINE-TUNED SPACY
B-PER	85.00	0.99	0.92	0.93
I-PER	0.88	0.94	0.80	0.79
B-AGE	0.82	0.84	0.90	0.85
B-DATE	0.63	0.90	0.91	0.84
I-DATE	0.94	0.98	0.80	0.75
B-MRID	0.82	0.95	0.93	0.85
B-ORG	0.00	0.91	0.76	0.83
I-ORG	0.00	1.00	0.00	0.89
B-LOC	0.00	0.67	0.33	0.67
I-LOC	0.60	1.00	0.00	1.00
B-PHONE	0.00	1.00	0.00	0.50
I-PHONE	0.00	1.00	0.00	1.00

The Table 2 highlights BiLSTM +CRF clear superiority in precision, making it the most reliable model for minimizing false positives across all named entity categories. The SPACY models exhibit mixed results, with the FINE-TUNED version providing some enhancements but still lagging behind BiLSTM+CRF.

Recall Comparison of Models Across Entities
Table 3 presents the recall scores for five models—BiLSTM, BiLSTM+CRF, BASE SPACY, FINE-TUNED SPACY —across various named entity categories. Recall measures the proportion of correctly identified entities out of all actual entities, providing insight into each model's ability to capture true positives effectively.

BiLSTM alone struggles significantly in several categories, particularly for tags like B-ORG, I-ORG, B-LOC, B-PHONE, and I-PHONE, where recall drops to 0.00. Adding the CRF layer (BiLSTM+CRF) improves recall across most tags, achieving notable

scores like B-PER (0.95) and B-AGE (0.99). However, it still underperforms for rare entities such as I-ORG (0.32) and I-LOC (0.71).

The BASE SPACY model shows moderate recall performance. It achieves decent scores for B-DATE (0.90) and B-MRID (0.98) but struggles significantly in other categories, such as B-ORG (0.68) and I-ORG (0.00).

FINE-TUNED SPACY improves slightly in some tags, such as B-DATE (0.95), but has lower recall for others, like I-DATE (0.77) and I-LOC (0.07). This indicates limited generalization for certain types of entities. However, categories like I-ORG, B-LOC, I-LOC, and B-PHONE remain challenging. For instance, the recall for I-ORG is 0.00 for BiLSTM and BASE SPACY, 0.26 for FINE-TUNED SPACY, and 0.32 for BİLSTM+CRF. Similarly, I-LOC recall is low across models, with BİLSTM+CRF achieving the highest recall (0.71).

Table 3 Recall Comparison of Models Across Entities

TAG	BiLSTM	BiLSTM+CRF	BASE SPACY	FINE-TUNED SPACY
B-PER	0.83	0.95	0.78	0.77
I-PER	0.83	0.97	0.70	0.78
B-AGE	0.87	0.99	0.78	0.81
B-DATE	0.62	0.85	0.90	0.95
I-DATE	0.86	0.87	0.79	0.77
B-MRID	0.80	0.90	0.98	0.98
B-ORG	0.00	0.82	0.68	0.66
I-ORG	0.00	0.32	0.00	0.26
B-LOC	0.00	0.50	0.12	0.50
I-LOC	0.21	0.71	0.00	0.07
B-PHONE	0.00	1.00	0.00	0.50
I-PHONE	0.00	0.83	0.00	0.17

The Table 3 highlights BİLSTM+CRF offers significant improvements over the basic BiLSTM, showing good recall in many categories but struggling with rare entities. SPACY models show moderate performance, with the FINE-TUNED variant offering some improvements but failing to match the recall levels of BERT. This table emphasizes the importance of advanced contextual embedding in ensuring comprehensive entity recognition.

F1-Score Comparison of Models Across Entities

Table 4 summarizes the F1-scores of four models—BiLSTM, BiLSTM+CRF, BASE SPACY and FINE-TUNED SPACY—for various named entity tags. The F1-score combines precision and recall, offering a balanced measure of a model's performance in identifying entities wholly and accurately.

Adding the CRF layer (BiLSTM+CRF) significantly improves the F1-scores over the standalone BiLSTM model. For instance: B-PER increases from 0.84 to 0.97. B-DATE improves from 0.63 to 0.88. However, both models perform poorly for rare or complex tags like B-LOC and I-ORG, where BiLSTM+CRF achieves only 0.57 and 0.49, respectively.

The BASE SPACY model struggles to balance precision and recall effectively, leading to lower F1-scores for tags like I-ORG (0.00) and I-LOC (0.00). However, it performs relatively well for more common entities such as B-DATE (0.90) and B-MRID (0.95). FINE-TUNED SPACY slightly improves challenging tags such as I-ORG (0.40) but fails to achieve competitive results compared to BiLSTM+CRF.

Table 4 F1-Score Comparison of Models Across Entities

TAG	BiLSTM	BiLSTM+CRF	BASE SPACY	FINE-TUNED SPACY
B-PER	0.84	0.97	0.84	0.84
I-PER	0.85	0.95	0.75	0.79
B-AGE	0.84	0.91	0.84	0.83
B-DATE	0.63	0.88	0.90	0.89
I-DATE	0.90	0.92	0.80	0.76
B-MRID	0.81	0.92	0.95	0.91
B-ORG	0.00	0.86	0.72	0.74
I-ORG	0.00	0.49	0.00	0.40
B-LOC	0.00	0.57	0.18	0.57
I-LOC	0.32	0.83	0.00	0.13
B-PHONE	0.00	1.00	0.00	0.50
I-PHONE	0.00	0.91	0.00	0.29

The F1-score analysis emphasizes the superiority BiLSTM+CRF demonstrates substantial improvements over the standalone BiLSTM model, achieving competitive scores for most entities except challenging ones like I-ORG and I-LOC. SPACY models offer reasonable performance for common tags but fall short for rare or complex entities. Overall, BiLSTM+CRF emerges as the most robust model, effectively balancing precision and recall for accurate and complete entity recognition.

4 CONCLUSION

In this study, we developed and manually annotated a novel dataset comprising 1,000 Turkish Electronic Health Record (EHR) documents, each containing sensitive entities such as personal names, dates, locations, and identification numbers. This dataset serves as a valuable resource for evaluating de-identification techniques tailored to the Turkish language, addressing its unique linguistic and structural challenges.

The principal aim of this study was to evaluate and compare traditional deep neural network-based approaches—specifically BiLSTM, BiLSTM+CRF, and the SpaCy NER model—for the task of de-identification in Turkish Electronic Health Records (EHRs). Utilizing these TEHRs, we trained and evaluated several models, including BiLSTM, BiLSTM+CRF, and SpaCy, to assess their effectiveness in de-identifying sensitive information. Our comparative analysis revealed that while the BiLSTM+CRF architecture achieved the highest accuracy, the fine-tuned SpaCy model offered comparable performance with significantly lower training costs. These findings suggest that fine-tuned models like SpaCy can be effective alternatives in low-resource clinical NLP settings, particularly when access to large, annotated datasets is limited.

Given the increasing prominence of Large Language Models (LLMs), which have demonstrated substantial improvements across a wide range of clinical natural language processing (NLP) tasks, we are extending this analysis as part of our ongoing work. To this end, we are incorporating transformer-based architectures such as BERTurk, ConvBERTurk, and ClinicalBERT. These models are currently being fine-tuned and evaluated on the same dataset to facilitate a comprehensive performance comparison between conventional deep learning methods and LLM-based approaches. The findings from these experiments are expected to inform future directions and contribute to the development of more robust and generalizable de-identification systems for Turkish clinical text.

References

[1] Kişisel verileri koruma kurumu, "KİŞİSEL VERİLERİN KORUNMASI KANUNU," 2016. Accessed: May 01, 2025. [Online]. Available: https://www.mevzuat.gov.tr/MevzuatMetin/1.5.6698.pdf

[2] M. Suzdaltseva, A. Shamakhova, and N. Dobrenko, "De-identification of Medical Information for Forming Multimodal Datasets to Train Neural Networks," 2021, doi: 10.5220/0010406001630170.

[3] J. Aberdeen *et al.*, "The MITRE Identification Scrubber Toolkit: Design, training, and assessment," *Int J Med Inform*, vol. 79, no. 12, pp. 849–859, Dec. 2010, doi: 10.1016/j.ijmedinf.2010.09.007.

[4] A. Stubbs, C. Kotfila, and O. Uzuner, "Automated systems for the de-identification of longitudinal clinical," *J Biomed Inform*, 2015, doi: 10.1016/j.jbi.2015.06.007.

[5] V. Yogarajan, M. Mayo, and B. Pfahringer, "A survey of automatic de-identification of longitudinal clinical narratives," Oct. 2018, Accessed: Aug. 19, 2021. [Online]. Available: http://arxiv.org/abs/1810.06765

[6] J. L. Leevy, T. M. Khoshgoftaar, and F. Villanustre, "Survey on RNN and CRF models for de-identification of medical free text," *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/S40537-020-00351-4.

[7] F. Dernoncourt, J. Y. Lee, O. Uzuner, and P. Szolovits, "De-identification of patient notes with recurrent neural networks," *Journal of the American Medical Informatics Association*, vol. 24, no. 3, pp. 596–606, May 2017, doi: 10.1093/JAMIA/OCW156.

[8] G. Akın¸seker, A. Akın¸seker, G. Gül¸sen, and E. Git, "Initial explorations on using CRFs for Turkish Named Entity Recognition," pp. 2459–2474, 2012.

Fall Risk Prediction from Clinical Patient Notes with Machine Learning

Mehmet YİĞİT [1][0000-0001-6115-0537], Hilal YİĞİT [2][0000-0002-9322-8033], Kemal TURHAN [3][0000-0001-7871-3025]

[1] Ordu üniversitesi Eğitim ve Araştırma Hastanesi, Ordu, Türkiye
[2] Ordu üniversitesi Eğitim ve Araştırma Hastanesi, Ordu, Türkiye
[3] İzmir Ekonomi Üniversitesi, İzmir, Türkiye
mehmettygt@gmail.com, hilalkirli.eu@gmail.com
kemal.turhan@ieu.edu.tr

Abstract Falls in hospitals, often underreported in structured records, pose significant challenges for patient safety. This study introduces a novel perspective for personalized fall risk prediction by leveraging unstructured Portuguese clinical notes from a tertiary public hospital in Porto Alegre, Brazil, using machine learning. The dataset comprises 1,078 annotated progress notes, with 32% containing fall-related content. Text processing incorporated Brazil-specific medical abbreviations and health expressions, followed by feature extraction based on the ITAKI Fall Risk Scale. Class imbalance was addressed using SMOTE, and a VotingClassifier integrating XGBoost, CatBoost, LightGBM, and Random Forest achieved 87% accuracy, 80% recall for fall cases, and a ROC-AUC of 0.937 on test data. These findings demonstrate that accurate and actionable fall risk prediction from unstructured clinical narratives can enhance patient safety.

Keywords: Machine Learning, Medical Informatics, Natural Language Processing

1 Giriş

Düşmeler, küresel olarak kasıtsız ölümlerin ikinci önde gelen nedenidir ve her yıl yaklaşık 684.000 ölümcül vaka meydana gelir [1]. Dünya genelinde 37,3 milyon tıbbi müdahale gerektiren düşme olayı, 38 milyonun üzerinde engelliliğe uyarlanmış yaşam yılı (DALY) kaybına yol açmaktadır [2]. Türkiye'de yaşlı nüfusun artmasıyla (2024'te 65 yaş üstü nüfus %10'u aşmıştır), düşme kaynaklı yaralanmalar sağlık sistemi üzerinde önemli bir yük oluşturmaktadır.

Düşme eylemi, hızlı ve kontrol edilemeyen bir şekilde aniden yüksek seviyeden alçak seviyeye doğru hareket etmektir [3]. Düşme sebebiyle oluşan yaralanmalar kalıcı sakatlığa sebep olabilir veya yüksek oranda ölümcüldür. Özellikle 60 yaş üzeri yaşlılarda, kişi kendine istemeden zarar vererek ölümcül düşmeye maruz kalabilmektedir [4]. Her seviyeden kazara düşmeler, hastanelerde mortalite ve morbidite sebebidir. 65 yaş ve üzeri bireylerde kas-iskelet sistemi zayıflığı, denge bozuklukları ve görme

azalması gibi fizyolojik değişikliklere ek olarak, reçete edilen bazı ilaçlar bu riski daha da artırmaktadır [5]. Özellikle nöroleptikler, benzodiazepinler, antidepresanlar ve antikonvülzanlar gibi nörolojik etkili ilaçlar; bilinç değişikliği, sedasyon ve postüral dengesizlik gibi etkiler nedeniyle düşme riskinin farmakolojik boyutunu oluşturmaktadır [6]. Düşme eyleminin gerçekleştiği yükseklik seviyesi, mortalite ve morbidite oranını etkilediği düşünülmektedir [7]. Bu nedenle, düşmelerin öngörülmesi ve önlenmesinde yalnızca fiziksel durum değil, aynı zamanda farmakoterapi profili ve yaşa bağlı fizyolojik özelliklerin de dikkate alınması büyük önem taşımaktadır. Düşmeler önlenebilir yaralanmalardır. Yatan hastalardaki düşmelerin yaklaşık %30'u yaralanmayla sonuçlanırken, %4 ila %6'sı ciddi yaralanmayla sonuçlanır [8-9]. Hastalarda düşme riskinin tespiti önemli bir klinik sorundur. Son yıllarda, yapay zekâ (YZ) ve makine öğrenimi (MÖ) teknikleri, bu riski değerlendirmek için etkili araçlar olarak öne çıkmaktadır. Elektronik Sağlık Kayıtlarından alınan anamnez verileriyle yapılan düşme riski hesaplamaları, klinik karar destek sistemlerinin geliştirilmesinde önemli bir rol oynamaktadır [9].

Geleneksel risk değerlendirme yöntemleri (ör. Morse, Hendrich ölçekleri) manuel, subjektif ve zaman alıcı olduğu düşünülmektedir. Elektronik sağlık kayıtlarındaki (EHR) yapılandırılmamış verilerden faydalanan makine öğrenimi (MÖ) teknikleri, bu sınırlılıkları aşmada umut vadetmektedir. Bu çalışma, Portekizce klinik notlardan kişiselleştirilmiş düşme riski tahmini yapmayı ve klinik karar destek sistemleri geliştirmeyi amaçlamaktadır.

2 Gereç ve Yöntem

2.1 Veri Seti

Çalışmada, Brezilya'nın Porto Alegre şehrindeki bir kamu hastanesinden toplanan veriler kullanılmıştır. Bu çalışma geriye dönük kohort çalışmasıdır. 2012-2017 yılları arasında düşme yaşayan 1.694 hastadan seçilen 367 hastaya ait 441 olay raporu ve 1.078 Portekizce serbest metinli ilerleme notunu içermektedir. Her olay raporu için ortalama 2,4 ilerleme notu analiz edilmiştir. Hasta demografisine göre düşme olaylarının dağılımı şu şekildedir: %87'si (316 hasta) tek düşme, %10,1'i (36 hasta) iki düşme, %3'ü (11 hasta) üç düşme, %0,3'ü (1 hasta) dört düşme, %0,5'i (2 hasta) beş düşme ve %0,3'ü (1 hasta) altı düşme yaşamıştır. Hasta anonimliğini sağlamak için tüm veriler kimliksizleştirilmiştir. Notların %68'i (723 not) düşme olayı içermezken, %32'si (355 not) düşme ile ilişkilidir.

2.2 Metin İşleme ve Özellik Çıkarımı

Portekizce klinik notlar, Brezilya'ya özgü tıbbi kısaltmalar ve sağlık ifadeleri dikkate alınarak işlenmiştir. Yerel tıbbi kısaltmalar (ör. LOC: Bilinç Seviyesi/Kaybı, SVA: Spontan İdrar Yapma Yeteneği, SVD: Kalıcı İdrar Sondası), Brezilya farmakolojik veri tabanları ve sağlık kaynaklarına dayanılarak genişletilmiş ve risk skorları hesaplanmıştır. Örneğin, LOC (bilinç kaybı) için 2 puan, SVA ve SVD için 1'er puan atanmıştır.

İlaç risk sınıflandırmaları, metinlerin tamamında geçen 100'den fazla ilacın (ör. diazepam, morfin: yüksek risk; amidaron, ondnasetron, omeprazol: düşük risk) yer aldığı bir veri tabanına dayandırılmıştır. Verilerin Türkçeleştirilmesi için Google Çeviri uygulaması kullanılmıştır. Tıbbi terminoloji ve kısaltmaların doğruluğunu sağlamak için Brezilya sağlık kaynaklarına dayandırılmıştır [10].

	Sütun1	Evrim	Hedef
0	1030	C23 Safra kesesinin kötü huylu neoplazmı Evrel...	0
1	591	R53 Halsizlik Yorgunluk Evrim NoInfo: s NoInfo...	0
2	269	R104 Karın ve pelvik ağrı Diğer ve belirtilmem...	0
3	186	J459 Astım Belirtilmemiş astım Akut şiddetli a...	1
4	772	Z000 Şikayeti veya tanısı olmayan kişilerin ge...	0
...	...	...	...
1073	260	I10 Primer esansiyel hipertansiyon Test Talebi...	0
1074	388	R51 Baş Ağrısı Evrimi NoInfo: KONUŞMA TERAPİS...	0
1075	715	R060 Solunum anormallikleri Dispne Evrim NoIn...	0
1076	654	R060 Solunum anormallikleri Dispne Evrim NoIn...	0
1077	744	J459 Astım Belirtilmemiş astım Evrim NoInfo: ...	0

İTAKİ Düşme Riski Ölçeği, hastaların düşme riskini değerlendirmek için minör (her biri 1 puan) ve majör (her biri 5 puan) risk faktörlerini puanlayan bir klinik araçtır. Ölçek, şu durumlarda uygulanır: hastanın kliniğe ilk yatışı, post-operatif dönem, düşme olayı, bölüm değişikliği veya genel durumda değişiklik. Minör risk faktörleri arasında 65 yaş üstü olmak, bilinç kapalılığı, son 1 ayda düşme öyküsü, kronik hastalıklar (ör. hipertansiyon, diyabet), fiziksel desteğe ihtiyaç, üriner/fekal inkontinans, zayıf görme, 4'ten fazla ilaç kullanımı, 3'ten az bakım ekipmanı, yatak korkuluklarının olmaması ve yürüme alanlarında engeller yer alır. Majör risk faktörleri arasında bilinç açık ancak koopere olmama, denge problemi, baş dönmesi, ortostatik hipotansiyon, görme engeli, bedensel engel, 3 veya daha fazla bakım ekipmanı ve son 1 haftada riskli ilaç kullanımı (ör. psikotroplar, benzodiazepinler) bulunur. Toplam puan 5'in altındaysa düşük risk, 5 veya üzerindeyse yüksek risk olarak sınıflandırılır. Yüksek riskli hastalar, oda kapısına veya yatak başına dört yapraklı yeşil yonca figürü ile işaretlenir [11].

Özellik çıkarımı, İTAKİ ölçeğine dayalı olarak Python tabanlı bir algoritmayla gerçekleştirilmiştir. Düzenli ifadeler ve kural tabanlı NLP teknikleri kullanılarak şu özellikler metinden çıkarılmıştır:

- Yaş: 65 yaş ve üstü hastalar için risk skoru, metinde yaş bilgisiyle eşleşen düzenli ifadelerle tespit edilmiştir.
- Düşme Öyküsü: Son bir ay içinde düşme öyküsü, metinde ilgili ifadelerin varlığına göre belirlenmiştir.
- Kronik Hastalık: Kronik hastalık öyküsü (ör. hipertansiyon, diyabet), metindeki ilgili terimlerle eşleştirilmiştir.
  - Tıbbi Kısaltmalar: Brezilya'ya özgü kısaltmalar (ör. LOC, SVA, SVD) için risk skorları hesaplanmış ve toplam risk skoru elde edilmiştir.

- İlaç Kullanımı: Yüksek riskli ilaçlar (ör. diazepam, morfin) ve ilaç sayısı, kapsamlı bir ilaç veri tabanına dayanılarak değerlendirilmiştir.
- Ek Özellikler: Düşme ile ilgili pozitif ve negatif ifadeler (ör. "düşme" kelimesi ile "yaşamamış" gibi olumsuzlayıcı kelimeler), fonksiyonel bağımsızlık (ör. "kendi başına") ve kognitif bozukluk (ör. "demans", "konfüzyon") gibi klinik göstergeler metinden çıkarılmıştır.

Bu özellikler, TF-IDF vektörleştirme ve sayısal özelliklerin ColumnTransformer ile birleştirilmesiyle modellenmiştir. Sınıf dengesizliği, SMOTE ile giderilmiştir.

2.3 Etiketleme Süreci

Düşme olaylarının etiketlenmesi, dört ay süren bir süreçte üç hemşirelik öğrencisi tarafından WebAnno sistemi kullanılarak gerçekleştirilmiştir. Her not, dikkatli bir okuma ve çifte kontrol yöntemiyle incelenmiş; tutarsızlıklar araştırma grubu toplantılarında tartışılmıştır. Etiketleme, Dünya Sağlık Örgütü'nün Hasta Güvenliği Teknik Raporları'ndaki düşme tanımlarına dayanır. Etiketlenen kavramlar arasında düşme sonrası prosedür, tıbbi değerlendirme, hasar düzeyi (yok, düşük, orta, yüksek, ölüm) ve hasar tipi (fiziksel, psikolojik, sosyal) yer alır.

2.4 Model Geliştirme

Veri analizleri Python'da Jupyter Notebook ortamında gerçekleştirilmiştir. TF-IDF ve sayısal özellikler ColumnTransformer ile birleştirilmiş, sınıf dengesi SMOTE ile sağlanmıştır. XGBoost, CatBoost, LightGBM ve Random Forest modellerini içeren bir VotingClassifier kullanılmıştır. Model performansı, recall ve ROC-AUC metrikleriyle değerlendirilmiştir.

3 Bulgular

Modelin test verisinde elde ettiği başarı ölçütleri aşağıda gösterilmiştir:

Table 1. Eğitim sonucu oluşan başarı metrikleri

Ölçüt	Değer
Doğruluk (Accuracy)	%87
Precision (Düşme riski sınıfı)	%80
Recall	%80
F1-Score	%80
ROC-AUC Skoru	0.937

Model, özellikle yüksek ROC-AUC değeri ile sınıflar arasında güçlü ayrım yapabilme becerisine sahiptir.

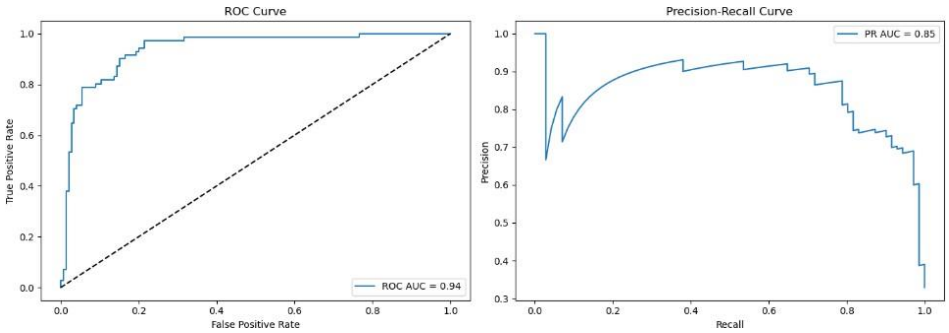


Fig. 1. ROC (Receiver Operating Characteristic) eğrisi

4 TARTIŞMA

Bu çalışma, 2012-2017 yıllarında toplanan 1.078 Portekizce klinik notla eğitilen modelin, klinik karar destek sistemlerine entegre edilebilecek düzeyde performans sunduğunu göstermektedir (ROC-AUC: 0.937). WebAnno kullanımı, çifte kontrol yöntemi ve İTAKİ ölçeğine dayalı kural tabanlı özellik çıkarımı, etiketleme ve modelleme kalitesini artırmıştır. Diğer modellerle karşılaştırıldığında, yapılandırılmamış metin verisiyle bu derece güçlü bir ayrım yapabilme kabiliyeti dikkat çekicidir.

Ayrıca, Sakane ve arkadaşları (2025) tarafından yürütülen bir çalışmada, erkek bireylerde düşme riski tahmini için video tabanlı MediaPipe + SVM modeliyle 0.909 AUC elde edilmiştir. Kadınlarda bu başarı daha düşüktür. Bu çeşitlilik, hem veri tipinin hem de model türünün performansa etkisini ortaya koymaktadır [12].

Sritharan ve arkadaşları (2025), COVID-19 hastalarında entübasyon ve mortalite gibi kritik sonuçları öngörmek üzere LASSO modelini kullanmış ve mortalite tahmini için 0.85 AUC skoru bildirmiştir [13]. Diğer yandan, Nedelec ve arkadaşları (2025), Birleşik Krallık ve Fransa verilerini kullanarak geliştirdikleri model ile %5 false positive oranı (FPR) altında %38.4 tespit oranı sağlamışlardır [14].

Son yıllarda makine öğrenmesi temelli düşme tahmin modelleri, farklı veri türleri ve hasta grupları üzerinde geniş çaplı olarak uygulanmıştır. Kore'de yapılan çalışmalardan biri [15], yaşlı bireylerin anket verileriyle geliştirilen Random Forest modelinde 0.98 gibi çok yüksek bir ROC-AUC skoruna ulaşmıştır. Aynı veri setinin farklı bir analizinde , Logistic Regression modeli daha düşük bir performans göstermiş (AUC: 0.76, F1: 0.26) ve modelin duyarlılığı (%17) oldukça sınırlı kalmıştır [16].

5 SONUÇ VE ÖNERİLER

Bu çalışma, serbest metinli klinik notlardan makine öğrenimiyle düşme riski tahmini yapılabileceğini gösteriyor. İTAKİ Düşme Riski Ölçeği'ne dayalı kural tabanlı özellik çıkarımı ve VotingClassifier modeli, yüksek doğruluk ve klinik karar destek

sistemlerine entegrasyon potansiyeli sunarak hasta güvenliğini artırmada yenilikçi bir bakış açısı sunuyor. Model, Brezilya'ya özgü tıbbi kısaltmalar (ör. LOC, SVA) ve ilaç risk sınıflandırmaları (ör. diazepam, morfin) ile zenginleştirilmiş, gerçek zamanlı tahmin yetkinlikleriyle klinik uygulamalara hazır hale getirilmiştir.

Gelecekte, ICD-10 kodları, RxNorm gibi ilaç veri tabanları ve SNOMED-CT gibi tıbbi ontolojilerle desteklenmesi halinde daha başarı olacağı varsayılmaktadır. Türkçe veya diğer dillerde veri setleriyle genellenebilirlik test edilmelidir. Brezilya'ya özgü tıbbi kısaltmaların uluslararası standartlara uyarlanması standardize edilmelidir. Ayrıca, modelin API tabanlı entegrasyonu ve hemşirelik ekiplerine düşme riski figürüyle eşleşen gerçek zamanlı uyarılar gönderen bir sistem geliştirilmesi, klinik uygulamalarda devrim yaratabileceği düşünülmektedir.

Kaynaklar

1. Türkiye İstatistik Kurumu (TÜİK): Ölüm ve Ölüm Nedeni İstatistikleri 2023. https://data.tuik.gov.tr/Bulten/Index?p=Olum-ve-Olum-Nedeni-Istatistikleri-2023-53709, son erişim tarihi 2025/06/09.
2. World Health Organization (WHO): Falls – Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/falls, son erişim tarihi 2025/06/09.
3. Oxford English Dictionary: Falling. http://oed.com/dictionary/falling_n1?tab=factsheet, son erişim tarihi 2025/06/09.
4. World Health Organization (WHO): Methods and data sources for global burden of disease estimates 2000–2019. https://cdn.who.int/media/docs/default-source/gho-documents/global-health-estimates/ghe2019_daly-methods.pdf, son erişim tarihi 2025/06/09.
5. Hsieh, T.M., et al.: Effect of Height of Fall on Mortality in Patients with Fall Accidents: A Retrospective Cross-Sectional Study. International Journal of Environmental Research and Public Health 17(11), 4163 (2020).
6. Rasu, R.S., Xavier, C., Rianon, N.: Dynamic changes in medication burden leading to fall and hospital readmissions in older adults: Toward a strategy for improving risk and managing costs. Journal of Managed Care & Specialty Pharmacy 31(1), 45–56 (2025).
7. Morse, J.M., et al.: Retrospective analysis of patient falls. Canadian Journal of Public Health 76, 116–118 (1985).
8. Ash, K.L., et al.: A case-control study of falls in the hospital setting. Journal of Gerontological Nursing 24(12), 7–15 (1998).
9. Hitcho, E.B., et al.: Characteristics and Circumstances of Falls in a Hospital Setting: A Prospective Analysis. Journal of General Internal Medicine 19(7), 732–739 (2004).
10. Dos Santos, H.D.P., et al.: Fall Detection in EHR using Word Embeddings and Deep Learning. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), s. 265–268. IEEE, Piscataway (2019). https://doi.org/10.1109/BIBE.2019.00054.
11. Barış, V.K., et al.: İTAKİ Düşme Riski Ölçeği'nin psikometrik özelliklerinin değerlendirilmesi. Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi 13(4), 214–221 (2020).
12. Sakane, M., et al.: Application of machine learning for detecting high fall risk individuals using marker-free gait analysis. Journal of Occupational Health 67(1), uiae075 (2025).
13. Nedelec, T., et al.: Machine learning prediction algorithms for 2-, 5- and 10-year risk of Alzheimer's, Parkinson's and dementia at age 65: A study using medical records from France and the UK General Practitioners. medRxiv. https://doi.org/10.1101/2025.01.01.23298710, son erişim tarihi 2025/06/09.
14. Cho, I., et al.: Machine learning approaches for fall prediction in Korean community-dwelling older adults. In: Innovation in Aging, cilt 8, ek 1, s. 1047 (2024).
15. Suh, S., et al.: Development of a fall prediction model for community-dwelling older adults in South Korea using machine learning: A secondary data analysis. Journal of Korean Biological Nursing Science 26(4), 288–299 (2024).
16. Carriero, C., et al.: The harms of class imbalance corrections for machine learning based prediction models: A simulation study. Statistics in Medicine 44(3–4), e10320 (2025).

Meme Kanseri Alt Tiplerinde Gen Ekspresyonu ve İlaç Duyarlılığı İlişkisinin Multiomik Analizi

Multiomics Analysis of Gene Expression and Drug Sensitivity Relationship in Breast Cancer Subtypes

Adem Sonuvar[1], Esra TOKUR SONUVAR[1], Uğur BİLGE[1]

[1] Biyoistatistik ve Tıp Bilişimi AD, Tıp Fakültesi, Akdeniz Üniversitesi Antalya.

Özet. Meme kanseri, moleküler heterojenitesi nedeniyle farklı klinik davranışlar sergileyen karmaşık bir hastalıktır. Bu çalışmada, 30 meme kanseri hücre hattı, 'Hormon/HER2 Pozitif' (n=12) ve 'TNBC' (üçlü negatif, meme kanseri) olmak üzere iki ana analiz grubuna ayrılarak incelenmiştir. Bu gruplar arasındaki gen ekspresyon profilleri (RNA-seq), proteomik veriler (RPPA) ve yedi farklı kinaz inhibitörüne karşı ilaç duyarlılıkları, E-MTAB-4801 veri seti kullanılarak analiz edilmiştir. t-SNE ve hiyerarşik kümeleme analizleri, bu iki grubun genel ekspresyon profillerine göre ayrıştığını göstermiştir. Diferansiyel gen ekspresyonu ve fonksiyonel zenginleştirme analizleri, Hormon/HER2 Pozitif grubunda östrojen yanıtı ve metabolik yolakların, TNBC grubunda ise immün yanıt ve epiteliyal-mezenkimal geçiş ile ilişkili yolakların aktive olduğunu belirlemiştir. Makine öğrenimi yöntemleri ile iki grubu ayırt etmede önemli olan potansiyel biyobelirteçler (FAM176A, CACNG1, GPR77) tanımlanmıştır. Karar ağacı modeli, özellikle FAM176A ve GREB1 genlerinin belirli ekspresyon eşik değerlerinin alt tip sınıflandırmasında yüksek doğruluk sağladığını göstermiştir. Ayrıca, RPPA verileri kullanılarak, 4E-BP1 protein ekspresyon seviyesinin mTOR inhibitörlerine duyarlılıkta önemli bir belirleyici olduğu doğrulanmıştır. Bu çalışma, multiomiks verilerin entegratif analizinin, meme kanserinde kişiselleştirilmiş tedavi stratejileri geliştirmede değerli bilgiler sağladığını göstermektedir.

Anahtar Kelimeler: Meme kanseri, Multiomiks, Gen ekspresyon profili, İlaç duyarlılığı, Biyobelirteçler, TNBC, Hormon Reseptörü

Abstract. Breast cancer is a complex disease exhibiting distinct clinical behaviors due to its molecular heterogeneity. In this study, 30 breast cancer cell lines were categorized into two main analysis groups: 'Hormone/HER2 Positive' (n=12) and 'TNBC' (triple-negative breast cancer). Gene expression profiles (RNA-seq), proteomic data (RPPA), and drug sensitivity to seven different

kinase inhibitors were analyzed using the E-MTAB-4801 dataset. t-SNE and hierarchical clustering demonstrated a separation based on the overall expression profiles of these two groups. Differential gene expression and functional enrichment analyses identified significant activation of estrogen response and metabolic pathways in the Hormone/HER2 Positive group, whereas immune response and epithelial-mesenchymal transition pathways were active in the TNBC group. Machine learning methods identified potential biomarker candidates (FAM176A, CACNG1, GPR77) crucial for discriminating between the two groups. The decision tree model demonstrated that specific expression threshold values of FAM176A and GREB1 genes provide high accuracy in subtype classification. Furthermore, analyses using RPPA data confirmed that the 4E-BP1 protein expression level is an important determinant of sensitivity to mTOR inhibitors. This study demonstrates that the integrative analysis of multi-omics data provides valuable insights for developing personalized treatment strategies in breast cancer.

Keywords: Breast cancer, Multi-omics, Gene expression profile, Drug sensitivity, Biomarkers, TNBC, Hormone Receptor.

Giriş

Meme kanseri, klinik seyir ve tedavi yanıtı açısından önemli farklılıklar gösteren moleküler alt tiplere sahip heterojen bir hastalıktır [12]. "Multiomiks" yaklaşımlar, genomik, transkriptomik ve proteomik verilerin entegrasyonu ile meme kanseri alt tiplerinin moleküler karakterizasyonunda önemli ilerlemeler sağlamıştır. Başlıca alt tipler olan hormon reseptörü pozitif (ER+), insan epidermal büyüme faktörü reseptörü 2 pozitif (HER2+) ve üçlü negatif (TNBC) sınıflandırması, tedavi stratejilerinin belirlenmesinde kritik rol oynar [16]. TNBC alt tipi, östrojen reseptörü (ER), progesteron reseptörü (PR) ve HER2 reseptörlerini eksprese etmemesiyle karakterizedir. Moleküler düzeydeki çalışmalar, bu alt tiplerin kendilerine özgü genetik imzalar taşıdığını, farklı sinyal yolaklarını aktive ettiğini ve farklı onkojenik sürücülere sahip olduğunu göstermiştir [2]. ER+ tümörler östrojen sinyaline bağımlıyken, HER2+ tümörler ERBB2 reseptörünün aşırı aktivasyonuna dayanır. TNBC alt tipi ise daha agresif bir fenotipe sahiptir ve hedefe yönelik tedavi seçenekleri sınırlıdır [10]. Bu çalışmadaki hücre hattı sınıflandırması, Jastrzebski ve ark. [8] tarafından tanımlanan alt tiplere dayanmaktadır.
Analizlerimiz, 'Hormon/HER2 Pozitif' ve 'TNBC' olmak üzere iki ana grup üzerinden yürütülmüştür. Alt tipler arasındaki moleküler farklılıklar, ilaç yanıtlarında da önemli değişkenliğe yol açmaktadır [7]. ER+ tümörler hormonal tedavilere, HER2+ tümörler anti-HER2 ajanlara duyarlıdır. TNBC için standart kemoterapiler ana tedavi seçeneğidir, ancak PI3K/AKT/mTOR ve MAPK yolaklarındaki değişiklikler hedefe yönelik tedaviler için potansiyel sunmaktadır [5, 6]. Güncel araştırmalar, TNBC'de immün kontrol noktası inhibitörleri ve antikor-ilaç konjugatlarının umut verici sonuçlar gösterdiğini ortaya koymaktadır [4, 9]. Yeni nesil genom dizileme teknolojileri, hasta-spesifik genetik varyantların tanımlanmasına katkıda bulunmaktadır [15]. İlaç direnci önemli bir klinik sorundur. İntegratif multiomiks yaklaşımlar, alt tiplere özgü moleküler özelliklerin ve ilaç duyarlılık belirteçlerinin tanımlanmasında önemli rol oynamaktadır.

Meme kanseri hücre hatları, farklı moleküler düzeylerdeki verilerin entegrasyonu için değerli modeller sunmaktadır [13].
Bu çalışmada, Jastrzebski ve arkadaşları [8] tarafından sağlanan 30 meme kanseri hücre hattına ait RNA-seq, RPPA ve yedi kinaz inhibitörüne karşı ilaç duyarlılık verilerini (ArrayExpress: E-MTAB-4801) entegre ederek, 'Hormon/HER2 Pozitif' ve 'TNBC' grupları arasındaki moleküler farklılıkları, potansiyel biyobelirteçleri ve mTOR inhibitörlerine duyarlılıkta 4E-BP1'in rolünü araştırdık. Çalışmamız, multiomiks veri entegrasyonunun meme kanserinde kişiselleştirilmiş tedavi yaklaşımlarının geliştirilmesindeki uygulamalarını göstermektedir.

Yöntem

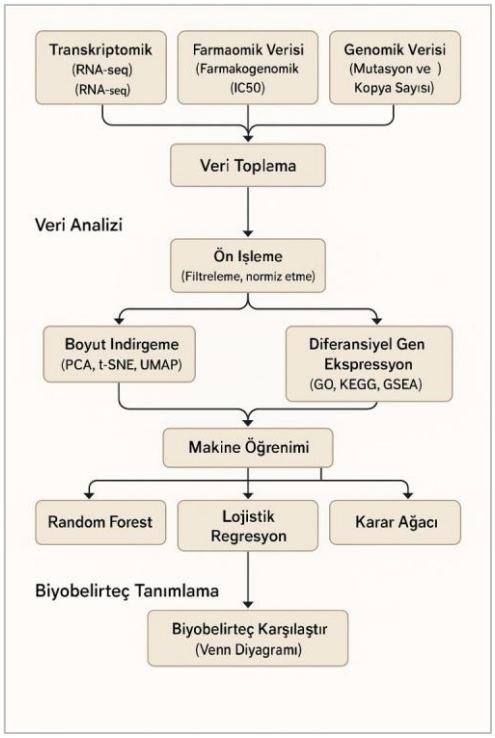
2.1. Veri Setleri ve Kaynaklar

Bu çalışmada, Jastrzebski ve arkadaşları [8] tarafından daha önce kapsamlı olarak karakterize edilmiş, 18 TNBC, 4 ER+, 4 HER2+ ve 4 ER+/HER2+ olmak üzere toplam 30 meme kanseri hücre hattını içeren bir panel kullanılmıştır. Bu panel, entegre multiomiks analizleri için zengin bir kaynak sağlamaktadır. Çalışmada kullanılan transkriptomik (RNA-seq; tüm genler için normalize edilmiş okuma sayıları), proteomik (RPPA; 152 farklı protein/fosfo-protein epitopu için ters faz protein dizisi verileri), farmakogenomik (İlaç Yanıtı; yedi farklı kinaz inhibitörüne [AZD8055, BEZ235, GDC0941, MK2206, PD0325901, lapatinib, foretinib] karşı 72 saatlik tedavi sonrası ölçülen ve IC50 değerleri hesaplanan göreceli canlılık verileri) ve genomik (Mutasyon ve Kopya Sayısı; hedefli DNA dizileme ile belirlenen mutasyonlar ve tüm genom kopya sayısı profilleri) veriler, ArrayExpress veri tabanından (Erişim Numarası: E-MTAB-4801) temin edilmiştir. Hücre hatları, Jastrzebski ve ark. [8] tarafından tanımlanan ER+, HER2+ ve TNBC (ER-/PR-/HER2-) alt tiplerine göre sınıflandırılmıştır. Ancak, bu çalışmada kullanılan ve 152 epitopu hedefleyen RPPA veri setinin PR proteinini içerip içermediği spesifik olarak incelenmemiş olup, analizler mevcut RNA-seq ve RPPA verileri üzerine yoğunlaşmıştır.

2.2. Veri Analizi

Çalışmamızda izlenen analitik yaklaşım ve multiomiks veri entegrasyonu iş akışı Şekil 1'de özetlenmiştir. Bu iş akışı, veri toplama, kalite kontrolü, boyut indirgeme, diferansiyel gen ekspresyonu analizi, fonksiyonel zenginleştirme, makine öğrenimi ile biyobelirteç tanımlama ve ilaç duyarlılığı analizi aşamalarını içermektedir. Analiz öncesinde veri setlerine çeşitli ön işleme adımları uygulanmıştır. RNA-seq okuma sayıları, düşük varyans gösteren veya düşük oranda eksprese edilen genler filtrelendikten sonra log2 transformasyonuna tabi tutulmuştur. RPPA verileri normalize edilerek teknik varyasyonların etkisi giderilmiştir. İlaç yanıtını gösteren IC50 değerleri log10 transformasyonu ile normalize edilmiştir. Tüm veri setlerindeki eksik değerler, k-en yakın komşu (k-NN) algoritması kullanılarak tamamlanmıştır. Hücre hatlarının moleküler profillerine göre sınıflandırılması ve ilişkilerinin görselleştirilmesi amacıyla boyut indirgeme teknikleri kullanılmıştır. Gen ekspresyon verisindeki temel varyans bileşenlerini belirlemek ve hücre hatlarının alt tiplere göre ayrışımını incelemek için Temel Bileşen

Analizi (PCA) uygulanmıştır. Yüksek boyutlu verilerin düşük boyutlu uzayda görselleştirilmesi için ayrıca t-Dağılımlı Stokastik Komşu Gömme (t-SNE) yöntemi kullanılmıştır. Hücre hatlarının gen ekspresyon profillerine dayalı olarak gruplandırılması amacıyla ise Ward.D2 bağlantı metodu ile hiyerarşik kümeleme analizi gerçekleştirilmiştir.



Şekil- 1 Meme kanseri hücre hatlarında multiomiks veri analizi iş akış şeması. Çalışmada kullanılan veri kaynakları, analitik yöntemler ve entegrasyon yaklaşımları gösterilmiştir.

Bu çalışmada, meme kanseri alt tipleri arasındaki temel moleküler farklılıkları ve ilaç yanıtı ile ilişkili biyobelirteçleri belirlemek amacıyla, danışmanımızın yönlendirmesi doğrultusunda hücre hatları iki ana grupta incelenmiştir. Jastrzebski vd. (2018) tarafından karakterize edilen 30 hücre hattı (18 TNBC, 4 ER+, 4 HER2+ ve 4 ER+/HER2+), hedefe yönelik tedavi yaklaşımlarına göre 'Hormon/HER2 Pozitif' (n=12, ER+, HER2+ ve ER+/HER2+ hatlarını içeren) ve 'TNBC' (n=18) olmak üzere iki ana grupta değerlendirilmiştir. Bu iki ana grup arasındaki istatistiksel olarak anlamlı gen ekspresyon farklılıkları, R/Bioconductor paketi DESeq2 (v1.34.0) kullanılarak belirlenmiştir [8]. Anlamlılık için FDR < 0.05 ve mutlak log2 kat değişim ($|\log2FC| > 1$ eşik değerleri kullanılmıştır. Bu gruplamanın temel gerekçesi, hormonal ve HER2 hedefli tedavilere yanıt verme potansiyeli olan alt tiplerin TNBC'den net bir şekilde ayrılması ve az sayıdaki örneklemlerin birleştirilerek analizlerin istatistiksel gücünün artırılmasıdır.

Diferansiyel olarak eksprese edilen genlerin fonksiyonel anlamını ve biyolojik rollerini değerlendirmek üzere fonksiyonel zenginleştirme analizleri yapılmıştır. Gene Ontology (GO) terimleri (biyolojik süreç, moleküler fonksiyon, hücresel komponent) ve KEGG (Kyoto Encyclopedia of Genes and Genomes) yolaklarındaki zenginleşmeler, R/Bioconductor paketi clusterProfiler (v4.2.2) ile analiz edilmiştir [19]. Ayrıca, tüm genlerin ekspresyon skorlarına dayalı olarak yolak aktivitesindeki değişimleri değerlendirmek amacıyla, MSigDB Hallmark (v7.4) gen seti koleksiyonu kullanılarak Gen Seti Zenginleştirme Analizi (GSEA) yapılmıştır.
Transkriptomik, proteomik ve ilaç yanıt verilerini entegre etmek ve katmanlar arası ilişkileri ortaya çıkarmak amacıyla çeşitli yaklaşımlar benimsenmiştir. Bunlar arasında, farklı omik katmanlarından elde edilen özelliklerin ağırlıklandırılarak entegre edildiği Çoklu Faktör Analizi (MFA), omik veri setleri arasındaki korelasyonları en üst düzeye çıkararak biyolojik ilişkileri belirlemeyi amaçlayan Özgül Özellik Analizi (DIABLO - mixOmics paketi içinde) ve moleküler etkileşim ağları kullanılarak çoklu omik verilerin ilişkilendirildiği ağ tabanlı entegrasyon yöntemleri bulunmaktadır.

2.3. Makine Öğrenimi ile Biyobelirteç Tanımlama:

Meme kanseri alt tiplerini ayırt eden ve ilaç yanıtını öngörebilen potansiyel biyobelirteçleri tanımlamak için üç farklı makine öğrenimi algoritması içeren bir analiz iş akışı uygulanmıştır. 'Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasındaki ayrımı modellemek için Random Forest, Lojistik Regresyon ve Karar Ağacı algoritmaları kullanılmıştır [15, 3, 14]. Random Forest, doğrusal olmayan bir makine öğrenim modeli olarak doğrusal ve lojistik regresyondan daha karmaşık veri yapılarını yakalayabilme özelliğiyle multiomik veri analizlerinde oldukça etkili olduğu gösterilmiştir [15].
Analizler için FDR değerlerine göre sıralanmış ve filtrelenmiş diferansiyel eksprese olan genler arasından seçilen özellik seti kullanılmıştır. Veri seti, eğitim (%70) ve test (%30) setlerine rastgele olarak ayrılmıştır. Random Forest modelinde genlerin önemini belirlemek için Mean Decrease Gini metriği hesaplanmıştır. Lojistik Regresyon modelinde LASSO (L1 regularizasyon, $\alpha=1$) yaklaşımı kullanılmış ve en uygun lambda değeri çapraz doğrulama ile belirlenmiştir.
Karar Ağacı modelinde rpart algoritması uygulanmıştır. Farklı bölünme kriterleri ve karmaşıklık parametreleri test edilerek, optimum model oluşturulmuştur. Modelin görselleştirilmesi için rpart.plot paketi kullanılarak, karar ağacı diyagramı oluşturulmuştur. Diyagramda, her düğümdeki örneklem sayıları ve her dal için bölünme kriterleri gösterilmiştir. Farklı modellerin belirlediği önemli genlerin kesişimlerini görselleştirmek için Venn diyagramı analizi yapılmıştır. Modellerin test seti üzerindeki performansı doğruluk, duyarlılık, özgüllük ve kappa metrikleri kullanılarak değerlendirilmiş ve karşılaştırılmıştır. Tüm modeller için 5 katlı çapraz doğrulama (5-fold cross-validation) yöntemi uygulanmıştır. Bu metodoloji, meme kanseri sağkalım tahmini ve biyobelirteç tanımlamasında etkinliği kanıtlanmış standart bir yaklaşımdır [14]. Analizler, R programlama dili ve ilgili paketler (randomForest, caret, glmnet, rpart, e1071, pROC, VennDiagram vb.) kullanılarak gerçekleştirilmiştir.
4E-BP1 protein ekspresyonu ve fosforilasyon durumunun mTOR inhibitörlerine (AZD8055, BEZ235) duyarlılıkla ilişkisini incelemek için ters faz protein dizisi (RPPA) verileri kullanılmıştır. 4E-BP1 protein ekspresyon seviyeleri ile ilaç yanıt

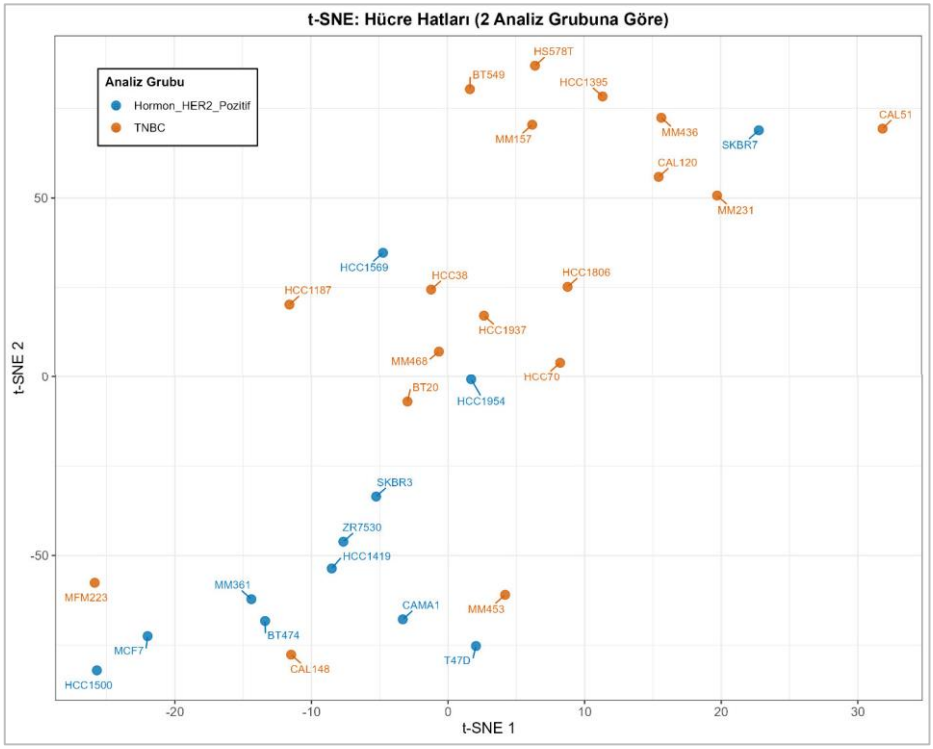
verileri (IC50 değerleri) arasındaki ilişki, Spearman korelasyon analizi ile değerlendirilmiştir. Protein ekspresyon seviyelerinin ilaç duyarlılığı üzerindeki etkisini daha ayrıntılı incelemek için hücre hatları, 4E-BP1 ekspresyon medyanına göre 'Yüksek' ve 'Düşük' ekspresyon gruplarına ayrılmış ve bu gruplar arasındaki ilaç duyarlılığı farkları t-testi ile analiz edilmiştir. Protein ekspresyon seviyesi ve ilaç yanıtı arasındaki ilişkinin gücünü ve yönünü belirlemek için doğrusal regresyon modelleri oluşturulmuştur. Bu modellerde, 4E-BP1 toplam protein ekspresyonu, fosforilasyon durumu ve diğer potansiyel değişkenlerin etkileri incelenmiştir. Alt tiplerin moleküler özelliklerini kapsamlı olarak karakterize etmek için GO terimleri (biyolojik süreç, moleküler fonksiyon, hücresel komponent) ve KEGG yolaklarındaki zenginleşmeler, R/Bioconductor paketi clusterProfiler (v4.2.2) ile analiz edilmiştir [19]. Yolak aktivitesindeki değişimleri tüm gen ekspresyon profillerine dayalı olarak değerlendirmek için, MSigDB Hallmark (v7.4) gen seti koleksiyonu kullanılarak GSEA (Gen Seti Zenginleştirme Analizi) yapılmıştır.

Bulgular

3.1. Gen Ekspresyon Profilleri ve Alt Tip Ayrımı

Hücre hatlarının gen ekspresyon profillerini analiz etmek için t-SNE (t-dağılımlı Stokastik Komşu Gömme) boyut indirgeme tekniği kullanılmıştır. Bu teknik, yüksek boyutlu ekspresyon verilerindeki karmaşık ilişkileri koruyarak, hücre hatları arasındaki moleküler benzerlikleri ve farklılıkları görselleştirmede etkindir. Şekil 2'de gösterildiği gibi, t-SNE analizi sonucunda 'Hormon/HER2 Pozitif' grubundaki hücre hatları (mavi noktalar) ve 'TNBC' grubu hücre hatları (turuncu noktalar) belirgin şekilde ayrışmaktadır. Hormon/HER2 Pozitif grup genellikle grafiğin alt ve sol bölgelerinde kümelenirken, TNBC hücre hatları ağırlıklı olarak üst ve sağ bölgelerde konumlanmıştır. Bu ayrışma, iki analiz grubu arasında transkriptomik seviyede önemli farklılıklar olduğunu doğrulamaktadır.

Dikkat çekici şekilde, TNBC grubundaki hücre hatları kendi içlerinde daha geniş bir dağılım göstermektedir, bu da bu alt tipin moleküler açıdan daha heterojen bir yapıya sahip olabileceğini düşündürmektedir. t-SNE analizinin bu sonuçları, önceki çalışmalarda gözlemlenen meme kanseri alt tiplerinin moleküler ayrımını desteklemekte ve özellikle TNBC'nin diğer alt tiplerden belirgin şekilde farklılaştığını göstermektedir. Bu bulgular, 'Hormon/HER2 Pozitif' ve 'TNBC' olarak tanımlanan iki ana analiz grubu arasındaki moleküler farklılıkları net bir şekilde ortaya koymakta ve sonraki diferansiyel ekspresyon ve fonksiyonel analizler için sağlam bir temel oluşturmaktadır.

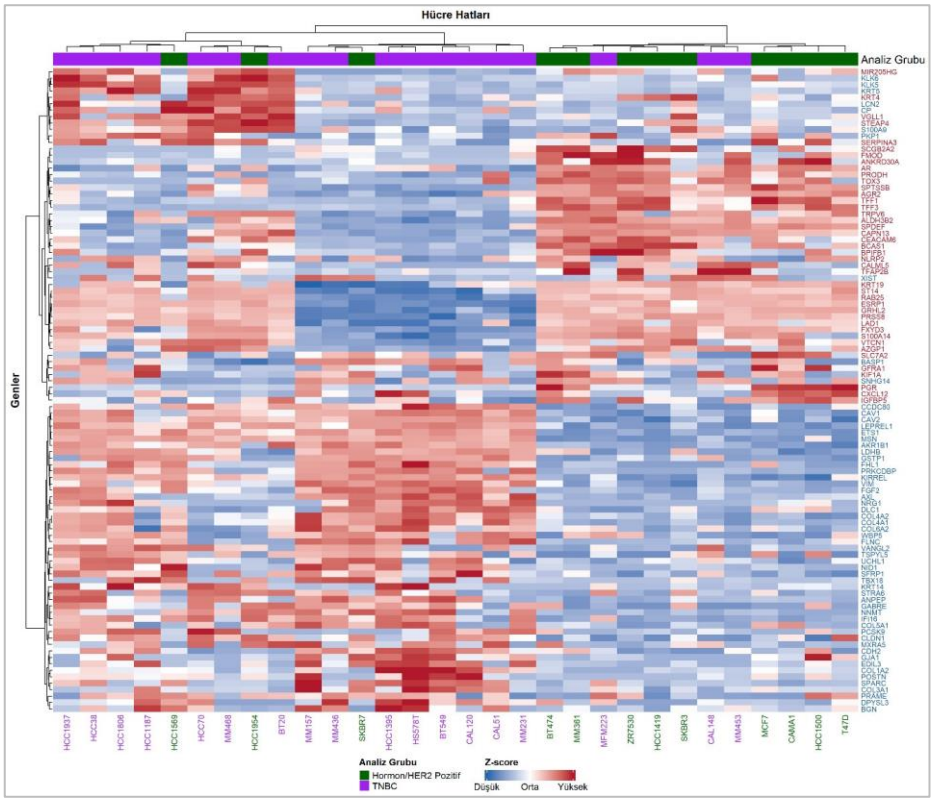


Şekil- 2 t-SNE: Hücre hatlarının gen ekspresyon profilleri. Analiz gruplarının ('Hormon/HER2 Pozitif' ve 'TNBC') t-SNE boyut indirgeme tekniği ile iki boyutlu uzaydaki dağılımı gösterilmiştir. Mavi noktalar Hormon/HER2 Pozitif grubu, turuncu noktalar ise TNBC grubunu temsil etmektedir. Gruplar arasındaki moleküler farklılıklar, t-SNE analizi ile belirgin bir şekilde ayrışmaktadır.

Şekil 3'te sunulan gen ekspresyon ısı haritası, hücre hatlarının transkriptomik profillerine dayalı kapsamlı bir görselleştirme sunmaktadır. Bu ısı haritasında, en yüksek varyansa sahip 100 genin ekspresyon seviyeleri Z-score ile normalize edilerek gösterilmiştir. Mavi renkli alanlar düşük ekspresyon, kırmızı alanlar ise yüksek ekspresyon seviyelerini temsil etmektedir. Hiyerarşik kümeleme sonucunda, hücre hatları net bir şekilde iki ana gruba ayrılmaktadır ve bu gruplar çalışmada tanımlanan analiz gruplarıyla ('Hormon/HER2 Pozitif' ve 'TNBC') büyük ölçüde örtüşmektedir. Üstteki renkli çubukta gösterildiği gibi, TNBC hücre hatları (mor) ve Hormon/HER2 Pozitif hücre hatları (yeşil) kendi içlerinde kümelenme eğilimindedir.

Isı haritasında, iki grup arasında farklı ekspresyon paternleri sergileyen gen kümeleri açıkça görülmektedir. Hormon/HER2 Pozitif grupta yüksek eksprese olan genler arasında TFF1, TFF3, GREB1, FOXA1 ve ALDH3B2 gibi genler yer alırken, TNBC grubunda AKR1B1, CAV1, CAV2, FABP5, EGFR, ROR1 ve VIM gibi genlerin yüksek ekspresyon gösterdiği tespit edilmiştir. Bu gen kümeleri, iki analiz grubu arasındaki temel biyolojik farklılıkları yansıtmaktadır. Özellikle haritanın üst bölgelerinde görülen gen kümelerinde Hormon/HER2 Pozitif hücre hatlarına özgü ekspresyon paternleri (kırmızı bölgeler) gözlenirken, haritanın alt bölgelerindeki gen kümeleri TNBC hücre

hatlarında daha yüksek ekspresyon (kırmızı bölgeler) sergilemektedir. Bu ekspresyon desenleri, meme kanseri alt tiplerinin farklı biyolojik özelliklerini ve potansiyel sinyal yolaklarını yansıtmaktadır. Bu ısı haritası, t-SNE ve diferansiyel gen ekspresyon analizlerinden elde edilen bulguları desteklemekte ve 'Hormon/HER2 Pozitif' ile 'TNBC' analiz grupları arasındaki moleküler farklılıkların kapsamlı bir görsel temsilini sunmaktadır.



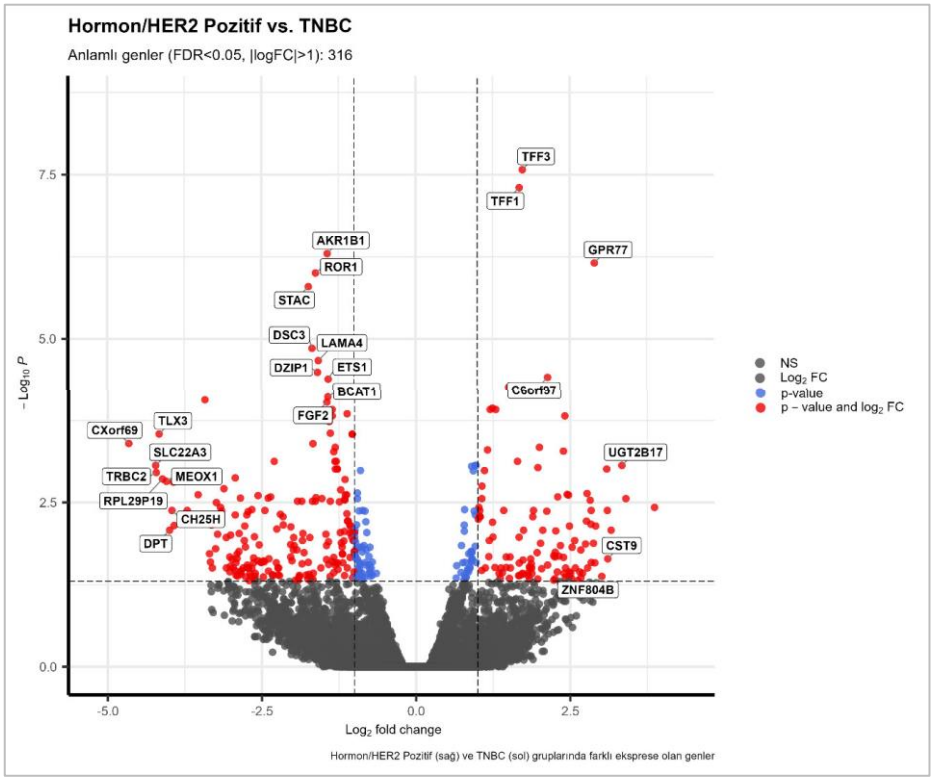
Şekil- 3 Gen ekspresyon profillerine dayalı kapsamlı ısı haritası. En değişken 100 genin Z-score normalizasyonu yapılmış ekspresyon değerleri gösterilmiştir. Hücre hatları ve genler hiyerarşik kümeleme (Öklid mesafesi ve ward.D2 metodu) ile gruplandırılmıştır. Üstteki çubuk, analiz gruplarını göstermektedir (Yeşil: Hormon/HER2 Pozitif, Mor: TNBC). Isı haritası, gruplar arasındaki ekspresyon paternlerindeki farklılıkları görselleştirmektedir.

3.2. Diferansiyel Gen Ekspresyonu ve Fonksiyonel Zenginleştirme

Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasında yapılan diferansiyel gen ekspresyon analizi, 316 genin istatistiksel olarak anlamlı düzeyde farklı eksprese edildiğini ortaya koymuştur (FDR < 0.05, |log2FC| > 1). Bu genlerden 114 tanesi Hormon/HER2 Pozitif grubunda, 202 tanesi ise TNBC grubunda daha yüksek ekspresyon göstermiştir. Şekil 4'teki volkan grafiğinde görüldüğü üzere, Hormon/HER2 Pozitif grubunda yukarı regüle olan genler arasında özellikle TFF3 ve TFF1 en yüksek anlamlılığa sahip genler olarak öne çıkmaktadır. Bu genler, meme dokusunda östrojen

tarafından indüklenen trefoil faktör ailesine ait proteinleri kodlar. Ayrıca, GPR77, C6orf97, UGT2B17 ve CST9 genleri de bu grupta anlamlı şekilde yüksek ekspresyon göstermektedir. TNBC grubunda yukarı regüle olan genler arasında ise AKR1B1 ve ROR1 genleri dikkat çekmektedir. AKR1B1, hücre stresine yanıtta ve tümör progresyonunda rol oynarken, ROR1 özellikle TNBC'de ekspresyonu artan ve terapötik hedef olarak araştırılan bir gendir. Diğer belirgin TNBC-spesifik genler arasında STAC, ETS1, BCAT1, FGF2, DSC3, LAMA4, DZIP1 ve TLX3 bulunmaktadır. Ayrıca, CXorf69, SLC22A3, TRBC2, MEOX1, RPL29P19, CH25H ve DPT genleri de TNBC grubunda belirgin şekilde yüksek ekspresyon göstermektedir

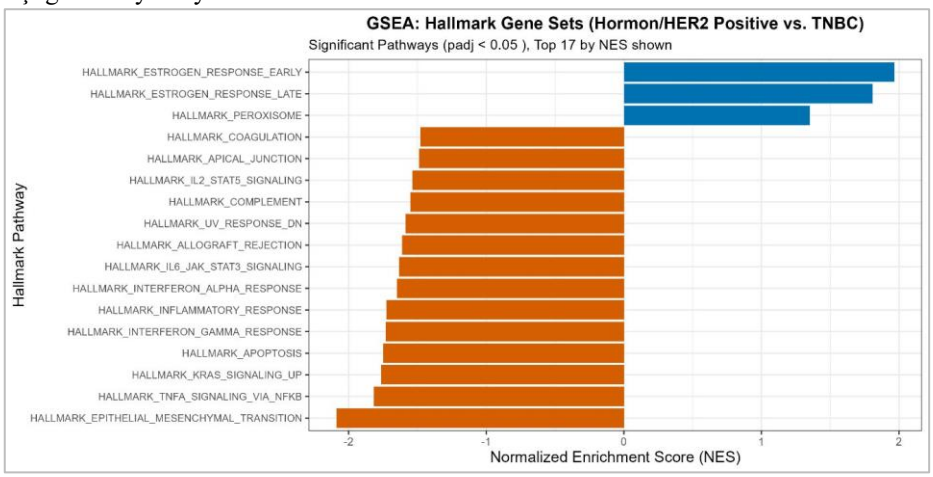
.



Şekil- 4 'Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasındaki diferansiyel gen ekspresyonu volkan grafiği. Kırmızı noktalar istatistiksel olarak anlamlı diferansiyel eksprese edilen genleri (FDR < 0.05 ve |log2FC| > 1) göstermektedir. X ekseni log2 kat değişimini, Y ekseni -log10(P değeri)'ni temsil eder.

Volkan grafiğinde ayrıca, hem istatistiksel anlamlılık (-log10P değeri) hem de ekspresyon değişiminin büyüklüğünü (log2FC) dikkate alarak, p-değeri anlamlı ancak log2FC eşik değerinin altında kalan genler de mavi ile gösterilmiştir. Bu genler, daha küçük ekspresyon değişimleri göstermelerine rağmen, potansiyel biyolojik öneme sahip olabilirler. 'Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasındaki diferansiyel gen ekspresyon farklılıklarının altında yatan biyolojik mekanizmaları anlamak amacıyla fonksiyonel zenginleştirme analizleri gerçekleştirildi. GO analizleri, Hormon/HER2

Pozitif grubunda yukarı regüle olan genlerin daha çok çeşitli metabolik süreçlerde (örn. hormon, amino asit metabolizması), TNBC grubunda yukarı regüle olan genlerin ise immün yanıt, hücre dışı matris organizasyonu ve hücre hareketi ile ilişkili süreçlerde yoğunlaştığını gösterdi. Benzer şekilde, KEGG yolak analizleri Hormon/HER2 Pozitif grupta Östrojen sinyal yolağını öne çıkarırken, TNBC grubunda Sitokin-sitokin reseptör etkileşimi ve JAK-STAT sinyal yolağı gibi immün/inflamatuar yolakların zenginleştiğini ortaya koydu.



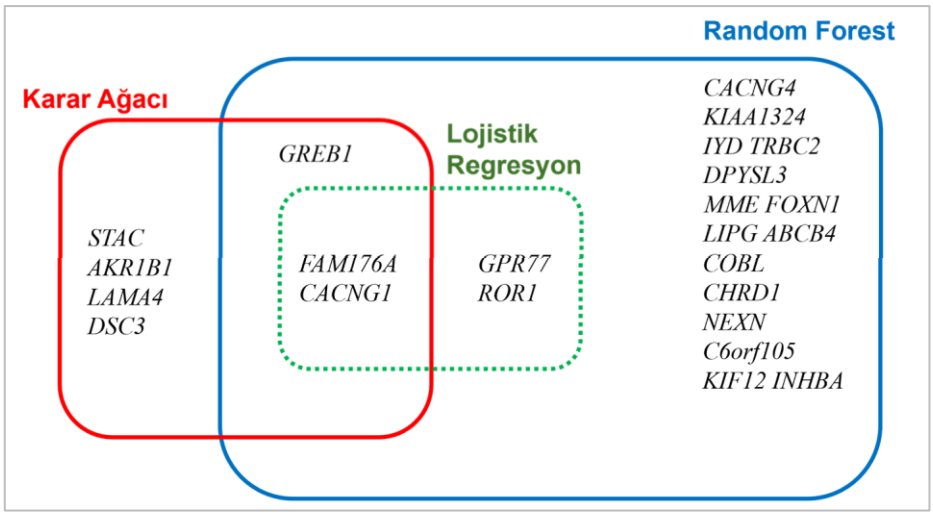
Şekil- 5 GSEA: Hallmark gen setleri analizi sonuçları. 'Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasında anlamlı (padj < 0.05) olarak zenginleşen Hallmark yolakları gösterilmiştir. Pozitif Normalize Edilmiş Zenginleşme Skorları (NES) Hormon/HER2 Pozitif grubunda, negatif skorlar ise TNBC grubunda zenginleşmeyi ifade eder.

Bu bulguları teyit eden ve daha geniş bir bakış açısı sunan Gen Seti Zenginleştirme Analizi (GSEA) sonuçları Şekil 5'te gösterilmektedir. Hallmark gen setleri kullanılarak yapılan GSEA, Hormon/HER2 Pozitif grubunda beklendiği gibi "Östrojen Yanıtı (Erken ve Geç)" yolaklarının anlamlı şekilde zenginleştiğini (pozitif NES) gösterdi. Buna karşılık TNBC grubunda (negatif NES) ise özellikle "İnflamatuar Yanıt", "İnterferon Alfa/Gama Yanıtı", "IL6-JAK-STAT3 Sinyali", "TNFa Sinyali (NFkB üzerinden)" gibi immün ve inflamatuar süreçler ile "Epitelyal Mezenkimal Geçiş (EMT)" ve "KRAS Sinyali Yukarı" gibi yolakların anlamlı düzeyde aktive olduğu belirlendi. Bu sonuçlar, iki analiz grubu arasında belirgin fonksiyonel farklılıklar olduğunu ve TNBC'nin immünolojik olarak aktif ve invaziv bir profille karakterize olduğunu desteklemektedir. Özellikle, TFF genlerinin Hormon/HER2 Pozitif grupta öne çıkması, östrojen sinyalinin bu alt tipteki önemini vurgularken, TNBC grubunda öne çıkan ETS1, ROR1 ve AKR1B1 gibi genler, hücre proliferasyonu, invazyon ve metastaz ile ilişkili sinyal yolaklarının aktivasyonunu göstermektedir. Bu moleküler imzalar, meme kanseri alt tiplerinin farklı biyolojik davranışlarını açıklamada ve potansiyel tedavi hedeflerinin belirlenmesinde önemli ipuçları sunmaktadır.

3.3. Potansiyel Biyobelirteçlerin Belirlenmesi ve İlaç Yanıtı İlişkisi

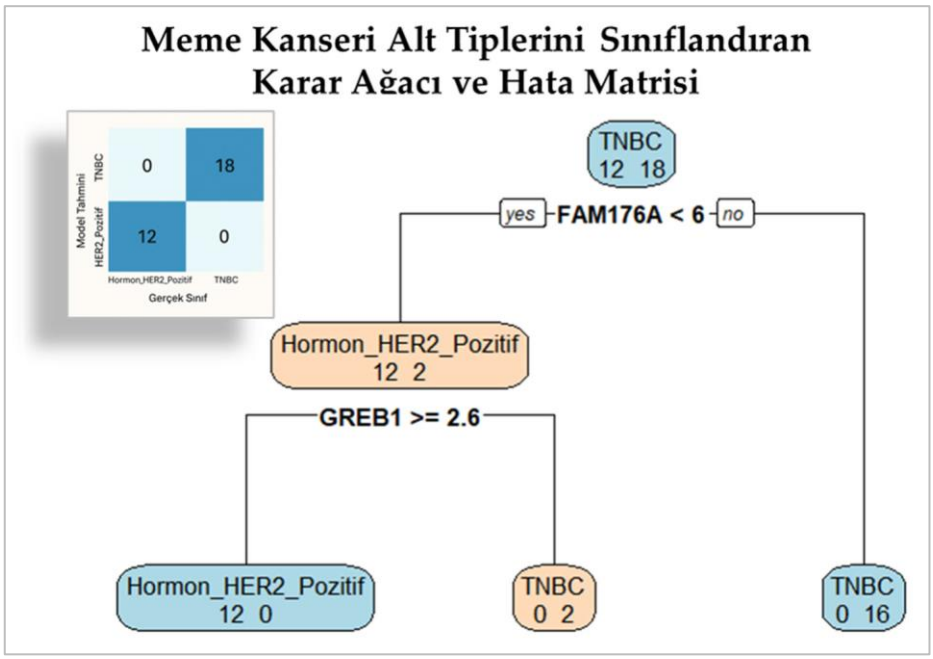
'Hormon/HER2 Pozitif' ve 'TNBC' gruplarını ayırt etmede kullanılabilecek potansiyel biyobelirteçleri tanımlamak için, Random Forest, Lojistik Regresyon ve Karar Ağacı

modelleri uygulanmıştır. Venn diyagramında (Şekil 6), Karar Ağacı modeli 4 özgün gen (STAC, AKR1B1, LAMA4, DSC3) belirlerken, Random Forest modeli 15 özgün gen tanımlamıştır. İki modelin kesişiminde GREB1 bulunmaktadır. Lojistik Regresyon ile Random Forest arasında 2 ortak gen (GPR77, ROR1) vardır. Üç modelin tümünde önemli bulunan 2 gen (FAM176A, CACNG1) dikkat çekmektedir. Random Forest'ta CACNG4, FAM176A, GPR77, KIAA1324 ve IYD; Lojistik Regresyon'da CACNG1, FAM176A, ROR1 ve GPR77; Karar Ağacı'nda ise STAC, AKR1B1, LAMA4 ve DSC3 öne çıkmıştır.



Şekil- 6 Modellerin Belirlediği Ortak Biyobelirteçler. Üç farklı makine öğrenimi modelinin (Karar Ağacı, Random Forest ve Lojistik Regresyon) belirlediği önemli genler ve kesişimleri. Üç modelin ortak bulduğu FAM176A ve CACNG1 genleri potansiyel biyobelirteç adayları olarak öne çıkmaktadır.

Şekil 7'deki karar ağacı modeli, 30 hücre hattı üzerinde meme kanseri alt tiplerinin ayrımını göstermektedir. FAM176A < 6 olduğunda, GREB1 ≥ 2.6 olan vakalar tamamen Hormon/HER2 Pozitif (12 örnek, 0 TNBC), GREB1 < 2.6 olanlar ise tamamen TNBC'dir (0 Hormon/HER2 Pozitif, 2 TNBC). FAM176A ≥ 6 olduğunda ise tüm örnekler TNBC'dir (0 Hormon/HER2 Pozitif, 16 TNBC). Bu bulgular, FAM176A'nın alt tip belirlemede ve GREB1'in Hormon/HER2 Pozitif grubu içindeki ayrımda önemli rol oynadığını göstermektedir



Şekil-7 Meme Kanseri Alt Tiplerini Sınıflandıran Karar Ağacı Modeli ve Hata Matrisi. Sol tarafta hata matrisi, sağda karar ağacı modellemesi ile elde edilen meme kanseri alt tipleri (TNBC ve Hormon/HER2 Pozitif) arasındaki ayrımı gösteren diyagram. Modelde FAM176A ve GREB1 genleri belirleyici değişkenler olarak kullanılmıştır. Kutulardaki değerler ilgili sınıflara ait örneklem sayılarını göstermektedir.

4. Tartışma

4.1. Alt Tipler Arasındaki Moleküler Farklılıklar

Bu çalışmada, 30 meme kanseri hücre hattının transkriptomik profilleri incelendiğinde, 'Hormon/HER2 Pozitif' ve 'TNBC' analiz grupları arasında belirgin moleküler farklılıklar saptanmıştır. t-SNE ve hiyerarşik kümeleme analizleri, bu iki grubun ekspresyon paternleri açısından net bir ayrışma gösterdiğini doğrulamaktadır (Şekil 2, Şekil 3). Bu bulgu, genom çapındaki moleküler profillemenin meme kanseri sınıflandırmasındaki önemini vurgulamaktadır [12, 16]. Fonksiyonel analizler, 'Hormon/HER2 Pozitif' grubunda Östrojen yanıtı ve metabolizma ile ilgili yolakların (GSEA analizi, Şekil 5), 'TNBC' grubunda ise immün yanıt ve inflamatuar süreçler ile invazyon ilişkili yolakların (Sitokin-sitokin reseptörü, JAK-STAT, EMT) aktive olduğunu göstermiştir.
Diferansiyel ekspresyon analizinde, Hormon/HER2 Pozitif grupta TFF1 ve TFF3 gibi östrojenle indüklenen genler, TNBC grubunda ise AKR1B1, ROR1, STAC ve ETS1 gibi genler yüksek ekspresyon göstermiştir (Şekil 4). GSEA analizinde TNBC grubunda "EMT" ve "KRAS Sinyali Yukarı" yolaklarının aktivasyonu, bu grubun daha agresif ve metastatik potansiyelini desteklemektedir [11].

Makine öğrenimi modelleri ile üç modelin kesişiminde yer alan FAM176A ve CACNG1 genleri potansiyel biyobelirteçler olarak belirlenmiştir (Şekil 6). Karar Ağacı modelinde (Şekil 7) FAM176A gen ekspresyonu alt tip ayrımında ilk belirleyici faktör, GREB1 ise Hormon/HER2 Pozitif grup içindeki sınıflandırmada önemli bulunmuştur. Bu bulgular, meme kanseri moleküler alt tiplerinin tanımlanmasında kullanılabilecek gen panellerine katkı sağlamakta ve literatürde bildirilen belirteçlere ek bilgiler sunmaktadır.

4.2. 4E-BP1'in mTOR İnhibitör Duyarlılığındaki Rolü

Çalışmamızın önemli bulgularından biri, orijinal çalışmadaki RPPA verilerine dayanarak, 4E-BP1 protein ekspresyon seviyesinin mTOR inhibitörlerine (AZD8055, BEZ235) duyarlılık için potansiyel bir belirteç olduğunun doğrulanmasıdır. Yüksek 4E-BP1 protein ekspresyonuna sahip hücre hatlarının bu inhibitörlere daha duyarlı olduğu, düşük ekspresyonlu hatların ise daha dirençli olduğu gözlenmiştir. Bu bulgu, Jastrzebski ve ark. [8] çalışmasıyla tamamen uyumludur ve translasyonel regülasyonun mTOR inhibisyonuna yanıtta kritik bir rol oynadığını göstermektedir. Önceki çalışmalarda genellikle 4E-BP1'in fosforilasyon durumuna odaklanılırken, bu veri seti üzerinde yapılan analizler ve referans çalışma, toplam 4E-BP1 protein miktarının da kritik önemde olduğunu vurgulamaktadır. Bu bulgu, RNA-seq verilerine dayalı gruplamamızdan bağımsız olarak, mTOR inhibitörlerinin klinik etkinliğini artırmak için hasta seçiminde kullanılabilecek pratik bir biyobelirteç potansiyeli taşımaktadır. EIF4EBP1 geninin amplifikasyonu veya yüksek protein ekspresyonu gösteren tümörlerin, aktif bölgeyi hedefleyen mTOR inhibitörleri ile tedaviden daha fazla yarar görebileceği hipotezini desteklemektedir.

4.3. Multiomiks Yaklaşımların Önemi

Bu çalışmada, Jastrzebski ve ark. [8] tarafından sağlanan çok katmanlı verilerin (transkriptomik, proteomik, farmakogenomik) analizi (Şekil 1), meme kanseri alt tiplerinin kompleks moleküler özelliklerini daha kapsamlı anlamamızı sağlamıştır. Farklı omik katmanlarından elde edilen bilgilerin birleştirilmesi, tek bir veri türünün sağlayabileceğinden daha zengin ve bütüncül bir bakış açısı sunmaktadır. Yaptığımız iki gruplu analiz ('Hormon/HER2 Pozitif' vs 'TNBC'), bu entegratif yaklaşımın, analiz grupları arasındaki temel moleküler farklılıkların ve potansiyel ilaç yanıtı belirteçlerinin (örn. 4E-BP1) daha net anlaşılmasını nasıl mümkün kıldığını göstermektedir. Belirlenen gen imzaları ve fonksiyonel yolak profilleri, klinik ortamda hastaların moleküler profillerine dayalı olarak kişiselleştirilmiş tedavi stratejileri geliştirmede ve tedavi yanıtını öngörmede kullanılma potansiyeli taşımaktadır.

4.4. Çalışmanın Kısıtlılıkları

Bu çalışmanın bazı kısıtlılıkları bulunmaktadır. Öncelikle, analizler in vitro hücre hattı modellerine dayanmaktadır ve elde edilen bulguların “in vivo” koşullarda ve klinik ortamlarda doğrulanması kritik öneme sahiptir. Hücre hatları, tümör mikro çevresinin karmaşık yapısını ve hücreler arası etkileşimleri tam olarak yansıtmayabilir. Ayrıca, hasta tümörlerinde gözlenen intratümoral heterojenite, hücre hattı modelleriyle tam olarak temsil edilemeyebilir [13]. Analizimizde ER+ ve HER2+ hücre hatlarının 'Hormon/HER2 Pozitif' adı altında tek bir grupta birleştirilmesi, istatistiksel gücü artırmış

ve TNBC ile olan temel farkları ortaya çıkarmış olsa da, ER+ ve HER2+ alt tipleri arasındaki potansiyel önemli farkları maskelemiş olabilir. Bu faktörler, bulguların doğrudan klinik uygulamaya aktarılmasında dikkatli olunmasını gerektirmektedir.

4.5. Gelecek Çalışmalar ve Perspektifler

Belirlenen potansiyel biyobelirteçlerin (4E-BP1 ekspresyonu, Şekil 6'daki FAM176A, CACNG1, GPR77 ve Şekil 7'deki karar ağacında öne çıkan FAM176A, GREB1 genleri) prognostik ve prediktif değerinin daha büyük hasta kohortlarında doğrulanması esastır. Bu biyobelirteçler ve modellerimizin kişiselleştirilmiş tedavi stratejileri için klinik karar destek sistemlerine entegre edilme potansiyeli araştırılmalıdır. Karar ağacımızın gösterdiği belirli gen ekspresyon eşik değerlerinin alt tip sınıflandırmasındaki yüksek doğruluğu, klinik uygulamalar için umut vericidir. Ayrıca, ER+ ve HER2+ gruplarının ayrıntılı incelenmesi ve tümör mikro çevresinin etkisini hesaba katan daha karmaşık modellerin geliştirilmesi, meme kanseri tedavisindeki direncin aşılması için önemli araştırma alanları olacaktır.

Sonuç

Sonuç olarak, bu çalışma 30 meme kanseri hücre hattının multiomiks verilerini kullanarak, 'Hormon/HER2 Pozitif' ve 'TNBC' grupları arasındaki belirgin transkriptomik ve fonksiyonel farklılıkları ortaya koymuştur. Hormon/HER2 Pozitif grup östrojen yanıtı ve metabolik süreçlerle karakterize olurken, TNBC grubu immün yanıtlar, EMT ve çeşitli sinyal yolaklarının aktivasyonu ile öne çıkmaktadır.

Çalışmamız, Jastrzebski vd. (2018) tarafından tanımlanan 4E-BP1 protein ekspresyonu ve mTOR inhibitör duyarlılığı arasındaki ilişkiyi doğrulamanın yanı sıra, üç farklı makine öğrenimi modelinin kesişiminde yer alan FAM176A ve CACNG1 genlerini potansiyel yeni biyobelirteçler olarak tanımlamıştır. Karar ağacı modelimiz (Şekil 7), özellikle FAM176A ve GREB1 genlerinin yüksek doğrulukla belirleyici olduğunu göstermiştir. FAM176A ekspresyonunun 6'nın altında veya üstünde olması örnekleri belirgin şekilde ayırmakta, GREB1 ise Hormon/HER2 Pozitif grubun sınıflandırmasında önem taşımaktadır. Bu transkriptomik imzalar ve biyobelirteçler, klinik örneklerde doğrulandığında, meme kanseri hastalarının moleküler sınıflandırılmasında ve tedavi kararlarında değerli araçlar sunabilir. Karar ağacı modelimizin basit ve yorumlanabilir yapısı, belirli gen ekspresyon eşik değerleriyle kolay uygulanabilir tanı algoritmaları oluşturulmasına olanak tanımaktadır. 4E-BP1 ekspresyonunun mTOR inhibitörlerine duyarlılıkla ilişkisi, hedefe yönelik tedavilerin etkinliğinin artırılmasında önemli bir bulgudur. TNBC alt tipinde immün yanıt ve EMT yolaklarının aktif olması, bu hasta grubunda immünoterapi ve EMT inhibitörlerinin potansiyel değerini göstermektedir. Bütünleşmiş multiomiks yaklaşımımız, kanser biyolojisini anlamak ve klinik karar verme süreçlerini iyileştirmek için güçlü bir yöntem sunmaktadır. Gelecek çalışmalarda bu bulguların daha geniş hasta kohortlarında doğrulanması, meme kanseri tedavisindeki translasyonel potansiyelini ortaya çıkaracaktır.

Kaynaklar

1. Anders, S., Huber, W., Love, M.I.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12), 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
2. Bertucci, F., Finetti, P., Rougemont, J., et al.: Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res. 65(6), 2170-2178 (2005). https://doi.org/10.1158/0008-5472.CAN-04-4115
3. Wei, Z., Han, D., Zhang, C., Wang, S., Liu, J., Chao, F., Song, Z., Chen, G.: Deep Learning-Based Multi-Omics Integration Robustly Predicts Relapse in Prostate Cancer. Front. Oncol. 12, 893424 (2022). https://doi.org/10.3389/fonc.2022.893424
4. Conway, M.E., McDaniel, J.M., Graham, J.D., et al.: STAT3 and GR cooperate to drive gene expression and growth of basal-like triple-negative breast cancer. Cancer Res. 80(21), 4663-4676 (2020). https://doi.org/10.1158/0008-5472.CAN-20-1379
5. Elkabets, M., Vora, S., Juric, D., et al.: mTORC1 inhibition is required for sensitivity to PI3K p110α inhibitors in PIK3CA-mutant breast cancer. Sci. Transl. Med. 5(196), 196ra99 (2013). https://doi.org/10.1126/scitranslmed.3005747
6. Huynh, M., Jayanthan, A., Pambid, M.R., et al.: RSK2: a promising therapeutic target for the treatment of triple-negative breast cancer. Expert Opin. Ther. Targets 24(2), 157-167 (2020). https://doi.org/10.1080/14728222.2020.1709824
7. Iorio, M.V., Ferracin, M., Liu, C.G., et al.: MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 65(16), 7065-7070 (2005). https://doi.org/10.1158/0008-5472.CAN-05-1783
8. Jastrzebski, K., Thijssen, B., Kluin, R.J.C., et al.: Integrative modeling identifies key determinants of inhibitor sensitivity in breast cancer cell lines. Cancer Res. 78(15), 4396-4410 (2018). https://doi.org/10.1158/0008-5472.CAN-17-2698
9. Jiang, Y.Z., Ma, D., Suo, C., et al.: Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 35(3), 428-440 (2019). https://doi.org/10.1016/j.ccell.2019.02.001
10. Lehmann, B.D., Bauer, J.A., Chen, X., et al.: Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121(7), 2750-2767 (2011). https://doi.org/10.1172/JCI45014
11. Nieto, M.A., Huang, R.Y., Jackson, R.A., Thiery, J.P.: EMT: 2016. Cell 166(1), 21-45 (2016). https://doi.org/10.1016/j.cell.2016.06.028
12. Perou, C.M., Sørlie, T., Eisen, M.B., et al.: Molecular portraits of human breast tumours. Nature 406(6797), 747-752 (2000). https://doi.org/10.1038/35021093
13. Polyak, K., Haviv, I., Campbell, I.G.: Co-evolution of tumor cells and their microenvironment. Trends Genet. 25(1), 30-38 (2009). https://doi.org/10.1016/j.tig.2008.10.012
14. Rodosthenous, T., Shahrezaei, V., Evangelou, M.: Integrating multi-OMICS data through sparse canonical correlation analysis for the prediction of complex traits: a comparison study, Bioinformatics, Volume 36, Issue 17, September 2020, Pages 4616–4625, https://doi.org/10.1093/bioinformatics/btaa530
15. Schmid, P., Adams, S., Rugo, H.S., et al.: Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. N. Engl. J. Med. 379(22), 2108-2121 (2022). https://doi.org/10.1056/NEJMoa1809615
16. Sørlie, T., Perou, C.M., Tibshirani, R., et al.: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA 98(19), 10869-10874 (2001). https://doi.org/10.1073/pnas.191367098
17. Subramanian, A., Tamayo, P., Mootha, V.K., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102(43), 15545-15550 (2005). https://doi.org/10.1073/pnas.0506580102
18. Yap, F.Y., Chong, P.F., Ahmad, K.A., et al.: Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med. Inform. Decis. Mak. 19(1), 48 (2019). https://doi.org/10.1186/s12911-019-0801-4
19. Yu, G., Wang, L.G., Han, Y., He, Q.Y.: clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16(5), 284-287 (2012). https://doi.org/10.1089/omi.2011.0118

Effect of FICZ-Induced AhR Activation on Lipid Profiles in Cancer Cells using Lipidomics

Sema Duruoğlu[1 [0009-0000-0143-7616]], Cemil Can Eylem[4[0000-0002-3353-2869]] , Emirhan Nemutlu[4[0000-0002-7337-6215]], Cansu Özdemir Saka[2,3[0000-0003-4095-7475]]* İdil Yet[1[0000-0002-1227-0716]]*

1. Institute of Health Sciences, Department of Bioinformatics, Hacettepe University, Ankara, Türkiye
2. Institute of Health Sciences, Department of Stem Cell Sciences, Hacettepe University, Ankara, Türkiye
3. Center for Stem Cell Research and Development (PEDI-STEM), Hacettepe University, Ankara, Türkiye
4. Faculty of Pharmacy, Department of Analytical Chemistry, Hacettepe University, Ankara, Türkiye

cansuozdemir@hacettepe.edu.tr , idil.yet@hacettepe.edu.tr

Abstract. Cancer is a complex disease characterized by abnormal cell behavior. The aryl hydrocarbon receptor (AhR) is a ligand-dependent transcription factor that regulates cellular responses to various signals. In this study, we investigated the lipidomic effects of AhR activation in HeLa cells after 24-hour treatment with 50 nM 6-formylindolo[3,2-b]carbazole (FICZ). Following lipid extraction, LC-qTOF-MS analysis was performed in both positive and negative ionization modes. Data processing and statistical analyses were carried out using MS-DIAL and R software. Notably, lipidomic analysis revealed significant alterations between the FICZ-treated and control groups. Specifically, levels of , TG 52:3, PC 32:0, and DG 32:1 were increased, whereas FA 40:5, DG 36:1, and CE 20:5 were decreased ($p < 0.05$, fold change > 1.5). These results suggest that AhR activation may affect lipid metabolism in cancer cells. Lipidomic analysis can contribute to understanding molecular mechanisms and support biomarker discovery in AhR-related cancer processes.

Keywords: AhR, Lipidomics, LC-qTOF-MS, 6-formylindolo[3,2-b]carbazole

1 Introduction

Cancer is defined as a complex disease that is characterized by uncontrolled cell growth and the immune system's inability to stop this abnormal growth. Cancer cells can eventually spread to other parts of the body, a process known as metastasis [1]. The aryl hydrocarbon receptor (AhR) is a ligand-dependent transcription factor that responds to both environmental (exogenous) and internal (endogenous) signals. It also plays an important role in recognizing foreign (xenobiotic) chemicals and starting cellular responses against them [2, 3]. Omics technologies have brought about a revolutionary change in the analysis of biological systems, enabling scientists to understand their structure and function, and supporting the development of big data-based approaches. These technologies allow researchers to collect and interpret large-scale data related to a specific level of a biological system in an organism [4]. In this context, lipidomics is one of the leading omics fields, particularly in disease diagnosis and the identification

of specific biomarkers. It has been shown that different lipid classes play critical roles in determining the severity of diseases. Therefore, lipidomic analyses contribute significantly to diagnostic and prognostic processes [5].

FA → Fatty acids function as both energy providers and essential components of cellular membranes. Additionally, they play active roles in modulating cellular and tissue metabolism, physiological functions, and responses to hormonal and other external stimuli [6].
TG → Triacylglycerols are the main form of energy storage in cells and play a critical role in energy homeostasis. They provide fuel during periods of fasting or increased energy demand and are involved in lipid transport and metabolism [7].
DG → Diacylglycerol is important for many cell signaling processes [8].
PC → Phosphatidylcholine serves as a direct substrate for sphingomyelin synthesis and also acts as a major source of second messengers such as DAG, phosphatidic acid, lysophosphatidic acid, and arachidonic acid, which can be further converted into other signaling molecules [9].
DMPE → Dimethylphosphatidylethanolamine (DMPE) is an intermediate in phosphatidylcholine biosynthesis and exhibits physicochemical properties between PE and PC. It also differentially interacts with amino acids such as lysine, influencing membrane protein function and transport [10].
CE → Cholesteryl esters are the main form of stored cholesterol and serve to transport excess cholesterol between organs and tissues, preventing cellular toxicity. They are mainly located in the hydrophobic cores of circulating lipoproteins and intracellular lipid droplets [11].

2 Methods

In this study, the human cervical adenocarcinoma cell line HeLa was used. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37°C in a humidified atmosphere with 5% CO_2. The experimental groups included a control group (UTC) and a group treated with 50 nM of the AhR agonist FICZ). All experiments were conducted using HeLa cells, seeded at a density of 1×10^6 cells per well. Cells used in this study were between passages 3 and 8 to minimize passage-related variation. Untreated cells served as the control group. Each experiment was performed in three independent biological replicates to ensure reproducibility and statistical power. The cells were exposed to the agent for 24 hours. At the end of the treatment, a Isopropanol:Water:Acetonitrile mixture (6:1:3) was added to each well. The plates were then quickly treated with liquid nitrogen to stop cellular metabolism at the same time. After freezing, the cells were scraped and transferred into tubes for further analysis. Samples were analyzed using Liquid Chromatography combined with hybrid Quadrupole Time-of-Flight Mass Spectrometry (LC-qTOF-MS) in both positive and negative ionization modes. Quality con-

trol (QC) pools were prepared for quality assurance. Quality control samples were prepared by pooling equal aliquots from each experimental group and were injected periodically throughout the analytical run to monitor instrument stability. For lipidomic analysis, steps such as data deconvolution, peak alignment, normalization, and data matrix construction were performed using the MS-DIAL software [12]. Group comparisons were conducted using a t-test. The significance threshold was set at $p < 0.05$ and Fold Change > 1.5 (approximately $\log_2 FC > \pm 0.58$). All statistical tests and visualizations were performed using the R programming language, with packages including "tidyverse", "ggplot2", and "ggrepel". Lipids showing significant increases were visualized in red, those with significant decreases in blue, and non-significant lipids in gray.

3 Results

To evaluate the lipidomic response of cancer cells to AhR activation, LC-qTOF-MS data obtained from HeLa cells treated with FICZ were analyzed. The lipidomic profiles of the control group and the group treated with 50 nM FICZ for 24 hours were compared, and differentially expressed lipids between the two groups were identified. The results showed that lipids such as triglycerides (TG 52:3, TG 55:2, TG 50:1, TG 48:3, TG 54:2 TG 46:2) and PC 32:0 , PC O-32:3 , DG 32:1 and DMPE 32:2 were upregulated, while others such as PC O-34:0 , FA 40:5 , DG 36:1 , CE 20:1 , FA 22:0 , CE 20:5 , TG 43:0 , CE 20:5 were significantly downregulated (Fig 1, Table 1). These findings suggest that AhR activation may influence lipid metabolism, and this effect could be related to cancer-associated biological processes at the lipidomic level.

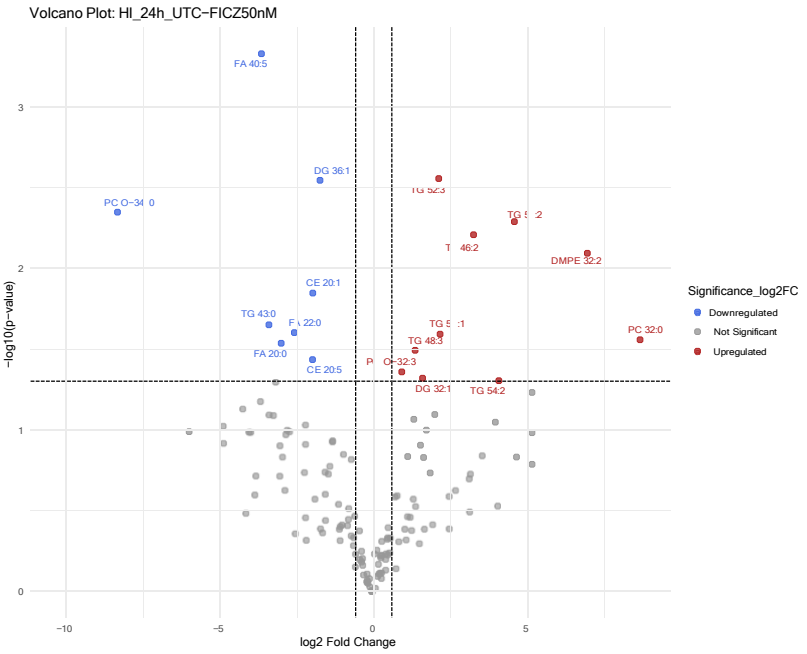


Fig. 1. Volcano plot showing differentially expressed metabolites in HeLa cells treated with 50 nM FICZ for 24 hours. Red points indicate significantly upregulated metabolites (fold change > 1.5 ($\log_2$FC = ±0.58), $p < 0.05$), and blue points represent significantly downregulated metabolites.

Table 1. Significantly altered lipids (FICZ 50 nM, 24-hour treatment vs. control).

Feature	log2FC	Fold Change	t_stat	p-value	Significance
FA 40:5	-3.6137	0.0817	15.7768	0.0005	Down
TG 52:3	2.1563	4.4576	-8.2255	0.0028	Up
DG 36:1	-1.7091	0.3059	6.5361	0.0028	Down
PC O-34:0	-8.2981	0.0032	14.8350	0.0045	Down
TG 55:2	4.5607	23.5998	-12.8852	0.0053	Up
TG 46:2	3.2312	9.3904	-6.4020	0.0064	Up
DMPE 32:2	6.9963	127.6729	-10.9942	0.0081	Up
CE 20:1	-1.9433	0.2600	4.3187	0.0142	Down
TG 43:0	-3.3723	0.0966	4.4296	0.0223	Down
FA 22:0	-2.5445	0.1714	5.4054	0.0250	Down
TG 50:1	2.2008	4.5975	-4.5188	0.0255	Up
PC 32:0	8.7091	418.4941	-5.8916	0.0276	Up
FA 20:0	-2.9723	0.1274	4.7531	0.0291	Down

TG 48:3	1.3905	2.6217	-3.6621	0.0321	Up
CE 20:5	-1.9517	0.2585	3.1640	0.0367	Down
PC O-32:3	0.9553	1.9389	-4.1876	0.0436	Up
DG 32:1	1.6318	3.0989	-3.4662	0.0477	Up
TG 54:2	4.1144	17.3205	-4.3194	0.0494	Up

4 Discussion

With the increasing availability of omics data, significant progress has been made in understanding biological processes at the cellular level. In our study, changes in lipid expression in HeLa cells were examined after activation of the AhR, and significant differences were observed in certain lipid classes. The decrease in fatty acids after AhR activation may lead to reduced energy production in cancer cells, which can suppress their proliferation. On the other hand, the increase in triglyceride levels may indicate that free fatty acids are converted into triglycerides for storage. Although our lipidomics analysis showed significant changes in certain lipid classes especially an increase in triglycerides following FICZ-induced AhR activation the exact biological mechanisms underlying these changes remain unclear. It is currently uncertain whether this increase reflects an adaptive cellular strategy to maintain lipid homeostasis or a pathological shift associated with disease progression. Further studies are required to validate and clarify the biological significance of these findings. In addition, the study was conducted using a single cell line (HeLa) and a single time point (24 hours), which may limit the generalizability of the findings. Although the findings suggest that AhR signaling functionally affects lipid balance, the molecular mechanisms behind these changes remain unclear. Therefore, future studies integrating transcriptomic and proteomic data will be important to uncover the regulatory genes and pathways involved in AhR-mediated lipid remodeling. Additionally, multi-omics approaches are needed to explore not only lipidomic changes but also the broader effects of AhR activation on inflammation, oxidative stress, and cellular differentiation. In this context, a systems biology-based approach will be essential to fully understand the context-specific roles of AhR in cancer biology. Therefore, to better understand the relationship between observed lipid changes and AhR signaling in the context of cancer, the integration of different types of data is necessary. Future studies may benefit from using different cell lines, time series, and the combination of various omics technologies. Such approaches could help to better understand the role of AhR in complex cellular processes.

References

1. Brown, J.S., Amend, S.R., Austin, R.H., Gatenby, R.A., Hammarlund, E.U., Pienta, K.J.: Updating the Definition of Cancer. Mol Cancer Res. 21, 1142–1147 (2023). https://doi.org/10.1158/1541-7786.MCR-23-0411.
2. Sondermann, N.C., Faßbender, S., Hartung, F., Hätälä, A.M., Rolfes, K.M., Vogel, C.F.A., Haarmann-Stemmann, T.: Functions of the aryl hydrocarbon receptor (AHR) beyond the canonical AHR/ARNT signaling pathway. Biochem Pharmacol. 208, 115371 (2023). https://doi.org/10.1016/j.bcp.2022.115371.

3. Mandal, A., Biswas, N., Alam, M.N.: Implications of xenobiotic-response element(s) and aryl hydrocarbon receptor in health and diseases. Hum Cell. 36, 1638–1655 (2023). https://doi.org/10.1007/s13577-023-00931-5.

4. Dai, X., Shen, L.: Advances and Trends in Omics Technology Development. Front Med (Lausanne). 9, 911861 (2022). https://doi.org/10.3389/fmed.2022.911861.

5. Wang, R., Li, B., Lam, S.M., Shui, G.: Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression. J Genet Genomics. 47, 69–83 (2020). https://doi.org/10.1016/j.jgg.2019.11.009.

6. Calder, P.C.: Functional Roles of Fatty Acids and Their Effects on Human Health. JPEN J Parenter Enteral Nutr. 39, 18S-32S (2015). https://doi.org/10.1177/0148607115595980.

7. Stone, S.J.: Mechanisms of intestinal triacylglycerol synthesis. Biochim Biophys Acta Mol Cell Biol Lipids. 1867, 159151 (2022). https://doi.org/10.1016/j.bbalip.2022.159151.

8. Barbernitz, X., Raben, D.M.: Phosphorylation of DGK. Adv Biol Regul. 88, 100941 (2023). https://doi.org/10.1016/j.jbior.2022.100941.

9. Gibellini, F., Smith, T.K.: The Kennedy pathway--De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. IUBMB Life. 62, 414–428 (2010). https://doi.org/10.1002/iub.337.

10. Kwarteng, D.O., Gangoda, M., Kooijman, E.E.: The effect of methylated phosphatidylethanolamine derivatives on the ionization properties of signaling phosphatidic acid. Biophys Chem. 296, 107005 (2023). https://doi.org/10.1016/j.bpc.2023.107005.

11. Gonen, A., Miller, Y.I.: From Inert Storage to Biological Activity-In Search of Identity for Oxidized Cholesteryl Esters. Front Endocrinol (Lausanne). 11, 602252 (2020). https://doi.org/10.3389/fendo.2020.602252.

12. Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., Higashi, Y., Okazaki, Y., Zhou, Z., Zhu, Z.-J., Koelmel, J., Cajka, T., Fiehn, O., Saito, K., Arita, M., Arita, M.: A lipidome atlas in MS-DIAL 4. Nat Biotechnol. 38, 1159–1163 (2020). https://doi.org/10.1038/s41587-020-0531-2.

Investigating Synthetic Lethal Interactions Using Machine Learning Approaches in Cancer Datasets

Mehmet Tevfik Kara [1] [0009-0009-2361-9049] and İdil Yet [2] [0000-0002-1227-0716]

[1] Bioinformatics Department, Graduate School of Health Sciences, Hacettepe University, 06100,
Ankara, Türkiye
idil.yet@hacettepe.edu.tr

Abstract. Cancer continues to pose a significant global health burden, with increasing incidence and mortality rates. Traditional treatment modalities such as surgical resection, radiotherapy, and chemotherapy often suffer from limitations, including non-specific toxicity, recurrence, and inaccessibility of tumor sites. Synthetic lethality (SL), defined as the concurrent inactivation of two genes leading to cell death, offers a targeted therapeutic avenue by exploiting genetic vulnerabilities in cancer cells. However, experimental identification of SL pairs remains costly and labor-intensive. To address this, computational approaches integrating biological date have gained prominence. The Early-Late Integrated Synthetic Lethality (ELISL) method leverages gene expression profiles and protein-protein interaction across tissue to predict SL interactions, even in context with limited data. This study employs the ELISL framework to identify candidate SL gene pairs in two common cancer (Papillary thyroid carcinoma and non-small lung cancer) and two rare cancers (acute myeloid leukemia and adrenocortical carcinoma).

Keywords: Cancer, Synthetic Lethality, Machine Learning.

1 Introduction

Cancer has become a major global health issue affecting millions of people in both developed and developing countries. According to World Health Organization (WHO), 20 million people were diagnosed with cancer in 2022, and 10 million died from it. These numbers are expected to double by 2050 [1] prompting national and international authorities to prioritize cancer treatment and invest heavily in related research. However, due to its complex and heterogenous nature, developing a definitive treatment for cancer remains challenge [2].

Various cancer treatment methods have been developed to date [3]. The primary method is surgical removes of the tumor. However, surgery can disrupt the structure of tissue and organs may not be prevent recurrence. Additionally, tumors located in risky anatomical regions can make surgery unfeasible. In such cases systemic and localized treatments like chemotherapy and radiotherapy are used. These aim to destroy cancer-

ous tissue without surgery but often damage healthy tissue as well, leading to side effects. Targeted therapies, which deliver drugs to specific molecular targets, can reduce these side effects. However, mutations in the target area may prevent the drug from being effective [4].

To overcome these limitations, an alternative approach has been developed to targets synthetic lethal (SL) gene pairs. A gene pair is defined as SL when the loss of function in both genes results in cell death, while the loss in only one does not [5]. The main challenge is detecting SL pairs. Since its not affordable that try every gene pair with laboratory experiments, computational approaches have been developed to prioritize potential SL pairs for experimental validation.

In common cancers like breast or lung cancer, most of the computational approaches effective but these methods are limited by data scarcity in rare cancer types, impacting model performance. The ELISL (Early-Late Integrated Synthetic Lethality) [6] approach addresses this issue by allowing gene data from unrelated tissues and incorporating protein sequence homology and protein-protein interaction data. This enables tissue-independent relationships to be identified and combined with tissue specific genes to predict SL pairs more effectively.

In this study, ELISL was used to predict SL interactions across two common cancers (papillary thyroid carcinoma and non-small cell cancer) and two are rare cancer (acute myeloid leukemia and adrenocortical carcinoma).

2 Methods

2.1 Amino acid Sequences

Protein sequences specific to *Homo Sapiens* were retrieved from the Uniprot [7] database and subsequently vectorized using the SeqVec [8] pretrained embedding model. Each protein sequence was represented as a 1024D embedding vector. For every gene pair, the absolute difference between their corresponding vectors was computed to serve as a feature.

2.2 Protein-Protein Interactions

Homo Sapiens protein-protein interaction data were obtained from the STRING [9] database. These interactions were vectorized using the PecanPy [10] embedding method, yielding 64D vectors. For each gene pair, the absolute difference between the corresponding vectors was calculated to capture interaction-based features

2.3 Cell Line Omics

Relevant cancer types were selected and filtered from the Broad 2019 dataset available on cBioportal [11]. Log transformed mRNA z-scores (RNA-seq RPKM) and copy number alterations were utilized to construct cell line omics feature.

Table 1. TCGA Code, Cancer Type, and Sample Count from Broad 2019 Dataset.

TCGA Code	Cancer Types	Samples
LUAD	Lung Adenocarcinoma	79
LAML	Acute Myeloid Leukemia	35
THCA	Thyroid Carcinoma	1
ACC	Adrenocortical Carcinoma	1

2.4 Tissue Omics

Cancer types were determined using selected studies from cBioportal, serving as the source repository. Log transformed mRNA z-scores relative to all samples (RNA-seq RPKM and RNA-seq RSEM) were used to detect and quantify transcriptomic alterations across tissue.

Table 2. TCGA Code, Cancer Type, and Sample Counts from TCGA Firehouse Legacy.

TCGA Code	Cancer Types	Samples
LUAD	Lung Adenocarcinoma	230
LAML	Acute Myeloid Leukemia	163
THCA	Thyroid Carcinoma	313
ACC	Adrenocortical Carcinoma	75

2.5 Building Machine Learning Model

Random Forest [12], Gradient Boosting [13], and XGBoost [14] algorithms were employed to construct the predictive models. Gene pairs were randomly split into training (%80) and testing (%20) sets. For each feature set, a probability score ranging between 0 and 1 was generated. These individual feature scores were subsequently aggregated to obtain a composite score for each gene pair.

3 Results

End of this study known synthetic lethal gene pairs were tested in selected cancer types. Three machine learning models were used. Then, based on the probability scores of known SL interactions, gene pairs were classified as synthetic lethal (SL Score > 0.6) or not (SL Score < 0.4). The hypothesis that the integration of tissue-specific and tissue-independent features would enhance synthetic lethality prediction in rare cancer types was supported by ELISL's performance in selected cancer datasets. This suggested that ELISL, using different machine learning approaches, has the potential to guide future SL-targeted therapies, especially in rare cancers.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024 May-Jun;74(3):229-263. doi: 10.3322/caac.21834. Epub 2024 Apr 4. PMID: 38572751.
2. Hanahan, D. (2022). Hallmarks of cancer: new dimensions. Cancer Discovery, 12(1), 31-46
3. Debela, D, T., Muzazu, S. G., Heraro, K. D., Ndalama, M. T., Mesele, B. W., Haile, D. C., … & Manyazewal, T. (2021). New approaches and procedures for cancer treatment: Current perspective. SAGE open medicine, 9, 20503121211034366.
4. Holohan, C., Van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. Nature Reviews Cancer, 13(10), 714-726.
5. Chan, D. A., & Giaccia, A. J. (2011). Harnessing synthetic lethal interactions in anticancer drug discovery. Nature reviews Drug discovery, 10(5), 351-364
6. Tepeli, Y. I., Seale, C., & Gonçalves, J. P. (2024). ELISL: early–late integrated synthetic lethality prediction in cancer. *Bioinformatics*, *40*(1), btad764.
7. UniProt: the Universal protein knowledgebase in 2025. *Nucleic Acids Research*, 2025, 53.D1: D609-D617.
8. Heinzinger, M., Elnaggar, A., Wang, Y., Dallago, C., Nechaev, D., Matthes, F., & Rost, B. (2019). Modeling aspects of the language of life through transfer-learning protein sequences. *BMC bioinformatics*, *20*, 1-17.
9. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., ... & Von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, *51*(D1), D638-D646.
10. Liu, R., & Krishnan, A. (2021). PecanPy: a fast, efficient and parallelized Python implementation of node2vec. *Bioinformatics*, *37*(19), 3377-3379.
11. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095
12. Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.
13. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
14. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

Prediction of Metastasis Risk Using Machine Learning Based on BRAF Mutation and miRNA Expression in Thyroid Cancer

Mehmet Erşan KALAYCI[1[0000-0002-4797-9020]], Kemal TURHAN[2[0000-0001-7871-3025]]

[1] Karadeniz Teknik Üniversitesi, Trabzon, Türkiye,
[2] İzmir Ekonomi Üniversitesi, İzmir, Türkiye,
mehmetersankalayci@ktu.edu.tr
kturhan.tr@gmail.com

Abstract. Accurate prediction of metastasis in thyroid cancer is critical for determining prognosis and developing effective treatment strategies. In this study, we aimed to predict metastasis status (M1 vs. M0) using clinical data, miRNA expression profiles, and BRAF mutation status. Mutation data, miRNA-seq, and clinical information were obtained and integrated from The Cancer Genome Atlas (TCGA) database. Differential expression analysis was performed to identify miRNAs significantly associated with metastatic and non-metastatic groups. Machine learning models, including Random Forest, XGBoost, and Decision Tree, were trained using the selected features. Among the models, XGBoost demonstrated the best performance with an ROC AUC of 0.993, a recall of 0.98, and an F1-score of 0.96. These findings suggest that miRNAs have strong potential as biomarkers for metastasis and that predictive models developed through integrated omics analysis can be utilized in clinical decision support systems. Early prediction of the risk of metastasis will provide a critical advantage for early diagnosis and timely intervention by allowing close follow-up of patients.

Keywords: Machine learning, metastasis risk, miRNA expression, thyroid cancer,

Predicting Type 1 Diabetes from Gut Microbiome Data: A Machine Learning Approach Combining Alpha/Beta Diversity and PICRUSt2-Inferred Metabolic Pathways

Açelya Dalgıç[1,2] and İdil Yet[1]

[1] Department of Bioinformatics, Hacettepe University, Ankara, Turkey 06100
[2] Department of AI Engineering, Hacettepe University, Ankara, Turkey 06100
idilyet@gmail.com

Abstract. Microbiome play essential roles in host homeostasis, influencing metabolism, immune function, and defense against pathogens, and are increasingly linked to the onset and progression of various diseases in human body (1). Microbiome research often uses alpha and beta diversity metrics to assess microbial composition-structure (2) and PICRUSt2 to predict functional pathway abundances within microbiome samples by leveraging 16S rRNA gene sequence data and reference genome databases to infer metabolic potential and molecular functions. (3). This study has two aims: (1) to compare gut microbiomes of Italian and Chinese cohorts (T1DM vs. healthy controls) and identify population- and disease-specific differences in microbial diversity (alpha/beta) and functional pathways as potential T1DM biomarkers; and (2) to evaluate these ecological metrics as machine learning (ML) input features for disease prediction. Integrating cross-population analysis with ML revealed distinct microbiome diversity and functional profiles between groups, demonstrating their predictive value for T1DM classification, even across ethnically diverse cohorts. As a result, we found clear differences in gut microbiome diversity and functional pathways between T1DM-control and Italy-China and we found that these metrics hold value in ML prediction of T1DM.

Keywords: Microbiome, Machine Learning, Microbial Ecology

1 Introduction

The term microbiota refers to the community of microorganisms inhabiting specific regions of the body, while the microbiome encompasses this microbial community along with its collective genetic material. Microbiomes exist across diverse ecosystems, from plants to animals, and are recognized for their dynamic and symbiotic relationships with their hosts. The microbiome can comprise a diverse range of microorganisms, including viruses, bacteria, archaea and protists (1). In humans, microbiomes are present in multiple anatomical sites, including the oral cavity, gastrointestinal tract, skin, and the vaginal environment. Studies have demonstrated that the human microbiome plays a vital role in maintaining host homeostasis by contributing to metabolic functions, immune defense, and protection against pathogens, as well as being implicated in the development and modulation of various diseases. The gut microbiome exhibits significant variation across human populations, influenced by factors such as geography, diet, ethnicity, and lifestyle. The Human Microbiome Project (HMP) revealed that while a core set of microbial functions is conserved across individuals, taxonomic composition can differ markedly between populations (4) . Microbiome studies employ alpha and beta diversity measures to characterize microbial communities at different ecological scales. Alpha diversity quantifies within-sample diversity using several key metrics such as The Shannon-Wiener index (5) which incorporates both species richness and evenness, is particularly sensitive to rare taxa and widely used in clinical microbiome studies and The Simpson index (5) emphasizes dominant species and provides complementary information about community stability. Beta diversity analysis compares microbial composition between samples through distance metrics such as Bray-Curtis dissimilarity (6) and Jaccard distance, a presence-absence metric, is useful for detecting community composition and similarity (6) . Another key analysis in microbiome studies is PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) which is used for the prediction of functional potential from 16S rRNA gene sequencing data. By leveraging phylogenetic placement and curated reference genomes, it infers the abundance of metabolic pathways and molecular functions (e.g., carbohydrate degradation, vitamin synthesis, or antibiotic resistance genes) that are likely present in a given

microbiome sample. While powerful for hypothesis generation, PICRUSt2 predictions are limited by the completeness of reference databases and the accuracy of taxonomic assignments, as they rely on evolutionary relationships rather than direct metagenomic sequencing [3] . Studies have shown that there are patterns in case and control groups of various diseases such as reduced alpha diversity characterizes dysbiotic states in obesity [7] while beta diversity calculations can help to distinguish disease states [8] . Modern pipelines like QIIME2 [9] implement these metrics while accounting for phylogenetic relationships and sequencing depth biases.

Machine learning models are widely employed in microbiome studies to analyze complex microbial datasets. Artificial Neural Networks (ANNs), inspired by the structure of biological neurons, consist of interconnected nodes with specific weights and are used for tasks such as image recognition and prediction. As described by researchers, these networks typically comprise an input layer, hidden layers for computations, and an output layer for classification or regression [10] . Support Vector Machines (SVMs), another important model, aim to identify the optimal hyperplane that separates different classes in multidimensional data, making them particularly useful for classification tasks [11] . Random Forests, developed as an ensemble method, mitigate overfitting by aggregating results from multiple decision trees while providing feature importance metrics [12] .

Model performance is evaluated using key metrics including accuracy, precision, recall, and AUC-ROC curves. SHAP (SHapley Additive exPlanations) values, introduced by Lundberg and Lee in their 2017 paper "A Unified Approach to Interpreting Model Predictions", provide a robust framework for quantifying feature importance. These evaluation methods, when combined, enable comprehensive assessment of model performance while accounting for different aspects of predictive accuracy [13] .

Type 1 Diabetes Mellitus (T1DM) is a form of diabetes characterized by immune-mediated destruction of pancreatic beta cells, resulting in insulin deficiency. It is most commonly diagnosed in children and adolescents, which is why it is also referred to as “juvenile diabetes.” T1DM accounts for approximately 10% of all diabetes cases. While its prevalence among children varies across populations, it is most frequently observed in Northern countries. Although T1DM is one of the most common chronic diseases in children, it can also occur in adults. First-degree relatives of T1DM patients are considered to be at increased risk. The progression of the disease typically occurs in three stages: an initial asymptomatic phase marked by the onset of immune-mediated pancreatic damage, a second phase characterized by beta cell dysfunction and dysglycemia, and a third phase defined by near-total beta cell failure and symptomatic hyperglycemia. Patients rely on insulin administration for survival [14] .

Type 1 Diabetes Mellitus (T1DM), which directly affects nutritional status, has also been shown to exert a significant influence on the gut microbiota. Studies have observed that individuals with T1DM exhibit an altered gut microbial composition compared to healthy individuals, characterized by reduced alpha diversity. Moreover, this dysbiosis has been associated with increased intestinal permeability, commonly referred to as Leaky Gut Syndrome. Changes in gut barrier function are thought to facilitate the translocation of infectious agents and other factors via the bloodstream to pancreatic lymph nodes, thereby triggering autoimmune responses associated with T1DM. Several studies have highlighted this link between gut microbiota and the pathogenesis of T1DM. For instance, beneficial bacterial species such as Akkermansia muciniphila and Faecalibacterium prausnitzii have been found in lower abundance in T1DM patients. In contrast, an increase in the abundance of Clostridium perfringens has been observed, which is associated with compromised gut integrity. Additionally, the ratio of the phylum Bacteroidetes to Firmicutes has been reported to increase both prior to and at the time of T1DM diagnosis [15] .

This study has two primary objectives: first, to conduct a cross-population case-control analysis of gut microbiome datasets from Italian and Chinese cohorts with Type 1 Diabetes Mellitus (T1DM) and healthy controls, in order to investigate how population-specific (ethnicity/diet/environment) and disease-associated (T1DM vs. healthy) differences manifest in microbial alpha/beta diversity and functional pathways, thereby assessing their potential as predictive biomarkers or mechanistic contributors to T1DM pathogenesis and second, to rigorously evaluate the utility of ecological diversity measures (such as alpha and beta diversity) as input features for machine learning models in disease prediction, motivated by the observed differences in functional

pathways and microbial community diversity between healthy and diseased states. By leveraging these ecological metrics, we aim to enhance predictive accuracy and uncover biologically meaningful patterns in disease mechanisms. The Italian and Chinese populations were chosen to represent different dietary patterns (Mediterranean vs. East Asian diets), and environmental exposures (urbanization, hygiene standards), which are known to shape gut microbiome composition and its functional role in autoimmune disease susceptibility.

1 METHOD

1.1 Data Selection and Information

This study leverages publicly available 16S rRNA gene sequencing data (V3-V4 region, Illumina) from three open-access projects to investigate gut microbiome signatures in type 1 diabetes mellitus (T1DM). Data were derived from fecal samples of patients and healthy controls across two Italian cohorts (PRJNA702261, n=52; PRJNA280490, n=107) [16] and one Chinese cohort (PRJNA766410, n=83) [17] .

1.2 Data Processing

Data quality control was performed using FastQC. Subsequent data processing was conducted with QIIME2. Sequencing reads were filtered based on a minimum Phred quality score of 30 (Q30), and denoising as well as amplicon sequence variant (ASV) extraction were carried out using the Deblur pipeline, with a trimming length of 250 base pairs applied uniformly across all datasets. Taxonomic classification was performed using a QIIME2 Naive Bayes classifier trained on the SILVA database, specifically targeting the V3–V4 region of the 16S rRNA gene.

1.3 Calculating the Input Features

Ecological diversity measures, including the Shannon–Wiener Index, Simpson Index, and Bray–Curtis dissimilarity-based Principal Coordinate Analysis (PCoA), were calculated at three taxonomic levels—Order, Family, and Genus—using the vegan package in R. Functional profiling was performed using PICRUSt2 via the QIIME2 plugin q2-picrust2 [3]. The resulting predicted metabolic pathways were further analyzed to identify those significantly associated with T1DM versus control groups, employing the LINDA statistical test within the ggpicrust2 R package.

1.4 Selecting Other Input Features

Age, sex, and country information of samples are given to the models as input features. Given the established higher prevalence of type 1 diabetes mellitus (T1DM) in childhood and adolescence, and the non-normal distribution of age across the dataset, age data for individuals under 18 were categorized in greater detail, considering the possible microbiome differences between the phase of the disease. We avoided including any information about which dataset each sample came from in our models due to inequal distribution of groups in Italy datasets.

1.5 Selecting the Models and Measuring Model Performances

Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) models were employed for classification tasks, with 25% of the data reserved as the test set. Model validation was conducted using 10-fold cross-validation (K=10). Performance metrics including F1-score, precision, recall, accuracy, and area under the receiver operating characteristic curve (AUC) were calculated to evaluate model efficacy. SHAP (SHapley Additive exPlanations) values were computed for each model to assess feature importance and visualized accordingly. The analyses were implemented using Python libraries such as scikit-learn, pandas, numpy, shap, and matplotlib.

2 RESULTS

In this study, alpha diversity indices (Shannon-Wiener and Simpson) were used as input features for the model. This approach aimed to eliminate taxonomic variations arising from confounding factors such as age and population differences across datasets, which could otherwise hinder prediction accuracy. Additionally, it addressed the high-dimensional structure of taxonomic tables, which may adversely affect model performance. The three taxonomic levels (order, family, and genus) yielded distinct relative results for alpha diversity indices. Shannon-Wiener and Simpson diversity indices exhibited statistically significant differences (ANOVA, $p < 0.05$) across three taxonomic levels (order, family, genus) when comparing disease states (T1DM vs. healthy controls) and populations (Italian vs. Chinese cohorts), indicating the influence of both dysbiosis and ethnic/environmental factors on microbial community structure.

Furthermore, Principal Coordinates Analysis (PCoA) of Bray-Curtis dissimilarity revealed differential clustering of samples across taxonomic resolutions, with finer scales (e.g., genus level) exhibiting clearer separation between datasets. Given that PCoA components represent the relative positioning of samples, they were deemed suitable as model input features. By incorporating diversity metrics calculated at order, family, and genus levels, the models were equipped with comprehensive ecological information to maximize predictive insight.

Pathway analysis revealed 168 pathways differentiating case and control groups, and 273 pathways distinguishing between two datasets. Comparative pathway analysis revealed statistically significant differences ($p.adj < 0.05$) between Italian and Chinese cohorts in 271 MetaCyc metabolic pathways, including key processes such as aromatic compound degradation (3-HYDROXYPHENYLACETATE-DEGRADATION-PWY), iron acquisition (AEROBACTINSYN-PWY), amino acid metabolism (ARG+POLYAMINE-SYN, ORNDEG-PWY), and central carbon metabolism (TCA, GLYCOLYSIS-TCA-GLYOX-BYPASS), highlighting population-specific microbial metabolic adaptations potentially linked to genetic, dietary, or environmental factors. Among these, 19 pathways were consistently associated with case-control status while remaining insignificant for geographic origin, including the superpathway of chorismate metabolism ("ALL-CHORISMATE-PWY"), amino acid degradation ("AST-PWY," "LEU-DEG2-PWY"), secondary metabolite biosynthesis ("ENTBACSYN-PWY"), and fucose/rhamnose degradation ("FUC-RHAMCAT-PWY"). Additionally, key metabolic processes such as the TCA cycle, respiration, and photosynthesis were implicated. These findings highlight both shared disease-associated pathways and region-specific metabolic variations, suggesting distinct yet overlapping biological mechanisms underlying the observed differences.

ANN, SVM, and RF models achieved moderate overall performance (AUC 0.70–0.77). SHAP analysis of SVM model revealed that country-specific differences (Italy/China), microbial energy metabolism (photosynthesis/respiration), and amino acid degradation pathways were the strongest drivers of predictions. We also found that both alpha diversity and beta diversity contribute meaningfully, but in distinct ways. Alpha diversity metrics like the Shannon and Simpson indices which are calculated at different taxonomic levels (genus, family, order) show moderate importance. However, beta diversity metrics, the PCoA coordinates of Bray Curtis Dissimilarity are found as stronger predictors with the importance order of with genus-level > family-level > order-level.

In this study, alpha diversity indices (Shannon-Wiener and Simpson) were employed as model input features to mitigate confounding effects from taxonomic variations (e.g., age, population differences) and high-dimensionality challenges in raw taxonomic tables. Distinct patterns emerged across three taxonomic levels (order, family, genus) for both alpha and beta diversity, while PICRUSt2-derived metabolic pathways (e.g., amino acid degradation, photosynthesis) complemented these metrics. Although ANN, SVM, and RF models achieved moderate performance (AUC 0.70–0.77) using these features, SHAP analysis highlighted country-specific differences (Italy/China) and microbial energy metabolism as key predictive drivers. To further refine these population-agnostic biomarkers, future work will apply PERMANOVA, ANOSIM, and differential abundance analyses to both taxonomic profiles and PICRUSt2-predicted pathways, with particular focus on identifying universally informative features of T1DM. We are also evaluating the implementation of alternative models.

References

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. Aralık 2015;3(1):31, s40168-015-0094-5.
2. Cassol I, Ibañez M, Bustamante JP. Key features and guidelines for the application of microbial alpha diversity metrics. Sci Rep. 03 Ocak 2025;15(1):622.
3. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, vd. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. Haziran 2020;38(6):685-8.
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. Haziran 2012;486(7402):207-14.
5. Nagendra H. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. Appl Geogr. Nisan 2002;22(2):175-86.
6. Ricotta C, Pavoine S. A new look at functional beta diversity. Ecol Indic. Haziran 2024;163:112136.
7. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. Aralık 2006;444(7122):1027-31.
8. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun. 05 Aralık 2017;8(1):1784.
9. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, vd. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. Ağustos 2019;37(8):852-7.
10. Grossi E, Buscema M. Introduction to artificial neural networks: Eur J Gastroenterol Hepatol. Aralık 2007;19(12):1046-54.
11. Cortes C, Vapnik V. Support-vector networks. Mach Learn. Eylül 1995;20(3):273-97.
12. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
13. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions.
14. Katsarou A, Gudbjörnsdottir S, Rawshani A, Dabelea D, Bonifacio E, Anderson BJ, vd. Type 1 diabetes mellitus. Nat Rev Dis Primer. 30 Mart 2017;3(1):17016.
15. Zheng P, Li Z, Zhou Z. Gut microbiome in type 1 diabetes: A comprehensive review. Diabetes Metab Res Rev. Ekim 2018;34(7):e3043.
16. Del Chierico F, Conta G, Matteoli MC, Fierabracci A, Reddel S, Macari G, vd. Gut Microbiota Functional Traits, Blood pH, and Anti-GAD Antibodies Concur in the Clinical Characterization of T1D at Onset. Int J Mol Sci. 06 Eylül 2022;23(18):10256.
17. Tan H, Shi Y, Yue T, Zheng D, Luo S, Weng J, vd. Machine learning approach reveals microbiome, metabolome, and lipidome profiles in type 1 diabetes. J Adv Res. Ekim 2024;64:213-21.

Retrosentez İçin Makine Öğrenimi Kullanımı

Güven ŞAHİN[1][0000-0003-3901-5819], Mehmet Erşan KALAYCI[2][0000-0002-4797-9020], Kemal TURHAN[3][0000-0001-7871-3025]

[1] Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye, [2] Karadeniz Teknik Üniversitesi, Trabzon, Türkiye, [3] İzmir Ekonomi Üniversitesi, İzmir, Türkiye

guven.sahin@erdogan.edu.tr, mehmetersankalayci@gmail.com,kemalturhan@ieu.edu.tr

GENİŞLETİLMİŞ ÖZET

Literatür Araştırması: Birçok hastalığın etkin tedavisinde farklı içeriklere sahip ürünlere ihtiyaç duyulmaktadır. Bu ürünlerin elde edilmesinde en sık tercih edilen yöntemlerden biri organik sentezdir [1]. Uzun bir geçmişe sahip olan organik sentez, yeni ilaçların ve malzemelerin keşfi konusunda hâlen beklenen hız ve verimliliği sağlayamamaktadır [2]. Hedef ürüne ulaşacak sentetik rotaların belirlenmesi sürecinde, büyük ölçüde uzman bilgi ve deneyimine ihtiyaç duyulmaktadır [3]. Durumun böyle olması keşif sürecinin daha fazla zaman almasına ve bu doğrultuda yapılan harcamaların da artmasına neden olmaktadır. İlaç üretimi için yapılacak harcamaları optimize etmesiyle ulusal ekonomiye olumlu yönde katkı sağlaması, üretim sürecini hızlandırarak hastaların ilaca erişim süresini kısaltması sebebiyle de toplumsal refaha katkı sağlayacağı düşünülmektedir. Belirtilen katkıları sağlaması için problem durumunun çözümüne yönelik kullanılan teknolojilerden biri olan makine öğrenimi teknolojileri sentetik rotaların keşfi için önem arz etmektedir. Yeni moleküllerin sentezinde kritik aşamalardan biri, sağlam bir şekilde planlanmış ve pratikte uygulanabilir retrosentetik rotaların oluşturulmasıdır [4]. Retrosentezin amacı, hedef molekülü daha basit başlangıç maddelerine indirgemek için uygun sentetik yollar geliştirmektir [5]. Literatür incelemesi sonucunda, makine öğrenimi tabanlı retrosentez yaklaşımlarının gelişim sürecinin henüz başlangıç safhasında olduğu gözlemlenmiştir. Bu doğrultu da yapılan çalışmada, kimyagerleri hedef ürüne ulaştıracak daha iyi sentez rotalarının daha hızlı bulunmasına yardımcı olacağı düşünülmektedir.

Çalışmanın amacı: Bu çalışmanın amacı ReactionT5V2-retrosynthesis modeline ilaç moleküllerinin reaktant tahmini için ince ayar yapmaktır.

Yöntem: ReactionT5 modeli ZINC ve Open Reaction Database (ORD) veri tabanındaki verilerle ön eğitilmiştir. Araştırmacılar kendi veri setleriyle ince ayar yapabilmektedir. Literatürdeki çalışmalara bakıldığında genellikle model USPTO veri setiyle eğitildikten sonra tekrar aynı veri setiyle top-k performansına bakılmıştır [6,7,8]. ReactionT5 modeli kapsamlı bir şekilde ön eğitilmiş olması, tahmin performansının başarılı olması ve ince ayar yapılabilmesinden dolayı ilaç moleküllerinin reaktant tahmininde tercih edilmiştir. Seq2Seq tabanlı ReactionT5 modeli iki aşamalı bir ön eğitim süreci ile geliştirilmiştir. İlk aşama olan Bileşen Ön Eğitimi'nde model, ZINC veri tabanından 23 milyon bileşik üzerinde

eğitim almıştır. Bu aşamada model, bileşenlerin temsil edilmesi ve işlenmesi için gerekli temel bilgileri öğrenmektedir. Aşamada sonunda elde edilen model "CompoundT5" olarak adlandırılır. İkinci aşama olan Reaksiyon Ön Eğitimi'nde CompoundT5 modeli, reaksiyon veritabanı olan Open Reaction Database (ORD) yer alan 1,5 milyon reaksiyon kullanılarak birden fazla bileşiği içeren kimyasal reaksiyon verilerini işlemiştir. Bu eğitim, modelin birden fazla molekül dahil olduğunda nasıl tepki vereceğini öğrenmesini sağlamıştır. Bu aşama sonunda elde edilen model ReactionT5 olarak adlandırılmıştır [9]. Model, kimyasal reaksiyonları ve bileşikleri metin tabanlı temsil etmek için SentencePiece unigram tokenizer kullanmaktadır. Bu tokenizer, kimyasal bileşikleri ve reaksiyonları küçük ve etkili token'lara bölmek amacıyla kullanılmıştır. Bu model, sentezin ürün veya verimini tahmin etmek için kullanılmaktadır. Modelin ikinci versiyonu olan "ReactionT5v2-retrosynthesis" ile reaktant tahmini görevleri için ince ayar yapılabilmektedir. 1,5 milyon ORD veri setiyle eğitilmiş ve USPTO veri seti veri sızıntısını önlemek için kullanılmıştır. Bu çalışmada modelin belirtilen ikinci versiyonu üzerinde ince ayar yapılmıştır.

Model için ince ayar yapılırken kullanılan Hiper parametreler,

```
"Python","finetune.py",
"--model_name_or_path=sagawa/ReactionT5v2-retrosynthesis",
"--epochs=20",
"--batch_size=32",
"--eval_beams=10",
"--train_data_path=../data/train.csv",
"--valid_data_path=../data/validation.csv",
"--output_dir=output"
```

Bu çalışmada Reaxy veri tabanından üç farklı hastalığa ait ilaçlardan elde edilen 500 sentez verisi ve USPTO-50K [10] veri setindeki 9 reaksiyon sınıfından elde edilen 2250 sentez verisi kullanılmıştır. 9 reaksiyon sınıfının her birinden 250 veri olacak şekilde toplamda 2750 sentez verisi çalışmanın veri seti olarak kullanılmıştır. Elde edilen bu verilere ön işlem uygulanmış ve Kemoinformatik için açık kaynak araç seti olan RDKit kütüphanesi kullanılarak sentezdeki tüm moleküllerin SMILES bilgileri kanonik hale getirilmiştir. İnce ayar için hazır hale getirilen veri seti %80 %10 %10 oranlarında eğitim, doğrulama ve test için ayrılmıştır. Modelin reaktant tahminindeki performansı için top-k (1,3,5,10) accuracy metriğine bakılmıştır. Eğitim ve performans testlerinde Türk Ulusal Bilim e-altyapısı (TRUBA) kullanılmıştır. İnce ayar sonucunda elde edilen model DRPred (DrugRetroPrediction) olarak adlandırılmıştır.

Tablo 1. Model (DRPred) Tahminleri

Model Girdisi (Ürün)	**Durum**
CN(C)S(=O)(=O)Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	
Model Çıktısı (Reaktant)	
CN(C)S(=O)(=O)Cl.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	DOĞRU
CN(C)S(=O)(=O)Cl.Cc1ccccc1.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	YANLIŞ
CN(C)S(=O)(=O)Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	YANLIŞ
CN(C)S(=O)(=O)Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1.OB(O)C=C1c2ccccc2CCc2ccccc21	YANLIŞ
CN(C)S(=O)(=O)Nc1cccc(B(O)O)c1.O=S(=O)(Oc1cccc(C=C2c3ccccc3CCc3ccccc32)c1)C(F)(F)F	YANLIŞ
CN(C)S(=O)(=O)Oc1cccc(C=C2c3ccccc3CCc3ccccc32)c1.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	YANLIŞ
CN(C)S(=O)(=O)Cl.Cc1cc(C)nc(C)c1.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	YANLIŞ
CN(C)S(=O)(=O)Cl.ClCCl.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1	YANLIŞ
CN(C)S(=O)(=O)Cl.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c1.c1ccncc1	YANLIŞ
CN(C)S(=O)(=O)Cl.ClC(Cl)Cl.Nc1cccc(C=C2c3ccccc3CCc3ccccc32)c	YANLIŞ

Tablo 1'de modelin yapmış olduğu tahmin sonuçlarından bir örnek verilmiştir. Örnekte, kimyasal bir sentezin ürünü SMILES formatında model girdisi olarak verilmiştir. Bu ürünü oluşturabilecek reaktantlar SMILES formatında modelin çıktısı olarak dönmektedir. Reaktantlar birbirinden nokta karakteri ile ayrılmaktadır. Her bir örnek için 1 ürüne karşılık 10 tahmin sonucu üretilmektedir.

Sonuç ve Değerlendirme: ReactionT5v2-retrosynthesis modeline ince ayar yapıldıktan sonra elde edilen DRPred modelinin reaktant tahmininde umut verici olduğu görülmüştür. Top-k accuracy metriği ile ilgili değerler Tablo 2'de verilmiştir.

Tablo 2. Top-k Accuracy (%)

Model	**Top-1**	**Top-3**	**Top-5**	**Top-10**
NAG2G	55.1	76.9	83.4	89.9
SemiRetro	54.9	75.3	80.4	84.1
MARS	54.6	76.4	83.3	88.5
RetroComposer	54.4	77.2	83.2	87.7
O-GNN	54.1	77.7	86.0	92.5
G2GT	54.1	69.9	74.5	77.7
DRPred (Bu çalışma)	56.0	66.0	68.0	72.0

Tablo 2'deki sonuçlar, tek adımlı reaktant tahmininde reaksiyon sınıfı bilinmediği durumda modellerin ulaştığı top-k performanslarıdır. Bu sonuca göre, modelin top-1 accuracy değeri %56 olarak tespit edilmiş ve model yapmış olduğu ilk tahminde doğru reaktantları bu oranda bulmuştur. Modelin yapmış olduğu en iyi üç tahmin içinde %66'lık Top-3 accuracy değeri doğru sonucu bulduğunu

göstermektedir. %68 oranıyla top-5 accuracy değeri modelin yapmış olduğu en iyi beş tahmin içinde doğru reaktantları bulduğunu, %72 oranıyla top-10 accuracy değeri modelin yapmış olduğu en iyi 10 tahmin içinde doğru reaktantları bulduğunu göstermektedir. Sonuç olarak ince ayar yapılarak elde edilen DRPred modelinin top-1 değeri diğer modellerle çok yakın olduğu görülürken top (3,5,10) değerlerinde geride kaldığı görülmüştür. Bu durumun, ince ayar yapılan veri sayısının kısıtlı olduğundan dolayı oluştuğu düşünülmektedir. Tabloda belirtilen diğer modeller 50 bin reaksiyon verisiyle eğitilmiştir. Model hedef moleküllerin reaktantlarını en yüksek olarak %72 oranında doğru bir şekilde tahmin etmiştir. Gelecek çalışmalarda daha yüksek sentez verisi kullanılarak ve ilaç sentezleri üzerine özelleştirilerek daha iyi top-k accuracy değerleri elde edilmesi, modelin nasıl öğrendiğini ve tahminde bulunduğunu farklı parametrelerle gösteren açıklanabilir yapay zekâ teknikleri kullanılması planlanmaktadır.

Anahtar Kelimeler: Makine Öğrenimi, Sentez Planı, Retrosentez

Kaynakça

1. Blakemore, D. C., Castro, L., Churcher, I., Rees, D. C., Thomas, A. W., Wilson, D. M., & Wood, A. (2018). Organic synthesis provides oppor- tunities to transform drug discovery. Nature chemistry, 10(4), 383- 394.
2. D. C. Blakemore , L. Castro , I. Churcher , D. C. Rees , A. W. Tho- mas , D. M. Wilson and A. Wood , *Nat. Chem.*, 2018, **10**, 383 —394
3. Guo, Z., Wu, S., Ohno, M., & Yoshida, R. (2020). Bayesian algorithm for retrosynthesis. Journal of Chemical Information and Modeling, 60(10), 4474-4486.
4. Lin, K., Xu, Y., Pei, J., & Lai, L. (2020). Automatic retrosynthetic route planning using template-free models. Chemical science, 11(12), 3355-3364.
5. Ishida, S., Terayama, K., Kojima, R., Takasu, K., & Okuno, Y. (2022). Ai- driven synthetic route design incorporated with retrosynthesis knowledge. Journal of chemical information and modeling, 62(6), 1357-1367.
6. Yao, L., Guo, W., Wang, Z., Xiang, S., Liu, W., & Ke, G. (2024). Node- aligned graph-to-graph: elevating template-free deep learning ap- proaches in single-step retrosynthesis, 4(3), 992-1003.
7. Gao, Z., Tan, C., Wu, L., & Li, S. Z. (2022). SemiRetro: Semi-template framework boosts deep retrosynthesis prediction. arXiv preprint arXiv:2202.08205.
8. Liu, J., Yan, C., Yu, Y., Lu, C., Huang, J., Ou-Yang, L., & Zhao, P. MARS: A Motif-based Autoregressive Model for Retrosynthesis Prediction. arXiv 2022. arXiv preprint arXiv:2209.13178.
9. Sagawa, T., & Kojima, R. (2023). ReactionT5: a large-scale pre-trained model towards application of limited reaction data. arXiv preprint arXiv:2311.06708.
10. Therapeutics Data Commons. Retrieved May 1, 2025, from https://tdcommons.ai/

FonokardiyogramVerileri ile Üfürüm Teşhisi: Derin Öğrenme Tabanlı Bir Yaklaşım

Yasin Görmez[1] [0000-0001-8276-2030], Fatma Hilal Yağın[2][0000-0002-9848-7958], **Abdulvahap Pınar**[3][0000-0002-3662-2579]

[1]Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Sivas Cumhuriyet University, Sivas 58140, Turkey

[2]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Ozal University, Malatya, Turkey.

[3]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye

ÖZET

Amaç: Bu çalışmanın temel amacı, kalbin ürettiği ses verilerini kullanarak (fonokardiyogram - PCG), üfürüm teşhisi yapan derin öğrenme tabanlı modellerinin geliştirilmesidir. Çalışmada yapılacak analizler sayesinde kardiyovasküler hastalıkların erken teşhisine yardımcı olabilecek otomatik ve yüksek doğruluklu modellerin elde edilmesi amaçlanmaktadır.

Yöntem: Çalışmada Kaggle platformundan elde edilmiş açık kaynak bir veri seti kullanılmıştır (https://www.kaggle.com/datasets/swapnilpanda/heart-sound-database/data). Veri setinde 816'sı sağlıksız ve 2725'i sağlıklı olmak üzere toplam 3541 örnek bulunmaktadır. Veri seti, indirilmiş olduğu platformda hali hazırda validasyon ve eğitim veri seti olmak üzere ikiye bölünmüştür. Validasyon veri setinde 301 (151 sağlıksız, 150 sağlıklı), eğitim veri setinde ise 3240 (665 sağlıksız, 2575 sağlıklı) örnek bulunmaktadır. Eğitim veri setinden rastgele %10 örnek seçilerek test veri seti oluşturulmuş bu işlem sonucunda 324 örnek ile test 2916 örnek ile eğitim veri seti oluşturulmuştur. Test ve eğitim veri seti oluşturma aşamasında sınıf dengesi gözetilmiştir. Evrişimsel sinir ağları, yineleyen sinir ağları ve dikkat ağları kullanılarak üç farklı derin öğrenme modeli geliştirilerek eğitim ve validasyon veri seti üzerinde hiper-parametre optimizasyon süreçleri yürütülmüştür. Daha sonra elde edilen optimum hiper-parametreler ile eğitilen modelin test veri seti üzerindeki performans skorları hesaplanmıştır.

Bulgular: Elde edilen sonuçlar kullanılarak doğruluk, kesinlik, duyarlılık ve auc metrikleri hesaplanmıştır. Bu kapsamda evrişimsel sinir ağı tabanlı modeli ile sırayla 93.51%, 94.96%, 96.83% ve 97.90%, yineleyen sinir ağı modeli ile sırasıyla 94.13%, 95.34%, 97.23% ve 98.10%, dikkat ağları modeli ile sırasıyla 95.06%, 96.12%, 97.63% ve 98.90% doğruluk, kesinlik, duyarlılık ve auc skorları elde edilmiştir.

Sonuç: Çalışmada, fonokardiyogram verileri kullanılarak geliştirilen derin öğrenme modelleri ile üfürüm teşhisi yüksek doğrulukla gerçekleştirilmiştir. Evrişimsel sinir ağları, yineleyen sinir ağları ve dikkat mekanizmalı modeller karşılaştırılmış, en başarılı sonuçlar dikkat ağları modelinde elde edilmiştir. sonuçlar, derin öğrenme tabanlı sistemlerin erken teşhiste doktorlara destek olabilecek etkili araçlar geliştirmek için büyük potansiyele sahip olduğunu ortaya koymaktadır.

Anahtar Kelimeler: Derin Öğrenme, Üfürüm Teşhisi, Fonokardiyogram, Kardiyovasküler Hastalıklar

Murmur Detection Using Phonocardiogram Data: A Deep Learning-Based Approach

Yasin Görmez[1][0000-0001-8276-2030],Fatma Hilal Yağın[2][0000-0002-9848-7958], **Abdulvahap Pınar**[3][0000-0002-3662-2579]

[1]Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Sivas Cumhuriyet University, Sivas 58140, Turkey

[2]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Ozal University, Malatya, Turkey.

[3]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye

ABSTRACT

Aim: The main objective of this study is to develop deep learning-based models for murmur detection using heart sound data (phonocardiogram - PCG). Through the analyses conducted in this study, it is aimed to obtain automatic and highly accurate models that can assist in the early diagnosis of cardiovascular disease.

Method: An open-source dataset obtained from the Kaggle platform was used in this study (https://www.kaggle.com/datasets/swapnilpanda/heart-sound-database/data). The dataset contains a total of 3541 samples, of which 816 are unhealthy and 2725 are healthy. The dataset was already divided into validation and training sets on the platform. The validation set includes 301 samples (151 unhealthy, 150 healthy), while the training set includes 3240 samples (665 unhealthy, 2575 healthy). A test dataset was created by randomly selecting 10% of the samples from the training dataset. As a result of this process, 324 samples were used for testing, and 2916 samples were used for training. Class balance was maintained during the creation of the test and training datasets. Three different deep learning models were developed using convolutional neural networks, recurrent neural networks, and attention networks. Hyperparameter optimization processes were conducted on the training and validation datasets. Subsequently, the performance scores of the model trained with the optimal hyperparameters were calculated on the test dataset.

Findings: Based on the obtained results, metrics such as accuracy, precision, recall, and AUC were calculated. In this context, the convolutional neural network model achieved accuracy, precision, recall, and AUC scores of 93.51%, 94.96%, 96.83%, and 97.90%, respectively. The recurrent neural network model achieved scores of 94.13%, 95.34%, 97.23%, and 98.10%, respectively. The attention network model achieved the highest scores with 95.06%, 96.12%, 97.63%, and 98.90%, respectively.

Conclusion: In this study, murmur detection was successfully performed with high accuracy using deep learning models developed based on phonocardiogram data. Convolutional neural networks, recurrent neural networks, and attention-based models were compared, and the best results were obtained from the attention-based model. The results reveal that deep learning-based systems have great potential to serve as effective tools in supporting physicians for early diagnosis.

Keywords: Deep Learning, Murmur Detection, Phonocardiogram, Cardiovascular Diseases

1. Giriş

Kalp üfürümleri, kalpteki yapısal veya işlevsel bozuklukların bir göstergesidir. Geleneksel olarak, bu sesler uzmanlar tarafından stetoskop ile tespit edilir. Ancak, fonokardiyogram (FKG) adı verilen kayıtlar, kalp seslerini hem zaman hem de frekans açısından daha ayrıntılı analiz etmeye imkân tanır ve bu da üfürüm tespitinde daha hassas sonuçlar elde edilmesini sağlamaktadır. Son yıllarda, bu alanda derin öğrenme yöntemleri, özellikle evrişimsel sinir ağları (CNN), kalp üfürümlerinin sınıflandırılmasında başarılı sonuçlar vermiştir. Bu çalışmada ise, kalp üfürümlerini tespit etmek amacıyla derin öğrenme tabanlı modeller geliştirilmiştir (1).

Veri Seti

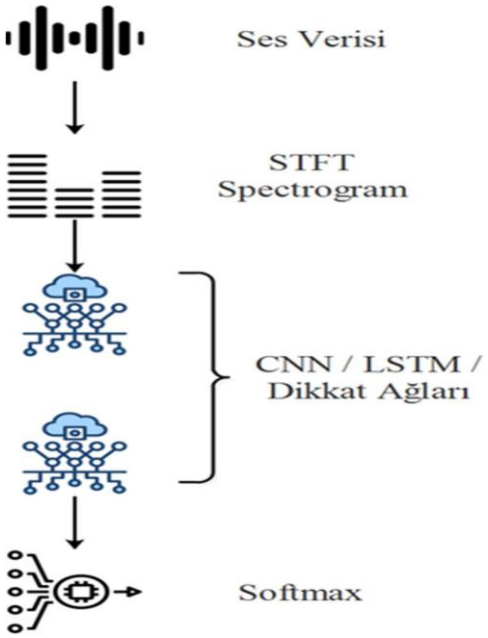
Fonokardiyogram çalışması için Kaggle platformundan (https://www.kaggle.com/datasets/swapnilpanda/heart-sound-database/data). elde edilmiş açık kaynak bir veri seti kullanılmıştır. Veri seti, indirilmiş olduğu platformda hali hazırda %20 test ve %80 eğitim veri seti olmak üzere ikiye bölünmüştür. Test ve eğitim veri seti oluşturma aşamasında sınıf dengesizliği gözetilmiştir. Evrişimsel sinir ağları, yineleyen sinir ağları ve dikkat ağları kullanılarak üç farklı derin öğrenme modeli geliştirilerek eğitim ve test veri seti üzerinde random search yöntemi ile hiper-parametre optimizasyon süreçleri yürütülmüştür.

Tablo 1. Fonokardiogram verilerinin hasta ve sağlıklı bireyler için bölütlenmiş dağılımı

Veri Seti	Hasta	Sağlıklı	Toplam
Eğitim	665	2575	3240
Test	151	150	301
Toplam	816	2725	3541

2. Materyal ve Metot

Derin öğrenme modellerini geliştirmek için Şekil 1'de gösterilen mimari model kullanılmıştır. Bu modelin derin öğrenme katmanında LSTM, CNN ve Dikkat ağları modelleri geliştirilmiştir (2).



Şekil 1. Derin öğrenme modelleri için kullanılan mimarinin yapısı

Spektrogram: Spektrogram, bir ses sinyalinin frekanslarının zaman içindeki değişimini gösteren bir grafiktir. Bu grafik; sesin hangi frekanslarda, hangi anda ve ne kadar güçlü olduğunu gösterir. Spektrogramlar, sesin ayrıntılı analizinde ve işlenmesinde sıkça kullanılır.

LSTM (Uzun Kısa Süreli Bellek Ağı): LSTM, tekrarlayan sinir ağlarının (RNN) özel bir türüdür. Özellikle sıralı verilerde (zaman serileri) güçlü sonuçlar verir. Bu başarısını, bilgi akışını yöneten özel yapılar sayesinde sağlar (uzun süreli bellek, yeni bilginin ne kadarının belleğe depolanacağı, hangi kısmının çıktı olarak kullanılacağı).

CNN (Evrişimli Sinir Ağı): CNN, görüntü verilerinde desenleri öğrenmek için kullanılan bir derin öğrenme yöntemidir. Sinir ağı katmanları kullanılarak özellikleri çıkarma, sınıflandırma ve nesne tanıma gibi aşamalarda yüksek başarı sağlar.

Dikkat Ağları: Dikkat ağları, bir yapay zeka modelinin veri içindeki en önemli kısımlara odaklanmasını sağlayan sistemlerdir. Modelin tüm bilgiye eşit şekilde bakmak yerine, önemli olan yerlere daha fazla "dikkat" vermesini sağlar. Bu sayede, özellikle doğal dil işleme ve görüntü işleme veri analizlerinde, model daha doğru ve verimli sonuçlar üretir.

3. Bulgular ve Sonuç

Tablo 2. Fonokardiyogram sinyallerinden üfürüm tespiti için model performanslarının karşılaştırılması

Model	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	AUC (%)
CNN	93.51	94.96	96.83	97.90
LSTM	94.13	95.34	97.23	98.10
Dikkat Ağları	**95.06**	**96.12**	**97.63**	**98.90**

Üfürüm tespiti model performansları Tablo 1'de sunulmuştur, sınıf dengesizliği bulunan veri yapısında Dikkat Ağları modelinin %98.90 AUC değeri ile en yüksek sınıflandırma performansına sahiptir. Ayrıca modelin yüksek duyarlılık (%97.63) ve kesinlik (%96.12) değerleri, özellikle azınlık sınıfın doğru tespitinde güçlü bir performansa sahip olduğunu göstermiştir. AUC ile birlikte kesinlik ve duyarlılık tamamlayıcı metrikleri dikkate alındığında, modelin dengesiz veri setlerinde güvenilir sonuçlar elde edildiği söylenebilir.

Kaynaklar

1. Abdelkarim, A., Deng, H., & Kumar, S. (2024). Heart murmur detection and clinical outcome prediction using time–frequency deep features and semi-supervised learning. Scientific Reports, 14, Article 12234. https://doi.org/10.1038/s41598-024-58274-6
2. TUTAR, Gökhan; AYDIN, Serdar. Kalp Atış Seslerinin Derin Öğrenme Kullanarak Sınıflandırılması. Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi, 2024, 13.4: 1927-1933.

Development of a Mobile Calculator for Mortality Prediction in Elderly Individuals Using the Multidimensional Prognostic Index

İhsan Doğan[1] (ORCID: 0000-0002-1234-5678), Pembe Keskinoğlu[2] (ORCID: 0000-0003-8765-4321)

[1]Dokuz Eylül Üniversitesi, Sağlık Bilimleri Enstitüsü, Medikal İnformatik A.B.D., İzmir, Türkiye

[2]Dokuz Eylül Üniversitesi, Tıp Fakültesi, Halk Sağlığı A.B.D., İzmir, Türkiye

Abstract

Elderly individuals face increased mortality risks due to age-related health issues and limited healthcare resources. The Multidimensional Prognostic Index (MPI) is a reliable tool used to assess mortality risk in geriatric patients [1]. However, its complexity and time-consuming nature limit its clinical use. In this study, we developed a mobile application with a user-friendly interface and secure backend to calculate MPI scores. The application was developed using Flutter (Dart) [2] and communicates with a Python-based server through RESTful APIs [3]. Usability was evaluated through Likert scale questionnaires completed by healthcare professionals [5]. The results showed high user satisfaction and suggest that the app may support clinical decision-making processes effectively.

Keywords: mortality, geriatrics, mobile application, prognostic index, Flutter, MPI

1 Giriş

Yaşlanan nüfus, sağlık sistemleri için yeni zorluklar ortaya çıkarmaktadır. Özellikle sağlık sonuçlarının öngörülmesi ve kaynak yönetimi büyük önem taşımaktadır. Çok Boyutlu Prognostik İndeks (ÇBPİ), fonksiyonel, bilişsel ve tıbbi parametreleri birleştirerek mortalite riskini öngörmede etkili bir araçtır [1]. Ancak manuel uygulaması zaman alıcı olup, yaygın kullanımını zorlaştırmaktadır. Bu çalışmanın amacı, klinik ortamlarda prognostik değerlendirmeyi kolaylaştırmak amacıyla bir mobil uygulama geliştirmektir.

2 Yöntem

Mobil uygulama, Google'ın açık kaynak kodlu arayüz geliştirme kiti olan Flutter ile Dart dili kullanılarak geliştirilmiştir [2]. Arka uç hizmetleri, Python dili ve Django REST Framework

kullanılarak yazılmıştır [3]. Sistem, Docker ile konteynerleştirilmiş [4] ve bir Linux sunucusunda barındırılmıştır. Uygulama, hasta kaydı, gerçek zamanlı ÇBPİ hesaplama ve istemci-sunucu arasında güvenli veri iletimini desteklemektedir. Kullanılabilirlik, 28 sağlık çalışanının katıldığı Likert ölçekli bir anket ile değerlendirilmiştir [5].

3 Bulgular

Uygulama Android cihazlara başarıyla yüklenmiş ve kullanıcılar tarafından olumlu geri bildirimler almıştır. Klinisyenler, kullanım kolaylığı, hız ve genel fayda açısından yüksek memnuniyet bildirmiştir. ÇBPİ değerlendirmesi için gereken süre önemli ölçüde azalmış ve karar verme süreci iyileştirilmiştir.

4 Tartışma

Bu çalışma, mobil sağlık teknolojilerinin klinik değerlendirme süreçlerini kolaylaştırma potansiyelini ortaya koymaktadır. ÇBPİ'nin mobil bir platforma entegrasyonu, geleneksel yöntemlerdeki zaman kaybı ve veri giriş hatalarını azaltmıştır. Uygulama şu anda yalnızca Android platformunda çalışmaktadır, ancak gelecekteki sürümlerle iOS desteği ve çok dilli arayüz eklenerek erişilebilirliği artırılması planlanmaktadır.

5 Sonuç

ÇBPİ değerlendirmesi için geliştirilen mobil uygulama, pratik, verimli ve erişilebilir bir araç sunmaktadır. Uygulama, klinik uygulamalarda ÇBPİ'nin faydasını artırmakta ve yaşlı hastalar için sağlık sonuçlarını iyileştirmeye katkıda bulunmaktadır. Geliştirme sürecinin devamı ve klinik iş akışlarına entegrasyon önerilmektedir.

6 Kaynakça

[1] Pilotto, A., et al. (2008). Kapsamlı geriatrik değerlendirmeye dayalı Çok Boyutlu Prognostik İndeks kısa ve uzun dönem mortaliteyi öngörür. Journal of Gerontology.

[2] Google Developers. (2023). Flutter Belgeleri. Erişim adresi: https://flutter.dev/

[3] Django REST Framework. (2023). Erişim adresi: https://www.django-rest-framework.org/

[4] Docker Inc. (2023). Docker Nedir? Erişim adresi: https://www.docker.com/

[5] Likert, R. (1932). Tutumların Ölçümü İçin Bir Teknik. Archives of Psychology.

Sağlıkta Biyoinformatik Dönüşümüne Yerli Bir Yanıt: Ilyome Deneyimi

Özet

Biyoinformatik, son yıllarda sadece akademik araştırmaların değil, aynı zamanda klinik pratiğin de ayrılmaz bir parçası haline gelmiştir. Genetik veri üretimindeki artış, bu verilerin yorumlanmasını ve klinik anlamda kullanılabilir hale getirilmesini stratejik bir gereklilik kılmıştır. Bu çalışmada, Türkiye merkezli bir biyoinformatik girişimi olan Ilyome'un geliştirdiği yerli varyant analiz platformunun; dışa bağımlılığı azaltma, analiz sürelerini kısaltma ve klinik karar destek süreçlerini güçlendirme potansiyeli ele alınmaktadır.

Giriş

Biyoinformatik, yüksek hacimli biyolojik verilerin işlenmesi ve yorumlanmasına olanak sağlayarak, sağlık hizmetlerinde dijital dönüşümün temel bileşenlerinden biri haline gelmiştir. Özellikle kişiselleştirilmiş tıp, tıbbi genetik, tıbbi biyoloji ve tanısal araştırmalarda biyoinformatik analizlerin kritik rolü giderek artmaktadır. Genetik sekanslama teknolojilerinin yaygınlaşması, biyoinformatik çözümlerin günlük klinik işleyişe entegrasyonunu kaçınılmaz kılmıştır.

Türkiye'deki Mevcut Durum ve Zorluklar

Türkiye'de genetik analiz süreçlerinin önemli bir bölümü hâlen dışa bağımlı yazılımlar ve altyapılar aracılığıyla yürütülmektedir. Bu durum, üç temel sorunu beraberinde getirmektedir:

1. Ülkeden döviz çıkışı,
2. Genetik analiz pazarında tekelleşme riski,
3. Yerli ve sürdürülebilir analiz altyapısının eksikliği.
4. Uzun analiz süreleri

Bu bağlamda, genetik verinin yalnızca analiz edilmesi değil, analiz sürecinin nerede ve hangi koşullarda gerçekleştirildiği de stratejik öneme sahiptir.

Yöntem: Ilyome Biyoinformatik Platformu

Ilyome, genetik merkezlerin ihtiyaç duyduğu varyant analiz süreçlerini sadeleştirmek ve hızlandırmak amacıyla geliştirilmiş bir yerli biyoinformatik platformdur. GATK (Genome Analysis Toolkit) standartlarıyla tam uyumlu çalışan Ilyome altyapısı, hatasız ve

tekrarlanabilir analiz akışları sunacak şekilde optimize edilmiştir. Platform, kullanıcı dostu arayüzü ile teknik uzmanlık gerektirmeden kullanılabilmekte, bu sayede klinik uzmanlara doğrudan analiz desteği sunmaktadır.

Bulgular

Ilyome platformu, bugüne kadar:
- 25.000'den fazla hastanın genetik analizini gerçekleştirmiştir.
- 80'i aşkın genetik merkezde, 600'den fazla uzman tarafından aktif olarak kullanılmaktadır.
- Türkiye'nin önde gelen kamu ve özel tıbbi genetik laboratuvarlarında entegre olarak çalışmaktadır.

Ayrıca platforma entegre edilen Variant Recruiter modülü ile klinik olarak anlamlı varyantlar önceliklendirilerek analiz süreçleri daha etkin hale getirilmiştir.

Ekip ve Geliştirme Süreci

Ilyome ekibi şu an itibarıyla 2 biyoinformatik uzmanı, 1 ön yüz geliştirici, 1 arka yüz geliştirici ve 1 iş geliştirme uzmanından oluşmaktadır. Tüm yazılım geliştirme süreci Türkiye'de, yerli kaynaklarla yürütülmektedir.

Uluslararası Yayılım ve Ekonomik Katkı

Ilyome yalnızca yurt içinde değil, Avrupa ve Orta Doğu pazarlarında da aktif olarak kullanılmaktadır. Böylece platform, yalnızca döviz çıkışını engellemekle kalmayıp, yazılım ihracatı yoluyla ülkeye döviz girdisi de sağlamaktadır. Geliştirilen RNA-Seq ve somatik varyant analiz pipeline'ları sayesinde, platform uluslararası araştırma ve klinik analiz taleplerini karşılayabilecek düzeyde yetkin hale gelmiştir.

Tartışma ve Sonuç

Ilyome, klinik biyoinformatik alanında stratejik bir yerli çözüm olarak konumlanmakta; dışa bağımlılığı azaltmakta ve katma değerli yazılım üretimiyle Türkiye'nin teknolojik kapasitesine katkıda bulunmaktadır. Aynı zamanda yüksek nitelikli insan kaynağının ülkede tutulmasına olanak sağlamakta, sağlık verisinin işlenmesi sürecini dijital ve güvenli bir temele oturtmaktadır.

Bu çalışma, biyoinformatiğin yalnızca teknolojik değil, aynı zamanda ekonomik ve stratejik bir alan olduğunu ortaya koymakta ve yerli yazılım girişimlerinin bu alandaki potansiyelini göstermektedir.

Anahtar Kelimeler

Biyoinformatik, Genetik Veri, Varyant Analizi, Yerli Yazılım, Klinik Karar Destek, GATK, Türkiye, Ilyome

Bazı Schiff Bazları ve Amin Türevlerinin in Siliko Değerlendirilmesi: Moleküler Kenetlenme ve ADME Analizi

Işık ÇAKMAK[1][0000-0003-4758-4077], Sinem ÖZDEMİR[1][0000-0002-6332-0891], Nurhan GÜMRÜKÇÜOĞLU [2][0000-0002-9669-6318]

[1] KTÜ Tıp Fakültesi, Temel Tıp, Biyoistatistik ve Tıp Bilişim, Trabzon, TÜRKİYE
[2] KTÜ Sağlık Hizmetleri Meslek Yüksekokulu, Trabzon, TÜRKİYE

Özet. Triazol türevleri, geniş bir biyolojik aktiviteye sahip olması yönüyle farmasötik araştırmalarda önemli bir yere sahiptir. Bu çalışmada, daha önce sentezlenmiş triazol schiff bazları ve amin türevi bileşiklerin potansiyel biyolojik aktiviteleri in siliko yöntemlerle değerlendirilmiştir. Moleküler kenetlenme çalışmaları, bileşiklerin hedef proteinlerle olan etkileşim potansiyellerini incelemek amacıyla gerçekleştirilmiştir. Çalışma kapsamında, antioksidan aktivite ile ilişkilendirilen hedef protein belirlenmiş ve bu protein üzerinde gerçekleştirilen analizler sonucunda, kenetlenme skorları, bağlanma konformasyonları değerlendirilmiştir. Bulgular bileşiklerin belirlenen hedef proteinle kristalize olan liganda (referans) kıyasla benzer veya daha yüksek bağlanma afinitesine sahip olduğunu ve antioksidan aktivite göstermeye yatkın olabileceğini ortaya koymaktadır. Ek olarak, ADME (emilim, dağıtım, metabolizma ve atılım) analizi ile farmakokinetik özellikler değerlendirilmiş ve bileşiklerin biyoyararlanım, karaciğer metabolizması ve hücresel geçirgenlik açısından uygun profiller sergilediği belirlenmiştir. İn siliko sonuçlar, bu triazol türevlerinin farmasötik uygulamalarda potansiyel adaylar olabileceğini göstermektedir. Çalışmalarımızın, gelecekteki in vitro ve in vivo araştırmalar için rehber niteliğinde olabileceği öngörülmekte olup, bu bulguları desteklemek amacıyla ek çalışmalar planlanmaktadır.

Anahtar Kelimeler: ADME, Moleküler Kenetlenme, Triazol Türevleri

In Silico Evaluation of Some Schiff Bases and Amine Derivatives: Molecular Docking and ADME Analysis

Işık ÇAKMAK[1][0000-0003-4758-4077] , Sinem ÖZDEMİR[1][0000-0002-6332-0891] , Nurhan GÜMRÜKÇÜOĞLU [2][0000-0002-9669-6318]

[1] KTU Faculty of Medicine, Basic Medicine, Biostatistics and Medical Informatics, Trabzon, TURKEY
[2] KTU Vocational School of Health Services, Trabzon, TURKEY

Abstract. Triazole derivatives have an important place in pharmaceutical research due to their wide range of biological activities. In this study, the potential biological activities of previously synthesised triazole schiff bases and amine derivative compounds were evaluated by in silico methods. Molecular docking studies were carried out to investigate the interaction potential of the compounds with target proteins. Within the scope of the study, the target protein associated with antioxidant activity was identified and the docking scores and binding conformations were evaluated as a result of the analyses performed on this protein. The findings reveal that the compounds have similar or higher binding affinity compared to the ligand (reference) crystallised with the identified target protein and may be prone to exhibit antioxidant activity. In addition, pharmacokinetic properties were evaluated by ADME (absorption, distribution, metabolism and excretion) analysis and the compounds were found to exhibit favourable profiles in terms of bioavailability, hepatic metabolism and cellular permeability. In silico results indicate that these triazole derivatives may be potential candidates for pharmaceutical applications. It is envisaged that our studies may serve as a guide for future in vitro and in vivo investigations and additional studies are planned to support these findings.

Keywords: ADME, Molecular Docking, Triazole Derivatives

Model Context Protocol is Necessary for AI Agents in the Healthcare Domain

Asli Boyraz

Recep Tayyip Erdog˘an University, Rize, Turkiye,
asli.yazagan@erdogan.edu.tr

Abstract. The use of artificial intelligence (AI) in the medical domain leads to a catastrophic revolution in the healthcare delivery system. The modern healthcare system must deal with diverse data types such as clinical notes, medical imaging, lab results, real-time monitoring, social determinants of health, genetic data, medication history, etc. Most Clinical Decision Support Systems (CDSS) still struggle with the complexity of medical data collected for a patient. The future of CDSS is modular, context-aware, and agent-driven. LLM-based AI agents designed to understand complex medical situations and offer evidence-based recommendations that ensure context-aware decision making for clinicians and healthcare providers to improve clinical tasks. An AI agent goes beyond just sending notifications; it acts as a reliable partner for healthcare providers. Through AI Agents, personalized medicine became accessible and affordable. However, data safety and ethical concerns are challenging problems. The safe and reliable integration of LLM-based agents with patient data is essential. Model Context Protocol (MCP) is a new standard for two-way connection between AI agents, data sources and tools. This study outlines a protocol tailored to the needs of the healthcare domain, with a focus on applying the Model Context Protocol (MCP) from a healthcare perspective.

Keywords: ai agents, model context protocol, llm-based ai, cdss, healthcare

1 A Protocol Need for LLM-based Agents

Large Language Model (LLM) is an artificial intelligence (AI) trained on large text data that could understand the language model and produce human-like text. This revolution has changed the way we interact with computers. Then, LLMs started to interact with other tools in order to do any task using tool calling mechanism. By this, LLMs invoke external tools and Application Programming Interface (APIs) to perform specific tasks. These tools can range from simple functions like retrieving data from a database to more complex operations such as sending emails or interacting with third-party services. For example, LLM can connect to a database to analyze some dataset and can prepare a report based on the data analysis. Although tool calling is powerful, it comes with

several challenges. Different tools have different APIs and this requires custom implementation to invoke specific tools to perform tasks. Without a common protocol, each AI application must implement its own integration logic and need to keep it up-to-date, leading to redundant efforts. This problem has led to the need to define a protocol between LLMs and other tools and services that eliminates the need for custom integrations. A protocol defines a set of rules or standards for systems or devices to communicate and interact with each other, ensures consistency, interoperability, and efficient data exchange between different platforms. In this way, standardization, security, control, and efficiency are achieved, which are crucial for the healthcare systems.

2 Model Context Protocol

Model Context Protocol (MCP) [1] is a new universal standard for how AI/LLMs applications interact with external systems, such as databases, tools, and resources. It is used to create a secure link between LLMs and data sources, integrating multiple data sources or services into an LLM workflow. AI Agents refers to artificial intelligence systems that are not only capable of responding but can also act autonomously [5]. These agents can analyze data, make decisions, and execute sequences of tasks without human input, making them a promising solution to bridge the gap between alert and action [2]. MCP allows AI agents to autonomously discover, select, and orchestrate tools based on task context. Thus, it fosters more adaptive and context-aware AI agents.

Designing an AI agent is a challenge providing real-time, relevant and structured information while maintaining security, privacy, and modularity. MCP is designed to solve this challenge. The most important advantage of MCP is the security, because the data do not need to be transferred to another computer. Moreover, only allowed operation can be applied, and only allowed data could be accessible to the tools. MCP acts like a universal adapter that allows AI models to connect to any system using a standard method [1]. MCP is fully supported by Google's ADK, which enables a wide range of MCP servers to seamlessly integrate with AI agents. MCP integration will be necessary to develop a secure, scalable, and legally compliant systems. As more services become MCP-aware, more sophisticated AI application will be available [3].

.

3 Transition of CDSS to AI Agents using MCP for Future Healthcare System

Clinical Decision Support Systems (CDSS) have long aimed to improve healthcare by offering alerts and recommendations [6]. In practice, alerts are ignored, recommendations lack specificity, and follow-up is entirely dependent on health care provides who are already overwhelmed [4]. As healthcare grows increasingly complex, the gap between receiving notifications and taking action becomes more

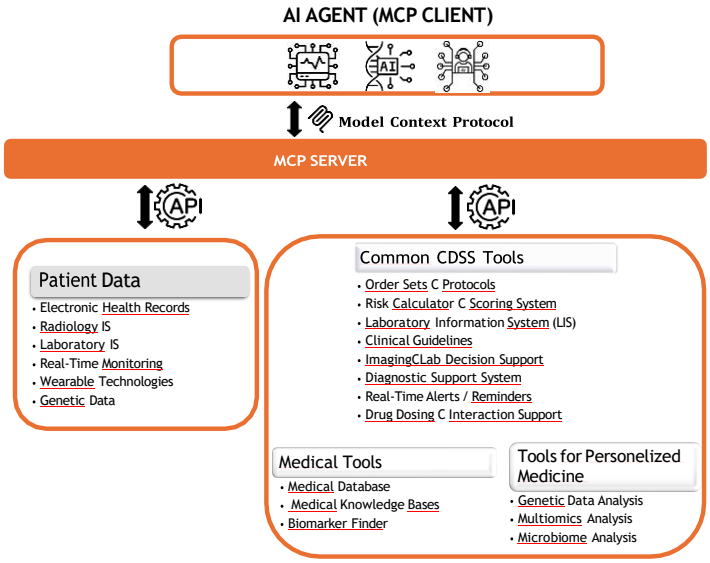


Fig. 1. Model Context Protocol Architecture for AI Agents in the Healthcare Domain

critical. AI agents are playing a growing role in assisting healthcare professionals by boosting diagnostic precision, enhancing patient care, and speeding up research across various areas of healthcare. AI agents in the healthcare domain are classified into four key categories: clinical diagnosis and decision support, mental health and therapy agents, general medical assistants for workflow optimization, and pharmaceutical and drug discovery agents [8]. An AI agent goes beyond just sending notifications; it comprehends the context, evaluates the urgency, and takes actionable steps in clinical decision-making, acting as a reliable partner for healthcare providers. Using MCP, it becomes easier to collect multiple clinical signals and combine them to infer a meaningful interpretation using external tools for AI Agent. For example, if a patient is admitted to the emergency department with complaints of high fever, a stiff neck, and altered mental status, the AI agent first collects the patient's EHR, medical history, vital data, and lab results. Then, it accesses external databases for the latest guidelines on diagnosis and management. After reasoning through the data, the AI may infer 'possible bacterial meningitis.' The infectious disease specialist is notified, isolation protocols are initiated, and the care team is updated in real-time. MCP makes AI Agent's actions safe and scalable by ensuring that decisions are grounded in accurate, real-time data and adhere to established clinical guidelines.

AI agents using MCP enable context-aware decision making in healthcare, bridging the gap between diagnosis and intervention. This capability can fast-track the advancement of personalized and preventive medicine. Healthcare providers

who embrace this shift have the opportunity to transform the future of medical care.

4 Conclusion and Discussion

AI Agent opens a new era for personelized medicine. As this technology expands, it is clear that the future of healthcare is not just intelligent, it also deliver concrete, specific treatment options based on a person's genetic profile, habits, and medical history. These systems combine contextual understanding with autonomous decision-making, allowing them to collaborate with clinicians, trigger real-time interventions, and coordinate workflows across care settings. MCP framework allows AI Agents to function reliably across diverse healthcare environments, fostering trust among clinicians while supporting the seamless integration of AI-driven solutions into everyday medical practice.

On the other hand, as MCP simplifies interaction with AI Agent, it simultaneously introduces new potential security problems. MCP servers gain deeper access to sensitive medical data and critical operations. Users must remain vigilant about the risks posed by unverified tools and misconfigured servers[7]. Improper installation or using untrusted sources can lead to data breaches, unauthorized system changes, or general system instability [7]. Deeper investigation into MCP security implications is an open research area.

The healthcare sector has been transforming in a way that is proactive, preventative and patient-centered with AI Agents using MCP to create a smarter, safer and more efficient healthcare system.

References

1. Introducing the Model Context Protocol https://www.anthropic.com/news/model-context-protocol
2. Wang, W., Ma, Z., Wang, Z., Wu, C., Chen, W., Li, X., Yuan, Y. (2025). A Survey of LLM-based Agents in Medicine: How far are we from Baymax?. arXiv preprint arXiv:2502.11211.
3. Luo, J., Zhang, W., Yuan, Y., Zhao, Y., Yang, J., Gu, Y., Zhang, M. (2025). Large Language Model Agent: A Survey on Methodology, Applications and Challenges. arXiv preprint arXiv:2503.21460.
4. Payne, T. H., Hines, L. E., Chan, R. C., Hartman, S., Kapusnik-Uner, J., Russ, A. L., Malone, D. C. (2015). Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. Journal of the American Medical Informatics Association, 22(6), 1243-1250.
5. Acharya, D. B., Kuppan, K., Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals–A Comprehensive Survey. IEEE Access.
6. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ digital medicine, 3(1), 17.
7. Hou, X., Zhao, Y., Wang, S., Wang, H. (2025). Model context protocol (mcp): Landscape, security threats, and future research directions. arXiv preprint arXiv:2503.23278.
8. Ferrag, M. A., Tihanyi, N., Debbah, M. (2025). From LLM Reasoning to Autonomous AI Agents: A Comprehensive Review. arXiv preprint arXiv:2504.19678.

Classification of Acute Myeloid Leukemia (AML) Gene Expression Dataset and Identification of Feature Importances Using Machine Learning Methods

Hüseyin KUTLU [1][0000-0003-0091-9984] and Cemil ÇOLAK[2][0000-0001-5406-098X]

[1] Adıyaman University, Adıyaman 02040, Türkiye
[2] İnönü University, Malatya 44280 Türkiye
hkutlu@adiyaman.edu.tr

Abstract. In this study, a high-dimensional gene expression dataset related to Acute Myeloid Leukemia (AML) was classified using various machine learning methods without applying dimensionality reduction techniques. The feature importances (probe set names) obtained from each classifier and the SHAP values were compared with 13 genes (BIK, CCNA1, FUT4, IL3RA, HOMER3, JAG1, WT1, ALDH1A1, PELO, PLXNC1, PRUNE, SERPINB9, TRIB2) identified as disease-related in the original study presenting the dataset. Then, feature selection was performed using LASSO, ElasticNet, and BORUTA methods. The selected features were also compared with the genes known to be associated with the disease in the referenced study. In the experimental phase, a 10-times repeated 10-fold cross-validation strategy and random search hyperparameter optimization were used to train 13 different machine learning models. Among them, Logistic Regression achieved the highest classification performance with an accuracy of 0.98, a recall of 0.96, and an F1-score of 0.97. However, among the top 20 feature importance values obtained with this algorithm, only two probe sets (HOMER3, WT1) matched the known disease-related genes. On the other hand, LightGBM, which showed lower classification performance, identified six probe sets corresponding to the disease-related genes. Among the feature selection methods, ElasticNet, LASSO, and BORUTA selected 89, 17, and 55 features respectively, which corresponded to 2, 6, and 10 of the known disease-related genes. Although the BORUTA feature selection method identified 10 of the disease-associated genes, only 6 of them were among the top 20 ranked probe sets based on importance. These results highlight a trade-off between classification performance and biological relevance in machine learning models and emphasize the importance of jointly evaluating different algorithms and feature selection methods in gene expression studies.

Keywords: Acute Myeloid Leukemia, Feature Importance, Machine Learning.

Akut Miyeloid Lösemi (AML) Gen İfade Veri Setinin Sınıflandırılması ve Makine Öğrenme Yöntemlerini Kullanarak Özellik Önemlerinin Belirlenmesi

Özet: Bu çalışmada, Akut Miyeloid Lösemi (AML) ile ilişkili yüksek boyutlu bir gen ekspresyon veri seti, boyut indirgeme teknikleri uygulanmadan çeşitli makine öğrenmesi yöntemleri ile sınıflandırılmıştır. Her bir sınıflandırıcıdan elde edilen öznitelik önemleri (prob seti adları) ve SHAP değerleri, veri setini sunan özgün çalışmada hastalıkla ilişkili olarak tanımlanan 13 gen (BIK, CCNA1, FUT4, IL3RA, HOMER3, JAG1, WT1, ALDH1A1, PELO, PLXNC1, PRUNE, SERPINB9, TRIB2) ile karşılaştırılmıştır. Ardından, LASSO, ElasticNet ve BORUTA yöntemleri kullanılarak öznitelik seçimi gerçekleştirilmiş; seçilen öznitelikler de referans çalışmadaki hastalıkla ilişkili genlerle karşılaştırılmıştır. Deneysel aşamada, 13 farklı makine öğrenmesi modeli, 10 tekrarlı 10 katlı çapraz geçerlik stratejisi ve rastgele arama hiperparametre optimizasyonu ile eğitilmiştir. Bu modeller arasında, Lojistik Regresyon algoritması %98 doğruluk, %96 duyarlılık ve %97 F1 skoru ile en yüksek sınıflandırma performansını göstermiştir. Ancak, bu algoritma ile elde edilen en önemli 20 öznitelik arasında yalnızca iki prob seti (HOMER3, WT1) bilinen hastalıkla ilişkili genlerle eşleşmiştir. Öte yandan, daha düşük sınıflandırma performansı sergileyen LightGBM algoritması, hastalıkla ilişkili genlerle eşleşen altı prob setini belirlemiştir. Öznitelik seçimi yöntemleri arasında ElasticNet, LASSO ve BORUTA sırasıyla 89, 17 ve 55 öznitelik seçmiş; bunlar da sırasıyla bilinen hastalıkla ilişkili genlerden 2, 6 ve 10'una karşılık gelmiştir. BORUTA yöntemi 10 hastalıkla ilişkili geni belirlemiş olsa da, bunların sadece 6'sı önem sıralamasında ilk 20'de yer almıştır. Bu sonuçlar, makine öğrenmesi modellerinde sınıflandırma performansı ile biyolojik anlamlılık arasında bir ödünleşim olduğunu ortaya koymakta ve gen ekspresyon çalışmalarında farklı algoritmaların ve öznitelik seçimi yöntemlerinin birlikte değerlendirilmesinin önemini vurgulamaktadır.

Anahtar Kelimeler: Akut Miyeloid Lösemi, Öznitelik Önemi, Makine Öğrenmesi.

1 INTRODUCTION

The information required to form and sustain a living organism is stored in its genome. The genomic DNA carries all of the genetic information of a living organism. However, the information necessary for protein production is first transcribed into mRNA (messenger RNA). mRNA levels reflect how active a gene is. Gene expression analyses are mostly performed by measuring mRNA molecules. Microarray technology is a method that can measure the expression levels of thousands of genes simultaneously. Determining gene expression levels in genomic research is of great importance for disease diagnosis and classification, biomarker discovery, and the identification of therapeutic targets [1].

In our body, blood cells are produced in the bone marrow. A healthy body produces and uses red blood cells to carry oxygen and white blood cells to fight microbes and stop bleeding. Acute Myeloid Leukemia (AML) occurs when the bone marrow malfunctions and produces non-functional white blood cells. These white blood cells multiply uncontrollably and fail to perform their functions, thereby weakening the immune system [2].

The GEO dataset with the accession code GDS3057 is a gene expression dataset related to Acute Myeloid Leukemia (AML). It consists of samples (from blood or bone marrow) collected from 64 individuals (26 AML patients and 38 healthy donors). In the collected samples, genetic information present in the cells is transferred from DNA to RNA. This RNA is the product of actively used genes in the cell. The RNAs are isolated from the cells. The extracted RNAs are loaded onto microarray chips, and a light (fluorescent) signal is applied. The brighter the light that binds to a gene on the chip, the more active that gene is considered to be. The brightness signals from the chip are transferred to a computer and digitally represented [3].

In this study, the Acute Myeloid Leukemia (AML) gene expression dataset with the accession code GDS3057, uploaded to the GEO repository, was used. The AML dataset includes expression levels of 22,283 genes from 64 individuals [4]. After downloading the dataset, each row was organized to represent one sample, each column to represent a gene expression level, and the last column to indicate the class as patient (1) or non-patient (0). Figure 1 shows a portion of the dataset used.

sample	1007_s_at	1053_at	117_at	121_at	1255_g_at	1294_at	1316_at	1320_at	1405_i_at	1431_at	...	AFFX-TrpnX-M_at	class
GSM239371	3.27542	6.69586	4.41097	7.62111	2.27194	7.90378	5.21323	2.80147	6.09898	3.21223	...	3.08049	1
GSM239487	2.96808	7.14802	6.83602	7.09940	2.17899	6.75748	4.47788	2.61972	5.57450	2.95902	...	2.85561	1
GSM239489	3.36886	6.97158	4.53694	7.94294	2.29210	8.41816	5.35851	2.91582	6.96664	3.35131	...	3.11268	1
GSM239492	2.99698	6.94238	4.49776	7.60755	2.22920	8.46694	4.66434	2.73466	6.30322	3.12166	...	2.96273	1
GSM239497	3.34631	6.26353	5.61793	7.90653	2.35270	7.90712	5.02168	2.98178	7.88112	3.47701	...	3.29020	1

5 rows × 22284 columns

Fig. 1. A portion of the dataset used

Machine learning techniques have been comprehensively applied to classify data from the GEO GDS3057 Acute Myeloid Leukemia (AML) dataset using various models. These techniques range from traditional machine learning algorithms to deep learning models, each offering unique advantages in terms of accuracy and computational efficiency.

Frasca et al. [5] applied classification models using a deep neural network (DNN) following a reduction in the number of variables, achieving approximately 92% classification accuracy. The results were then compared to a support vector machine (SVM)-based approach, which achieved an accuracy of 87.39%. In a study conducted by Francese et al. [6], deep learning and genetic algorithms were used for feature selection and classification of leukemia data. An unsupervised deep learning model based on autoencoders was used for dimensionality reduction, achieving approximately 92% accuracy with a DNN and 87.39% with an SVM.

Susmi et al. [7] developed a framework using a column-based feature selection algorithm and a feedforward backpropagation neural network, achieving 85.6% classification accuracy in distinguishing between ALL and AML. Khanam et al. [8] proposed a convolutional neural network model called ANCOM to classify AML subtypes and healthy leukocytes. Their model achieved an overall detection accuracy of 91.7% and classification accuracy of 92.5%, outperforming other models such as VGG16 and ResNet50 without the need for image segmentation.

In this study, 13 different machine learning algorithms were used to classify the data without applying any dimensionality reduction techniques, and important features were identified using SHAP values. Feature selection methods such as ElasticNet, LASSO, and BORUTA were applied. The probe set names representing the important genes mentioned in the article [4], which presented the dataset with selected features, were identified. The success of machine learning algorithms' feature importance and feature selection methods in identifying these genes was compared.

Highlights of the Study

- **Analysis on the Entire Dataset**: The study conducted analyses on the high-dimensional gene expression dataset without initially applying any dimensionality reduction techniques. This approach prevented data loss and allowed for the evaluation of all potentially important biomarkers.
- **Comprehensive Machine Learning Models**: A total of 13 different machine learning algorithms were utilized to perform comparative analyses. This wide range of algorithms enabled the assessment of various approaches in terms of their performance in AML classification.
- **Comparison of Various Feature Selection Methods**: Feature selection methods based on different principles—such as LASSO, ElasticNet, and BORUTA—were applied, and their effectiveness in identifying disease-associated genes was compared.
- **Balance Between Model Performance and Biological Relevance**: The study demonstrated that algorithms with the highest classification performance do not necessarily identify biologically meaningful features. This finding underscores that focusing solely on model performance is insufficient in bioinformatics studies.

2 Materials and Methods

2.1 Dataset

The dataset identified by the code GDS3057 in the NCBI GEO (Gene Expression Omnibus) database is a microarray-based gene expression dataset associated with Acute Myeloid Leukemia (AML). The dataset is pre-normalized in its original form and consists of a total of 64 samples (26 AML patients and 38 healthy controls). It was created to identify the genetic subtypes of AML in children. The dataset contains expression levels (mRNA amounts) of 22,283 different genes. Each row in the dataset represents an individual patient or healthy subject (identified by a GSM code), and each column represents the expression level of a specific gene. Genes and Probe Set Names Associated with Acute Myeloid Leukemia (AML) [9].
BIK: This gene is known for its role in promoting apoptosis (programmed cell death). In AML, it shows increased expression, indicating a possible response to the disease.
CCNA1: This gene plays a role in regulating the cell cycle, which is critical for cell division. Elevated levels of CCNA1 in AML patients have been associated with a favorable prognosis.
FUT4: Although the exact role of FUT4 in AML is not well understood, it is overexpressed in certain myeloid cell lineages.
IL3RA: This gene plays a key role in hematopoiesis (formation of blood cells) and has been linked to leukemogenesis (development of leukemia).

HOMER3: The specific function of HOMER3 in AML is not fully determined, but it is thought to play a role in transcriptional regulation.
JAG1: This gene is part of the Notch signaling pathway, which is important for cell communication and may influence leukemia development.
WT1: This gene has been directly linked to leukemogenesis and is associated with poor clinical outcomes in AML.
ALDH1A1: This gene is involved in retinoic acid metabolism and has shown decreased expression in AML.
PELO, PLXNC1, PRUNE, SERPINB9, and TRIB2: These genes also show decreased expression in AML, although their specific roles are less well understood.
The mapping of probe set names to their corresponding genes was obtained using Python code from the dataset.

2.2 Classification Metrics

The classification algorithms used in the study are listed below.

Logistic Regression Used in binary classification problems. It performs probability estimation and provides class prediction through the sigmoid function. It is interpretable and computationally fast [10].

Random Forest Constructs numerous decision trees and makes classification decisions based on majority voting. It is robust against overfitting [11].

Gradient Boosting Trains weak learners (usually decision trees) sequentially to reduce errors. It offers high accuracy but requires attention to overfitting [12].

XGBoost An optimized version of Gradient Boosting. It supports parallel computing and yields fast and high-accuracy results on large datasets [13].

LightGBM A boosting algorithm developed by Microsoft. It is efficient in terms of memory and speed. It works very fast with large-scale data [14].

CatBoostA boosting algorithm developed by Yandex, effective with categorical data. It automatically handles categorical variables and reduces overfitting [15].

Decision Trees Classifies data by dividing it into branches. These models are simple but interpretable [16].

SVM (Support Vector Machines) Separates data by projecting it into a high-dimensional space. Kernels are used for non-linear problems. Especially powerful for small sample datasets [17].

Naive Bayes Based on Bayes' theorem. It assumes that features are independent from each other [18].

K-Nearest Neighbors (KNN)A sample is classified based on the class of its k closest neighbors[19].

LDA (Linear Discriminant Analysis) Finds linear combinations that maximize class separability. It can be used for both dimensionality reduction and classification [20].

MLP (Multi-Layer Perceptron)A simple artificial neural network model composed of input, hidden, and output layers [21].

AdaBoost An ensemble learning algorithm that combines simple models like decision trees to build a strong classifier. It is more robust against overfitting and works well especially with low-variance, high-bias models [22].

2.3 Performance Metrics

The number of instances where the model correctly classifies a truly positive case—i.e., identifying an AML patient as "AML"—is represented by TP (True Positive). The number of instances where the model correctly classifies a truly negative case—i.e., identifying a healthy individual as "healthy"—is represented by TN (True Negative). The number of instances where the model incorrectly classifies a negative (healthy) case as positive (AML), meaning a healthy individual is predicted as having AML, is represented by FP (False Positive). Conversely, the number of cases where the model incorrectly classifies a positive (AML) case as negative (healthy), meaning an AML patient is predicted as healthy, is represented by FN (False Negative).
The overall accuracy of the model, which measures the proportion of correctly classified instances among all instances, is calculated as Accuracy = (TP + TN) / (TP + TN + FP + FN). However, the accuracy metric can be misleading in imbalanced datasets like AML.
The proportion of actual positive cases (AML) that are correctly predicted by the model is measured by Sensitivity (Recall / True Positive Rate) = TP / (TP + FN), which is used to evaluate the model's ability to detect AML patients. The proportion of actual negative cases that are correctly identified is measured by Specificity = TN / (TN + FP), used to assess the model's performance in correctly detecting healthy individuals. To evaluate how many of the predicted positive cases are actually positive, the Precision = TP / (TP + FP) metric is used. Given that our dataset is imbalanced, the F1-score is used to balance Precision and Recall, and is calculated as: F1 = 2 × (Precision × Recall) / (Precision + Recall) [23].

2.4 Experimental Environment

A resampling method was used in the modeling analyses. Specifically, a 10-times repeated 10-fold cross-validation technique was applied as the resampling method. Random search was used for hyperparameter optimization.

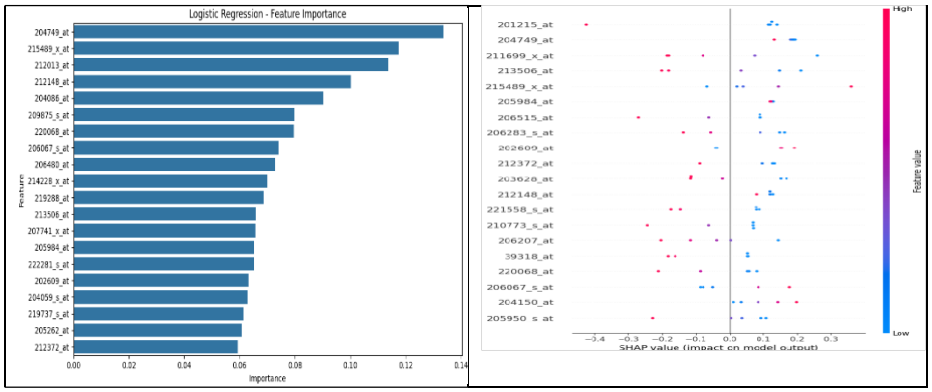
3 Experimental Results

The dataset was classified using different machine learning algorithms under the same conditions, and the classification metrics shown in Table 1 were obtained.

Table 1. Classification Results.

Classification Algorithm	Accuracy	Sensitivity (Recall)	Specificity	Precision	F1
Logistic Regression	0.98	0.96	1	0.98	0.97
Random Forest	0.93	0.87	0.86	0.95	0.89
Gradient Boosting	0.92	0.84	0.97	0.95	0.88
XGBoost	0.95	0.89	0.99	0.97	0.92
LightGBM	0.97	0.94	0.99	0.98	0.94
Catboost	0.92	0.84	0.98	0.95	0.88
Decision Trees	0.90	0.86	0.92	0.91	0.87
SVM	0.97	0.93	1	1	0.96
Naive Bayes	0.98	0.99	0.97	0.97	0.98
KNN	0.84	0.64	0.97	0.92	0.73
LDA	0.87	0.72	0.97	0.93	0.79
AdaBoost	0.97	0.93	1	0.99	0.95
MLP	0.97	0.93	1	0.99	0.95

Feature importance and SHAP values for some algorithms are presented in the figure 2.



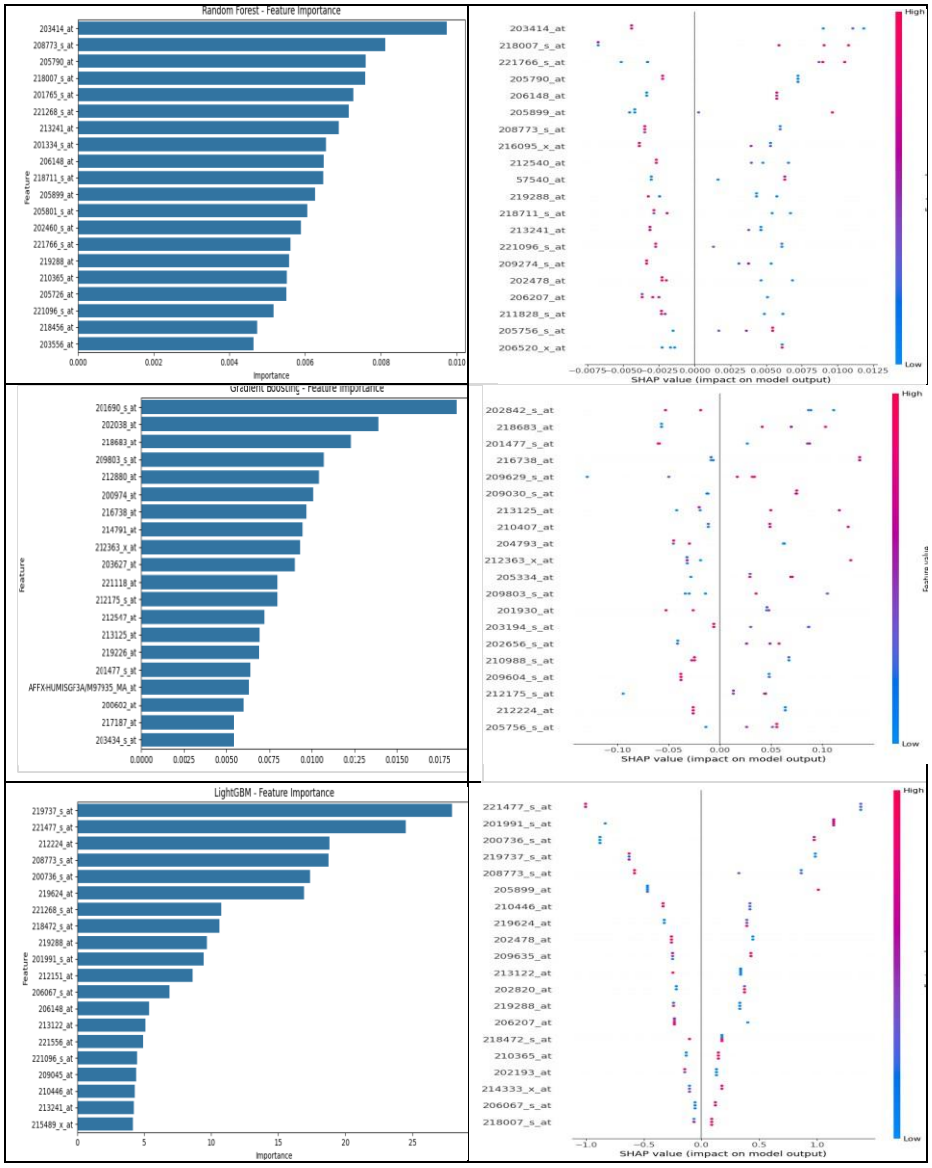


Fig. 2. Feature importance and SHAP values for some algorithms

In the classification performed with LightGBM, 6 out of the 13 clinically known genes were identified among the top 20 most influential features (i.e., probe set names or features). When it comes to SHAP values, this number was found to be 5. LightGBM was followed by Random Forest, which identified 3 genes; Logistic Regression, which identified 2 genes; and Naive Bayes, which identified 1 gene. The top 20 probe set names determined by the other classifiers did not correspond to any of the known genes.

Regarding the feature selectors, Elastic Net identified 2 clinically relevant genes out of 89 selected features. LASSO identified 6 clinically known genes among the 17 features it selected. BORUTA, on the other hand, 10 clinically known genes among the 55 features it selected. Among the top 20 features ranked by feature importance based on these genes, 6 were found to correspond to clinically known genes.
The results of the analysis performed on the gene expression dataset using various feature selection algorithms can be summarized as follows:
The Lasso algorithm successfully identified two key genes from the referenced study: TRIB2 (202478_at) and PRUNE (210988_s_at).
The Boruta algorithm provided a much more comprehensive outcome and successfully identified all 10 important genes mentioned in the study: TRIB2, HOMER3, WT1, IL3RA, PLXNC1, JAG1, PRUNE, FUT4, ALDH1A1, and PELO. Boruta detected 13 probes corresponding to these genes (some genes had multiple probes): 202478_at, 204647_at, 206067_s_at, 206148_at, 206470_at, 209099_x_at, 209586_s_at, 209893_s_at, 210988_s_at, 212224_at, 213241_at, 216268_s_at, and 218472_s_at.
The ElasticNet algorithm was more selective and identified only 6 important genes: WT1, IL3RA, PLXNC1, PRUNE, ALDH1A1, and PELO. The corresponding probes were: 206067_s_at, 206148_at, 206470_at, 209586_s_at, 212224_at, and 218472_s_at.
LightGBM also identified 6 genes, though the composition slightly differed: ALDH1A1, HOMER3, IL3RA, PELO, PLXNC1, and WT1. The probes corresponding to these genes were: 212224_at, 215489_x_at, 206148_at, 218472_s_at, 213241_at, and 206067_s_at.
Random Forest identified fewer genes: PLXNC1 (213241_at), IL3RA (206148_at), and CCNA1 (205899_at).
The Logistic Regression (LR) algorithm only identified HOMER3 (215489_x_at) and WT1 (206067_s_at).
Naive Bayes identified just one gene: HOMER3 (222222_s_at).
These results demonstrate that different feature selection methods highlight different genes, and Boruta provides the most comprehensive feature selection for this dataset.

4 Discussion

In this study, classification was performed on high-dimensional gene expression data associated with Acute Myeloid Leukemia (AML) using various machine learning methods without applying dimensionality reduction techniques. In addition, the results obtained from feature selection methods (LASSO, ElasticNet, and BORUTA) were compared with the 13 genes identified as disease-associated in the original study that presented the dataset.
The results of the study indicate a trade-off between classification performance and the selection of biologically meaningful features. Although the Logistic Regression algorithm achieved the highest classification performance with 0.98 accuracy, 0.96 sensitivity, and 0.97 F1-score, only two of the top 20 features identified by this algorithm (HOMER3 and WT1) were among the known disease-associated genes. This suggests

that high classification performance does not necessarily imply the identification of biologically relevant features.
On the other hand, LightGBM, which exhibited lower classification performance, successfully identified probe sets corresponding to six disease-related genes (ALDH1A1, HOMER3, IL3RA, PELO, PLXNC1, and WT1). This underscores that algorithm selection should be evaluated not only based on model performance but also on biological relevance.
Among the feature selection methods, BORUTA yielded the most comprehensive results by successfully identifying 10 of the 13 genes known from the literature. However, only 6 of these genes were among the top 20 probe sets in terms of importance ranking. This suggests that although BORUTA is highly capable of detecting biologically significant genes, it may not rank them among the top features.
Regularization-based methods such as ElasticNet and LASSO selected 89 and 17 features, respectively, and identified 2 and 6 of the known disease-associated genes. The fact that LASSO identified more relevant genes despite selecting fewer features than ElasticNet suggests that LASSO may be more suitable for gene selection in diseases like AML.
The results of this study indicate that instead of relying on a single feature selection method, combining multiple approaches may be more beneficial when working with gene expression data. Moreover, evaluating both model performance and biological significance is crucial for reaching accurate conclusions in bioinformatics studies.
In future studies, ensemble approaches that integrate the results of different feature selection methods may be explored. Furthermore, a more detailed investigation of the selected genes through biological pathway and functional analyses could contribute to a better understanding of disease mechanisms.

5 RESULTS

In this study, classification and biomarker identification were performed on Acute Myeloid Leukemia gene expression data using various machine learning algorithms and feature selection methods. Logistic Regression achieved the highest classification performance (0.98 accuracy) but was limited in identifying genes known to be associated with the disease. In contrast, the LightGBM algorithm identified more disease-associated genes despite having lower classification performance. Among the feature selection methods, BORUTA showed the best performance by successfully identifying 10 out of 13 disease-associated genes. These results highlight a trade-off between classification performance and biological relevance in machine learning models and emphasize the importance of jointly evaluating different algorithms and feature selection methods in gene expression studies.

References

[1] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray," *Science (80-.).*, vol. 270, no. 5235, pp. 467–470, Oct. 1995, doi: 10.1126/science.270.5235.467.

[2] H. Döhner, D. J. Weisdorf, and C. D. Bloomfield, "Acute Myeloid Leukemia," *N. Engl. J. Med.*, vol. 373, no. 12, pp. 1136–1152, Sep. 2015, doi: 10.1056/NEJMra1406184.

[3] C. Evers *et al.*, "Molecular definition of chromosome arm 5q deletion end points and detection of hidden aberrations in patients with myelodysplastic syndromes and isolated del(5q) using oligonucleotide array CGH," *Genes, Chromosom. Cancer*, vol. 46, no. 12, pp. 1119–1128, Dec. 2007, doi: 10.1002/gcc.20498.

[4] D. L. Stirewalt *et al.*, "Identification of genes with abnormal expression changes in acute myeloid leukemia," *Genes, Chromosom. Cancer*, vol. 47, no. 1, pp. 8–20, Jan. 2008, doi: 10.1002/gcc.20500.

[5] M. Frasca, R. Francese, M. Risi, and G. Tortora, "A Deep Learning and Genetic Algorithm Based Feature Selection Processes on Leukemia Data," Jan. 25, 2022. doi: 10.21203/rs.3.rs-1247212/v1.

[6] R. Francese, M. Frasca, M. Risi, and G. Tortora, "A deep learning and genetic algorithm based feature selection processes on Leukemia Data," in *2022 26th International Conference Information Visualisation (IV)*, IEEE, Jul. 2022, pp. 412–417. doi: 10.1109/IV56949.2022.00074.

[7] S. Jacophine Susmi, H. Khanna Nehemiah, A. Kannan, and Jabez Christopher, "Relevant Gene Selection and Classification of Leukemia Gene Expression Data," in *Emerging Research in Computing, Information, Communication and Applications*, Singapore: Springer Singapore, 2016, pp. 503–510. doi: 10.1007/978-981-10-0287-8_47.

[8] A. S. Khanam *et al.*, "Classification of acute myeloid leukemia using convolutional neural network," *IET Conf. Proc.*, vol. 2023, no. 11, pp. 253–258, Sep. 2023, doi: 10.1049/icp.2023.1788.

[9] A. Magenta, P. Fasanaro, S. Romani, V. Di Stefano, M. C. Capogrossi, and F. Martelli, "Protein Phosphatase 2A Subunit PR70 Interacts with pRb and Mediates Its Dephosphorylation," *Mol. Cell. Biol.*, vol. 28, no. 2, pp. 873–882, Jan. 2008, doi: 10.1128/MCB.00480-07.

[10] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, 2000. doi: 10.1002/0471722146.

[11] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Syst. Appl.*, vol. 237, p. 121549, Mar. 2024, doi: 10.1016/j.eswa.2023.121549.

[12] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Stat.*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.

[13] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[14] M. R. Machado, S. Karray, and I. T. de Sousa, "LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry," in

2019 14th International Conference on Computer Science & Education (ICCSE), IEEE, Aug. 2019, pp. 1111–1116. doi: 10.1109/ICCSE.2019.8845529.

[15] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Jun. 2017, doi: 1706.09516.

[16] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[17] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[18] Vikramkumar, V. B, and Trilochan, "Bayes and Naive Bayes Classifier," Apr. 2014, doi: https://doi.org/10.48550/arXiv.1404.0933.

[19] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

[20] N. A. A. Shashoa, N. A. Salem, I. N. Jleta, and O. Abusaeeda, "Classification depend on linear discriminant analysis using desired outputs," in *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, IEEE, Dec. 2016, pp. 328–332. doi: 10.1109/STA.2016.7952041.

[21] A. H. A. Abdel-aziem and T. H. M. Soliman, "A Multi-Layer Perceptron (MLP) Neural Networks for Stellar Classification: A Review of Methods and Results," *Int. J. Adv. Appl. Comput. Intell.*, vol. 3, no. 2, pp. 29–37, 2023, doi: 10.54216/IJAACI.030203.

[22] P. Beja-Battais, "Overview of AdaBoost : Reconciling its views to better understand its dynamics," Oct. 2023, doi: 10.48550/arXiv.2310.18323.

[23] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," Apr. 2024, doi: 10.1162/tacl_a_00675.

Önceden Eğitilmiş Yapay Zeka Modelleri ile Miyokard Enfarktüsünün Tahmini ve Sınıflandırılması

Cemil Çolak[1][0000-0001-5406-098X], **Abdulvahap Pınar**[2][0000-0002-3662-2579], Ahmet Kadir Arslan[1][0000-0001-8626-9542], Fatma Hilal Yağın[3][0000-0002-9848-7958]

[1]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya 44280, Turkey
[2]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye
[3]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Ozal University, Malatya, Turkey.

ÖZET

Amaç: Bu çalışma, miyokard enfarktüsünün tahmini ve sınıflandırılmasında önceden eğitilmiş yapay zeka modellerinin kullanımına ilişkin mevcut araştırmaları kapsamlı bir şekilde incelemeyi amaçlamaktadır. Çalışmada, literatürde belgelenen çeşitli modeller, metodolojiler ve bulgular vurgulanmıştır.

Gereç ve Yöntem: Bu araştırma, miyokard enfarktüsü tahmini ve sınıflandırmasında önceden eğitilmiş yapay zeka modellerinin kullanımını inceleyen bir literatür taramasıdır. Akademik kaynaklardaki miyokard enfarktüsü ve yapay zekâ uygulamalarına ilişkin araştırma yayınları kullanılmıştır. Yöntem, yapay zekâ modellerini (özellikle BERT, ResNet ve XGBoost), veri kaynaklarını (EHR, ECG ve CMR), teknikleri (transfer öğrenme, derin öğrenme ve makine öğrenimi) kapsamıştır. Literatür taramasının kalitesi, "Narrative Review Articles Değerlendirme Ölçeği (SANRA)" ile değerlendirilmiştir.

Bulgular: Önceden eğitilmiş yapay zekâ modelleri, özellikle CNN ve transformer mimarileri, miyokard enfarktüsünün tahmini ve sınıflandırılmasında önemli potansiyele sahiptir. Bu modeller, yüksek doğruluk, hızlı tespit ve kişiselleştirilmiş yaklaşımlar sunarken, veri ve hesaplama kaynaklarının kullanımını optimize etmektedir.

Sonuç: Bu çalışma, miyokard enfarktüsünün tahmini ve sınıflandırılmasında önceden eğitilmiş yapay zekâ modellerinin kullanımını detaylı bir şekilde incelemiştir. Bulgular, özellikle CNN ve transformer tabanlı modellerin, erken teşhis ve kişiselleştirilmiş tedavi yaklaşımlarıyla miyokard enfarktüsü tanı ve tedavisinde önemli avantajlar sunduğunu göstermektedir. Ancak, veri kalitesi, model yorumlanabilirliği ve kapsamlı doğrulama gereksinimi gibi sorunların klinik uygulamada ele alınması gerekmektedir. Gelecek araştırmalar, bu kısıtlamaları ele alarak yapay zekâ tabanlı çözümlerin kardiyolojide pratik ve etik uygulanmasına odaklanırsa, hasta sonuçlarını iyileştirebilir ve hassas tıp alanında yeni bir çağ başlatabilir. Bu inceleme, SANRA kriterleri kullanılarak oluşturulmuştur.

Anahtar kelimeler: Miyokard Enfarktüsü, Transfer Öğrenme, Yapay Zekâ, Anlatıya Dayalı Derleme Makaleleri Değerlendirme Ölçeği

Pre-trained Artificial Intelligence Models in the Prediction and Classification of Myocardial Infarction

Cemil Çolak[1[0000-0001-5406-098X]], **Abdulvahap Pınar**[2[0000-0002-3662-2579]], Ahmet Kadir Arslan[1[0000-0001-8626-9542]], Fatma Hilal Yağın[3[0000-0002-9848-7958]]

[1]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya 44280, Turkey

[2]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye

[3]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Ozal University, Malatya, Turkey.

ABSTRACT

Aim: This work seeks to provide a thorough examination of the existing research concerning the use of pre-trained artificial intelligence models for the prediction and categorization of myocardial infarction, emphasizing the many models, methodology, and findings documented in the literature.

Materials and Methods: This research is a literature review that looks at the use of pre-trained artificial intelligence models for myocardial infarction prediction and classification. We utilized research publications from academic sources dealing with myocardial infarction and artificial intelligence applications. The technique covered artificial intelligence models (particularly BERT, ResNet, and XGBoost), data sources (EHR, ECG, and CMR), methods (transfer learning, deep learning, and machine learning), and the quality of the literature review was evaluated using the Scale for the Assessment of Narrative Review Articles (SANRA).

Results: Pre-trained artificial intelligence models, particularly CNN and transformer architectures, have significant promise in the prediction and categorization of myocardial infarction. These models provide elevated accuracy, prompt detection, and tailored methodologies, while optimizing data and computing resource use.

Conclusion: This research thoroughly studied the use of pre-trained artificial intelligence models to predict and classify myocardial infarction. Our findings indicate that models based on CNN and transformer topologies, in particular, provide considerable benefits in the diagnosis and treatment of myocardial infarction, with the possibility for early detection and a tailored strategy. However, concerns like as data quality, model interpretability, and the requirement for thorough validation must be addressed in clinical practice. Future research addressing these constraints and concentrating on the practical and ethical implementation of AI-based solutions in cardiology has the potential to enhance patient outcomes and usher in a new age of precision medicine. This review was created using the Scale for the Assessment of Narrative Review Articles (SANRA) criteria, which included the topic's relevance, defined goals, a literature search, and acceptable source citing.

Keywords: Myocardial İnfarction, Transfer Learning, Artificial İntelligence, Scale For The Assessment of Narrative Review Articles,

1. Giriş

Genellikle kalp krizi olarak bilinen miyokard enfarktüsü (MI), kalp kasına yetersiz kan akışının olduğu kritik bir tıbbi durumdur. Genellikle kalbi besleyen koroner arterlerden birinde meydana gelen ani bir tıkanıklıktan kaynaklanır. Bu tıkanıklığın başlıca nedeni, zaman içinde arter duvarlarında biriken ve "plak" olarak bilinen yağ birikintilerinin kırılmasıyla oluşan bir trombüstür. Bunun sonucunda kalp dokusu oksijensiz kalır ve potansiyel olarak onarılamaz doku hasarına ve nekroza yol açar. Bireyler arasında değişken ilerleme gösteren bu hastalık bazen orta şiddette semptomlarla kendini gösterirken, bazen de ani ölüme yol açacak kadar şiddetli olabilen bir hastalıktır (1). Kardiyovasküler Hastalık Yönetiminde Yapay Zekanın (YZ) Artan Rolü ise, Son yıllarda YZ, özellikle Makine Öğrenmesi (ML) ve Derin Öğrenme (DÖ) yöntemleri sayesinde, kardiyovasküler hastalıkların tanı ve tedavisinde uzman hekimlere destek sağlamıştır. YZ sistemleri, büyük ve karmaşık tıbbi veri kümelerini inceleyerek, klinisyenlerin gözden kaçırabileceği belirli kalıpları ortaya çıkarma potansiyeline sahiptir. Kalp krizi riskinin tahmin edilmesi, koroner arter hastalığının non-invaziv yöntemlerle teşhisi, ritim bozukluklarının tespiti ve kalp yetmezliği olan bireyler için kişiselleştirilmiş tedavi seçeneklerinin belirlenmesi gibi uygulamalarda YZ sistemleri kullanılır. Ayrıca, elektronik sağlık kayıtları, tıbbi görüntüler ve giyilebilir teknolojilerden elde edilen verilerin YZ destekli analizleri, daha isabetli ve etkili klinik kararların alınmasına katkı sağlamaktadır (2). Önceden eğitilmiş YZ modelleri, geniş veri kümeleriyle eğitildikleri için MI tahmini ve sınıflandırması gibi özel alanlarda yeniden kullanılabilir. Buna göre, az sayıda etiketli veriyle bile yüksek doğrulukta sonuçlar elde etmek mümkündür. Bu yaklaşım, transfer öğrenmesi olarak bilinir ve sağlık alanında zaman ve kaynak tasarrufu sağlar. MI tahmini ve sınıflandırılmasında önceden eğitilmiş YZ modellerinin kullanımı, klinik uygulamalarda daha doğru ve etkili karar destek sistemlerinin geliştirilmesine katkı sağlar. Bu modeller, özellikle sağlık alanında sıkça karşılaşılan veri eksikliği ve sınırlı hesaplama kaynakları gibi zorlukların aşılmasında önemli bir avantaj sunmaktadır (3). Bu derleme, mevcut literatürde yer alan model türlerini, kullanılan yöntemleri ve elde edilen sonuçları ele alarak, MI tahmini ve kategorizasyonunda önceden eğitilmiş YZ modellerinin potansiyelini kapsamlı biçimde ortaya koymayı amaçlamaktadır.

2. Bulgular

Miyokard Enfarktüsü Tahmini için Önceden Eğitilmiş Yapay Zeka Modelleri

ME çalışmaların temel hedefi, kalp krizi meydana gelmeden önce yüksek risk taşıyan bireyleri belirleyebilmektir.

- **Transformer Tabanlı Modeller**
- **Gradient Boosting Decision Trees (GBDT)**
- **XGBoost ve Sinir Ağları**
- **Maskeli Otomatik Kodlayıcı ve Çok Modlu Kontrastif Öğrenme**

Tablo 1. Miyokard enfarktüsü tahmini için eğitilmiş modeller

Model / Yaklaşım	Açıklama	Kaynak (Ref)
Convolutional Neural Networks (CNNs)	Geniş resim veri setleri üzerinde önceden eğitilmiş; EKG görüntüleri ve kardiyak görüntüleme verileri ile ME sınıflamasında etkili bulunmuştur.	(Khan E,2024)
ECGNet (Attention Mekanizmalı ResNet)	Çok ölçekli ResNet ve dikkat mekanizması içerir; Elektrokardiyografi (EKG) verileriyle ME'nün tespiti ve lokalizasyonunu yüksek doğrulukla yapar.	(Cao Y,2022)
Derin Segmentasyon Ağı (LGE-CMR görüntüleri için)	Klinik Geç Gadolinyum Tutulumu-Kardiyak Manyetik Rezonans Görüntüleme ME ve mikrovasküler tıkanıklıkların otomatik segmentasyonu için geliştirilmiş, yüksek düzeyinde doğruluk sağlamaktadır.	(Schwab M,2025)
Transformer Modelleri	ME sınıflandırması için kullanılmakta; metin tabanlı sorgularla modellenmiştir, özellikle yapılandırılmamış verilerde (örn. hasta anlatıları) "zero-shot" sınıflama sunar.	(Cai M,2025)
SuPreME (Multimodal EKG Temsil Öğrenimi)	Denetimli ön eğitim (supervised pre-training) çerçevesi; çok modlu EKG verileriyle farklı ME türlerini yapılandırılmamış metinlerden sıfır atışla sınıflar.	(Cai M,2025)
XGBoost + Derin Öğrenme Modelleri	4. Evrensel ME Tanımı'na göre ME türlerini sınıflar; yüksek AUC değerleri elde edilmiştir.	(Khan E,2024)
Sınıflama ve Regresyon Modelleri	Kalp krizi ve mikrovasküler hasarın varlığını belirlemek ve hasar oranını tahmin etmek amacıyla kullanılmıştır; kısa vadeli risk değerlendirmesinde de etkili olduğu ifade edilmiştir.	(Khan E,2024)
Topluluk Öğrenme (Ensemble Learning) Yöntemleri	Klinik verilerle birleştirilerek ME türlerinin ve sonuçlarının daha hassas sınıflandırılması hedeflenmiştir.	(Rahman SSMM)

3. Sonuç

Miyokard enfarktüsü tahmini için önceden eğitilmiş yapay zeka modelleri, modelin kullanım amacı, ilgili performans metrikleri ve modellere ilişkin ana bulgular tablo 2'de sunulmuştur.

Tablo 2. Miyokard enfarktüsü tahmin ve sınıflandırmasında önceden eğitilmiş yapay zeka modellerinin karşılaştırmalı analizi

Alıntı yapılan çalışmalar	Model Tipi	Eğitim Öncesi Veri	ME Görevi için Veri Kaynakları	Metodoloji	Görev	Temel Performans Metrikleri	Ana Bulgular
(Khera R,2021)	XGBoostve Sinir Ağları	Belirtilmemiş	Ulusal kayıt verileri	ML modellerini geliştirdi ve doğruladı	Tahmin	C-istatistiği 0.90 (en iyi ML modeli)	ML modelleri AME sonrası hastane içi mortalite ayrımını önemli ölçüde iyileştirmedi
(Choi A,2022)	Lojistik Regresyon ve XGBoost	Belirtilmemiş	Ulusal EMS kayıt verileri	Tahmin modellerini geliştirdi ve karşılaştırdı	Tahmin	Daha yüksek tahmin performansı	ML modelleri zamanında AMI yönetimi gerektiren hastaların tanımlanmasını iyileştirebilir
(Alzubaidi L,2021)	CNN	30 milyon tıbbi görüntü üzerinde eğitilmiş	Tıbbi görüntüleme verileri	Önceden eğitilmiş bir CNN modeli önerdi	Sınıflandırma	Tıbbi görüntü verilerinde genelleme hedefi	MedNet tıbbi görüntü sınıflandırmasında gelecekteki araştırmalar için bir temel görevi görebilir
(Khan E, 2021)	Makine Öğrenimi	Elektronik Sağlık Kaydı (EHR) verileri	ED hasta verileri	ME ve miyokardiyal yaralanmayı fenotiplendiren ML	Sınıflandırma ve Tahmin	Tanısal ve prognostik	ML modelleri ME'yı fenotiplendirebilir ve kısa vadeli olayları tahmin edebilir
(Choi A,2022)	XGBoost ve Çok Katmanlı Perseptron	Ulusal Acil Tıbbi Hizmetler kayıt verileri	Hastane öncesi veriler	Tahmin modelleri geliştirdi	Tahmin	Daha yüksek tahmin performansı	Hastane öncesi verilere dayalı tahmin modelleri AME tanımlamasını iyileştirebilir
(Joloudari JH, 2022)	KNN, DT, SVM, NB, RF, ANN, BPNN, HMM, GMM, PNN, MLP	Physikalisch-Technische Bundesanstalt (PTB) veritabanı diğer veri setleri	EKG sinyalleri	ML yöntemlerinin sistematik literatür taraması	Sınıflandırma	Doğruluk oranları %76.67'den %99.72'ye	Çeşitli ML ve DL modelleri değişen doğruluk ile ME tanısı için kullanılmıştır
(Baloglu UB, 2019)	CNN (11-katmanlı)	Belirtilmemiş	EKG sinyalleri	ME tespiti için bir CNN önerildi.	Tespit	ME tespiti için sonuçlar rapor edildi.	CNN, ME tespiti için kullanıldı.
(Acharya UR,2018)	CNN (10-katmanlı)	Belirtilmemiş	EKG sinyalleri	ME tespiti için bir CNN önerildi.	Tespit	ME tespiti için sonuçlar rapor edildi.	CNN, ME tespiti için kuldllanıı.
(Reasat T, 2017)	CNN	Belirtilmemiş	Üç derivasyonlu EKG sinyal segmentleri	Bir indüksiyon modülüne girdi.	Tespit	ME tespiti için sonuçlar rapor edildi.	CNN, ME tespiti için kullanıldı.
(Fatimah B, 2021)	Dalgacık Dönüşümü + KNN	Belirtilmemiş	EKG sinyali	Sinyali alt bantlara ayrıştırdı ve özellikler çıkarıldı.	Tanı	ME tespiti için sonuçlar rapor edildi.	KNN, ME tanısı için kullanıldı.

Bu çalışma neticesinde, CNN ve transformer tabanlı mimarilerle geliştirilen öncede eğitilmiş yapay zeka modelleri, ME'nin erken tanı ve tahmininde yüksek doğruluk düzeyleri sergileyerek kayda değer bir potansiyel ortaya koymaktadır. Veri ve hesaplama kaynaklarının etkin kullanımı sayesinde bu modeller, kişiselleştirilmiş tanı ve tedavi yaklaşımlarına imkân sağlamaktadır. Bununla birlikte, model yorumlanabilirliği, veri kalitesi ve erişilebilirliği, söz konusu teknolojilerin sağlık hizmetlerine entegrasyonu açısından önemlidir.

Kaynaklar

1. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). Circulation. 2018;138(20):e618- e51.
2. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. Journal of the American College of Cardiology. 2018;71(23):2668-79.
3. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine. 2019;25(1):65-9.
4. What is a Heart Attack? | American Heart Association, April 1, 2025 [Available from: https://www.heart.org/en/health-topics/heart-attack/about-heart-attacks.
5. Ojha N DA. Myocardial Infarction 2023 Aug 8 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK537076/.
6. Liu R, Wang M, Zheng T, Zhang R, Li N, Chen Z, et al. An artificial intelligence-based risk prediction model of myocardial infarction. BMC bioinformatics. 2022;23(1):217.
7. Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA cardiology. 2021;6(6):633-41.
8. Khan E, Lambrakis K, Liao Z, Gerlach J, Briffa T, Cullen L, et al. Machine-learning for phenotyping and prognostication of myocardial infarction and injury in suspected acute coronary syndrome. JACC: Advances. 2024;3(9_Part_2):101011.
9. Cao Y, Liu W, Zhang S, Xu L, Zhu B, Cui H, et al. Detection and localization of myocardial infarction based on multi-scale resnet and attention mechanism. Frontiers in Physiology. 2022;13:783184.
10. Schwab M, Pamminger M, Kremser C, Haltmeier M, Mayr A. Deep Learning Pipeline for Fully Automated Myocardial Infarct Segmentation from Clinical Cardiac MR Scans. arXiv preprint arXiv:250203272. 2025.
11. Cai M, Jiang J, Huang W, Liu C, Arcucci R. SuPreME: A Supervised Pre-training Framework for Multimodal ECG Representation Learning. arXiv preprint arXiv:250219668. 2025.
12. Rahman SSMM, Chen Z, Lalande A, Decourselle T, Cochet A, Pommier T, et al.
13. Choi A, Kim MJ, Sung JM, Kim S, Lee J, Hyun H, et al. Development of prediction models for acute myocardial infarction at prehospital stage with machine learning based on a nationwide database. Journal of Cardiovascular Development and Disease. 2022;9(12):430.
14. Alzubaidi L, Santamaría J, Manoufali M, Mohammed B, Fadhel MA, Zhang J, et al. MedNet: pre-trained convolutional neural network model for the medical imaging tasks. arXiv preprint arXiv:211006512.2021
15. Khan E, Lambrakis K, Liao Z, Gerlach J, Briffa T, Cullen L, et al. Machine-learning for phenotyping and prognostication of myocardial infarction and injury in suspected acute coronary syndrome. JACC: Advances. 2024;3(9_Part_2):101011.
16. Joloudari JH, Mojrian S, Nodehi I, Mashmool A, Zadegan ZK, Shirkharkolaie SK, et al. Application of artificial intelligence techniques for automated detection of myocardial infarction: a review. Physiological Measurement. 2022;43(8):08TR1
17. Baloglu UB, Talo M, Yildirim O, San Tan R, Acharya UR. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern recognition letters. 2019;122:23-30.
18. Fatimah B, Singh P, Singhal A, Pramanick D, Pachori RB. Efficient detection of myocardial infarction from single lead ECG signal. Biomedical Signal Processing and Control. 2021;68:102678.
19. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in biology and medicine. 2018;100:270-8.
20. Reasat T, Shahnaz C, editors. Detection of inferior myocardial infarction using shallow convolutional neural networks. 2017 IEEE region 10 humanitarian technology conference (R10-HTC); 2017: IEEE.

Biostatapps: Bir Kullanıcı Dostu Veri Bilimi ve Yapay Zekâ Temelli Yazılım Platformu

Ahmet Kadir ARSLAN, Cemil ÇOLAK, Hasan UCUZAL, Emek GÜNDOĞAN, Şeyma YAŞAR, Fatma Hilal YAĞIN, Zeynep KÜÇÜKAKÇALI, İpek BALIKÇI ÇİÇEK, Harika G.G. BAĞ

İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi AD., Malatya, Türkiye E-posta: arslan.ahmet@inonu.edu.tr

Amaç: Gününüzde veri bilini ve yapay zekâ yaklaşımları, akademik araştırmalar ve klinik uygulanalar için kritik bir rol oynamaktadır. Ancak, bu alanlarda geliştirilen araçların karmaşıklığı ve teknik gereksinimleri, özellikle kodlana deneyini sınırlı olan kullanıcılar için erişin engeli oluşturmaktadır. Bu çalışma, İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişini Anabilin Dalı bünyesinde geliştirilen BiostatApps platformunu tanıtmakta ve bu platformun veri bilini ve yapay zekâ tabanlı analizlerde sunduğu olanakları ortaya koymaktadır. Amaç, kullanıcıların kolay erişebileceği, acık kaynaklı ve pratik araçlar vasıtasıyla sağlık bilimleri alanında yürütülen araştırmaların kalitesini ve verimliliğini artırmaktır.

Yöntem: BiostatApps (https://biostatapps.inonu.edu.tr) platformunda bulunan yazılınlar, R (Shiny) ve Python (Dash, Streanlit) tabanlı acık kaynaklı kütüphaneler kullanılarak geliştirilmiştir. Kullanıcı arayüzleri, teknik bilgisi sınırlı olan araştırmacıların bile etkileşimli grafikler oluşturabilmesi, istatistiksel testler uygulayabilmesi ve makine öğrenmesi modellerini eğitebilmesi için optimize edilmiştir. Öne çıkan uygulanalar arasında, örneklen büyüklüğü ve güç analizi, bilgi keşfi süreci, çeşitli tematik istatistiksel/veri bilini araçları, klinik karar destek sistemleri ve görüntü/video işlene/sınıflandıran vb. bulunaktadır.

Bulgular: Platform, başta sağlık alanındaki araştırmacılar olmak üzere tün bilin insanlarının analiz/modellene ihtiyaçlarına çözüm sunmakta ve öğrencilerin uygulanalı veri bilini/yapay zekâ eğitimlerini desteklemektedir. Özellikle programlana bilgisi olmayan kullanıcılar için sezgisel ve rehberli bir analiz deneyini sağlamaktadır. Kullanıcı geri bildirimlerine göre, tez danışmanlık süreçlerinde, klinik araştırma projelerinde ve bilimsel yayın hazırlıklarında etkin bicinde kullanılmaktadır.

Sonuç: BiostatApps, çağdaş veri bilini uygulamalarını ve yapay zekâ yaklaşımlarını bir araya getiren, erişilebilir ve kullanıcı dostu bir platformdur. Sağlıkta dijital dönüşün sürecine katkı sağlamakla birlikte, istatistik/bilişin okuryazarlığın artırılması ve araştırma kalitesinin iyileştirilmesi anacıyla önemli bir eğitin ve analiz aracıdır. Akademik ve klinik ortamlarda kullanılabilirliği ile fark yaratan bu platform, aynı zamanda acık bilin ilkeleri doğrultusunda sürekli geliştirilmekte ve genişletilmektedir.

Anahtar Kelimeler: Biostatapps, Biyoistatistik, Veri Bilini, Yapay Zekâ, Tıp Bilişini.

Kaynaklar

1. Arslan, A. K., Yasar, S, Colak, C., C Yologlu, S. (2018). WSSPAS: An Interactive Web Application for Sanple Size and Power Analysis with R Using Shiny. Turkiye Klinikleri Journal of Biostatistics, 10(3), 224-246. doi:10.5336/biostatic.2018- 62787.
2. Ucuzal, H., YAŞAR, Ş., C Çolak, C. (2019, October). Classification of brain tunor types by deep learning with convolutional neural network on nagnetic resonance inages using a developed web-based interface. In 2019 3rd International Synposiun on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-5). IEEE.
3. Arslan, A. K., Erdil, N., Guldogan, E., Colak, C., Akca, B., C Colak, M. C. (2023). Prediction of postcoronary artery bypass grafting atrial fibrillation: POAFRiskScore tool. The Thoracic and cardiovascular surgeon, 71(04), 282-290.
4. Yaşar, Ş., Arslan, A. K., Polat, B. B., Melekoğlu, R., Çolak, C., C Yoloğlu, S. (2024). A novel software for nethod conparison: MCS (nethod conparison software)— assessing agreenent between estinated fetal weights calculated by Hadlock I–V fornulas and birth weight. Archives of Gynecology and Obstetrics, 310(5), 2439- 2452.
5. Arslan, A. K., Yaşar, Ş., C Çolak, C. (2019, Septenber). An Intelligent Systen for the Classification of Lung Cancer Based on Deep Learning Strategy. In 2019 International Artificial Intelligence and Data Processing Synposiun (IDAP) (pp. 1- 4). IEEE.

Biostatapps: A User-Friendly Software Platform Based on Data Science and Artificial Intelligence

Ahmet Kadir ARSLAN, Cemil ÇOLAK, Hasan UCUZAL, Emek GÜLDOčAN, Şeyma YAŞAR, Fatma Hilal YAčIN, Zeynep KÜÇÜKAKÇALI, İpek BALIKÇI ÇİÇEK, Harika G.G. BAč

İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye

E-mail: arslan.ahmet@inonu.edu.tr

Aim: Nowadays, data science and artificial intelligence approaches play a critical role in acadenic research and clinical applications. However, the conplexity and technical requirenents of the tools developed in these fields often constitute a barrier to access, especially for users with linited coding experience. This study introduces the BiostatApps platforn developed within the Departnent of Biostatistics and Medical Infornatics at İnönü University Faculty of Medicine and presents the opportunities it offers for data science and AI-based analyses. The ain is to enhance the quality and efficiency of research conducted in health sciences through easily accessible, open-source, and practical tools.

Methods: The software tools available on the BiostatApps platforn (https://biostatapps.inonu.edu.tr) were developed using open-source libraries based on R (Shiny) and Python (Dash, Streanlit). The user interfaces have been optinized to enable researchers with linited technical knowledge to generate interactive graphics, perforn statistical tests, and train nachine learning nodels. Notable applications include sanple size and power analysis, knowledge discovery processes, various thenatic statistical/data science tools, clinical decision support systens, and inage/video processing/classification, anong others.

Results: The platforn provides solutions to the analysis/nodeling needs of all scientists, particularly researchers in the health donain, and supports students in practical data science and artificial intelligence education. It offers an intuitive and guided analytical experience, especially for users without progranning skills. According to user feedback, it has been effectively utilized in thesis supervision processes, clinical research projects, and scientific publication preparation.

Conclusion: BiostatApps is an accessible and user-friendly platforn that brings together contenporary data science applications and AI approaches. In addition to contributing to the process of digital transfornation in healthcare, it serves as a significant educational and analytical tool for inproving statistical/infornatics literacy and enhancing research quality. With its usability in acadenic and clinical environnents, the platforn stands out and continues to be developed and expanded in line with the principles of open science. **Keywords:** Biostatapps, Biostatistics, Data Science, Artificial Intelligence, Medical Infornatics.

References

1. Arslan, A. K., Yasar, S, Colak, C., C Yologlu, S. (2018). WSSPAS: An Interactive Web Application for Sanple Size and Power Analysis with R Using Shiny. Turkiye Klinikleri Journal of Biostatistics, 10(3), 224-246. doi:10.5336/biostatic.2018- 62787.
2. Ucuzal, H., YAŞAR, Ş., C Çolak, C. (2019, October). Classification of brain tunor types by deep learning with convolutional neural network on nagnetic resonance inages using a developed web-based interface. In 2019 3rd International Synposiun on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-5). IEEE.
3. Arslan, A. K., Erdil, N., Guldogan, E., Colak, C., Akca, B., C Colak, M. C. (2023). Prediction of postcoronary artery bypass grafting atrial fibrillation: POAFRiskScore tool. The Thoracic and cardiovascular surgeon, 71(04), 282-290.
4. Yaşar, Ş., Arslan, A. K., Polat, B. B., Melekoğlu, R., Çolak, C., C Yoloğlu, S. (2024). A novel software for nethod conparison: MCS (nethod conparison software)— assessing agreenent between estinated fetal weights calculated by Hadlock I–V fornulas and birth weight. Archives of Gynecology and Obstetrics, 310(5), 2439- 2452.
5. Arslan, A. K., Yaşar, Ş., C Çolak, C. (2019, Septenber). An Intelligent Systen for the Classification of Lung Cancer Based on Deep Learning Strategy. In 2019 International Artificial Intelligence and Data Processing Synposiun (IDAP) (pp. 1- 4). IEEE.

Key Challenges in Developing Explainable Ensemble Learning Models for Classification of Atherosclerotic Heart Diseases

Ahmet Kadir ARSLAN, Hasan UCUZAL, Cemil ÇOLAK

İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye
E-mail: arslan.ahmet@inonu.edu.tr

Abstract

Atherosclerotic heart disease (ASHD) as one of the global causes of morbidity and mortality has a pressing need for valid and interpretable diagnostic models. Elegant ensemble learning techniques like XG Boost, and Random Forest help in significantly increasing classification accuracy, however they operationalize as "black box" models, lacking clinical trust and adoption. This article presents a close look at some key issues in building explainable ensemble models for ASHD; namely, balancing accuracy with interpretability, use of methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), and population specific discrepancies. New studies report models with accuracy up to 97%, but those complexities make it difficult to understand decision making processes and methods such as Born-Again Tree Ensembles are used to keep the accuracy level but at the same time increase the transparency. Practical implementations used in early diagnosis include electronic health records and socio-demographic data, while stacking ensembles report 91% accuracy. SHAP-driven insights inform more personalized medicine; predictors such as hemoglobin levels would otherwise be selected arbitrarily for the sake of intervention. Clinical decision support systems bring these models to life by providing explainable predictions which promise to increase trust. Nevertheless, difficulties remain in adapting models to sex-based risk factors, flaws of racial/ethnic nature, and age-related risks. Comparative analyses indicate that boosting methods such as XGBoost perform better than bagging, but at increased cost in process power. Future research will need emphasis on frameworks that balance between accuracy and interpretability, healthcare disparities and promote clinical applicability. Solving these challenges will make it possible for ensemble models to transform ASHD diagnostics as well make possible early interventions and equitable care for patients.

Keywords:
Atherosclerotic heart disease, Ensemble learning, Explainable AI, SHAP values, Clinical decision support.

Acknowledgement: This research was supported by TUBITAK project number 323S042. We thank TUBITAK for this support.

References

Miao, K. H., Miao, J. H., & Miao, G. J. (2016). Diagnosing Coronary Heart Disease using Ensemble Machine Learning.*International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2016.071004

Gulhane, M., & Sajana, T. (2024). Integrating optimized ensemble learning and explainable AI for accurate heart disease diagnosis.*Journal of Information and Optimization Sciences*. https://doi.org/10.47974/jios-1832

Magalhães, L., & Gonçalves, E. C. (2024).*Enhancing the Interpretability of Cardiovascular Disease Classifiers using Born-Again Tree Ensembles*. https://doi.org/10.5753/kdmile.2024.243749

Mienye, I. D., & Jere, N. (2024). Optimized Ensemble Learning Approach with Explainable AI for Improved Heart Disease Prediction.*Information*. https://doi.org/10.3390/info15070394

Rezk, N. G., Alshathri, S., Sayed, A., Hemdan, E. E., & El-Behery, H. (2024). XAI-Augmented Voting Ensemble Models for Heart Disease Prediction: A SHAP and LIME-Based Approach.*Bioengineering*. https://doi.org/10.3390/bioengineering11101016

Sultan, S., Javaid, N., Alrajeh, N., & Aslam, M. (2025). Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross- Validation: A Symmetric Approach.*Symmetry*. https://doi.org/10.3390/sym17020185

Hayat, M. T., Javaid, N., Nazir, A., Alrajeh, N., Shafiq, M., & Choi, J.-G. (2024).*A Deep Learning Approach for Prediction of Heart Disease using Ensemble Learning Techniques*. https://doi.org/10.21203/rs.3.rs-4526000/v1

Khan, H., Javaid, N., Bashir, T., Akbar, M., Alrajeh, N., & Aslam, S. (2024). Heart disease prediction using novel Ensemble and Blending based Cardiovascular Disease Detection Networks: EnsCVDD-Net and BlCVDD-Net.*IEEE Access*. https://doi.org/10.1109/access.2024.3421241

Miranda, E., Adiarto, S., Bhatti, F. M., Zakiyyah, A. Y., Aryuni, M. Kom. M., & Bernando, C. (2023). Understanding Arteriosclerotic Heart Disease Patients Using Electronic Health Records: A Machine Learning and Shapley Additive exPlanations Approach.*Healthcare Informatics Research*. https://doi.org/10.4258/hir.2023.29.3.228

Bye, A., Wiig-Fisketjon, A., Topranin, V. de M., Dalen, H., & Langaas, M. (2024). Improved sex-specific cardiovascular disease risk prediction using machine learning and explainable artificial intelligence.*European Journal of Preventive Cardiology*. https://doi.org/10.1093/eurjpc/zwae175.304

Nagra, L. S., Thapa, S., Rahi, P., Sharma, T., Ashok, H., & Pathania, V. (2024).*An Accurate Ensemble Machine Learning Model with Precision Engineering for Chronic Coronary Artery Disease Prognosis*. https://doi.org/10.1109/amathe61652.2024.10582249

Mehmood, W., Khaliq, A., Qamar, Z., Qudsia, M., & Raza, S. (2024).*Improved Coronary Artery Disease (CAD) Prediction using Machine Learning Method with Ensemble Technique*. https://doi.org/10.1109/icodt262145.2024.10740225

Atherosclerotic Cardiovascular Disease Prediction and Categorization Using Pre-trained Artificial Intelligence Models

Assist. Prof. Dr. Furkan ŞAKİROĞLU [1][0000-0002-7163-5658], Prof. Dr. Cemil ÇOLAK [2][0000-0001-5406-098X], Prof. Dr. Mehmet Cengiz ÇOLAK [3][0000-0002-0914-4719]

[1]Atatürk University, Biostatistics and Medical İnformatics, Yakutiye, Erzurum, Türkiye;

fsakiroglu@gmail.com

[2]İnönü University, Biostatistics and Medical İnformatics, Battalgazi, Malatya, Türkiye;

cemil.colak@inonu.edu.tr

[3]İnönü University, Cardiovascular Surgery, Battalgazi, Malatya, Türkiye;

mehmet.colak@inonu.edu.tr

Abstract

Atherosclerotic cardiovascular disease (ASCVD) is one of the leading causes of global morbidity and mortality. The current study provides a systematic review of the use of AI technologies applied to the prediction and management of ASCVD. Traditional risk assessment approaches have their restrictions, leading to a growing preference for artificial intelligence and machine learning techniques in risk assessment.

First, this study tackles the complex pathophysiology of ASCVD and the problems associated with the current diagnosis. Followed by an in-depth analysis of the wide variety of AI models that can be applied to electronic health records, medical imaging data, and other biomarkers. Special attention would be paid toward the potential of natural language processing models like BERT in predicting risk from textual clinical data, and the overwhelming success of convolutional neural networks such as ResNet and VGG in plaque-based analysis through imaging modalities.

Although the research results show that these models have a lot to offer in the clinical world, the authors also describe some serious disadvantages: data bias, interpretability of the model, and computational needs. Highlights, in particular, the need for multicenter validation studies as well as develop explainable AI (XAI) techniques.

Overall, AI-based approaches may pave the way for a new paradigm in ASCVD management. Nevertheless, deploying these technologies in everyday clinical practice will require overcoming technical, ethical and regulatory challenges. As such, interdisciplinary collaboration and thorough clinical validation studies are essential for fulfilling the promise of these novel strategies to enhance patient outcomes.

Keywords: Atherosclerotic cardiovascular disease, pre-trained artificial intelligence, machine learning, risk prediction, deep learning, medical imaging

1. Introduction

Atherosclerotic cardiovascular disease (ASCVD) remains a major global health burden. Traditional risk models like QRISK3 often fail to capture the complex, nonlinear relationships between risk factors. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), offers promising advances in ASCVD prediction and diagnosis. Techniques such as histopathology-based segmentation and transfer learning with pre-trained convolutional neural networks (CNNs) have significantly improved plaque detection in various imaging modalities [Holmberg et al., 2021]. Hybrid and ensemble models further enhance accuracy by integrating data from ECGs, imaging, clinical variables, and even genetic or microbiota information [Majumder et al., 2025; Cyriac et al., 2022]. AI-driven analysis of electronic medical records (EMRs) and time-series data also improves risk stratification [Kasartzian & Tsiampalis, 2025]. In imaging, AI aids lesion detection in CT, MRI, OCT, and retinal

scans [Abdollahi et al., 2024; Klüner et al., 2024]. Pre-trained models like BERT, ResNet, and VGG enable accurate predictions even with limited data [Sunilkumar & Kumaresan, 2024]. These innovations collectively support earlier diagnosis, personalized risk assessment, and more precise cardiovascular care.

2. Literature Search Methodology and Selection Criteria

This narrative review will conduct a systematic literature search using three major academic databases: Google Scholar, PubMed, and Scopus. The search will target peer-reviewed articles, theses, books, and grey literature related to the application of pre-trained AI models (e.g., BERT, ResNet, VGG) in the prediction and classification of atherosclerotic cardiovascular disease (ASCVD). A core search string—"Pre-trained Artificial Intelligence Models in Atherosclerotic Cardiovascular Disease"—will be used, supplemented by related keywords and Boolean combinations (e.g., "deep learning," "transfer learning," "cardiovascular risk," "coronary artery disease," "stroke," "prediction," "classification"). The review will include English-language studies published within the last 5 to 10 years, focusing on those that apply pre-trained models specifically to ASCVD risk prediction or disease characterization. Inclusion criteria require studies to report methods, results, and discussion in sufficient detail. Excluded will be non-English studies, inaccessible full texts, editorials, abstracts without full articles, and papers not directly addressing ASCVD with pre-trained models. The screening will follow a two-step process: title/abstract screening followed by full-text review based on predefined eligibility criteria to ensure inclusion of only high-quality and relevant studies.

3. Using Pre-trained AI Models to Predict ASCVD

Pre-trained AI models, especially those based on transformer architectures like BERT, have shown strong potential in predicting ASCVD risk and related outcomes. Med-BERT, which adapts BERT to structured Electronic Health Record (EHR) data, outperforms non-pre-trained models with minimal fine-tuning, addressing the challenge of limited medical data [Rasmy et al., 2021]. Similarly, Clinical BERT has been used to identify reasons for statin nonuse, supporting personalized ASCVD interventions [Chang et al., 2023]. BERT has also been applied to predict secondary cardiovascular diseases in patients with metabolic disorders, achieving impressive results in healthcare utilization pattern modeling [Chang et al., 2023]. In comparative studies, XLNet has even outperformed BERT in predicting mortality among cardiac patients, suggesting potential advantages for specific tasks [Kim et al., 2024]. These models work with both structured and unstructured EHR data—such as clinical notes, diagnoses, and prescriptions—and are typically evaluated using metrics like AUC, precision, and recall [Kasartzian & Tsiampalis, 2025]. Despite high performance, challenges remain, including computational demands, interpretability, and biases in clinical data [Parsa et al., 2024].

4. Training Data On ASCVD Classification Using Pre-Trained AI Models

Pre-trained convolutional neural networks (CNNs), especially ResNet and VGG architectures, are increasingly used for classifying ASCVD through medical imaging and other data sources. These models help detect plaques, assess stenosis, and identify high-risk lesions across modalities such as CCTA, ultrasound, and mammography [Laidi et al., 2024; Danilov et al., 2021]. ResNet-50, fine-tuned on CCTA data, outperforms models like Inception-v3 and VGG19, thanks to its residual connections that enhance feature extraction [Danilov et al., 2021]. Hybrid models combining ResNet-50 and VGG-16 have achieved diagnostic accuracies as high as 99.45% for coronary artery disease (CAD) [Barnawi, 2023]. ResNet-50 backbones in Faster RCNNs also perform well in classifying carotid plaques as stable or vulnerable using ultrasound [Zhang & Zhao, 2024]. VGG16 has shown effectiveness in detecting CAD from phonocardiograms and identifying breast arterial calcifications in mammograms, surpassing more complex models in accuracy [Martua et al., 2023; Khan & Masala, 2023]. While ResNet-50 typically outperforms VGG19, the latter is still widely used in ASCVD detection on CT images [Laidi et al., 2024]. Common CNNs like ResNet-50, VGG16, and VGG19, along with other models like Inception and DenseNet, support both binary and multi-class classification tasks [Barnawi, 2023]. Despite high accuracy, challenges such as limited labeled datasets, inter-patient variability, and high computational demands remain obstacles to clinical adoption [Moso et al., 2024].

5. Comparative Synthesis of Ascertainment Models for Pre-trained Artificial Intelligence in ASCVD

The strengths and weaknesses of pre-trained AI models like BERT, ResNet, and VGG are frequently compared in ASCVD prediction and classification. BERT, designed for natural language processing (NLP), excels at analyzing sequential data such as EHRs and clinical notes, uncovering hidden risk factors and aiding in treatment adherence prediction [Rasmy et al., 2021]. However, it is not suitable for image analysis and has limited interpretability due to its complex architecture [Parsa et al., 2024]. ResNet, optimized for image analysis, uses residual connections to prevent vanishing gradients, making it ideal for tasks like plaque detection and stenosis quantification [Danilov et al., 2021]. It requires large datasets and has high computational demands, though transfer learning can address data limitations [Sunilkumar & Kumaresan, 2024]. VGG, a simpler CNN, is effective for detecting CAD from phonocardiogram signals and identifying breast arterial calcifications from mammograms but struggles with more complex features due to its shallower structure [Martua et al., 2023; Khan & Masala, 2023]. Alternative models such as Inception, DenseNet, and EfficientNet offer advantages like lightweight design (Inception), efficiency with smaller datasets (DenseNet), and optimized accuracy with fewer resources (EfficientNet) [Barnawi, 2023]. For ASCVD risk prediction, BERT is best for EHR data, ResNet is ideal for image-based assessments like CCTA, and VGG offers a balance of simplicity and performance [Kasartzian & Tsiampalis, 2025]. Successful clinical adoption depends on data quality, interpretability, and regulatory compliance [Parsa et al., 2024].

6. Discussion

Pre-trained AI models, particularly transformer-based models like BERT for text and CNNs such as ResNet and VGG for images, show great potential in ASCVD prediction and classification [Chowdhury et al., 2025]. However, their clinical application is hindered by challenges like high dimensionality, data sparsity, and heterogeneity [Kasartzian & Tsiampalis, 2025]. Fine-tuning these models is gaining popularity, especially due to the scarcity of well-labeled data in medical tasks [Suneetha & Mahalngam, 2023]. These models have already outperformed conventional methods in tasks such as EHR-based risk prediction and plaque identification, demonstrating higher diagnostic accuracy [Holmberg et al., 2021; Rasmy et al., 2021]. Machine learning models can quickly detect high-risk individuals, leading to earlier interventions and personalized treatments [Bhatt et al., 2023]. AI applications in medical imaging can enhance decision-making, reduce heart disease-related events, and automate administrative tasks, adding significant value to clinical workflows [Abdollahi et al., 2024]. AI also offers the potential for more uniform and less-biased evaluations, which could result in better care [Sirisena & Okeahialam, 2024]. However, biases in training data and the need for diverse, representative datasets remain concerns [Parsa et al., 2024]. Interpretability is another challenge, as clinicians must trust AI outputs. Explainable AI (XAI) techniques are essential for improving transparency and usability [Parsa et al., 2024]. Additionally, most studies are single-center trials without external validation, impacting generalizability. Multicenter research is necessary for broader validation, and ethical concerns related to patient privacy, accountability, and regulation need to be addressed [Kasartzian & Tsiampalis, 2025].

7. Conclusion

Pre-trained artificial intelligence (AI) models, including BERT for natural language processing and convolutional neural networks (CNNs) such as ResNet and VGG for medical imaging, have shown considerable promise in improving risk stratification, disease classification, and clinical workflow automation in atherosclerotic cardiovascular disease (ASCVD). These models can contribute to more accurate and personalized cardiovascular care. However, their clinical adoption faces significant challenges, including the need for large, diverse, and high-quality datasets, ensuring model interpretability to build clinical trust, and meeting stringent external validation and regulatory requirements. Addressing these challenges is critical to translating AI advancements into practice. Future research should emphasize not only technical improvements but also evaluation of clinical utility, cost-effectiveness, and equitable application across populations. Successful integration of pre-trained AI models into healthcare will require close collaboration between researchers, clinicians, and policymakers.

8. References

Abdollahi, M., Jafarizadeh, A., Ghafouri-Asbagh, A., Sobhi, N., Pourmoghtader, K., Pedrammehr, S., . . . Acharya, U. R. (2024). Artificial intelligence in assessing cardiovascular diseases and risk factors via retinal fundus images: A review of the last decade. *WIREs Data Mining and Knowledge Discovery, 14*(6), e1560. doi:https://doi.org/10.1002/widm.1560

Awasthi, S., Sachdeva, N., Gupta, Y., Anto, A. G., Asfahan, S., Abbou, R., . . . Vellanki, N. (2023). Identification and risk stratification of coronary disease by artificial intelligence-enabled ECG. *EClinicalMedicine, 65*.

Barnawi, S. A. A. (2023). Evaluation of Pre-Trained CNN Models for Cardiovascular Disease Classification: A Benchmark Study.

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms, 16*(2), 88.

Chang, H.-J., Song, S., Kim, G., Kim, T., Shin, D., Kim, S., . . . Sung, J. (2023). BERT-based model to predict cardiovascular disease by analyzing healthcare utilization behavior of patients newly diagnosed with metabolic diseases.

Chowdhury, M. A., Rizk, R., Chiu, C., Zhang, J. J., Scholl, J. L., Bosch, T. J., . . . Chen, W. C. W. (2025). The Heart of Transformation: Exploring Artificial Intelligence in Cardiovascular Disease. *Biomedicines, 13*(2), 427. Retrieved from https://www.mdpi.com/2227-9059/13/2/427

Cyriac, S., R, S., & Raju, N. (2022). Cardiovascular Disease Prediction through Ensembled Transfer Learning on Cardiac Magnetic Resonance Imaging. *Journal of Computer Science, 18*(11). doi:10.3844/jcssp.2022.1005.1020

Danilov, V. V., Klyshnikov, K. Y., Gerget, O. M., Kutikhin, A. G., Ganyukov, V. I., Frangi, A. F., & Ovcharenko, E. A. (2021). Real-time coronary artery stenosis detection based on modern neural networks. *Sci Rep, 11*(1), 7582. doi:10.1038/s41598-021-87174-2

Girlanda, F., Demler, O., Menze, B., & Davoudi, N. (2024). Enhancing Cardiovascular Disease Prediction through Multi-Modal Self-Supervised Learning. *arXiv preprint arXiv:2411.05900*.

Holmberg, O., Lenz, T., Koch, V., Alyagoob, A., Utsch, L., Rank, A., . . . Kufner, S. (2021). Histopathology-based deep-learning predicts atherosclerotic lesions in intravascular imaging. *Frontiers in Cardiovascular Medicine, 8*, 779807.

Kasartzian, D.-I., & Tsiampalis, T. (2025). Transforming Cardiovascular Risk Prediction: A Review of Machine Learning and Artificial Intelligence Innovations. *Life, 15*(1), 94. Retrieved from https://www.mdpi.com/2075-1729/15/1/94

Khan, M. R., Haider, Z. M., Hussain, J., Malik, F. H., Talib, I., & Abdullah, S. (2024). Comprehensive Analysis of Cardiovascular Diseases: Symptoms, Diagnosis, and AI Innovations. *Bioengineering, 11*(12), 1239. Retrieved from https://www.mdpi.com/2306-5354/11/12/1239

Khan, R., & Masala, G. L. (2023). Detecting Breast Arterial Calcifications in Mammograms with Transfer Learning. *Electronics, 12*(1), 231. Retrieved from https://www.mdpi.com/2079-9292/12/1/231

Kim, Y., Kang, H., Seo, H., Choi, H., Kim, M., Han, J., . . . Kim, Y. H. (2024). Development and transfer learning of self-attention model for major adverse cardiovascular events prediction across hospitals. *Sci Rep, 14*(1), 23443. doi:10.1038/s41598-024-74366-9

Klüner, L. V., Chan, K., & Antoniades, C. (2024). Using artificial intelligence to study atherosclerosis from computed tomography imaging: A state-of-the-art review of the current literature. *Atherosclerosis*, 117580.

Laidi, A., Ammar, M., Daho, M. E. H., & Mahmoudi, S. (2024). Deep learning models for coronary atherosclerosis detection in coronary CT angiography. *Current Medical Imaging, 20*(1), e211222212072.

Majumder, S., Sen, K., & Karanjai, R. (2025). Artificial Intelligence-Based Target for Personalized Interventions of Atherosclerosis from Gut Microbiota Signature. *SynBio, 3*(1), 2.

Martua B, R. B., Bustamam, A., & Hermawan. (2023). Arrhythmia Classification Using the Deep Learning Visual Geometry Group (VGG) Model. *Nusantara Science and Technology Proceedings, 2023*(36), 7-18. doi:10.11594/nstp.2023.3702

Moso, J. C., Bennai, M. T., Guessoum, Z., Cormier, S., & Telalović, J. H. (2024). Detection of Atherosclerosis Using Deep Learning.

Naghavi, M., Reeves, A., Budoff, M., Li, D., Atlas, K., Zhang, C., . . . Levy, D. (2024). AI-powered Coronary Artery Calcium Scans (AI-CAC) Enables Prediction of Heart Failure and Outperforms NT-proBNP: The Multi-Ethnic Study of Atherosclerosis. *medRxiv*, 2024.2001. 2004.23300650.

Parsa, S., Somani, S., Dudum, R., Jain, S., & Rodríguez, F. (2024). Artificial Intelligence in Cardiovascular Disease Prevention: Is it Ready for Prime Time? *Current Atherosclerosis Reports, 26*, 1-10. doi:10.1007/s11883-024-01210-w

Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine, 4*(1), 86.

Sirisena[1], A., & Okeahialam, B. (2024). Harnessing artificial intelligence (AI) for early detection of atherosclerotic cardiovascular disease (ASCVD) in sub-Saharan Africa. *JOURNAL OF ARTIFICIAL INTELLIGENCE, 1*(4), 3-6.

Suneetha, A., & Mahalngam, T. (2023). Fine tuning bert based approach for cardiovascular disease diagnosis. *International Journal of Intelligent Systems and Applications in Engineering, 11*(6s), 59-66.

Sunilkumar, G., & Kumaresan, P. (2024). Deep Learning and Transfer Learning in Cardiology: A Review of Cardiovascular Disease Prediction Models. *IEEE Access, 12*, 193365-193386. doi:10.1109/ACCESS.2024.3514093

Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukçu, E., & Leiva, V. (2023). Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology, 12*(1), 117.

Umar, B. U., Ajao, L. A., Dogo, E. M., Ajao, F. J., & Atama, M. (2023). Artificial intelligence model for prediction of cardiovascular disease: An empirical study. *Artificial Intelligence in Health, 1*(1), 42-56.

Zhang, H., & Zhao, F. (2024). Deep Learning-Based Carotid Plaque Ultrasound Image Detection and Classification Study. *RCM, 25*(12). doi:10.31083/j.rcm2512454

Zhang, J., Han, R., Shao, G., Lv, B., & Sun, K. (2022). Artificial Intelligence in Cardiovascular Atherosclerosis Imaging. *J Pers Med, 12*(3). doi:10.3390/jpm12030420

Sağlık Hizmetlerinde Dijital Onam ve Hukuki Çerçevesi

Bilgenur KARA[1], Filiz İŞLEYEN[1]

[1] Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü, ANKARA 06800, TÜRKİYE
nurbilgekara@gmail.com

Özet. Sağlık hizmeti sunumu gelişen teknolojilerle beraber dijitalleşme sürecine dahil olan hizmetlerden biridir. Bu dönüşüm sağlık hizmeti sunumunda dijitalleşme adaptasyonunu beraberinde getirmektedir. Tıbbi müdahalenin hukuka uygunluk şartlarından biri olan aydınlatılmış onamın, ıslak imza ile alınmasının dijital dünyaya taşınması konusu da bu sürecin bir parçası olarak karşımıza çıkmaktadır. Ancak söz konusu işlemler hem operasyonel hem hukuki birtakım sorunları beraberinde getirmektedir. Dijital imza konusu kişisel verilerin korunmasına ilişkin mevzuat doğrultusunda özellikle biyometrik verilerin işlenmesi ve tıbbi müdahalelerde rızanın geçerlilik koşulları bakımından çeşitli hukuki konuları gündeme taşımaktadır. Bu çalışmada dijital onam ve biyometrik imza süreçlerinin hukuki boyutunun incelenmesi amaçlanmıştır.

Anahtar Kelimeler: Dijital Onam , Sağlık Hukuku, Aydınlatılmış Onam

Digital Consent in Healthcare Services and Its Legal Framework

Bilgenur KARA[1], Filiz İŞLEYEN1[1]

[1]Ministry of Health General Directorate of Health Information Systems, ANKARA 06800, TÜRKİYE
nurbilgekara@gmail.com

Abstract. The provision of healthcare services is among the domains increasingly integrated into the process of digital transformation driven by technological advancements. This evolution necessitates the adaptation of healthcare delivery to digital systems. One significant component of this transformation is the digitisation of informed consent an essential legal requirement for the legitimacy of medical interventions which has traditionally been obtained via wet signatures. However, the transition of informed consent procedures into digital formats brings forth various operational and legal challenges. In particular, the use of digital signatures raises important legal concerns under data protection legislation, especially regarding the processing of biometric data and the validity requirements of consent in medical procedures. This study aims to examine the legal dimensions of digital consent and biometric signature processes within the context of healthcare law.

Keywords: Digital Consent, Health Law, Informed Consent

1 Giriş

Sağlık hizmeti sunumu süreçlerinde, sağlık hizmeti sunucularında gerçekleşen ameliyat, yatış, tedavi vb. hizmetlerin tamamında sağlık hizmeti alan kişiler ile yakınlarından söz konusu tıbbi müdahalelere ilişkin ıslak imzalı aydınlatılmış onam formu alınmaktadır. Onam süreçleri, hasta haklarını korumak ve etik ilkelerin bütünlüğünü sağlamak için tasarlanmıştır. Sağlık hizmetlerinin sürekliliği ve sıklığı göz önünde bulundurulduğunda, ıslak imza ile alınan bu formlar, sağlık profesyonelleri ve hizmet sunucuları açısından ciddi bir operasyonel yük teşkil etmektedir. Örneğin, kağıt formlar okunaksız veya eksik olabilmekte; bu durum da işlemlerin tekrarlanmasına ve gecikmelere neden olabilmektedir. Ayrıca kağıt formlar fiziki olarak bir yerde bulunduğundan uzaktan veya hibrit bakım modelleri için bir engel teşkil edebilir. Gelişen teknolojiler çerçevesinde onamın dijital ortamda hukuka uygun şekilde alınması hem arşiv süreçlerinde yaşanan maliyet sorunlarına hem de kamu kaynaklarının etkin kullanımına engel teşkil edebilen hususlara (kağıt tüketimi sebebiyle çevreye verilen zararlar dahil olmak üzere) çözüm sağlayabilecektir.

Tıbbi müdahalenin hukuka uygun hale gelmesinin bir koşulu olan rıza, sağlık hizmeti sunumunda hali hazırda ıslak imza ile hasta ve hasta yakınlarından alınmaktadır. Gelişen teknolojiler doğrultusunda bu işlemin elektronik ortamda örneğin tablet üzerine atılacak biyometrik imza vb. yöntemler ile yapılması gündeme geldiğinde ise hem biyometrik verilerin işlenecek olması sebebiyle 6698 sayılı Kişisel Verilerin Korunması Kanunu ve sair mevzuat hem de onamın geçerlilik şartları kapsamında değerlendirme yapılması gerekmektedir.

Bu çalışmanın amacı sağlık hizmeti sunumu süreçlerinde aydınlatılmış onam formu ve imza gerektiren yatış formu vb. belgelerde onamın elektronik yöntemler kullanılarak (biyometrik imza, sms vb. yöntemler) alınması ile ilgili yasal mevzuat açısından konunun incelenmesidir.

2 Aydınlatılmış Onamın Hukuki Niteliği

Hasta Hakları Yönetmeliği'nin 4 üncü maddesinin (g) bendinde tıbbi müdahale; *"tıp mesleğini icraya yetkili kişiler tarafından uygulanan sağlığı koruma, hastalıkların tedavisi ve teşhisi için ilgili mesleki yükümlülükler ve standartlara uygun olarak tıbbın sınırları içinde gerçekleştirilen fizikî ve ruhî girişimi"* olarak tanımlanmaktadır. [1]

Hukuka uygun bir tıbbi müdahale için; tıbbi müdahalenin tıp mesleğini icra etmeye hukuken yetkilendirilen taraflarca yerine getirilmesi, tıbbi müdahalenin yasal bir amaç doğrultusunda yapılması, tıbbi müdahalenin standart ve mevcut yöntemlerle yerine getirilmesi ve bu süreçlerde hastanın aydınlatılması ve rızasının alınması olarak karşımıza çıkmaktadır [2].

Tıbbi müdahaleye ilişkin aydınlatılmış onamın alınmaması ya da hukuka aykırı şekilde alınması halinde sağlık personelinin sorumluluğu ortaya çıkacaktır [3].

3 Tıbbi Müdahale ve Rızaya İlişkin İlgili Mevzuat

Anayasa'nın 17'nci maddesinin ikinci fıkrasında *"Tıbbî zorunluluklar ve kanunda yazılı haller dışında, kişinin vücut bütünlüğüne dokunulamaz; rızası olmadan bilimsel ve tıbbî deneylere tâbi tutulamaz."* hükmü [4], İnsan Hakları ve Biyotıp Sözleşmesi'nin 5'inci maddesinde *"Sağlık alanında herhangi bir müdahale, ilgili kişinin bu müdahaleye özgürce ve bilgilendirilmiş bir şekilde muvafakat etmesinden sonra yapılabilir."* ifadesi [5], Türk Medeni Kanunu'nun 23'üncü maddesinde *"Yazılı rıza üzerine insan kökenli biyolojik maddelerin alınması, aşılanması ve nakli mümkündür."* hükmü [6], 1219 sayılı Tababet ve Şuabatı San'atlarının Tarzı İcrasına Dair Kanun'un 70'inci maddesinde *Tabipler, diş tabipleri ve dişçiler yapacakları her nevi ameliye için hastanın, hasta küçük veya tahtı hacirde ise veli veya vasisinin evvelemirde muvafakatını alırlar.* ***Büyük ameliyei cerrahiyeler için bu muvafakatin tahriri olması lazımdır."*** hükmü [7] bulunmaktadır.

Yine Hasta Hakları Yönetmeliği'nin 5'inci maddesinin (d) bendinde *"Tıbbi zorunluluklar ve kanunlarda yazılı haller dışında, rızası olmaksızın kişinin vücut bütünlüğüne ve diğer kişilik haklarına dokunulamaz."* hükmü, (e) bendinde *"Kişi, rızası ve Bakanlığın izni olmaksızın tıbbi araştırmalara tabi tutulamaz."* hükmü, 22'nci maddesinde *"Kanunda gösterilen istisnalar hariç olmak üzere, kimse, rızası olmaksızın ve verdiği rızaya uygun olmayan bir şekilde tıbbi ameliyeye tabi tutulamaz"* hükmü, 24 üncü maddesinin birinci fıkrasında *"Tıbbi müdahalelerde hastanın rızası gerekir."* hükmü, 25'inci maddesinde *"Kanunen zorunlu olan haller dışında ve doğabilecek olumsuz sonuçların sorumluluğu hastaya ait olmak üzere; hasta kendisine uygulanması planlanan veya uygulanmakta olan tedaviyi reddetmek veya durdurulmasını istemek hakkına sahiptir. Bu halde, tedavinin uygulanmamasından doğacak sonuçların hastaya veya kanuni temsilcilerine veyahut yakınlarına anlatılması ve bunu gösteren yazılı belge alınması gerekir."* hükmü, 26'ncı maddesinde *"Mevzuatta öngörülen durumlar ile uyuşmazlığa mahal vermesi tıbben muhtemel görülen tıbbi müdahaleler için sağlık kurum ve kuruluşunca 15 inci maddedeki bilgileri içeren rıza formu hazırlanır. Rıza formunda yer alan bilgiler; sözlü olarak hastaya aktarılarak rıza formu hastaya veya kanuni temsilcisine imzalatılır. Rıza formu iki nüsha olarak imza altına alınır ve bir nüshası hastanın dosyasına konulur, diğeri ise hastaya veya kanuni temsilcisine verilir. Acil durumlarda tıbbi müdahalenin hasta tarafından kabul edilmemesi durumunda, bu beyan imzalı olarak alınır, imzadan imtina etmesi halinde durum tutanak altına alınır. Rıza formu bilgilendirmeyi yapan ve tıbbi müdahaleyi gerçekleştirecek sağlık meslek mensubu tarafından imzalanır. Verilen bilgilerin doğruluğundan ilgili sağlık meslek mensubu sorumludur."* hükmü, 28'inci maddesinin birinci fıkrasında "*Mevzuatın öngördüğü istisnalar dışında, rıza herhangi bir şekle bağlı değildir.*" hükmü, ikinci fıkrasında *"Hukuka ve ahlaka aykırı olarak alınan rıza hükümsüzdür ve bu şekilde alınan rızaya dayanılarak müdahalede bulunulamaz."* hükmü, 29'uncu maddesinde *"18 yaşından küçük ve mümeyyiz olmayanlardan organ ve doku alınamaz. Bu şartları tamam olanlardan teşhis, tedavi ve bilimsel amaçlar ile organ veya doku alınması, 2238 sayılı Organ ve Doku Alınması, Saklanması ve Nakli Hakkında Kanun'un 6 ncı maddesinde öngörülen yazılı şekil şartına tabidir."* hükmü [8] ve 2238 sayılı Organ ve Doku Alınması, Saklanması, Aşılanması ve Nakli Hakkında Kanun'un 6'ncı maddesinde

"Onsekiz yaşını doldurmuş ve mümeyyiz olan bir kişiden organ ve doku alınabilmesi için vericinin en az iki tanık huzurunda açık, bilinçli ve tesirden uzak olarak önceden verilmiş yazılı ve imzalı veya en az iki tanık önünde sözlü olarak beyan edip imzaladığı tutanağın bir hekim tarafından onaylanması zorunludur." hükmü [9] bulunmaktadır.

Yukarıda yer alan açıklamalar kapsamında; tıbbi müdahalenin hukuka uygun olması ancak tıbbi müdahalede bulunulacak hastanın veya yasal temsilcisinin bu müdahaleye kural olarak rıza göstermiş olması ile mümkündür [10].

Aydınlatılmış onam konusunu rıza ve rızanın şekli açısından incelemek gerekirse; tıbbi müdahaleye rıza açık olabileceği gibi örtülü yani zımni olarak da açıklanabilmektedir. Hastanın zımni rızasının olduğu kabul edilen durumlar, genellikle basit tıbbi müdahaleler olarak karşımıza çıkmaktadır.[1] Ancak yukarıda yer alan mevzuat kapsamında tıbbi müdahalede hukuka uygunluk şartlarından biri olan rızanın, yazılı şekle uyulmasının zorunlu tutulduğu durumlarda yazılılık şartına uyulmadan açıklanması, rızanın hukuka uygun hale gelmesine engel teşkil edebilmektedir [11].

4 İmzanın Dijital Olarak Alınması Hususu

Aydınlatılmış onam formlarının, teknolojik gelişmeler doğrultusunda dijital ortamda alınması esasında imzanın dijital olarak alınması konusunu gündeme getirmektedir. İmza ile ilgili hem Türk Borçlar Kanunu'nda hem Elektronik İmza Kanunu'nda hükümler bulunduğundan konuyu bu minvalde değerlendirmek gerekmektedir.

Türk Borçlar Kanunu'nun 14'üncü maddesinde *"Yazılı şekilde yapılması öngörülen sözleşmelerde borç altına girenlerin imzalarının bulunması zorunludur."* hükmü, 15'inci maddesinde *"İmzanın, borç altına girenin el yazısıyla atılması zorunludur. Güvenli elektronik imza da, el yazısıyla atılmış imzanın bütün hukuki sonuçlarını doğurur."* hükmü bulunmaktadır. 5070 sayılı Elektronik İmza Kanunu'nun 3'üncü maddesinin birinci fıkrasının (a) bendinde *"Elektronik veri, elektronik, optik veya benzeri yollarla üretilen, taşınan veya saklanan kayıtları ifade eder."* hükmü ile (b) bendinde *"Elektronik imza, başka bir elektronik veriye eklenen veya elektronik veriyle mantıksal bağlantısı bulunan ve kimlik doğrulama amacıyla kullanılan elektronik veriyi ifade eder."* hükmü, "Güvenli elektronik imza" başlıklı 4'üncü maddesinde *"Güvenli elektronik imza; münhasıran imza sahibine bağlı olan, sadece imza sahibinin tasarrufunda bulunan güvenli elektronik imza oluşturma aracı ile oluşturulan, nitelikli elektronik sertifikaya dayanarak imza sahibinin kimliğinin tespitini sağlayan, imzalanmış elektronik veride sonradan herhangi bir değişiklik yapılıp yapılmadığının tespitini sağlayan elektronik imzadır"* hükmü, 5'inci maddesinde *"Güvenli elektronik imzanın*

[1] Zımni rıza, üstü kapalı, açık olmayan şekilde verilmiş rıza anlamına gelmektedir (Hukuk Sözlüğü, "Zımnı Rıza", Ankara: Yetkin Yayınları, 2010, 842).

elle atılan imza ile aynı hukukî sonucu doğurur. Kanunların resmî şekle veya özel bir merasime tabi tuttuğu hukukî işlemler ile banka teminat mektupları ve Türkiye'de yerleşik sigorta şirketleri tarafından düzenlenen kefalet senetleri dışındaki teminat sözleşmelerinin güvenli elektronik imza ile gerçekleştirilemez" hükmü bulunmaktadır.

Yukarıda anılan mevzuat hükümleri çerçevesinde; elektronik ortamda oluşturulan herhangi bir elektronik imzanın, elle atılan imza ile aynı hukuki geçerliliğe sahip olması için ülkemizde Kanun seviyesindeki düzenlemelere göre, nitelikli elektronik sertifikaya dayanan Güvenli Elektronik İmza ile oluşturulması gerekmektedir. Mevzuat kapsamında ıslak imza ile aynı hukuki geçerliliğe sahip, kimlik tespitine imkan veren ve elektronik ortamda oluşturulan imza örneği olarak sadece güvenli elektronik imzanın kabul edilebileceği değerlendirilmektedir.

5 Kişisel Verilerin Korunması Kanunu Kapsamında Dijital İmza ve Biyometrik Verilerin İşlenmesi

Tablet, cep telefonu gibi dijital ortamlarda kişilerin ıslak imzaya benzer şekilde el yazısı ile attığı imzanın kişinin fiziksel ve davranışsal özelliklerini içerdiğinden biyometrik veri niteliği taşıdığı görülmektedir [12].

"Biyometrik imza verisinin kullanılmasına ilişkin görüş talebi" ile ilgili olarak Kişisel Verileri Koruma Kurulunun 27.08.2020 tarihli ve 2020/649 sayılı Karar (Karar) özetinde *"Biyometrik veriler, herhangi bir müdahaleye gerek olmaksızın zahmetsiz bir şekilde elde edilen ve genel itibariyle ömür boyu değişmeden kalan veriler olup, bu verilerin değiştirilmesi veya unutulması mümkün değildir; çünkü birey ona ait özellikleri bizzat kendisi taşımaktadır."* ifadesi ile biyometrik veriler açıklanmıştır. Karar'da biyometrik imzanın, imza sahiplerinin belirli biyometrik verilerini kullanarak imzalarını özel bir tablet/ped üzerinde oluşturmaları ve genellikle bu verilerin imzalanan belgeye çözülemez biçimde bağlanmasıyla elde edildiği, her ne kadar biyometrik imza ile elle atılan ıslak imza arasında benzer yönler bulunsa da her ikisinin de farklı kavramlar olduğu, biyometrik imza çözümlerinin belirli bir standart çerçevesinde tanımlanmadığından farklı kurgusal özelliklere sahip olduğu ve ıslak imza ile denk sayılmadığı, biyometrik imza ile elle atılan imzanın bütün fonksiyonları bakımından aynı olmadığı hususuna Avrupa Sayısal Tek Pazarı'ndaki (Digital Single Market) elektronik ortamdaki işlemler için elektronik kimlik tanımlama ve güven hizmetleriyle ilgili Avrupa Birliği düzenlemesi standartı olan "Elektronik Kimlik Belirleme ve Güven Hizmetleri Düzenlemesi"nde (eIDAS) de değinildiği, ıslak, elle atılan klasik imza, imzanın görselliğine dayanarak o imzanın statik veya geometrik özelliklerini dikkate almakta iken (imzanın nasıl göründüğü ile alakalıdır); biyometrik imza ise imzanın dinamik özelliklerini (imzanın nasıl oluştuğu) dikkate aldığı ve bu doğrultuda, biyometrik imzanın analizinde, biyometrik imza esnasında uygulanan basıncın miktarı, yazma açısı, kalemin hızı ve ivmesi, harflerin oluşumu, imzanın yönü ve benzer diğer kişinin sahip olduğu benzersiz dinamik özelliklerin kullanılmakta olduğu ifade edilmiştir [13].

6698 sayılı Kişisel Verilerin Korunması Kanunu'nun 6'ncı maddesine göre özel nitelikli kişisel verilerden olan biyometrik veriler; "*İlgili kişinin açık rızasının*

olması",*"Kanunlarda açıkça öngörülmesi"*,*"Fiili imkânsızlık nedeniyle rızasını açıklayamayacak durumda bulunan veya rızasına hukuki geçerlilik tanınmayan kişinin, kendisinin ya da bir başkasının hayatı veya beden bütünlüğünün korunması için zorunlu olması"*,*"İlgili kişinin alenileştirdiği kişisel verilere ilişkin ve alenileştirme iradesine uygun olması"*,*"Bir hakkın tesisi, kullanılması veya korunması için zorunlu olması"*,*"Sır saklama yükümlülüğü altında bulunan kişiler veya yetkili kurum ve kuruluşlarca, kamu sağlığının korunması, koruyucu hekimlik, tıbbi teşhis, tedavi ve bakım hizmetlerinin yürütülmesi ile sağlık hizmetlerinin planlanması, yönetimi ve finansmanı amacıyla gerekli olması" "İstihdam, iş sağlığı ve güvenliği, sosyal güvenlik, sosyal hizmetler ve sosyal yardım alan hukuki yükümlülüklerin yerine getirilmesi için zorunlu olması"*,*"Siyasi, felsefi, dini veya sendikal amaçlarla kurulan vakıf, dernek ve diğer kâr amacı gütmeyen kuruluş ya da oluşumların, tâbi oldukları mevzuata ve amaçlarına uygun olmak, faaliyet alanlarıyla sınırlı olmak ve üçüncü kişilere açıklanmamak kaydıyla; mevcut veya eski üyelerine ve mensuplarına veyahut bu kuruluş ve oluşumlarla düzenli olarak temasta olan kişilere yönelik olması"* halinde işlenebilecektir. Bu kapsamda biyometrik imza kullanılarak onam alınması ancak mezkûr Kanun'da sayılan hallerden birinin varlığı halinde ve Kanun'un 4'üncü maddesinde yer alan ilkelere uyumlu olmak koşuluyla, Kanun'un 12'nci maddesi kapsamında uygun güvenlik önlemlerinin alınması şartıyla mümkün olabilecektir.[2]

Biyometrik imzanın sahip olduğu teknik ve teknolojik özellikler ile ilgili kişiye ait olup olmadığını ayırt etmeye yarayan hususların "ISO/IEC 19794-7:2014 Biometric data interchange formats — Part 7: Signature/sign time series" ve "ISO/IEC 19794-11:2013 Information technology -- Biometric data interchange formats -- Part 11: Signature/sign processed dynamic data" standardında düzenlendiği görülmektedir [14].

Elle atılan ıslak imzada yapıldığı gibi biyometrik imzanın da gerçekten imza sahibi tarafından atılıp atılmadığının belirlenebileceği görüşüne [15] katılmaktayız.

Kişisel Verileri Koruma Kurumu tarafından yayımlanan "Biyometrik Verilerin İşlenmesinde Dikkat Edilmesi Gereken Hususlara İlişkin Rehber"'de biyometrik veri işleme hususunda "Özel Nitelikli Kişisel Verilerin İşlenmesinde Veri Sorumlularınca Alınması Gereken Yeterli Önlemler"e ilişkin Kişisel Verileri Koruma Kurulunun 2018/10 sayılı kararında belirtilen tedbirlerin alınmasının zorunlu olduğu, ilaveten Rehber'de belirtilen; "türetilmiş biyometrik verilerin, orijinal biyometrik özelliğin yeniden elde edilmesine izin vermeyecek biçimde saklanması, biyometrik çözümü kullanamayan veya kullanmaya açık rızası olmayan ilgili kişiler için herhangi bir kısıtlama veya ek maliyet olmaksızın alternatif bir sistem sağlanması vb." gibi tedbirlerin de alınması gerektiği belirtilmiştir [16].

---

[2] 6698 sayılı Kanun'un 4'üncü maddesinde kişisel verilerin, hukuka ve dürüstlük kurallarına uygun, doğru ve gerektiğinde güncel olma, belirli, açık ve meşru amaçlar için işlenme, işlendikleri amaçla bağlantılı, sınırlı ve ölçülü olma ve ilgili mevzuatta öngörülen veya işlendikleri amaç için gerekli olan süre kadar muhafaza edilme ilkelerine uygun şekilde işlenmesinin zorunlu olduğu; 12'nci maddesinde veri sorumlusunun; kişisel verilerin hukuka aykırı olarak işlenmesini ve erişilmesini önlemek ile kişisel verilerin muhafazasını sağlamak amacıyla uygun güvenlik düzeyini temin etmeye yönelik gerekli her türlü teknik ve idari tedbirleri almak zorunda olduğu hüküm altına alınmıştır.

6 Elektronik Onamda Alternatif Yöntemler

Yukarıdaki başlıklarda detaylı açıklanan onam formlarında biyometrik imza kullanımına ek olarak aydınlatılmış onam formunun ve doğrulama kodunun kısa mesaj hizmeti ile ilgili kişiye gönderilerek kişiden alınacak doğrulama kodu ile onam işleminin gerçekleştirilmesi gibi yöntemler ile söz konusu işlemlerin yapılıp yapılamayacağı hususunda yani imza gerektiren dokümanların biyometrik imza dışında başka dijital yöntemler aracılığıyla (kişilerin telefonuna SMS ile gönderilen kodun ilgili personele iletilmesi vb.) yapılması hususunda ise; Türk Borçlar Kanunu'nda imza dışında "kişiye münhasır nitelikte ve güvenilir biçimde kimlik doğrulaması sağlayan bir onay sisteminin" yer almadığı görülmektedir.

7 Tartışma ve Sonuç

Dijital onam süreçlerine ilişkin yapılan çalışmalarda genellikle elektronik sistemlerin kâğıt tabanlı sistemler ile karşılaştırıldığı görülmüştür [17-19]. Bu çalışmalarda yapılan kıyaslamalar sonucunda ditijal onamın beklendiği üzere, veri tamlığı, okunabilirlik ve doğruluk açısından önemli iyileştirmeler sağladığını göstermiştir.

Gelişen teknolojiler karşısında sağlık hizmeti sunumu süreçleri de dijitalleşmekte ve bu konuda geliştirilen uygulamaların büyük çoğunluğu kullanıcılar tarafından geçmişe kıyasla daha hızlı kabul edilmektedir. Yine de bu kabul ediş halinin hukuki dayanaklarla desteklenmesi gerekmektedir. Sağlık bilişiminde veriye erişim yöntemi, saklama koşulları veya paylaşımı vb. hususlarında her ülkenin ayrı yasal düzenlemeleri olduğu için genel bir uygulama bulunamamıştır. Sağlık hizmeti süreçlerinde tıbbi müdahalenin hukuka uygun hale gelmesi için alınan onamların vatandaşlara kesintisiz, hızlı ve güvenilir bir sağlık hizmeti sunulabilmesi amacıyla dijital ortama taşınması operasyonel yüklerin azaltılmasına imkan sağlayacaktır. Ülkemizde birinci, ikinci ve üçüncü basamak sağlık hizmeti sunan tüm sağlık hizmeti sunucularında Sağlık Bilgi Yönetim Sistemleri kullanılmakta olup bu sistemlerin uygun geliştirmelerle elektronik onam uygulamalarını destekleyebilecek potansiyele sahip olduğu değerlendirilmektedir. Ancak konuya dair mevzuat incelendiğinde ıslak imza ve güvenli elektronik imza ile aynı hukuki sonucu doğuracak bir dijital onayın olmadığı görüldüğünden söz konusu düzenlemelerle alakalı yapılacak teknik çalışmalar ile konunun yasal olarak düzenlenmesinin sağlık hizmetinin ülkemizde çağın gereklerine uygun şekilde sunulmasına katkı sağlayacağı değerlendirilmektedir.

Kaynakça

1. Hasta Hakları Yönetmeliği, 01.08.1998 tarihli ve 23420 sayılı Resmî Gazete.
2. Çakal, A. (2018). Türk Özel Hukukunda Tıbbi Müdahaleye Rıza (Yüksek Lisans Tezi, Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü).
3. Kurt, M. G. (2017). Tıbbi Müdahalelerde Aydınlatılmış Onam. Türkiye Barolar Birliği Dergisi, (130), 401–422. https://tbbdergisi.barobirlik.org.tr/m2020-146-1897
4. Türkiye Cumhuriyeti Anayasası (Kanun No. 2709, Kabul Tarihi: 18/10/1982). Resmî Gazete, (Tarih: 9/11/1982, Sayı: 17863 Mükerrer)
5. Biyoloji ve Tıbbın Uygulanması Bakımından İnsan Hakları ve İnsan Haysiyetinin Korunması Sözleşmesi: İnsan Hakları ve Biyotıp Sözleşmesinin Onaylanmasının Uygun Bulunduğuna Dair Kanun (Kanun No. 5013, Kabul Tarihi: 3.12.2003).
6. Türk Medeni Kanunu, 08.12.2001 tarihli ve 24607 sayılı Resmî Gazete.
7. Tababet ve Şuabatı San'atlarının Tarzı İcrasına Dair Kanun, 14.04.1928 tarihli ve 863 sayılı Resmî Gazete.
8. Hasta Hakları Yönetmeliği, 01.08.1998 tarihli ve 23420 sayılı Resmî Gazete.
9. Organ ve Doku Alınması, Saklanması, Aşılanması ve Nakli Hakkında Kanun, 03.06.1979 tarihli ve 16655 sayılı Resmî Gazete.
10. Tanık, S. (2023). Tıbbi Müdahalelerde Aydınlatılmış Onam (Yüksek lisans tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, Özel Hukuk Anabilim Dalı). İzmir.
11. Kocaoğlu, B. (2020). Tıbbi Müdahaleye Rıza. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü; Yılmaz, H. (2021). Tıbbi Müdahalelerde Aydınlatılmış Onamın Hukuki Niteliği. Türkiye Biyoetik Dergisi, 8(2), 45-52.
12. Leyla Keser Berber, Biyometrik İmza ve Türk Borçlar Kanunu'ndaki Yazılı Şekil Şartı ile Hukuk Muhakemeleri Kanunundaki İmza Açısından Yeri, sayfa:9, Çevrimiçi: https://itlaw.bilgi.edu.tr/media/document/2019/08/biyometrik-imza.pdf (Erişim Tarihi 12.05.2024)
13. Kişisel Verileri Koruma Kurulunun 27.08.2020 tarihli ve 2020/649 sayılı "Biyometrik imza verisinin kullanılmasına ilişkin görüş talebi" ile ilgili olarak Kararı Çevrimiçi: https://www.kvkk.gov.tr/Icerik/6815/2020-649 (Erişim Tarihi 12.05.2024)
14. Leyla Keser Berber, Biyometrik İmza ve Türk Borçlar Kanunu'ndaki Yazılı Şekil Şartı ile Hukuk Muhakemeleri Kanunundaki İmza Açısından Yeri, sayfa:9, Çevrimiçi: https://itlaw.bilgi.edu.tr/media/document/2019/08/biyometrik-imza.pdf (Erişim Tarihi 12.05.2024)
15. Leyla Keser Berber, Biyometrik İmza ve Türk Borçlar Kanunu'ndaki Yazılı Şekil Şartı ile Hukuk Muhakemeleri Kanunundaki İmza Açısından Yeri, sayfa:9, Çevrimiçi: https://itlaw.bilgi.edu.tr/media/document/2019/08/biyometrik-imza.pdf (Erişim Tarihi 12.05.2024)
16. Kişisel Verileri Koruma Kurumu tarafından yayımlanan "Biyometrik Verilerin İşlenmesinde Dikkat Edilmesi Gereken Hususlara İlişkin Rehber" Çevrimiçi: https://kvkk.gov.tr/SharedFolderServer/CMSFiles/bd06f5f4-e8cc-487e-abe1-d32dc18e2d7e.pdf (Erişim Tarihi 11.05.2024)
17. Chimonas, S., Lipitz-Snyderman, A., Matsoukas, K., & Kuperman, G. (2023). Electronic consent in clinical care: An international scoping review. BMJ Health & Care Informatics, 30(e100726). https://doi.org/10.1136/bmjhci-2022-100726
18. Obaidi, H., Elkhyatt, Y., Alzubaidi, M., & Househ, M. (2024). Use of E-Consent in Healthcare Settings: A Scoping Review. Studies in health technology and informatics, 316, 1064–1068. https://doi.org/10.3233/SHTI240594
19. Dyke R, St-John E, Shah H, Walker J, Loughran D, Anakwe R, et al. Comparing shared decision making using a paper and digital consent process. A multi-site, single centre study in a trauma and Orthopaedic Department. The Surgeon. 2023;21(4):235–41. doi:10.1016/j.surge.2022.05.004

Kimyasal Bileşiklerin Antimikrobiyal Özelliğinin Makine Öğrenmesi Yöntemleriyle Belirlenmesi

Sinem ÖZDEMİR[1][0000-0002-6332-0891], Işık ÇAKMAK[1][0000-0003-4758-4077], Nurhan GÜMRÜKÇÜOĞLU[2][0000-0002-9669-6318]

[1] KTÜ, Tıp Fakültesi, Temel Tıp, Biyoistatistik ve Tıp Bilişimi, Trabzon, TÜRKİYE
[2]KTÜ, Sağlık Hizmetleri Meslek Yüksekokulu, Trabzon, TÜRKİYE

Özet. Çalışma kapsamında antimikrobiyal özelliği olduğu ve antimikrobiyal özelliği olmadığı bilinen bileşikler ile daha önce sentezlenen hidrazon, oksadiazol, amino, schiff bazları ve açillendirme bileşikleri kullanılmıştır. Bu tür bileşiklerin çeşitli biyolojik aktivitelere sahip olduğu bilinmektedir. Ancak, bu biyolojik aktivitelerin belirlenmesine yönelik geleneksel yöntemler maliyetli ve zaman alıcıdır. Bu bağlamda, makine öğrenmesi (ML) ve derin öğrenme (DL) yöntemleri, hem maliyeti azaltmak hem de süreçleri hızlandırmak adına önemli bir potansiyel sunmaktadır. Bu çalışmada, sentezlenen moleküllerin antimikrobiyal potansiyelinin belirlenebilmesi amacıyla, biyolojik testler öncesinde in silico yapısal öngörüler gerçekleştirilmiştir. SVM model eğitiminde kimyasal yapılara ait 2B yapı bilgisi kullanılarak temsil vektörleri oluşturulmuştur. Elde edilen bulgular, AUC (Area Under Curve) değerinin istenilen düzeyde (AUC:0.85) olduğunu göstermiştir; bu da modelin ayırt edici kapasitesinin yeterli olduğunu işaret etmektedir. Bununla birlikte, recall (duyarlılık) değerinin beklenenden düşük çıktığı (Recall: 0.58) gözlemlenmiştir. Bu durumun olası sebepleri arasında sınıf dengesizliği, modelin pozitif sınıfları yeterince iyi tanıyamaması veya bazı biyolojik aktivitelerin yapısal özelliklerle yeterince ilişkilendirilememesi yer almaktadır. Bu sorunun üstesinden gelmek için, veri ön işleme adımlarının iyileştirilmesi, gelişmiş özellik seçimi tekniklerinin kullanılması modelin pozitif sınıfları daha iyi öğrenmesini sağlayacak stratejilerin eklenmesi önerilmektedir. Gelecek çalışmalar, tahmin performansını artırmak amacıyla model optimizasyonu ve iyileştirmeleri üzerine odaklanılması planlanmaktadır.

Anahtar Kelimeler : Biyolojik Aktivite, Derin Öğrenme, Makine Öğrenmesi

Determination of the Antimicrobial Properties of Chemical Compounds Using Machine Learning Methods

Sinem ÖZDEMİR[1][0000-0002-6332-0891], Işık ÇAKMAK[1][0000-0003-4758-4077], Nurhan GÜMRÜKÇÜOĞLU[2][0000-0002-9669-6318]

[1] KTU, Faculty of Medicine, Department of Basic Medicine, Biostatistics and Medical Informatics, Trabzon, TURKEY
[2] KTU, Vocational School of Health Services, Trabzon, TURKEY

Abstract. Within the scope of the study, compounds known to have antimicrobial and non-antimicrobial properties and previously synthesized hydrazone, oxadiazole, amino, schiff bases and acylating compounds were used. It is known that such compounds have various biological activities. However, traditional methods for determining these biological activities are costly and time-consuming. In this context, machine learning (ML) and deep learning (DL) methods offer significant potential both in terms of reducing costs and accelerating processes. In this study, in silico structural predictions were performed before biological tests in order to determine the antimicrobial potential of the synthesized molecules. In SVM model training, representation vectors were created using 2D structure information of chemical structures. The findings obtained showed that the AUC (Area Under Curve) value was at the desired level this indicates that the discriminatory capacity of the model is sufficient. However, it was observed that the recall (sensitivity) value was lower than expected. Possible reasons for this situation include class imbalance, the model's inability to recognize positive classes well enough, or the inability to adequately associate some biological activities with structural features. To overcome this problem, it is suggested to improve data preprocessing steps, use advanced feature selection techniques, and add strategies that will allow the model to learn positive classes better. Future studies are planned to focus on model optimization and improvements in order to increase prediction performance. Keywords: Biological Activity, Deep Learning, Machine Learning

Keywords: Biological Activity, Deep Learning, Machine Learning

Tıp Bilişimi Araştırmalarında Tematik Eğilimler ve Yapay Zekanın Etkisi: 2018-2024 Dönemi Bibliyometrik Analizi

Thematic Trends in Medical Informatics Research and the Impact of Artificial Intelligence: A Bibliometric Analysis of the 2018-2024 Period

Adem SONUVAR[1], Ahmet YARDIMCI[1], Kemal Hakan GÜLKESEN[1], Esra TOKUR SONUVAR[1]

[1] Biyoistatistik ve Tıp Bilişimi AD, Tıp Fakültesi, Akdeniz Üniversitesi Antalya.

Özet. Bu çalışmada, 2018-2024 yılları arasında tıp bilişimi alanındaki araştırma eğilimlerini ve yapay zeka gelişiminin bu alana etkisini incelemek amacıyla kapsamlı bir bibliyometrik analiz gerçekleştirilmiştir. Web of Science veri tabanından "Medical Informatics" kategorisinde yayınlanan 36,277 makale, R programlama dilindeki bibliometrix paketi kullanılarak analiz edilmiştir. Anahtar kelime eş-oluşum ağları ve tematik haritalar oluşturularak dört zaman diliminde (2018-2019, 2020-2021, 2022, 2023-2024) araştırma temalarının evrimi incelenmiştir. Analizler, bu dönemde "derin öğrenme", "makine öğrenmesi", "dijital sağlık" ve "COVID-19" konularının baskın temalar olduğunu göstermiştir. 2018-2019 döneminde temel metodolojilerin geliştirildiği, 2020-2021 döneminde pandemi yanıtı ve NLP ilerlemeleri, 2022'de etik ve uygulanabilirlik konularının öne çıktığı, 2023-2024 döneminde ise dağıtık mimariler ve büyük dil modellerinin belirgin hale geldiği gözlenmiştir. Çalışma, tıp bilişiminin teknolojik yeniliklere ve sağlık hizmetleri ihtiyaçlarına hızla uyum sağlayabilen dinamik bir disiplin olduğunu, ancak uygulama odaklı yaklaşımlar ile metodolojik gelişmeler arasında dalgalanmalar gösterdiğini ortaya koymuştur. Bu bulgular, gelecek araştırmaların yönlendirilmesi ve tıp bilişiminin sağlık hizmetlerinin dönüşümündeki rolünün anlaşılması açısından değerli içgörüler sunmaktadır.

Anahtar Kelimeler: Tıp bilişimi, Bibliyometrik analiz, Yapay zeka, Makine öğrenmesi, Derin öğrenme, Dijital sağlık, Büyük dil modelleri

Abstract. This study conducted a comprehensive bibliometric analysis to examine research trends in medical informatics from 2018 to 2024 and the impact of

artificial intelligence developments on this field. 36,277 articles published in the "Medical Informatics" category from the Web of Science database were analyzed using the bibliometrix package in R programming language. Keyword co-occurrence networks and thematic maps were created to investigate the evolution of research themes across four time periods (2018-2019, 2020-2021, 2022, 2023-2024). Analyses showed that "deep learning," "machine learning," "digital health," and "COVID-19" were dominant themes during this period. The 2018-2019 period focused on developing foundational methodologies, 2020-2021 on pandemic response and NLP advancements, 2022 highlighted ethics and applicability concerns, while 2023-2024 saw distributed architectures and large language models become prominent. The study reveals that medical informatics is a dynamic discipline capable of rapidly adapting to technological innovations and healthcare needs, but shows fluctuations between application-focused approaches and methodological developments. These findings provide valuable insights for directing future research and understanding the role of medical informatics in healthcare transformation.

Keywords: Medical informatics, Bibliometric analysis, Artificial intelligence, Machine learning, Deep learning, Digital health, Large language models

1. Giriş

Tıp bilişimi, "tıp ve sağlık hizmetlerinde veri, bilgi ve bilginin sistematik işlenmesi ile ilgilenen disiplin" olarak tanımlanmaktadır [1]. 1950'lerden günümüze sürekli evrim geçiren bu alan, teknolojik ilerlemeler, sağlık hizmetleri gereksinimleri ve bilimsel keşiflerle şekillenmiştir. Tıp bilişimi tarihi iyi belgelenmiş olup, 1959'da Ledley ve Lusted'in tıbbi karar vermeye yönelik ilk hesaplamalı yaklaşımları sunmasından [3], 1970'lerde ayrı bir disiplin olarak kurulmasına [4], 1980'lerde klinik karar destek sistemlerine [5] ve 1990'larda yapay zeka uygulamalarına [6] uzanan önemli dönüm noktaları içermektedir. 2010'da Reinhold Haux, tıp bilişiminin tarihsel gelişimini incelemiş ve geleceğe dair projeksiyonlar sunmuştur [1]. Bu çalışmayı temel alan Gülkesen ve Haux (2019), 2006-2017 tıp bilişimi literatürünün bibliyometrik analizini gerçekleştirmiş, mobil sağlık, organizasyonel sistemler, elektronik sağlık kayıtları, biyomedikal veri analizi ve klinik bilişim olmak üzere beş temel araştırma kümesi tanımlamıştır [7].

Mevcut çalışmamız, yapay zeka ve makine öğrenmesi alanlarındaki önemli gelişmelerin yaşandığı 2018-2024 dönemindeki tıp bilişimi araştırmalarını analiz etmeyi amaçlamaktadır. Araştırma sorularımız: 2018-2024 tıp bilişimi literatüründeki baskın araştırma konuları ve eğilimler nelerdir? Bu konular, yapay zeka gelişiminin farklı aşamalarında nasıl evrilmiştir? Bu eğilimler, Haux'un 2010'daki projeksiyonlarıyla ne ölçüde uyumludur? Bu soruları yanıtlamak için, dört belirgin dönemi (2018-2019, 2020-2021, 2022, 2023-2024) kapsayan kapsamlı bir bibliyometrik analiz gerçekleştirdik. Çalışmamız, son gelişmeleri yansıtan güncel bir tıp bilişimi araştırması analizi sunmakta, yapay zeka ve makine öğrenmesi gelişmelerinin araştırma önceliklerini nasıl etkilediğini incelemekte ve önceki projeksiyonların doğruluğunu değerlendirmektedir.

2. Yöntem

2.1 Veri Toplama ve İşleme

Web of Science Core Collection veri tabanında "Medical Informatics" kategorisindeki Ocak 2018-Aralık 2024 arasında yayınlanan, İngilizce lisanında, orijinal makale (article) doküman türü tarandı. Bildiri metinleri, kitap bölümleri ve diğer doküman türleri dışlandı. Veri analizi için bibliometrix R paketi (versiyon 4.1.4) [7] kullanıldı. Araştırma temalarını ve evrimlerini belirlemek için yazar anahtar kelimeleri (DE) eş-kelime analizi kullanıldı. Kavramsal parçalanmayı azaltmak için eş anlamlı terimleri birleştiren bir sözlük oluşturuldu. Temizlenen anahtar kelime verisi, eş-oluşum ağlarının oluşturulmasında kullanıldı.

2.3 Dönem Analizi ve Tematik Haritalama

Araştırma temalarının yapay zeka gelişmelerine bağlı evrimini incelemek için analiz dört kronolojik döneme bölündü: 2018-2019 (derin öğrenme uygulamalarının sağlık alanında ivme kazandığı erken dönem), 2020-2021 (BERT gibi transformer tabanlı modellerin ortaya çıktığı doğal dil işleme ilerlemeleri dönemi), 2022 (ChatGPT ve benzer büyük dil modellerinin tanıtıldığı önemli dönüm noktası) ve 2023-2024 (çeşitli sağlık alanlarında üretken yapay zeka uygulamalarının yaygınlaştığı dönem). Her dönem için tematik kümeler, merkezilik (diğer araştırma alanlarıyla bağlantıları gösteren) ve yoğunluk (temanın iç bağlantılarının gücünü yansıtan) olmak üzere iki temel metriğe göre stratejik diyagramlarda konumlandırıldı. Bu çerçeve, araştırma temalarını dört farklı kategoride sınıflandırmamızı sağladı: motor temalar (yüksek merkezilik, yüksek yoğunluk), niş temalar (düşük merkezilik, yüksek yoğunluk), ortaya çıkan veya azalan temalar (düşük merkezilik, düşük yoğunluk) ve temel temalar (yüksek merkezilik, düşük yoğunluk).

3. Bulgular

3.1 Yayın Eğilimleri ve Tematik Analiz

Yıllık bilimsel üretim analizi, tıp bilişimi yayınlarının 2018'de 3,615'ten 2024'te 6,170'e istikrarlı bir artış gösterdiğini ortaya koydu. 2020-2021 döneminde belirgin artışlar gözlendi (5,809 ve 5,767 yayın), ardından 2022'de hafif bir düşüş (5,224 yayın) ve 2023-2024'te tekrar büyüme gerçekleşti. Tıp bilişiminde incelenen makalelerin %28,23'ünün uluslararası iş birlikleri ile yazıldığı ve özellikle yapay zeka araştırmalarında teknoloji şirketlerinin katılımı gözlendi. En sık kullanılan anahtar kelimeler "derin öğrenme" (2,765 makalede), "makine öğrenmesi" (2,601), "mhealth" (2,485), "dijital sağlık" (2,444) ve "COVID-19" (2,175) olarak belirlendi. Gülkesen ve Haux'un [6] 2006-2017 analizinde "ortaya çıkan konular" olarak tanımlanan yapay zeka terimleri, 2018-2024 döneminde temel araştırma odaklarına dönüşmüştür.

3.2 Dönemsel Tematik Evrim

Tematik evrim haritası, dört zaman diliminde araştırma temalarının belirgin bir gelişim gösterdiğini ortaya koydu. **2018-2019 Dönemi:** "Elektronik sağlık kaydı" ve "sosyal medya" motor temalar olarak öne çıkarken, "makine öğrenmesi" ve "derin öğrenme" henüz ortaya çıkan temalar kadranında yer aldı. **2020-2021 Dönemi:** "COVID-19" ve pandemiyle ilgili konular motor temalar olarak belirginleşti. "Özellik çıkarma" ve "segmentasyon" niş temalar olarak yapay zeka metodolojisindeki uzmanlaşmayı yansıttı. **2022 Dönemi:** "Dijital sağlık" ve "mhealth" motor temalardaki pozisyonlarını güçlendirirken, "derin öğrenme" niş temalar kadranına kaydı, "makine öğrenmesi" haritanın merkezine yaklaştı. **2023-2024 Dönemi:** "Dijital sağlık" ve mobil sağlık uygulamaları temel temalar kadranına kaydı, "makine öğrenmesi" ve "yapay zeka" merkezi bir küme oluşturdu, "derin öğrenme" niş tema olarak uzmanlaştı. Bu dönemde belirgin motor temalar görülmedi.

4. Tartışma

Analizimiz, 2018'den 2024'e tıp bilişimi yayınlarında önemli bir büyüme gösterdi. 2020-2021'deki yükseliş COVID-19 pandemisiyle çakışarak dijital sağlık araştırmalarında benzersiz bir artışa neden oldu [8, 9]. 2022'deki geçici düşüş, pandemi sonrası bilgi konsolidasyonu ve yeni teknolojilerin keşfi dönemini temsil ederken, 2023-2024'teki toparlanma ChatGPT gibi büyük dil modellerinin klinik uygulamalarına odaklanan yayınlardan kaynaklandı [8, 9].

Anahtar kelime analizi, tıp bilişimi literatüründeki yapay zeka ile ilgili terimlerin ("derin öğrenme", "makine öğrenmesi", "yapay zeka") baskınlığı, önceki çalışmalara göre önemli bir değişimi temsil etmektedir. Gülkesen ve Haux, 2006-2017 analizlerinde bu terimleri ortaya çıkan konular olarak tanımlarken [7], bulgularımız 2018-2024 döneminde bunların temel araştırma odaklarına dönüştüğünü göstermektedir. Mobil ve dijital sağlıkla ilgili terimlerin ("mhealth", "dijital sağlık", "mobil telefon") yüksek sıklığı, hasta merkezli, teknoloji destekli bakım sunumu modellerine artan vurguyu yansıtmaktadır. "COVID-19"un çalışma dönemimizin yalnızca ortalarında ortaya çıkmasına rağmen öne çıkan bir anahtar kelime olması, pandeminin araştırma öncelikleri üzerindeki derin etkisini ve tıp bilişimi topluluğunun bu küresel sağlık krizine hızlı tepkisini vurgulamaktadır.

Her dönem belirgin karakteristikler göstermiştir: 2018-2019'da temel metodolojik gelişmeler, 2020-2021'de pandemi yanıtı ve yapay zeka ilerlemeleri, 2022'de artan uzmanlaşma ve etik farkındalık, 2023-2024'te ise dağıtık hesaplama ve üretken yapay zeka uygulamaları. 2023-2024 döneminde belirgin motor temaların olmaması, üretken yapay zeka teknolojilerinin yol açtığı bir geçiş dönemini yansıtabilir. Haux'un 2010 projeksiyonlarıyla karşılaştırıldığında, kişiselleştirilmiş sağlık hizmeti beklenenin ötesinde çeşitlilik gösterirken, klinik-moleküler bilgi entegrasyonu daha az belirgin olmuş,

kullanıcı deneyimi ise daha fazla önem kazanmıştır. Haux'un çerçevesi, derin öğrenme ve büyük dil modellerinin etkisini tam olarak öngörememiştir.

Çalışmamızın sınırlılıkları (yalnızca Web of Science kategorisi kullanımı ve bibliyometrik analizin doğal kısıtları) ve gelecek araştırma yönlerine (büyük dil modellerinin klinik uygulamaları, federe öğrenme ve etik boyutlar) rağmen, bulgularımız tıp bilişiminin teknolojik gelişmelere hızla uyum sağlayan dinamik bir disiplin olduğunu göstermektedir.

5. Sonuç

2018-2024 tıp bilişimi araştırmalarının bibliyometrik analizimiz, yapay zeka uygulamaları (özellikle makine öğrenmesi ve derin öğrenme), mobil sağlık çözümleri, elektronik sağlık kayıtları ve veri analiz metodolojilerinin dönem boyunca tutarlı odak alanları olduğunu ortaya koymuştur. Bu temalar, genel yaklaşımlardan uzmanlaşmış uygulamalara doğru belirgin bir evrim göstermiştir. Zamansal haritalama, teknolojik ilerlemelerle uyumlu net geçişler sergilemiştir: 2018-2019 temel metodolojileri kurarken, 2020-2021 pandemi yanıtı ve yapay zeka ilerlemelerine odaklanmış, 2022 etik farkındalığı öne çıkarmış, 2023-2024 ise dağıtık hesaplama ve üretken yapay zeka uygulamalarına yönelmiştir. Bulgularımız, tıp bilişiminin teknolojik gelişmelere hızla uyum sağlayabilen, metodolojik ilerlemeler ile uygulama odağı arasında dalgalanan dinamik bir disiplin olduğunu göstermektedir. Alan, temel odağını korurken yeni teknolojileri bünyesine katma, sağlık hizmeti ihtiyaçlarına yanıt verme ve etik hususlara artan dikkat gösterme konusunda dikkat çekici bir uyarlanabilirlik sergilemiştir.

Kaynaklar

1. Haux, R.: Medical Informatics: Past, present, future. International Journal of Medical Informatics. 79(9), 599-610 (2010).
2. Shortliffe, E.H., Cimino, J.J.: Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Springer Science & Business Media (2006)
3. Ledley, R.S., Lusted, L.B.: Reasoning Foundations of Medical Diagnosis. Science. 130(3366), 9-21 (1959).
4. Collen, M.F.: A History of Medical Informatics in the United States, 1950 to 1990. American Medical Informatics Association (1995)
5. Buchanan, B.G., Shortliffe, E.H.: Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley (1984)
6. de Dombal, F.T., et al.: Computer-aided Diagnosis of Acute Abdominal Pain. British Medical Journal. 2(5804), 9-13 (1972).
7. Gülkesen, K.H., Haux, R.: Research subjects and research trends in medical informatics. Methods of Information in Medicine. 58(S01), e1-e13 (2019).
8. Aria, M., Cuccurullo, C.: bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 11(4), 959-975 (2017).
9. Wang, J., et al.: COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. Journal of Medical Internet Research. 23(5), e19458 (2021).

Bibliometric Analysis of Machine Learning Applications in Psoriasis Treatment: Health Informatics and Economics Perspectives

Mehtap ÇAKMAK BARSBAY[1][0000-0001-7609-6254]

[1] Ankara University, Ankara 06560, Türkiye

Abstract. Machine learning is revolutionizing dermatology, particularly in diagnosing chronic inflammatory diseases such as psoriasis, predicting patient outcomes, and optimizing therapy selection. This study investigates publication trends, influential peer-reviewed manuscripts, collaborative networks, citation impact, and thematic evolution in machine learning applications for psoriasis treatment, with a focus on health informatics, healthcare resource optimization, and cost-effectiveness. Data is extracted from PubMed using a combination of search queries, processed through bibliometric analysis, and visualized using the Bibliometrix tool. Findings highlight the intersection between artificial intelligence, dermatology, and health economics, emphasizing the role of electronic health records based artificial intelligence models in personalized treatment strategies.

Keywords: Chronic inflammatory diseases, Bioinformatics, Health economics.

Akciğer Kanserinin Sınıflandırılmasında Makine Öğrenmesi Modellerinin Karşılaştırmalı Performans Analizi

Funda İpekten[1][0000-0002-6916-9563], Abdulvahap Pınar[2][0000-0002-3662-2579], Elif Çelik Gürbulak[3][0000-0002-5073-1907]

[1]Adıyaman Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Adıyaman, Türkiye

[2]Adıyaman Üniversitesi Rektörlük Birimi, Adıyaman, Türkiye

[3]Erciyes Üniversitesi Veteriner Fakültesi Biyometri Anabilim Dalı, Kayseri, Türkiye

ÖZET

Amaç: Bu çalışmada, Akciğer kanseri veri seti kullanılarak Lojistik Regresyon (LR), eXtreme Gradient Boosting (XGBoost), Elastic-Net ve Neural-Net (NN) Makine Öğrenimi (ML) algoritmalarının performansları karşılaştırılması amaçlanmıştır.

Yöntem: Analizlerde kullanılan veri seti, Kaggle platformundan elde edilmiş olup 309 bireye ait 15 biyobelirteç bilgisini içermektedir. Veriler, %80 eğitim ve %20 test seti olarak ikiye bölünmüştür. Değişken seçimi, Random Forest algoritması ile gerçekleştirilmiş ve en yüksek önem düzeyine sahip 14 değişken modele dahil edilmiştir. Sınıflandırma modellerinin performansı; Doğruluk, F1 skor, Ağırlıklı F1 skor, Matthews Korelasyon Katsayısı (MKK), Eğri altındaki alan (EAA), Duyarlılık, Seçicilik, Pozitif Kestirim Değeri (PKD), Negatif Kestirim Değeri (NKD) metrikleri kullanılarak değerlendirilmiştir.

Bulgular: Akciğer kanseri sınıflandırma performanslarına ilişkin LR, XGBoost, Elastic Net ve NN modelleri kullanılmıştır. İlgili modellere ilişkin metriklerin performans ölçütleri karşılaştırılmış olup XGBoost ve NN modelleri diğer modellere göre üstün başarı göstermiştir. XGBoost modeline ilişkin metrikler; Doğruluk (%95.1), F1 skor (%97.3), Ağırlıklı F1 skor (%94.5), MKK (%73.6), EAA (%87.7), Duyarlılık (%100), Seçicilik (%94.7), PKD (%100) ve NKD (%57.1) elde edilmiştir. NN modeline ilişkin metrikler ise, Doğruluk (%95.1), F1 skor (%97.3), Ağırlıklı F1 skor (%94.5), MKK (%73.6), EAA (%86.1), Duyarlılık (%100), Seçicilik (%94.7), PKD (%100) ve NKD (%57.1) olarak hesaplanmıştır. Elde edilen bulgular incelendiğinde, XGBoost ve NN modelleri EAA metrikleri açısından farklılık gözlemlense de bu iki modelin diğer ölçütleri benzerlik göstermektedir. Modelin en belirleyici özellikleri alerji, yorgunluk, sararmış parmaklar, nefes darlığı ve alkol tüketimi gibi değişkenler yer almaktadır.

Sonuç: XGBoost ve NN modelleri, ağırlıklı F1 skor ve MKK ölçütleri bakımından akciğer kanseri sınıflandırmasında en başarılı ML yöntemleri olarak bulundu. Pozitif sınıfları doğru bir şekilde ayırt etmede oldukça başarılı yöntemler olduğu söylenebilir. Ayrıca değişken önem düzeyleri, alerji, yorgunluk, sararmış parmaklar, nefes darlığı ve alkol tüketiminin sınıflandırma doğruluğu üzerinde belirleyici olduğunu ortaya koymuştur. Elde edilen sonuçlar, ML tabanlı modellerin hem tanı sürecinde destekleyici karar sistemleri olarak kullanılabileceğini hem de akciğer kanserine ilişkin biyobelirteçlerin tanımlanmasında anlamlı katkı sunduğunu göstermektedir.

Anahtar Kelimeler: Makine öğrenmesi, sınıflandırma, performans metrikleri, akciğer kanseri

Comparative Performance Analysis of Machine Learning Models for Classification of Lung Cancer

Funda İpekten[1][0000-0002-6916-9563], Abdulvahap Pınar[2][0000-0002-3662-2579], Elif Çelik Gürbulak[3][0000-0002-5073-1907]

[1]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Adıyaman University, Adıyaman, Türkiye

[2]Rectorate Unit, Adıyaman University, Adıyaman, Türkiye

[3] Department of Biometrics, Erciyes University, Faculty of Veterinary Medicine, Kayseri, Türkiye

ABSTRACT

Aim: The goal of this study is to assess the performance of Logistic Regression (LR), eXtreme Gradient Boosting (XGBoost), Elastic-Net, and Neural-Net (NN) Machine Learning (ML) algorithms on a lung cancer dataset.

Method: The dataset used in the analyses was obtained from the Kaggle platform and contains 15 biomarker information of 309 individuals. The data was divided into 80% training set and 20% test set. Variable selection was performed with the Random Forest algorithm and 14 variables with the highest level of importance were included in the model. The performance of the classification models was evaluated using the metrics of Accuracy, F1 score, Weighted F1 score, Matthews Correlation Coefficient (MCC), Area under the Curve (AUC), Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV).

Results: For lung cancer classification, Logistic Regression (LR), XGBoost, Elastic Net, and Neural Network (NN) models were employed. Among these, XGBoost and NN outperformed the others based on multiple performance metrics. XGBoost achieved the following results: Accuracy (95.1%), F1 Score (97.3%), Weighted F1 Score (94.5%), MCC (73.6%), AUC (87.7%), Sensitivity (100%), Specificity (94.7%), PPV (100%), and NPV (57.1%). The NN model yielded identical values across all metrics except AUC (86.1%). Although AUC differed slightly, both models demonstrated comparable performance across other metrics. Key predictive features included allergy, fatigue, yellow fingers, shortness of breath, and alcohol consumption.

Conclusion: XGBoost and Neural Network models were identified as the most effective ML approaches for lung cancer classification based on Weighted F1 Score and MCC. Both models demonstrated high accuracy in detecting positive cases. Feature importance analysis highlighted allergy, fatigue, yellow fingers, shortness of breath, and alcohol consumption as key predictors. These findings suggest that ML-based models can support diagnostic decision-making and contribute to the identification of potential biomarkers for lung cancer.

Keywords: Machine learning, classification, performance metrics, lung cancer

1. Giriş

Akciğer kanseri, tüm dünyada hem erkeklerde hem de kadınlarda kansere bağlı ölümler içinde ilk sıralarda yer almakta ve ciddi bir halk sağlığı problemi oluşturmaktadır **(Sung et al., 2021)**. Dünya Sağlık Örgütü (WHO) verilerine göre, 2022 yılında yaklaşık 2,2 milyon yeni akciğer kanseri vakası bildirilmiş ve bu hastalığa bağlı 1,8 milyon ölüm gerçekleşmiştir **(WHO, 2023).** Bu yüksek mortalite oranının temel nedenlerinden biri, hastalığın çoğunlukla ileri evrede teşhis edilmesidir. Erken evrelerde semptomların ya hiç görülmemesi ya da spesifik olmaması, hastalığın geç tanınmasına yol açmaktadır **(ACS, 2022).** Erken tanı konan hastalarda sağkalım oranı önemli ölçüde yükselirken, ileri evre vakalarda bu oran dramatik biçimde düşmektedir. Bu durum, akciğer kanseriyle mücadelede erken tanının hayati önem taşıdığını açıkça ortaya koymaktadır.

Geleneksel tanı yöntemleri, genellikle görüntüleme teknikleri, biyopsi ve patolojik inceleme gibi invaziv yaklaşımlara dayanmakta olup, bu yöntemler zaman alıcı, maliyetli ve hasta açısından yorucu olabilmektedir. Buna ek olarak, risk gruplarının belirlenmesi ve erken tanı için biyobelirteçlerin kullanımı henüz yaygın klinik pratikte standartlaşmamıştır. Bu nedenle, özellikle düşük maliyetli, daha az invaziv ve daha hızlı tanı imkanı sağlayabilecek yeni teknolojilere ve hesaplama yöntemlerine duyulan ihtiyaç her geçen gün artmaktadır. Bu ihtiyaca yanıt olarak, son yıllarda yapay zeka (YZ) ve makine öğrenmesi (ML) temelli yaklaşımlar tıp ve biyomedikal alanlarda yaygın biçimde kullanılmaya başlanmıştır. ML algoritmaları, özellikle yüksek boyutlu, çok değişkenli ve gürültülü verilerle baş edebilme kapasitesiyle klasik istatistiksel yöntemlere alternatif olarak öne çıkmaktadır. Bu algoritmalar, veri içindeki karmaşık ilişkileri modelleyebilme ve öngörü yapabilme yetenekleri sayesinde, erken tanı, risk sınıflandırması ve kişiselleştirilmiş tedavi planlarının geliştirilmesi gibi alanlarda umut verici çözümler sunmaktadır **(Deo, 2015; Esteva et al., 2017).** Özellikle sınıflandırma problemlerinde yaygın olarak kullanılan makine öğrenmesi yöntemleri arasında Lojistik Regresyon (LR), eXtreme Gradient Boosting (XGBoost), Elastic Net ve Yapay Sinir Ağları (Neural Network, NN) gibi modeller yer almaktadır. LR, basitliği ve yorumlanabilirliği ile dikkat çekerken; XGBoost, güçlü ağaç tabanlı topluluk (ensemble) yöntemi olarak hem doğrusal hem de doğrusal olmayan örüntüleri öğrenmede başarılı sonuçlar vermektedir **[Chen, 2016].** Elastic Net ise L1 (lasso) ve L2 (ridge) cezalandırmalarını birleştirerek değişken seçimini dengeleyen hibrit bir regresyon yöntemidir **(Zou & Hastie, 2005)**. NN algoritmaları ise, özellikle derin öğrenme uygulamaları sayesinde büyük veri setlerinde oldukça karmaşık örüntüleri öğrenme kapasitesiyle dikkat çekmektedir **(LeCun, Bengio, & Hinton, 2015).** Bu tekniklerin hastalık sınıflandırması, biyobelirteç tanımlaması ve karar destek sistemlerinin geliştirilmesinde kullanılabilirliği, özellikle akciğer kanseri gibi karmaşık hastalıkların tanısında önemli fırsatlar sunmaktadır. Literatürde ML algoritmalarının akciğer kanseri tanısında kullanılmasına dair çeşitli çalışmalar yapılmış olup, bu çalışmalar bazı algoritmaların

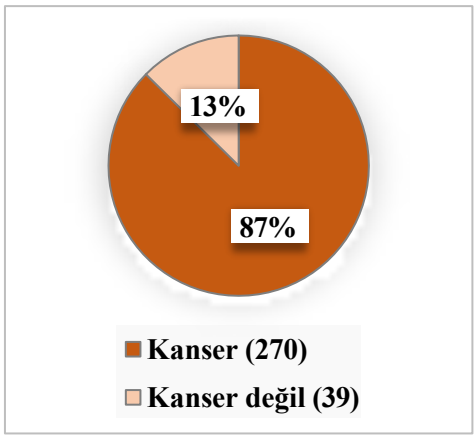
diğerlerine göre daha yüksek sınıflandırma performansı gösterdiğini bildirmiştir **(Radhika, Nair, & Veena, 2019).**

Bu çalışmanın temel amacı, akciğer kanseri veri seti üzerinde dört farklı makine öğrenmesi algoritmasının (Lojistik Regresyon, XGBoost, Elastic Net ve Yapay Sinir Ağı) sınıflandırma başarılarının karşılaştırmalı olarak değerlendirilmesidir. Ayrıca modellerin karar süreçlerine etki eden değişkenlerin belirlenmesiyle, sınıflandırma doğruluğu üzerinde etkili biyobelirteçlerin analizi yapılmıştır. Böylece, hem en etkili sınıflandırma modelinin belirlenmesi hem de potansiyel tanısal göstergelerin ortaya konulması hedeflenmiştir. Elde edilen sonuçların, ileri tanı destek sistemlerinin geliştirilmesi ve klinik karar verme süreçlerinin desteklenmesi açısından literatüre katkı sağlaması beklenmektedir.

2. Yöntem

2.1. Veri Seti ve Özellikler

Analizlerde kullanılan veri seti, çevrimiçi açık veri platformundan **(Kaggle, 2023)** alınmış olan bir gerçek veri olup, akciğer kanseri tanısı almış (270) veya almamış (39) toplam 309 bireyin verilerini içermektedir **(Şekil 2.1).** Veri setinde yer alan 15 bağımsız değişken, bireylerin semptomlarına (örneğin; yorgunluk, hırıltı, öksürük, nefes darlığı, yutma güçlüğü, göğüs ağrısı), yaşam tarzı alışkanlıklarına (örneğin; sigara ve alkol kullanımı), psikolojik durumlarına (anksiyete, sosyal baskı), alerjik reaksiyonlara ve kronik hastalık varlığına ilişkin bilgileri içermektedir. Veri seti eğitim (%80) ve test (%20) seti olarak ikiye ayrılmıştır. Modelde kullanılacak en anlamlı değişkenlerin belirlenmesi amacıyla Random Forest algoritması ile özellik seçimi uygulanmış ve en yüksek önem düzeyine sahip 14 değişken modele dahil edilmiştir. Özellikle sınıf dengesizliği problemi nedeniyle, modellerin performansını daha objektif değerlendirmek amacıyla Matthews Korelasyon Katsayısı ve ağırlıklı F1 skor metrikleri öncelikli olarak dikkate alınmıştır.



Şekil 2.1: Akciğer kanseri sınıf dağılımı ve yüzdeleri (Bağımlı değişken)

2.2. Uygulanan Makine Öğrenmesi Modelleri

Çalışmada dört farklı makine öğrenmesi algoritması karşılaştırılmıştır. Lojistik Regresyon modeli, ikili sınıflandırma problemlerinde yaygın olarak kullanılan ve açıklanabilirliği yüksek olan doğrusal bir yöntemdir **(Hosmer Jr, Lemeshow, & Sturdivant, 2013)**. Elastic Net, L1 (lasso) ve L2 (ridge) cezalarını birleştirerek yüksek korelasyonlu değişkenlerin olduğu durumlarda modelin aşırı öğrenmesini engelleyen bir düzenlileştirme yöntemidir **(Zou & Hastie, 2005).** XGBoost, karar ağaçlarına dayalı gradyan artırmalı bir topluluk öğrenme algoritması olup, sınıflandırma problemlerinde yüksek performansıyla öne çıkmaktadır **(Chen & Guestrin, 2016).** Yapay Sinir Ağı (NN) ise çok katmanlı yapısıyla özellikle doğrusal olmayan ilişkileri modelleyebilme açısından avantaj sağlamaktadır **(LeCun et al., 2015)**.

2.3. Model Performans Değerlendirmesi

Tablo 2.1. Sınıflandırma tablosu (Hata matrisi)

		Gerçek Sınıf	
		Pozitif	**Negatif**
Tahmin Edilen Sınıf	**Pozitif**	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

Veri ön işleme ve model eğitimi, makine öğrenmesi süreçlerinin temel adımlarından biri olmakla birlikte; modelin güvenilirliğini ve genellenebilirliğini belirlemede performans değerlendirmesi en az bu süreçler kadar belirleyici bir rol oynamaktadır. Bu kapsamda, model başarımı genellikle doğruluk (accuracy), F1 skoru, ağırlıklı F1 skoru, Matthews korelasyon katsayısı (MCC), ROC eğrisi altındaki alan (AUC), duyarlılık (sensitivity), seçicilik (specificity), pozitif kestirim değeri (PPV) ve negatif kestirim değeri (NPV) gibi çeşitli ölçütler kullanılarak çok boyutlu olarak değerlendirilmektedir. Veri setindeki gerçek hastalık durumu ile sınıflandırma modelimiz ile tahmin edilen hastalık durumu arasındaki sayıların belirlenmesi ile bu ölçütler hesaplanmaktadır. Gerçek durum ile tahmin edilen durumun sayısal gösterimine iliskin tablo 2.1'de gösterilmektedir.

Tablo 2.2. Model Performans değerlendirme metrikleri

Kullanılan Metrikler	Formüller
Doğruluk	$\frac{DP + DN}{DP + DN + YP + YN}$
F1 skor	$2 * \left(\frac{DP}{2} * DP + YP + YN\right)$
Ağırlıklı F1 skor	Ağırlıklı $F1 = \frac{\sum_{i=1}^{K} n_i * F1_i}{\sum_{i=1}^{K} n_i}$
Matthews korelasyon katsayısı	$(DP * DN) - \frac{YP * YN}{DP + YP} * (DP + YN) * (YP + DN) * (YN + DN)$
Eğri Altında kalan Alan	$\int_0^1 TPR(x)d(FPR(x))$
Duyarlılık	$\frac{DP}{DP + YN}$
Seçicilik	$\frac{DN}{DN + YP}$
Pozitif kestirim değeri	$\frac{DP}{DP + YP}$
Negatif kestirim değeri	$\frac{DN}{DN + YN}$
TPR: True Positive Rate (Duyarlılık), FPR: False Positive Rate (1-seçicilik)	

Modellerin sınıflandırma başarımları, test verisi üzerinde çeşitli ölçütler ile değerlendirilmiştir. Bu ölçütler arasında sınıflandırma başarısının genel göstergesi olan doğruluk (Accuracy), sınıflar arası dengenin sağlanmasında etkili olan F1 skoru ve ağırlıklı F1 skoru yer almıştır. Bunun yanında, özellikle dengesiz sınıf yapılarında daha güvenilir bir değerlendirme sağlayan Matthews korelasyon katsayısı (MCC) kullanılmıştır. Modelin ayırt edici gücünü ölçmek amacıyla ROC eğrisi altındaki alan (AUC), pozitif vakaları saptama başarısını ölçmek üzere duyarlılık (Recall/sensitivity) ve pozitif kestirim değeri, negatif sınıfları doğru tanımlamadaki başarı için ise seçicilik (Specificity) ve negatif kestirim değeri hesaplanmıştır **(Tablo 2.2).**

3. Bulgular

Analiz sürecinde elde edilen bulgulara aşağıda yer verilmiş; her bir makine öğrenmesi algoritmasının, sınıflandırma metrikleri ve değişken önem düzeyleri ayrı ayrı raporlanmıştır.

Tablo 3.1. Sınıflandırma Modellerine Ait Performans Değerlendirme Metrikleri

Model	Doğruluk	F1 skor	Ağırlıklandırılmış F1 skor	MKK	EAA	Duyarlılık	Seçicilik	PKD	NKD
LR	91.8	95.5	90.8	52.8	90.6	75.0	93.0	98.1	42.9
XGBoost	**95.1**	**97.3**	**94.5**	**73.6**	**87.7**	**100.0**	**94.7**	**100.0**	**57.1**
Elastic Net	93.4	96.4	92.2	63.2	92.5	100.0	93.1	100.0	42.9
Neural Network	**95.1**	**97.3**	**94.5**	**73.6**	**86.1**	**100.0**	**94.7**	**100.0**	**57.1**

LR: Lojistik Regresyon, MKK: Matthews korelasyon katsayısı, EAA: Eğri altında kalan alan, PKD: Pozitif kestirim değeri, NKD: Negatif kestirim değeri. Veriler yüzde (%) olarak ifade edilmiştir. Yüksek değerler kalın yazıyla belirtilmiştir.

Modellerin test veri seti üzerindeki başarımlarına göre, XGBoost ve Yapay Sinir Ağı (NN) modelleri en yüksek performans düzeylerine ulaşmıştır. Her iki model de %95.1 doğruluk, %97.3 F1 skoru, %94.5

ağırlıklı F1 skoru ve %73.6 Matthews korelasyon katsayısı elde etmiştir. Ayrıca ROC eğrisi altında XGBoost modeli için %87.7, NN modeli için ise %86.1'lik alan hesaplanmıştır. Duyarlılık ve pozitif kestirim değeri her iki modelde de %100 olup, bu modellerin pozitif (kanserli) sınıfları ayırt etmede oldukça başarılı olduğu görülmüştür. Seçicilik değeri %94.7, negatif kestirim değeri ise %57.1 olarak hesaplanmıştır. Lojistik Regresyon ve Elastic Net modellerinin doğruluk düzeyleri sırasıyla %86.8 ve %88.7 olarak bulunmuştur. Ancak bu modeller, pozitif sınıfları tanılamada XGBoost ve NN modellerine kıyasla daha düşük başarı sergilemiştir **(Tablo 3.1).** Özellik önem düzeylerine göre en etkili ilk beş değişkenler sırasıyla alerji, yorgunluk, sararmış parmaklar, nefes darlığı ve alkol tüketimi olduğu belirlenmiştir. Elde edilen sonuca göre, bu değişkenlerin akciğer kanseri sınıflandırmasında güçlü belirleyiciler olduğu söylenebilir.

4. Tartışma

Çalışma sonuçları, özellikle XGBoost ve Yapay Sinir Ağı (YSA) modellerinin akciğer kanseri sınıflandırmasında üstün performans gösterdiğini ortaya koymuştur. Çalışma sonuçlarımız literatürle benzer sonuçlar göstermiş olup özellikle XGBoost modeli, kanser sınıflandırmasında sıkça tercih edilen ve başarılı sonuçlar veren bir yöntem olarak öne çıkmaktadır. **Guan ve arkadaşlarının (2023)** metabolik indekslere dayalı çalışmasında da XGBoost, geleneksel yöntemlere kıyasla daha başarılı sonuçlar vermiştir **(Guan et al., 2023). Benveniste ve arkadaşları (2023)** ise PLCO ve NLST veri setleri üzerinde geliştirdikleri XGBoost tabanlı modelle 5 yıllık akciğer kanseri risk tahmininde yüksek AUC değerleri elde etmişlerdir **(Benveniste, Alberge, Xing, & Bibault, 2023).** Mevcut çalışmamızda XGBoost ve Neural Network algoritmalarının %95.1 doğruluk oranı ile öne çıkması, bu yöntemlerin farklı veri türlerinde ve farklı örneklem gruplarında istikrarlı biçimde yüksek performans sunduğunu göstermektedir. Elde ettiğimiz sonuçlar, pozitif sınıfların (kanser vakalarının) doğru sınıflandırılmasında bu iki modelin özellikle etkili olduğunu ortaya koymuştur. %100 duyarlılık değeri, hiçbir kanser vakasının gözden kaçırılmadığını göstererek bu modellerin erken teşhis potansiyeline güçlü bir katkı sunduğunu ortaya koymaktadır. Literatürde rapor edilen derin öğrenme temelli çalışmalar da benzer şekilde başarılı sonuçlar bildirmektedir. Örneğin, **Khattar ve arkadaşları (2024)** tarafından geliştirilen konvolüsyonel sinir ağı (CNN) tabanlı model, yüksek duyarlılık ve seçicilik oranları ile dikkat çekmiştir **(Khattar et al., 2024).** Benzer şekilde, **Nasrullah ve arkadaşları (2019),** derin öğrenme temelli sistemlerinin akciğer nodülü tespiti ve sınıflandırmasında yüksek performans sergilediğini rapor etmiştir **(Nasrullah et al., 2019).** Ancak, çalışmamızda ulaşılan %100 duyarlılık değeri, bu modellerle karşılaştırıldığında daha yüksek bir başarı düzeyine işaret etmektedir. Çalışmada, Random Forest algoritması ile gerçekleştirilen değişken seçimi sonucunda 15 özellikten 14'ünün belirlenmiş olması, modelin açıklayıcılığını korumuştur. Seçilen değişkenlerin arasında öne çıkanlar; alerji, yorgunluk, sararmış parmaklar, nefes darlığı ve alkol tüketimidir.

Bu bulgular, literatürdeki klinik gözlemlerle büyük ölçüde uyumludur. Rivera ve arkadaşları (2013), nefes darlığını akciğer kanserinin erken belirtilerinden biri olarak tanımlamış; Mock ve arkadaşları (2015) ise yorgunluğun, kanserle ilişkili sistemik bir semptom olduğunu belirtmiştir **(Mock et al., 2007; Rivera, Mehta, & Wahidi, 2013).** Alkol tüketimiyle akciğer kanseri arasındaki ilişki, Fehringer ve arkadaşları (2017) tarafından bazı durumlarda bağımsız bir risk faktörü olarak değerlendirilmiştir **(Fehringer et al., 2017).** Çalışmamızda bu değişkenin yüksek önem düzeyiyle öne çıkması, alkol tüketiminin göz ardı edilmemesi gereken bir faktör olabileceğini düşündürmektedir. Sararmış parmaklar (clubbing) ise, Spicknall ve arkadaşlarının (2005) vurguladığı üzere akciğer kanserinin fiziksel ipuçlarından biri olarak değerlendirilmektedir **(Spicknall, Zirwas, & English III, 2005).** Bu tür semptomların modelde yüksek önem derecesine sahip olması, fiziksel muayene bulgularının makine öğrenmesi tabanlı sınıflandırma sistemlerine entegre edilmesinin değerini göstermektedir. Öte yandan, çalışmamızda kullanılan Lojistik Regresyon (LR) ve Elastic Net modelleri, açıklanabilirlik açısından avantaj sunmakla birlikte, sınıflandırma başarımı bakımından daha sınırlı kalmıştır. Bu durum literatürde de benzer şekilde vurgulanmaktadır. Örneğin, Hosmer et al. (2013), LR'nin özellikle küçük veri setlerinde ve doğrusal ilişkilerin baskın olduğu durumlarda başarılı olabildiğini ancak karmaşık örüntülerde sınırlı kaldığını ifade etmiştir **(Hosmer Jr et al., 2013).** Çalışmamızda ulaşılan bulgular, makine öğrenmesi tabanlı algoritmaların özellikle erken tanı süreçlerinde güçlü karar destek sistemleri olarak kullanılabileceğini göstermektedir. Yüksek duyarlılık, tarama programlarında kanser vakalarının gözden kaçma olasılığını minimize ederken, nispeten düşük negatif kestirim değeri, yanlış pozitif sonuçlara bağlı olarak klinik kaynakların gereksiz kullanım riskini gündeme getirmektedir. Bu nedenle, modelin gerçek negatifleri daha iyi ayırt edebilecek şekilde optimize edilmesi, pratik uygulamalarda daha dengeli sonuçlar elde edilmesine katkı sağlayacaktır. Sonuç olarak, bu çalışma, XGBoost ve Neural Network gibi güçlü algoritmaların akciğer kanseri teşhisinde yüksek başarı potansiyeline sahip olduğunu ortaya koymakta; özellik seçimi ve modelleme stratejilerinin dikkatli şekilde yapılandırılması durumunda sınıflandırma doğruluğunun artırılabileceğini göstermektedir. Bulgularımız, biyobelirteç tanımlama süreçlerine katkı sağlayarak, bireye özgü tedavi planlamalarının geliştirilmesinde önemli bir temel oluşturabilir.

5. Sonuç

Bu çalışma, akciğer kanseri sınıflandırmasında makine öğrenmesi (ML) tabanlı modellerin etkili bir şekilde kullanılabileceğini göstermiştir. Dört farklı algoritmanın karşılaştırıldığı analizde, XGBoost ve Yapay Sinir Ağı (NN) modelleri özellikle pozitif sınıfları doğru tespit etme açısından üstün performans sergilemiş; F1 skoru, ağırlıklı F1 skoru ve Matthews Korelasyon Katsayısı gibi metriklerde diğer modellere kıyasla daha başarılı sonuçlar elde etmiştir. Değişken önem düzeyi analizinde; alerji, yorgunluk, sararmış parmaklar, nefes darlığı ve alkol kullanımı gibi semptomatik ve davranışsal değişkenlerin sınıflandırma kararında

belirleyici rol oynadığı görülmüştür. Bu bulgu, söz konusu değişkenlerin potansiyel biyobelirteç olarak değerlendirilmesine olanak sağlamakta ve literatürde tanımlanan klinik göstergelerle uyum göstermektedir.

Elde edilen sonuçlar, ML algoritmalarının yalnızca tanı sürecinde değil, aynı zamanda klinik karar destek sistemlerinin geliştirilmesinde ve yüksek risk taşıyan bireylerin erken dönemde saptanmasında etkili araçlar olduğunu ortaya koymaktadır. Bu tür yaklaşımlar, bireyselleştirilmiş erken tanı stratejilerinin oluşturulmasına katkı sağlayarak, hastalıkla mücadelede önemli bir rol üstlenebilir. İlerleyen çalışmalarda bu modellerin daha büyük, dengeli ve çok merkezli veri setleriyle doğrulanması, tanısal doğruluğun artırılması açısından kritik bir adım olacaktır.

6. Kaynaklar

ACS. (2022). Lung cancer early detection. *American Cancer Society*. Retrieved from https://www.cancer.org

Benveniste, P.-L., Alberge, J., Xing, L., & Bibault, J.-E. (2023). Development and external validation of a lung cancer risk estimation tool using gradient-boosting. *arXiv preprint arXiv:2308.12188*.

Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature, 542*(7639), 115-118.

Fehringer, G., Brenner, D. R., Zhang, Z. F., Lee, Y. C. A., Matsuo, K., Ito, H., . . . Overvad, K. (2017). Alcohol and lung cancer risk among never smokers: A pooled analysis from the international lung cancer consortium and the SYNERGY study. *International journal of cancer, 140*(9), 1976-1984.

Guan, X., Du, Y., Ma, R., Teng, N., Ou, S., Zhao, H., & Li, X. (2023). Construction of the XGBoost model for early lung cancer prediction based on metabolic indices. *BMC medical informatics and decision making, 23*(1), 107.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*: John Wiley & Sons.

Kaggle. (2023). Lung Cancer Prediction Dataset. Retrieved from https://www.kaggle.com/datasets/thedevastator/diagnosis-of-lung-cancer

Khattar, S., Aftaab, M., Verma, T., Patial, D., Kaur, B., San, H. T. S., & Kaur, B. (2024). *Deep learning-based lung cancer detection using convolutional neural networks.* Paper presented at the AIP Conference Proceedings.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

Mock, V., Abernethy, A. P., Atkinson, A., Barsevick, A. M., Berger, A. M., Cella, D., . . . Escalante, C. P. (2007). Cancer-related fatigue clinical practice guidelines in oncology. *JNCCN Journal of the National Comprehensive Cancer Network, 5*(10), 1054-1078.

Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors, 19*(17), 3722.

Radhika, P., Nair, R. A., & Veena, G. (2019). *A comparative study of lung cancer detection using machine learning algorithms.* Paper presented at the 2019 IEEE international conference on electrical, computer and communication technologies (ICECCT).

Rivera, M. P., Mehta, A. C., & Wahidi, M. M. (2013). Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest, 143*(5), e142S-e165S.

Spicknall, K. E., Zirwas, M. J., & English III, J. C. (2005). Clubbing: an update on diagnosis, differential diagnosis, pathophysiology, and clinical relevance. *Journal of the American Academy of Dermatology, 52*(6), 1020-1028.

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

WHO. (2023). Cancer fact sheet. *World Health Organization*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cancer

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 67*(2), 301-320.

State Detection of Brain Tumors and Patient Life Expectancy Prediction with Explainable Artificial Intelligence-Based Multimodal System

Ahmet Caner Tat[1][0009-0004-4403-8251], Emirhan Yıldız[2][0009-0003-6066-268X] and Salih Kızılışık[3][0009-0007-2340-3610]

[1,2,3] Gazipaşa Mahallesi, Eskişehir Teknik Üniversitesi İki Eylül Kampüsü, 26555, Eskişehir, Türkiye
tatahmetcaner@gmail.com

Abstract. Accurate and early diagnosis of brain tumors is vital for improving patient outcomes and guiding treatment. This study presents an explainable, AI-driven multimodal system that integrates MRI, histopathology images, and clinical data to detect glioma states and predict life expectancy. Using techniques like SHAP, the system offers transparent decision-making support for clinicians. The modular architecture allows flexibility in clinical use while enhancing interpretability. The model achieves strong performance and provides actionable insights for oncologists, radiologists, and pathologists. Accurate and early diagnosis of brain tumors is vital for improving patient outcomes and guiding treatment. This study presents an explainable, AI-driven multimodal system that integrates MRI, histopathology images, and clinical data to detect glioma states and predict life expectancy. Using techniques like SHAP, the system offers transparent decision-making support for clinicians. The modular architecture allows flexibility in clinical use while enhancing interpretability. The model achieves strong performance and provides actionable insights for oncologists, radiologists, and pathologists.

Keywords: Brain cancer, Artificial Intelligence, Multimodal models,

1 Introduction

The healthcare sector is undergoing a significant transformation due to rising costs and a global shortage of healthcare professionals. These pressures have accelerated the adoption of IT-based solutions aimed at improving the quality, accessibility, and efficiency of healthcare services [1]. The COVID-19 pandemic further exposed vulnerabilities in healthcare systems, including shortages of diagnostic tools, overburdened staff, and fragmented data sharing [2]. During this period, artificial intelligence (AI) demonstrated its potential across various domains, ranging from diagnosis and treatment decision support to contact tracing and pandemic response systems [3].

Despite these advances, many existing AI-assisted diagnostic systems rely on single-modality data, either text-based or image-based, which limits their ability to capture the complex and multidimensional nature of medical conditions [4]. Recent research has shown that integrating multiple data types through multimodal learning improves predictive performance by 1.2% to 27.7% compared to unimodal approaches [5].

Gliomas, which originate from glial cells, represent about one-third of all brain tumors and can significantly impair neurological function. Their classification includes glioblastoma, astrocytoma, and oligodendroglioma, each with differing prognoses and treatment strategies. Accurate diagnosis and survival prediction are critical, especially since gliomas can exhibit aggressive behavior and resistance to treatment. Prior studies have attempted to classify gliomas using CNNs, achieving promising but limited results. For example, Ertosun et al. reported 96% accuracy in binary classification between glioblastoma and low-grade gliomas, but only 71% between Grade II and III [6]. Jin et al. achieved 87.5% accuracy by incorporating histological and molecular features [7]. More recently, a study aligned with the 2021 WHO classification achieved 86.1% accuracy and an AUC of 0.961 using molecular markers like IDH mutation and 1p/19q codeletion [8].

However, these approaches often focus solely on classification and do not address survival prediction. Few studies, such as a 2021 clinical cohort analysis, attempted to predict survival time but were limited to clinical data [9]. Others explored imaging-based survival models using modalities like MRI and PET but noted limitations in reflecting the full biological complexity of gliomas [10].

To overcome these challenges, this study presents a multimodal deep learning system that integrates MRI, histopathology, and clinical data for comprehensive glioma analysis. Our approach complies with the 2021 WHO classification and expands the scope of previous models by not only classifying tumors into six categories but also predicting both overall and postoperative survival times. Through this model, we aim to offer clinicians a more reliable and efficient AI-powered tool for informed decision-making in glioma diagnosis and management [11].

2 Methodology

Multi-modal systems inherently involve high complexity due to the integration of diverse data types and processing techniques. To facilitate a clearer understanding of our system, we divide it into three distinct components: the MRI module, the Histopathology module, and an MLP-based integration module that combines these with clinical data.

2.1 Datasets

In the literature, there are few datasets available for diffuse-type glioma due to its rarity. During the research, we identified publicly available datasets suitable for our study from various data collection studies. We have identified four datasets that we

can utilize for our project: UPENN-GBM [12], TCGA-GBM[13], TCGA-LGG[14], and CPTAC-GBM[15]. Some data in these collections contains images that could potentially be used to reconstruct a human face. To safeguard the privacy of participants, we have signed a TCIA Restricted License Agreement to access datasets. The type and amount of data contained in each of the data sets are detailed in Table 1.

	UPENN-GBM	TCGA-GBM	TCGA-LGG	CPTAC-GBM
Clinicial Data	Full	Full	Full	Full
MRI Segmented	611	-	-	-
MRI Raw data	630	262	199	70
Histopathology	71	262	199	200

Table 1. Types and numbers of data from datasets

Magnetic resonance imaging (MRI) plays a critical role in radiotherapy by providing superior soft tissue contrast compared to computed tomography (CT), the standard imaging modality in the field. Multiparametric MRI (mpMRI), which integrates anatomical and functional imaging, enables advanced disease detection, localization, and treatment response assessment. This facilitates personalized radiotherapy by capturing both structural and biochemical tissue changes [16]. MRI data is typically stored in two main format categories: standard formats such as DICOM , and research-oriented formats including Analyze, NIfTI, and MINC, which are optimized for post-processing [17]. The dataset comprises multiple MRI sequences (e.g., DWI, DCE, T1, T1CE, T2, Fluid) acquired under varying scanner parameters. Some images are already in NIfTI format and require minimal preprocessing, while others are raw DICOM images.

Histopathology involves microscopic examination of tissue samples obtained via biopsy, providing detailed insight into tumor morphology. The dataset includes high-resolution digital histopathology images, allowing for the visualization of individual tumor cells and nuclei. These images occupy substantial storage space and are primarily in DICOM format, with a subset available in OpenSlide (NDPI) format.

Clinical data collected during patient diagnosis and treatment, with appropriate consent, comprise both numerical and categorical variables. Despite challenges such as institutional privacy constraints and missing features across datasets, essential information, including patient age, gender, survival status, and glioma grade is available. This data supports model development by informing patient stratification during training, validation, and testing, and facilitates survival analysis and outcome prediction.

2.2 MRI Module

MR images from mpMRI machines, typically in DICOM format, were converted to the standardized NIfTI format to facilitate preprocessing. Using SimpleITK, we handled image metadata, resampled scans to a uniform resolution, and normalized intensities. The preprocessing pipeline (Fig. 3) included several key steps: bias field correction for intensity uniformity, B-spline interpolation to standardize voxel spacing (3 mm), and skull stripping via the nnU-NetV2's Brain Extraction Tool to remove non-brain tissues. Histogram matching aligned intensity distributions across scans, while Curvature Flow filtering denoised images without blurring anatomical boundaries. Spatial cropping removed irrelevant regions, and images were finally resized to 128×128×52 to meet CNN input requirements.

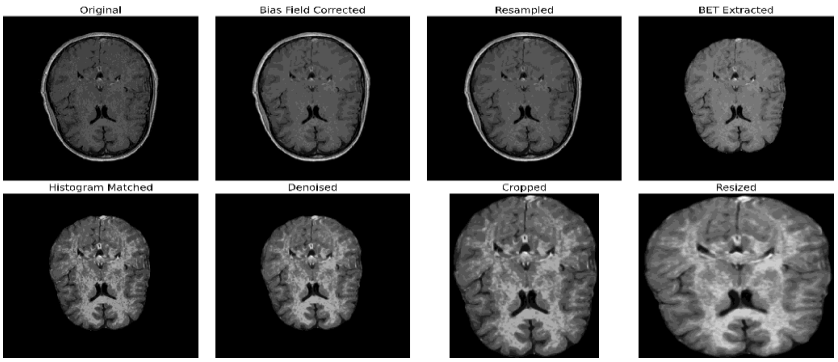


Fig. 1. Preprocessing pipeline for MR Images

For feature extraction from MRI slices, the InceptionV3 convolutional neural network (CNN) architecture was employed. Originally developed by Google for image classification tasks, InceptionV3 offers improved computational efficiency and accuracy compared to earlier Inception models. Since ImageNet pre-trained weights are not directly applicable to brain imaging, the model was fine-tuned using two public brain MRI datasets from Kaggle [18,19], allowing for better adaptation to domain-specific features. Each MRI slice passed through the fine-tuned model yields a 2048-dimensional feature vector. Given that each patient scan contains 52 slices, this process was repeated for each slice. The dataset was divided into training (80%), validation (10%), and test (10%) subsets to ensure robust model evaluation.

To model the sequential nature of MRI slices, a Bidirectional Long Short-Term Memory (BiLSTM) network was integrated. While CNNs effectively capture spatial features from individual slices, RNNs—particularly LSTM and GRU architectures—enable temporal context modeling across the sequence. The BiLSTM processes MRI volumes in both forward and backward directions, allowing the model to incorporate context from both preceding and succeeding slices. This dual perspective improves the model's ability to recognize complex anatomical and pathological patterns.

Prior to integration into the main model, the LSTM was pretrained on a separate dataset of 797 brain MR images lacking histopathological data. During pretraining, two

dense layers (32 neurons each) were appended to the LSTM for feature learning, then removed after pretraining. The dataset was split into 70% training and 30% validation. Following pretraining, the BiLSTM layers were frozen to preserve learned representations of brain structure. The final BiLSTM module outputs two sets of 64 features, one from each direction, which are concatenated into a single 128-dimensional feature vector representing the entire MRI volume.

2.3 Histopathology Module

The preprocessing pipeline standardizes raw histopathology images from multiple formats (e.g., whole-slide, DICOM, NIfTI, NumPy) into clean, analysis-ready image tiles. This involves center cropping and intensity-based filtering to remove background and irrelevant regions. For whole-slide images, the tissue area is partitioned into fixed-size tiles (e.g., 240×240 pixels), retaining only those with adequate tissue content. Tiles are optionally converted to grayscale and normalized to reduce variation across samples and enhance consistency.

To capture intra-patient variability and identify discriminative tissue patterns, a clustering approach was employed. Thirty patients from each of the three glioma subtypes (Glioblastoma, Astrocytoma, Oligodendroglioma) were selected. Patient-level image tiles were clustered using the K-means algorithm, which assigns tiles to K groups based on feature similarity. K-means iteratively updates cluster centroids by minimizing intra-cluster distances. Despite limitations such as sensitivity to initial centroid placement and the need to predefine K, K-means offers computational efficiency, making it suitable for high-dimensional histopathology data. To manage memory and computation, clustering was conducted in mini-batches of 10,000 tiles.

A VGG-16 CNN, pre-trained on ImageNet and fine-tuned using a subset of 90 patients, was used to evaluate the discriminative power of the resulting clusters. Cluster performance was assessed based on predictive accuracy for glioma subtype classification. Based on the evaluation, the three most informative clusters (Clusters 2, 6, and 8) were selected. A balanced sampling strategy was applied to select 2,500 frames proportionally from these clusters for further modeling.

Feature extraction from the selected 2,500 frames was performed using the convolutional layers of the fine-tuned VGG-16 benchmark model. Each frame was transformed into a 512-dimensional feature vector, resulting in a feature matrix of shape (2500, 512). To avoid information leakage, feature extraction and benchmarking were performed solely on the training set, ensuring the integrity of the test set throughout the modeling pipeline.

A transformer-based model was employed to process the sequential histopathology features. Pretraining was conducted on 979 histopathology slides not paired with MRI data, using a 70/30 train-validation split. During pretraining, two dense layers (32 neurons each) were added to aid in learning and later removed when the transformer was integrated into the main model. The pretrained model was frozen to preserve learned tissue structure representations. This transformer setup enables the model to learn both fine-grained morphological and broader contextual features from histo-

pathology, enhancing its ability to support glioma subtype classification when fused with other modalities.

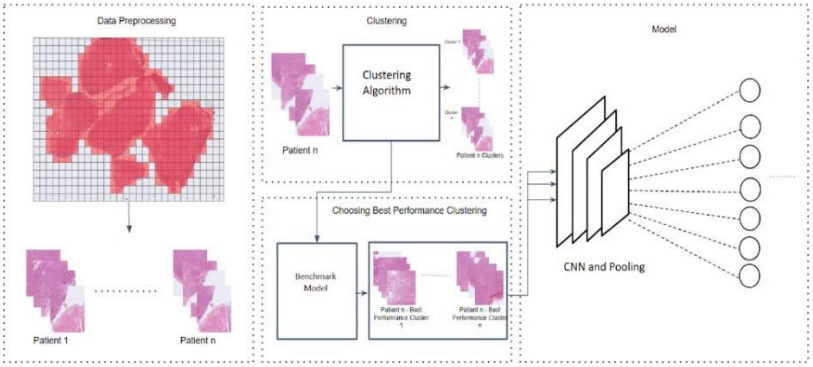


Fig. 2. Histopathology module structure

2.4 Clinical Data and MLP Integration

A Multilayer Perceptron (MLP) is employed as the final decision-making module in the multimodal pipeline. The MLP is a fully connected neural network architecture consisting of an input layer, one or more hidden layers, and an output layer. Each neuron in these layers performs a linear transformation followed by a non-linear activation (e.g., ReLU or sigmoid), enabling the network to model complex, non-linear relationships across different data modalities. MLPs are particularly well-suited for structured and low-dimensional data and are commonly used in multimodal fusion tasks to learn joint representations across input domains.

In this framework, high-level features extracted independently from the MRI and histopathology modules are concatenated with patient-specific clinical data (such as age and gender) to form a unified, comprehensive multimodal feature vector. This integrated representation captures both anatomical structures and demographic context, offering a holistic view of the patient's condition. The combined feature vector is then input to the MLP, which learns interdependencies between modalities and contributes to the final classification.

All data modalities (MRI, histopathology, and clinical data) were rigorously matched at the patient level to ensure consistency and prevent label leakage. After data matching, a total of 289 complete patient samples were retained. The dataset was subsequently split into training (60%), validation (20%), and test (20%) sets. As shown in Figure 3, the MLP model provides predictions for postoperative survival time, Mortality, Tumor Grade, and Glioma Type.

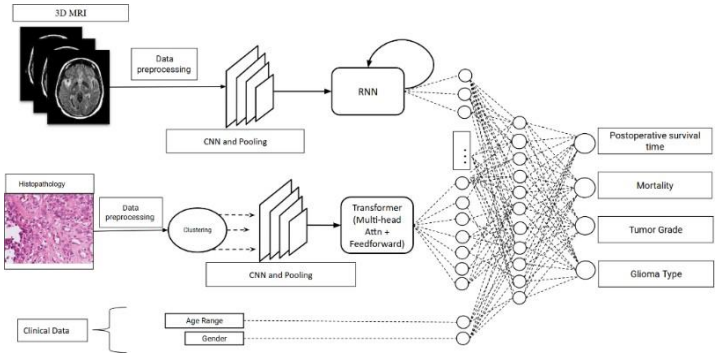


Fig. 3. Whole system structure

2.5 Explainability Stage

Deep neural network-based architectures are complex and often have a high number of hidden layers, making them difficult to interpret [20]. Hence, they are referred to as black-box models. Despite achieving high success rates in diagnosing diseases, the reliability of these systems' decisions is questionable due to their black-box nature.

SHAP (SHapley Additive exPlanation) is used to understand the extent to which features contribute to the decision. In our project, SHAP helps us to determine the overall contribution percentages in prediction between clinical data, MRI, and Histopathology data. SHAP also helps us identify the most contributing histopathology image segments and present them to the user.

2.6 Website Integration

A web application that allows users to upload MRI and histopathology image files, along with patient age and gender, through a web interface (React frontend). When a user submits the form, the backend receives the files and information, saves the MRI and histopathology files into organized folders on the server, and logs the details. It then processes the gender input into a numeric format and calls a prediction function that analyzes the uploaded data and clinical information. The backend responds to the frontend with confirmation that the upload and processing were successful, along with the submitted details. The site is configured to accept requests only from the designated frontend securely. It serves uploaded files as static content, making it a bridge between the user-facing website and the machine learning inference system.

3 Results

3.1 Single Modality Model Evaluations

To express the effectiveness of the multimodal approach, we conducted an experiment in which one of the primary data modalities was removed by setting its input

values to zero while keeping the overall model architecture unchanged. By evaluating the model's performance without this modality, we aimed to gain insights into the contribution and significance of each data type.

The results of the model illustrated in Figure 4 and Table 3, the contribution of MRI images to the outcome was 97.91%, while clinical data contributed 2.09%, and histopathology provided no measurable contribution (0.00%) according to Shap analysis.

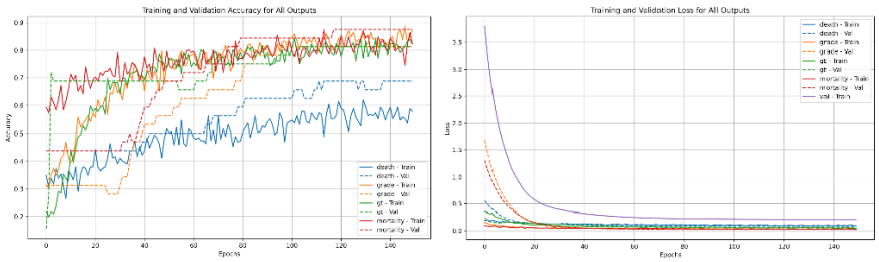


Fig. 4. Accuracy and loss graphs of MRI + Clinical

	ACC	PREC	RECALL	F1 Score
GRADE	0.875	0.9545	0.875	0.9133
MORTALITY	0.9063	0.9565	0.9167	0.9359
DEATH (MACRO AVG)	0.7188	0.7690	0.7460	0.7424
TYPE (MACRO AVG)	0.875	0.895	0.708	0.772

Table 2. MRI+Clinical Test Metrics

The results of the model are presented in Figure 5 and Table 3. In this case, MR images contributed 0%, clinical data contributed 2.08%, and histopathology images contributed 97.92% to the outcome according to Shap analysis.

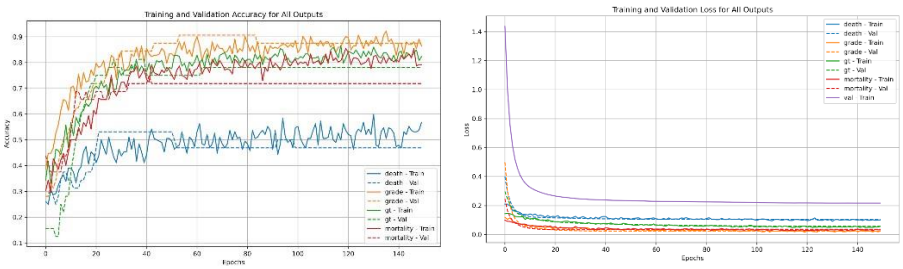


Fig. 5. Accuracy and Loss graphs of Histopathology + Clinical

	ACC	PREC	RECALL	F1 Score
GRADE	0.9063	0.9565	0.9167	0.9364
MORTALITY	0.9063	0.9167	0.9565	0.9362
DEATH (MACRO AVG)	0.6250	0.6389	0.6482	0.6339
TYPE (MACRO AVG)	0.8750	0.8949	0.7083	0.7711

Table 3. Histopathology + Clinical Test Metrics

As can be seen from the results, MR images provide valuable insights into the patient's prognosis and future outcomes. However, histopathology images contain more detailed and usable information for determining the type and grade of glioma.

3.2 Multimodal Results

As shown in Figure 6, and Table 4, the multimodal model's predictions were primarily influenced by histopathology images (53.31%), followed by MRI images (44.89%) and clinical data (1.80%), as determined by SHAP analysis.

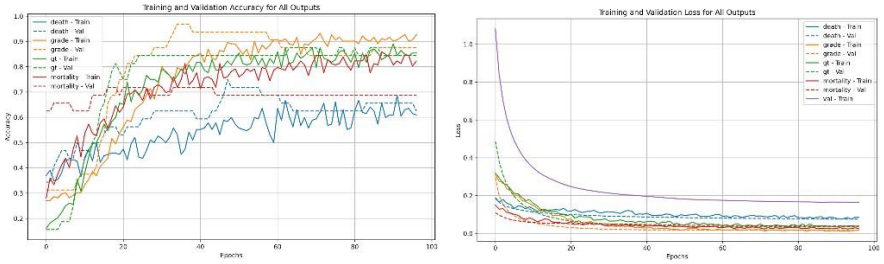


Fig. 6. Accuracy and loss graphs of MRI + Histopathology + Clinical

	ACC	PREC	RECALL	F1 Score
GRADE	0.9375	0.9583	0.9583	0.9583
MORTALITY	0.9688	0.96	1	0.9796
DEATH (MACRO AVG)	0.8125	0.7897	0.8307	0.8082
TYPE (MACRO AVG)	0.9394	0.9524	0.8333	0.8633

Table 4. MRI + Histopathology + Clinical Test Metrics

Compared to single-modality models, the multimodal approach leverages the strengths of both data types. From histopathology images, it primarily extracts information related to glioma type and grade, while MRI images contribute prognostic information such as mortality risk and expected survival. By combining these complementary sources, the main model outperforms the individual models by effectively integrating and processing both types of information. Although this integration introduces greater complexity and potential instability, these challenges can be mitigated using appropriate modeling techniques and careful hyperparameter tuning.

3.3 Confusion Matrices

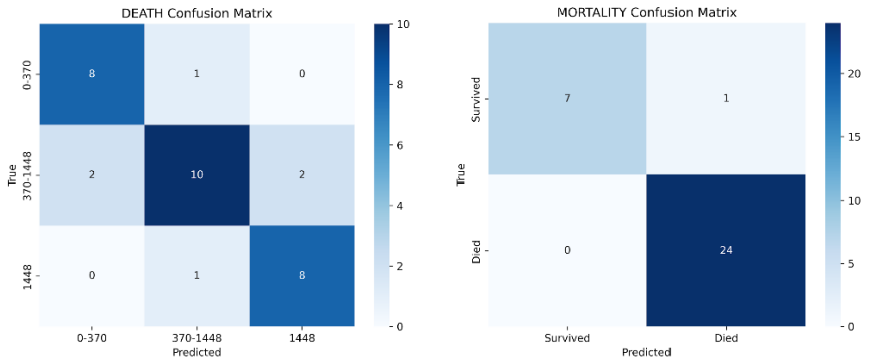


Fig. 7. Confusion matrices of life expectancy and mortality

As shown in Figure 7, our system demonstrated high performance in predicting mortality but struggled slightly in predicting life expectancy. This showed us that life expectancy is much more reliant on multiomic data (the integration of multiple layers of biological information and other patient factors to provide a comprehensive view of biological systems and disease mechanisms).

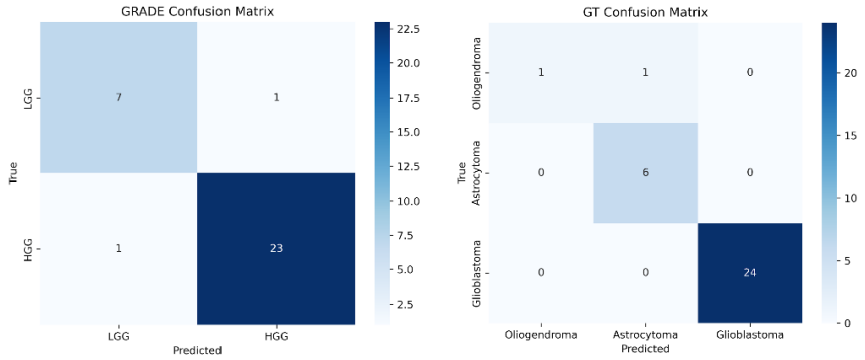


Fig. 8. Confusion matrices of tumor grade and tumor type

4 Discussion and conclusions

A key challenge in this study was the class imbalance among glioma types, particularly the low number of Astrocytoma and Oligodendroglioma cases, which limited the model's ability to generalize. This imbalance affected both training and testing sets, reducing the reliability of tumor-type classification metrics. To address a similar issue in tumor grade prediction, Grades 2 and 3 were merged into a 'low-grade glio-

ma' category, and the rest as 'high-grade', improving classification stability.

Future work should explore data augmentation and additional MRI modalities like diffusion-weighted imaging (DWI). Incorporating tumor segmentation may add spatial biomarkers, while expanding the dataset to include genomic and proteomic profiles could align the model more closely with precision medicine goals.

This thesis proposed an explainable, AI-based multimodal system for diagnosing glioma types and predicting patient life expectancy. By integrating MRI, histopathology, and clinical data, the model overcame the limitations of unimodal approaches. It aligns with the 2021 WHO classification and uses SHAP for transparency. Results showed that combining modalities improved both classification and survival predictions. The system's modular design and web interface highlight its clinical applicability. Key contributions include a multimodal deep learning pipeline, histopathology preprocessing via clustering, and integrated explainability—offering a robust and interpretable tool for medical diagnostics.

Acknowledgement

This work is supported by the TÜBİTAK 2209A Research Project Support Programme for Undergraduate Students grant 1919B012466872. Thanks to our advisor, Dr. Sema Candemir, for her collaboration.

References

1. Shaheen, M.Y. Applications of Artificial Intelligence (AI) in healthcare: A review. ScienceOpen Preprints (2021). doi:10.14293/S2199-1006.1.SOR-.PPVRY8K.v1
2. Pavli, A., Theodoridou, M., & Maltezou, H.C. Post-COVID syndrome: incidence, clinical spectrum, and challenges for primary health care professionals. Arch. Med. Res. 52(6), 575–581 (2021). doi:10.1016/j.arcmed.2021.09.013
3. van der Schaar, M. et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Mach. Learn. 110, 1–14 (2021). doi:10.1007/s10994-021-05931-6
4. Luo, N., Zhong, X., Su, L., Cheng, Z., Ma, W., & Hao, P. Artificial intelligence-assisted dermatology diagnosis: From unimodal to multimodal. *Comput. Biol. Med.* 165, 107413 (2023). doi:10.1016/j.compbiomed.2023.107413

5. Huang, S.C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M.P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit. Med. 3, 136 (2020). doi:10.1038/s41746-020-00356-1
6. Ertosun, M.G. & Rubin, D.L. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. AMIA Annu. Symp. Proc. (2015), 1899–1908.
7. Jin, L. et al. Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. Neuro-Oncol. 23(1), 44–52 (2021). doi:10.1093/neuonc/noaa100
8. José, L. et al. Artificial Intelligence–Assisted Classification of Gliomas Using Whole Slide Images. Arch. Pathol. Lab. Med. 147(8), 916–924 (2023). doi:10.5858/arpa.2022-0110-OA
9. Kudulaiti, N. et al. A nomogram for individualized prediction of overall survival in patients with newly diagnosed glioblastoma: a real-world retrospective cohort study. BMC Surg. 21, 238 (2021). doi:10.1186/s12893-021-01233-z
10. García-García, S. et al. Current evidence, limitations and future challenges of survival prediction for glioblastoma based on advanced non-invasive methods: a narrative review. Medicina 58(12), 1746 (2022). doi:10.3390/medicina58121746
11. Louis, D.N. et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro-Oncol. 23(8), 1231–1251 (2021). doi:10.1093/neuonc/noab106
12. Bakas, S. et al. Multi-parametric magnetic resonance imaging (mpMRI) scans for de novo glioblastoma (GBM) patients from the University of Pennsylvania Health System (UPENN-GBM) (Version 2) [Data set]. Cancer Imaging Archive (2021). doi:10.7937/TCIA.709X-DN49
13. Scarpace, L. et al. The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM) (Version 5) [Data set]. Cancer Imaging Archive (2016). doi:10.7937/K9/TCIA.2016.RNYFUYE9
14. Pedano, N. et al. The Cancer Genome Atlas Low Grade Glioma Collection (TCGA-LGG) (Version 3) [Data set]. Cancer Imaging Archive (2016). doi:10.7937/K9/TCIA.2016.L4LTD3TK
15. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme Collection (CPTAC-GBM) (Version 16) [Data set]. Cancer Imaging Archive (2018). doi:10.7937/K9/TCIA.2018.3RJE41Q1
16. Wang, C. et al. Multi-parametric MRI (mpMRI) for treatment response assessment of radiation therapy. Med. Phys. 49(4), 2794–2819 (2022). doi:10.1002/mp.15535
17. Bidgood Jr, W.D. et al. Understanding and using DICOM, the data interchange standard for biomedical imaging. J. Am. Med. Inform. Assoc. 4(3), 199–212 (1997). doi:10.1136/jamia.1997.0040199
18. Bhuvaji, S. et al. Brain Tumor Classification (MRI) [Data set]. Kaggle (2020). doi:10.34740/KAGGLE/DSV/1183165
19. Nickparvar, M. Brain Tumor MRI Dataset [Data set]. Kaggle (2021). doi:10.34740/KAGGLE/DSV/2645886
20. Adadi, A. & Berrada, M. Peeking inside the black box: A survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160 (2018). doi:10.1109/ACCESS.2018.2870052

The Impact of Large Language Models on Health and Medical Informatics

Ruken ERDEM DEMİR[1][0000-0002-5966-7454], Mehmet Erşan KALAYCI[1][0000-0002-4797-9020],

Sinem ÖZDEMİR[1][0000-0002-6332-0891] and Kemal TURHAN[2][0000-0001-7871-3025]

[1] Department of Biostatistics and Medical Informatics, Graduate School of Health Sciences, Karadeniz Technical University, 61080 Trabzon, Türkiye.
[2] Department of Biostatistics and Medical Informatics, Graduate School of Health Sciences, Izmir University of Economics, 35330 Izmir, Türkiye.

erdemdemirruken@gmail.com (Corresponding Author)
mehmetersankalayci@ktu.edu.tr
sinemozdemir@ktu.edu.tr
kturhan.tr@gmail.com

Abstract

In recent years, Large Language Models (LLMs) such as GPT, PaLM, and LLaMA have begun to reshape the landscape of health and medical informatics. With their ability to understand and generate human language at an advanced level, these models are being increasingly explored for applications like clinical documentation, decision support systems, and biomedical literature analysis. Their potential to ease clinicians' workload and enhance patient care is significant. However, their adoption also brings important concerns—especially around accuracy, bias, privacy, and explainability. Ethical and legal questions remain, particularly when it comes to aligning these systems with evidence-based medical practices and determining responsibility in clinical use cases. The aim of this study is to comprehensively examine the transformative impact of LLMs on health and medical informatics by reviewing current applications, ethical concerns, limitations, and future prospects. Looking ahead, future research is likely to focus on fine-tuning LLMs for specific medical domains, integrating them with multimodal data sources, and designing tools that can be used directly by patients. While the promise of LLMs is clear, realizing their full potential will require careful validation, close collaboration between disciplines, and adaptive regulatory approaches.

Keywords: Large Language Models, Medical Informatics, Clinical Decision Support, Artificial Intelligence in Healthcare, Ethical AI

1. Introduction

In recent years, large language models (LLMs) like OpenAI's GPT series [1], Google's PaLM [2], and Meta's LLaMA [3] have come to the forefront as game-changing technologies across various fields, including healthcare and medical informatics [4]. Trained on massive text datasets, these models can understand, interpret, and generate human language with impressive fluency and contextual sensitivity [5]. Their ability to tackle complex, language-driven tasks has made them especially appealing in healthcare settings, where unstructured data, linguistic variability, and the need for nuanced language processing are everyday challenges [6]. This paper aims to provide a structured and in-depth review of how LLMs are being utilized in medical informatics, to discuss their strengths and vulnerabilities, and to outline pathways for safe and effective integration into healthcare systems. By doing so, it seeks to inform

researchers, clinicians, and policymakers about the responsible development and deployment of LLM-based technologies in healthcare.

The healthcare sector generates enormous volumes of textual data daily, ranging from electronic health records (EHRs) and radiology reports to scientific articles and clinical guidelines. Making sense of this unstructured information and turning it into actionable insights has long been a complex task, often reliant on time-consuming manual efforts. Large language models (LLMs) offer a promising alternative by automating many of these language-heavy processes. Their use could help ease the burden on clinicians, support better decision-making, and ultimately lead to improved patient care [7]. This paper examines the diverse ways LLMs are shaping medical informatics, highlighting their applications, current limitations, ethical implications, and potential future developments.

2. Applications of LLMs in Medical Informatics

2.1 Clinical Documentation and Electronic Health Records

One of the most promising uses of LLMs in healthcare is the automation of clinical documentation—a task widely recognized as a major contributor to physician burnout [8]. These models can help by transcribing voice dictations, turning free-text notes into structured formats, and even generating documentation templates tailored to specific clinical scenarios. According to recent findings, LLM-powered systems have been shown to cut documentation time by as much as 30%, allowing clinicians to dedicate more attention to direct patient care [7].

Beyond documentation, LLMs also show strong potential in summarizing patient records. This capability helps healthcare providers quickly grasp critical details about a patient's medical history, medications, and test results—particularly useful during consultations or when care is being handed off between teams [8]. In fast-paced environments like emergency or intensive care units, access to concise and accurate summaries can play a crucial role in improving clinical decisions and outcomes.

2.2 Clinical Decision Support

Large language models are increasingly being incorporated into clinical decision support systems (CDSS) to assist with tasks like diagnosis, treatment planning, and risk assessment [9]. By drawing on clinical guidelines, medical literature, and real-world patient data, LLMs can offer evidence-based suggestions that are customized to individual patient profiles [10]. Early pilot studies have demonstrated that these models can list possible diagnoses, explain the reasoning behind their suggestions, and even flag potential drug interactions before they become harmful.

That said, using LLMs in clinical decision-making is not without its challenges. Medical decisions often involve high-stakes consequences, and relying on AI systems whose inner workings are not easily interpretable can pose serious risks. Because of this, most current CDSS applications use LLMs as supportive tools—augmenting the expertise of clinicians rather than replacing it [10].

2.3 Biomedical Literature Mining and Question Answering

With the rapid and continuous expansion of biomedical literature, LLMs have become valuable tools for literature mining and evidence retrieval. Models such as SciBERT and BioGPT are now being used to allow clinicians and researchers to ask questions in natural language and receive clear, contextually appropriate answers in return [11]. This eliminates the need for complex Boolean queries or the time-consuming process of manually reviewing hundreds of abstracts, significantly accelerating the process of synthesizing medical evidence [12].

For instance, a clinician might ask, "What are the latest treatment guidelines for triple-negative breast cancer?" and instantly receive a summary drawn from recent publications, clinical guidelines, and trial results. This ability to distill vast, unstructured scientific literature into actionable insights is essential for staying up to date in fast-changing areas of medicine.

3. Challenges and Limitations

Despite their considerable promise, large language models face several critical challenges when applied in healthcare. One of the foremost concerns is accuracy—specifically, the issue of hallucination, where models produce information that appears credible but is factually incorrect [13]. In a clinical context, such errors can have serious consequences, potentially leading to misdiagnoses, inappropriate treatments, or delays in care [14]. Reducing these risks requires robust validation strategies, including human oversight and systematic post-processing reviews.

Bias and fairness represent another pressing concern. Because LLMs are trained on vast datasets that often mirror existing societal inequities, they can unintentionally perpetuate or even amplify disparities in healthcare delivery [15]. For example, underrepresentation of minority groups in training data can result in less reliable predictions or recommendations for those populations. Tackling these issues calls for more than just technical fixes—it requires thoughtful dataset design and active involvement from diverse stakeholders throughout the development process.

Data privacy and security are also major limitations, especially given the strict regulatory landscape surrounding patient information [16]. Implementing LLMs in clinical settings demands careful handling of data sources, strong de-identification protocols, and secure deployment environments to prevent breaches or unauthorized access.

Finally, the infrastructure demands of these models pose practical barriers. Cutting-edge LLMs require substantial computational resources—including high-end memory, storage, and processing capacity—which may not be feasible for smaller healthcare organizations. This raises valid questions about the cost, scalability, and accessibility of deploying such systems across different clinical environments.

4. Ethical and Regulatory Considerations

The integration of large language models into healthcare workflows brings with it a host of complex ethical challenges. Chief among these is the issue of explainability: for clinicians to make informed decisions, they must be able to understand the reasoning behind AI-generated recommendations. Yet, LLMs often function as "black boxes," offering limited insight into how outputs are derived [17]. To address this, researchers are investigating explainable AI

methods—such as attention maps and counterfactual explanations—to increase transparency without compromising performance.

Another major concern involves the alignment of LLM outputs with evidence-based medicine (EBM). Although these models can generate summaries or even suggest treatments, the provenance of their content is not always clear or verifiable. This becomes particularly problematic when their outputs are interpreted as clinical guidance. Maintaining consistency with established clinical standards is critical, but most current LLMs lack built-in mechanisms to validate their responses against formal medical guidelines [18,19].

To bridge this gap, emerging approaches like retrieval-augmented generation (RAG) are being developed. These frameworks combine LLMs with trusted external databases—such as PubMed or UpToDate—to anchor responses in verified sources. Similarly, citation-aware LLMs and prompting strategies like DocPrompting aim to make the origins of generated content more traceable, allowing users to verify the source and credibility of specific claims. Despite these advances, there is still no universally accepted method for treating LLMs as reliable certifiers of factual accuracy in clinical settings [18,19].

Regulatory oversight, while progressing, remains incomplete. Organizations like the U.S. Food and Drug Administration (FDA) have begun to issue guidance on AI and machine learning in medical software, but the pace of LLM innovation frequently outstrips the development of formal regulatory structures [20]. Moving forward, it will be essential to establish clear standards for evaluation, documentation, and post-deployment monitoring to safeguard patient safety and build institutional trust.

Lastly, accountability remains a critical and unresolved issue. When LLM-generated outputs influence clinical decisions, it is still unclear who bears responsibility in the event of an error—the clinician, the healthcare institution, or the technology provider [21]. Until such legal ambiguities are addressed, large-scale deployment of LLMs in healthcare will remain a cautious endeavor.

5. Future Directions

Looking ahead, the future of large language models in medical informatics is likely to focus on domain-specific fine-tuning and multimodal integration. By training LLMs on specialized biomedical datasets and combining them with structured data, imaging, and genomic information, researchers aim to build AI tools that are not only more accurate but also more aligned with clinical language, terminologies, and workflows [22]. These tailored models have the potential to better understand medical abbreviations, interact with knowledge graphs, and support highly specific use cases in healthcare.

Another promising direction is the development of collaborative AI systems that complement, rather than replace, healthcare professionals [23]. Integrating LLMs into electronic health record (EHR) systems, designing interfaces with clinician input, and enabling real-time feedback loops can significantly improve both usability and clinical trust. In such partnerships, LLMs can serve as intelligent assistants—augmenting human expertise while keeping decision-making in professional hands.

Patient-facing applications are also on the rise. AI-driven tools such as chatbots for symptom checking, medication adherence, and health education have the potential to empower

individuals and ease the burden on already strained primary care systems [24]. However, for these technologies to serve diverse populations effectively, issues such as digital literacy, accessibility, and cultural competence must be addressed from the outset to avoid deepening existing health disparities.

In Turkey, significant strides are being made in integrating large language models into healthcare applications. A notable example is the "SağlıkGPT" project, spearheaded by Dr. İsmail İşeri at Ondokuz Mayıs University. This initiative, supported by TÜBİTAK TEYDEB 1505, aims to develop a domestically trained LLM-based patient guidance system tailored for the Turkish healthcare context. The project focuses on analyzing patient symptoms and directing them to appropriate healthcare services, thereby enhancing the efficiency of the healthcare system and providing digital consultancy for international patients within the scope of health tourism. This endeavor exemplifies the potential of LLMs to address specific linguistic and systemic needs in Turkish healthcare, marking a significant step towards localized AI-driven solutions [25-26].

Finally, the road ahead must include ongoing, rigorous evaluation of LLMs in real-world clinical environments. Beyond initial pilot studies, large-scale, multicenter trials will be necessary to assess long-term outcomes, cost-effectiveness, and any unintended consequences. Establishing evidence-based guidelines for responsible deployment will be key to realizing the full potential of LLMs in healthcare.

6. Conclusion

Large language models mark a significant turning point in the evolution of medical informatics, offering new ways to tackle persistent challenges in data interpretation, knowledge access, and clinical decision-making. Their capacity to understand and generate human language at scale introduces remarkable opportunities to enhance healthcare delivery, advance medical research, and deepen patient engagement.

However, these opportunities come with important caveats. Issues surrounding accuracy, fairness, explainability, data privacy, and regulatory oversight remain major obstacles to widespread adoption.

Unlocking the full potential of LLMs in healthcare will depend on close collaboration across disciplines—bringing together technologists, healthcare professionals, ethicists, legal scholars, and policymakers. With a commitment to responsible innovation, rigorous validation, and inclusive design, LLMs can serve not just as technological tools, but as catalysts for building more effective, equitable, and patient-centered health systems.

References

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
2. Elnaggar, A. (2023, July). A journey into the Generative AI and large language models: From NLP to BioInformatics. In International Conference on Grammatical Inference (pp. 7-7). PMLR.
3. Xie, Q., Chen, Q., Chen, A., Peng, C., Hu, Y., Lin, F., ... & Bian, J. (2024). Me-llama: Foundation large language models for medical applications. Research square, rs-3.
4. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2022). Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138.
5. Yang, R., Tan, T. F., Lu, W., Thirunavukarasu, A. J., Ting, D. S. W., & Liu, N. (2023). Large language models in health care: Development, applications, and challenges. Health Care Science, 2(4), 255-263.
6. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
7. Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., ... & Wu, Y. (2022). A large language model for electronic health records. NPJ digital medicine, 5(1), 194.
8. Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J. B., Aali, A., Bluethgen, C., ... & Chaudhari, A. S. (2023). Clinical text summarization: adapting large language models can outperform human experts. Research square, rs-3.
9. Yesankar, P., Puri, C., & Gote, P. M. (2025, March). AI-Powered Clinical Decision Support Systems (CDSS): Challenges, Benefits, Applications, and Future Directions. In 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS) (pp. 1192-1197). IEEE.
10. Karttunen, P. (2023). Large Language Models In Healthcare Decısıon Support. Tampere University.
11. Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. arXiv preprint arXiv:1903.10676.
12. Baksi, K. D. (2021). Recent advances in automated question answering in biomedical domain. arXiv preprint arXiv:2111.05937.
13. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., … & Gabriel, I. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.
14. Kim, Y., Jeong, H., Chen, S., Li, S. S., Lu, M., Alhamoud, K., ... & Breazeal, C. (2025). Medical Hallucinations in Foundation Models and Their Impact on Healthcare. arXiv preprint arXiv:2503.05777.
15. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623.
16. Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507.
17. Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Precise4Q Consortium. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC medical informatics and decision making, 20, 1-9.

18. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.
19. Lehman, E., Jain, S., Pichotta, K., Goldberg, Y., & Wallace, B. C. (2021). Does BERT pretrained on clinical notes reveal sensitive data?. arXiv preprint arXiv:2104.07762.
20. FDA, U. (2021). Artificial intelligence and machine learning (AI/ML) software as a medical device action plan. US Food Drug Administration, Silver Spring, MD, accessed May, 4.
21. Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37-43.
22. Wang, C., Li, M., He, J., Wang, Z., Darzi, E., Chen, Z., ... & Shen, Y. (2024). A survey for large language models in biomedicine. arXiv preprint arXiv:2409.00133.
23. Chen, I. Y., Szolovits, P., & Ghassemi, M. (2019). Can AI help reduce disparities in general medical and mental health care?. AMA journal of ethics, 21(2), 167-179.
24. Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digital health, 5, 2055207619871808.
25. Ondokuz Mayıs Üniversitesi. (2025). SağlıkGPT Projesi, TÜBİTAK TEYDEB 1505 kapsamında desteklenecek. Ondokuz Mayıs Üniversitesi. https://www.omu.edu.tr/tr/icerik/haber/saglikgpt-projesi-tubitak-teydeb-1505-kapsaminda-desteklenecek
26. Demirören Haber Ajansı. (2025). Sağlık sisteminde verimliliği arttıracak yapay zeka tabanlı dil modeli projesi. DHA. https://www.dha.com.tr/saglik-yasam/saglik-sisteminde-verimliligi-arttiracak-yapay-zeka-tabanli-dil-modeli-projesi-2584685

The Role of Radiomics in Dentistry in the Age of Digital Transformation and Artificial Intelligence

[1]Ali RASAT, [2]Selmi TUNÇ

[1]Specialist Dentist, DMD

Akdeniz University Faculty of Dentistry, Departmant of Oral and Maxillofacial Radiology, Antalya, Turkey, 07058. alicanras07@hotmail.com.

[2]Associate Professor, PhD, DMD

Akdeniz University Faculty of Dentistry, Departmant of Oral and Maxillofacial Radiology, Antalya, Turkey, 07058. selmiyard@gmail.com.

ABSTRACT

Recently, the beneficial impacts of digitalization in healthcare have facilitated the advancement of innovative methodologies in diagnosis and treatment planning through the integration of medical data with artificial intelligence algorithms. Radiomics is a pioneering discipline that underpins image-based decision support systems, which seek to extract quantitative data from medical images and analyze this data using statistical and artificial intelligence methodologies. The radiomics workflow, encompassing image acquisition, model building, and clinical application, necessitates certain quality and accuracy standards at every stage. In dentistry, the applications of radiomics and artificial intelligence utilizing cross-sectional three-dimensional imaging techniques, such as computed tomography (CT) and cone beam computed tomography (CBCT), have garnered heightened interest in recent years, particularly owing to the intricate anatomy of the maxillofacial region. For clinical use, models must adhere to stringent standards of accuracy, repeatability, and generalizability.

Future research anticipates that radiomics will significantly enhance dentistry in diagnosis, treatment planning, and disease prognosis prediction.

Keywords: Artificial intelligence, dentistry, maxillofacial, dentistry.

1. Introduction

Over the past two decades, digitalization has permeated several aspects of life, particularly evident in the health sector through the digitization of medical data acquired during clinical practice. The capacity to transform medical data into digital format has resulted in the creation of advanced tools for data analysis. Originally conceived in engineering, these programs have subsequently been incorporated into clinical practice by physicians. These clinical software programs not only facilitate the production of solutions in a more efficient and expedited manner but also substantially enhance the development of the 'big data' idea by augmenting the continuously growing data repository. In 2012, the Federal Big Data Commission of the TechAmerica Foundation characterized big data as "extensive volumes of rapid, intricate, and variable data necessitating sophisticated techniques and technologies for the capture, storage,

distribution, management, and analysis of information" [1]. Medical imaging is a fundamental diagnostic procedure that facilitates the visualization of tissues through various kinds of energy utilizing non-invasive techniques in the diagnosis and treatment of diseases. Currently, medical imaging facilitates comprehensive assessment of the morphological, functional, and metabolic characteristics of tissues. The advancement of sophisticated imaging techniques, including computed tomography (CT), cone beam computed tomography (CBCT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasonography (USG), has transformed diagnosis and treatment planning. Nonetheless, the assessment of anatomical images relying exclusively on physician observation may be influenced by individual experience, expertise, and personal characteristics (e.g., weariness, psychological state), hence amplifying the demand for more objective and quantitative analytical tools.

Radiomics is an innovative methodology designed to obtain deeper insights into biological processes by the extraction of extensive quantitative data from medical imaging. This approach involves extracting numerous parameters, including tissue structure, shape, intensity, and signal variation, from medical images and analyzing these data using statistical methods and artificial intelligence/machine learning algorithms. Consequently, clinical deductions regarding diagnosis, prognosis, and treatment response can be established [2]. In 2017, a research team lead by Grossmann announced the application of an image analysis technique known as 'radiomics' to enhance customized medicine in the realm of diagnostic imaging. The word, while not precisely defined, typically denotes that digital medical images transcend mere 'pictures' and encompass intricate data that can be examined by mathematical methodologies or quantitative perspectives [3]. Radiomics is a nascent translational domain in quantitative imaging associated with machine learning. It can quantify the textural information of designated regions and volumes of interest (ROI: region of interest, VOI: volume of interest) in digital diagnostic images by mathematically inferring signal intensity distribution and pixel/voxel relationships that are undetectable to the human eye [4].

Radiomics employs comprehensive analysis of imaging characteristics to enhance tissue comprehension, aiding in the prediction of disease progression, treatment response, and recurrence likelihood. The radiomics workflow comprises a series of sequential procedures, commencing with the gathering of digital radiography data. The sources of radiomics data are varied; many imaging devices with distinct imaging parameters can generate quantitative values [6, 7]. The objective of radiomics is to transform pictures into quantifiable data with high precision and efficiency, comprising various stages with distinct inputs and outputs, each presenting unique problems; discrepancies in outcomes may arise from inaccuracies in any of these processes (Figure 1) [8].

1.1 Radiomix Procedure

1.1.1 Image Acquisition

The initiation of radiologic analysis involves the selection of a high-quality and suitable imaging protocol. Imaging modalities including CBCT, CT, MRI, nuclear medicine, PET, and USG [9]. The quality of the image is crucial, as substandard or artifact-laden photographs might negatively impact the entire process.

1.1.2 Segmentation

Following the gathering of picture data, segmentation techniques may be employed manually, semi-automatically, or entirely automatically. Regions of Interest (ROI) and Value of Interest (VOI) are extracted from the image and segmented. At this juncture, the delineations of the target lesion, tumor, tissue, or anatomical structure are precisely ascertained [10]. Every segmentation method possesses distinct advantages and downsides. Manual and semi-automatic picture segmentation, typically involving manual correction, are the predominant techniques; nonetheless, they possess numerous drawbacks. Initially, manual segmentation is labor-intensive. The duration needed for manual segmentation may fluctuate depending on the dimensions of the target area and the volume of data. Secondly, manual and semi-automated segmentation may be subject to individual errors based on the user [11]. Automatic segmentation is the optimal choice to reduce human errors. Nonetheless, the efficacy of machine learning-based segmentation algorithms is contingent upon the quality of the training data; insufficient data diversity may render the model ineffective. Additionally, clinicians may experience skepticism towards certain automatic segmentation algorithms due to a lack of comprehension regarding their operational mechanisms (the “black box” issue), necessitating a final verification by the user post-procedure. Nonetheless, additional research is required to create strong and generalizable algorithms for automatic segmentation techniques [12].

1.1.3 Feature Extraction

The quantitative analysis of segmented data computes radiomix characteristics. Shape-based characteristics of first, second, or higher orders can theoretically be computed with infinite probability [13]. Shape-based evaluation encompasses characteristics of the ROI, including morphology, dimensions, positioning, vascularization, and spiculation. First-order statistics analyze the intensity values of the pixels or voxels directly. Second-order features examine the spatial relationships among pixels or voxels, focusing on the interactions between neighboring pixels rather than their intensity values. The aim is to comprehend texture patterns and textural features [15]. Higher-order features are extracted from the image through the application of several filters and transformations, including Wavelet, Laplacian of Gaussian (LoG), and local binary patterns (LBP), followed by image preprocessing.

Subsequently, first- or second-order features are reextracted to unveil concealed patterns and intricate linkages [8].

1.1.4 Data Preprocessing

At this juncture, normalization and standardization procedures are employed to render the obtained pictures and extracted characteristics appropriate for analysis, which is essential for the accuracy and reproducibility of the characterisation [2]. In high-dimensional data sets, enhancing the accuracy, consistency, and generalizability of the model is essential.

Consequently, issues such as noise and artifacts in the photos are identified. The creation of radiological feature values at various scales substantially influences the parameter stability of artificial intelligence algorithms. Consequently, by synchronizing the images acquired by various devices or protocols, the comparability of the results is guaranteed [16].

1.1.5 Features Selection

This extensive dataset should ideally be supplemented by a model incorporating genomic

profiles, histological characteristics, serum biomarkers, patient history, and diagnostic tests, hence ensuring the reproducibility and reliability of the data quality and the accuracy of the outcome. Among the myriad features extracted, only those that are significant are retained, while redundant, repetitive, or noise-inducing features are discarded. This level employs statistical approaches and machine learning algorithms [17, 18].

1.1.6 Development of the Model

A multitude of algorithms exists for model development. It is essential to choose a model that is suitable for the categorization procedure. Techniques like as logistic regression, support vector machines (SVM), random forests, and gradient boosting machines (XGBoost) are employed to differentiate between benign and malignant tumors. Linear regression and survival trees are employed for the prediction of survival time. Furthermore, convolutional neural networks (CNN) can be employed with extensive datasets. Despite the seemingly haphazard selection of algorithms in the literature, optimal practice entails choosing the algorithm through extensive experimentation [16, 18-20].

1.1.7. Training and Validation

Prior to advancing to the validation phase of the data, it is bifurcated into a training set and a test set. The training set comprises the majority group (70-80%), while the remaining portion constitutes the test set (20-30%). A validation test is conducted on the data segmented into clusters at a specified rate. For clinical acceptance, the results must be confirmed with independent datasets, ideally sourced from a different institution [10, 21]. Consequently, independent external validation is regarded as the most effective approach for model validation. In small-scale pilot or early research, independent validation data is not always attainable. Internal validation approaches may be employed in these instances. The predominant internal validation strategies identified in the literature are k-fold cross-validation and leave-one-out cross-validation (LOOCV) [20].

1.1.8 Assessment of Performance

The assessment of categorization performance is typically conducted using the area under the curve (AUC) [22]. Note that in the presence of class imbalance within the dataset, AUC may not serve as a reliable performance metric. It is essential to consider additional performance metrics like as accuracy, sensitivity, specificity, ROC curve, and Matthews correlation coefficient for a comprehensive evaluation [16]. Following the assessment of classifications, regression and clinical validation analyses are performed and tested on real patient data for inclusion in the clinical decision support system [14].

1.1.9 Readying the Model for Clinical Utilization

Regardless of the thoroughness of prior steps, the models must be evaluated using independent datasets from several centers or devices. The developed model must demonstrate generalizability not just on the training dataset but also in clinical applications [23].
Standardize the imaging parameters, segmentation methodologies, and feature extraction criteria [24]. The model's diagnostic precision, therapeutic efficacy, and prognosis prediction

capacity are assessed prior to its application in clinical practice. Ultimately, we must evaluate the model prospectively in novel patient cohorts and within real-time clinical settings [25].

1.2. Radiomics and Artificial Intelligence

Radiomics and artificial intelligence encompass extensive domains characterized by diverse methodological components. This variability results in a lack of consensus on numerous phases, potentially causing a deficiency in standardization and yielding disparate outcomes [20]. Nevertheless, due to the continuous advancements in effective integration, computing capabilities, and the accessibility of extensive data sets, it appears unavoidable that humans and machines will collaborate in forthcoming therapeutic applications.

1.3 Applications in Dentistry

Radiography examination is essential for diagnosing and treating most dento-maxillofacial illnesses, resulting in a substantial collection of digital radiography pictures stored in the archiving systems and databases of numerous hospitals and clinics. In diagnosing diseases in the maxillofacial region, the anatomical complexity and proximity to vital vascular and neural structures necessitate the use of 3D imaging (such as CT/CT/CBCT) for effective diagnostic and treatment planning, making it a crucial element of digital workflows in patient management. A diverse array of deep learning and radiomics applications utilizing CT scans has been established for the diagnosis, treatment planning, and prognostication of numerous maxillofacial disorders [26]. Currently, the findings of artificial intelligence research are highly promising; nonetheless, the field remains in its nascent stages. For routine application in clinical environments, it must be clinically realistic and yield precise findings for the clinician. Up until 2019, dentistry was just beginning to recognize the concept of radiomics.
After this period, there has been an increase in studies on this field. The introduction of three-dimensional imaging technologies such as CBCT into clinical use and the widespread use of artificial intelligence/machine learning algorithms in healthcare have also supported this increase (Figure 2). Collaboration between radiologists and artificial intelligence specialists is advocated to create radiomics models, leading to a strong and therapeutically relevant use of radiomics and artificial intelligence in dentistry.

1.3.1 Assessment of Cysts and Neoplasms in the Mandibular Bones

Typically, routine radiographic assessment using a panoramic radiograph aligns with the identification of cysts and malignancies in the jaws. While CBCT is the predominant imaging modality in dentistry for comprehensive evaluation, it is insufficient for establishing a definitive diagnosis only from the radiographic image, despite the existence of distinct clinical, radiographic, and histopathologic characteristics of various illnesses. Indeed, distinguishing between benign and malignant tumors on CBCT is quite challenging. In response to the necessity for precise diagnosis of maxillofacial cysts and tumors, numerous investigations have established deep learning models utilizing 2D or 3D radiographic pictures for the automated identification of diverse jaw cysts and cancers [27-30]. Xu et al. [31] developed artificial intelligence models to differentiate between odontogenic cysts and ameloblastomas by utilizing particular characteristics from CBCT images, achieving

remarkably high accuracy rates. Sha et al. [32] created a CBCT-based radiomix model to

distinguish between odontogenic cysts, odontogenic keratocysts, and ameloblastomas, with great diagnostic precision. Additional imaging in this domain may facilitate the identification of various jaw cysts and cancers, hence tremendously aiding maxillofacial radiologists and significantly enhancing early diagnosis and prognosis [33].

1.3.2 Assessment of Salivary Gland Pathologies

Salivary gland illnesses encompass inflammatory, infectious, or neoplastic conditions primarily affecting the parotid glands, subsequently involving the submandibular-sublingual and minor salivary glands. The diagnosis of salivary gland illnesses presents a significant problem in dentistry and otolaryngology, as it is predominantly dependent on clinical expertise and diagnostic imaging techniques (MRI, CT, USG) [33]. Radiomics analysis enables the automatic and objective differentiation of benign and malignant tumors by assessing lesion features, including tissue heterogeneity, form, and margins [34]. Deep learning models for diagnosing salivary gland illnesses have predominantly been built using MRI scans, owing to their superior capability in illustrating soft tissues compared to other imaging modalities [35]. Nonetheless, given that MRI is hardly utilized in dentistry and that radiomics research pertaining to ultrasound is limited, the models created were predominantly founded on CT images [33]. Yuan et al. [36] created a CNN model utilizing CT images to differentiate between pleomorphic adenoma and malignant parotid gland tumors; Zhang et al. [37] established a CNN model on CT images to distinguish between benign and malignant parotid gland tumors; Kise et al. [38] developed a CNN model on CT images to autonomously identify fatty degeneration of the salivary gland, a significant diagnostic indicator for Sjögren's syndrome, achieving high accuracy rates.

1.3.3 Assessment of Pulpal and Periapical Pathologies

Pulpal tissue is typically viable but may sustain damage due to inflammation or trauma. This injury may occur inside the pulpal tissue and spread to the periapical region, resulting in indications of inflammation in the adjacent bone. Alterations in the internal imaging characteristics of alveolar bone linked to inflammatory periapical lesions indicate either a reduction or augmentation in bone structure, or a mix thereof. In endodontics, CBCT is frequently employed to examine alterations in the root canal, including obstructions, dilations, contractions, perforations, as well as the morphology of the canal and modifications in the adjacent alveolar bone. Rosa et al. [39] distinguished between periapical granulomas and periapical cysts by utilizing textural radiomic characteristics derived from CBCT images, achieving superior accuracy relative to traditional radiologic assessment. Gonzalez et al. [40] investigated the variations in specific radiomic characteristics for identifying periapical lesions according to volumetric size, cortical expansion, resorption, and form in CBCT images, revealing distinctions in textural and first-order radiomic features.

1.3.4 Assessment of Maxillary Sinus Pathologies

The maxillary sinus is the largest facial sinus and frequently participates in dental surgeries, such as treatments involving the posterior teeth and surgeries for dental implant preparation. Precise diagnosis and categorization of maxillary sinus diseases before surgical interventions

in the sinus area are essential for ensuring a reliable treatment outcome. CT, CBCT, or MRI is frequently employed to initially detect conditions such as mucoceles, sinusitis, mucus retention cysts, or cysts and tumors in the maxillary sinus, and to assess their relationship with adjacent anatomical structures. Precise identification of maxillary sinus diseases on conventional imaging modalities can occasionally be challenging. Ito et al. [41] investigated the tissue properties of mucosal thickening in the maxillary sinus mucosa via CT to distinguish odontogenic from non-odontogenic maxillary sinusitis, revealing substantial differences between the two groups. Shaujan et al. [42] employed CT scans to differentiate inverted papilloma from chronic rhinosinusitis polyps, and their radiomic models demonstrated high accuracy in this classification. Moreover, literature also presents CNN-based sinus segmentation [43, 44] and morphological alterations in sinus mucosa [45, 46].

1.3.5 Assessment of Temporomandibular Joint (TMJ) Disorders

TMJ disorders encompass concerns with the joint itself (including disc displacement, partial dislocation, and stiffness), conditions affecting the joint (such as arthritis and certain cancers), fractures, developmental abnormalities, and complications with the masticatory muscles (such as discomfort and tension). Typically, inflammatory alterations in the synovial membrane of the joints or masticatory muscles result in modifications to the joint complex generated by inflammatory mediators. This results in degenerative and deformative alterations in the cartilage matrix, subchondral bone, and associated joint tissues. The diagnosis and treatment of TMJ disorders rely heavily on extensive clinical knowledge and experience, leading to varying perspectives among clinicians regarding both diagnostic and treatment strategies.
Recent studies demonstrate that radiomics algorithms, particularly those based on MRI and CBCT data, yield significant insights in the assessment of TMJ disorders. Orhan et al. [47] categorized alterations in the condyle and disc dislocations using radiomixing with MRI data and saw a significant concordance in their results. Ogawa et al. [48] observed substantial alterations in tissue characteristics associated with condylar flatness and erosion in CT and MRI images. The findings indicate that tissue analysis could be beneficial for the early identification of degenerative alterations in the TMJ.

1.3.6 Differentiation Between Benign and Malignant

Specific clinical and radiographic indicators are commonly employed to distinguish between benign and malignant disorders in the craniofacial area. Nonetheless, certain benign formations may deceive the clinician during the first diagnosis by presenting pictures that are consistent with malignancy on radiographic assessment. Indeed, early-stage cancers may not exhibit any discernible abnormalities on radiographic assessment. Consequently, radiomics can discern the unique characteristics distinguishing benign from malignant lesions by examining microscopic textural variations (heterogeneity, edge patterns, density distributions) that are imperceptible to the naked eye, converting them into mathematical data for analysis using artificial intelligence algorithms. Iqbal et al. [49] shown that the amalgamation of clinical examinations, imaging modalities, and histopathologic assessments enhances the precision of oral lesion diagnoses, whereas integrated methodologies facilitate earlier malignancy detection and diminish misdiagnosis rates. Yuan et al. [50] created an MRI-based

radiomics algorithm and nomogram to forecast preoperative outcomes in patients with squamous cell carcinoma (SCC) in the head and neck area, demonstrating significant predictive capability. Advancements in research and technology suggest that radiomixin will diminish the necessity for invasive treatments in distinguishing between benign and malignant conditions, enhance prognosis, and offer significant diagnostic ease for physicians.

1.3.7 Fractures

The mandible is the sole movable bone of the face and the most commonly fractured bone in maxillofacial trauma [33]. Despite the significant prevalence of fracture cases observed nowadays, there is an absence of radiomics-based research in this domain. Literature includes studies on the automatic detection of fracture lines utilizing artificial intelligence algorithms [51-53], although there is an absence of radiomix-based analysis. This is attributed to the user's ability to easily and reliably diagnose fracture lines visually on radiographic images, along with the superior efficacy of radiomics in identifying early or microscopic alterations that are often imperceptible to the naked eye.

2. Conclusion

Radiomix is a domain that requires further development and investigation, as its application in dentistry and related research remains nascent. Nevertheless, it is evident that highly promising results have been achieved in this domain. Future research and the incorporation of artificial intelligence models with radiomics data are expected to yield substantial advancements in disease diagnosis, personalized treatment planning, and outcome prediction.

3. References

[1] TechAnerica Foundation. *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*. Washington, DC: TechAnerica Foundation's Federal Big Data Connission (2012). https://www.gove:ec.con/pdfs/072512bb1.pdf.

[2] P. Lanbin *et al.*, "Radionics: E:tracting nore infornation fron nedical inages using advanced feature analysis," *Eur J Cancer*, vol. 48, no. 4, pp. 441–446, (2012). doi: 10.1016/j.ejca.2011.11.036.

[3] P. Grossnann *et al.*, "Defining the biological basis of radionic phenotypes in lung cancer," *Elife*, vol. 6, (2017). doi: 10.7554/ELIFE.23421.

[4] J. E. van Tinneren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, "Radionics in nedical inaging—'how-to' guide and critical reflection," *Insights Imaging*, vol. 11, no. 1, pp. 1–16, (2020). doi: 10.1186/S13244-020-00887-2/TABLES/3.

[5] A. F. Leite, K. de F. Vasconcelos, H. Willens, and R. Jacobs, "Radionics and Machine Learning in Oral Healthcare," *Proteomics Clin Appl*, vol. 14, no. 3, p. 1900040, (2020). doi: 10.1002/PRCA.201900040;WGROUP:STRING:PUBLICATION.

[6] W. Rogers *et al.*, "Radionics: Fron qualitative to quantitative inaging," *British Journal of Radiology*, vol. 93, no. 1108, (2020). doi: 10.1259/BJR.20190948/7451999.

[7] C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, "A deep look into radionics," *Radiologia Medica*, vol. 126, no. 10, pp. 1296–1311, (2021). doi: 10.1007/S11547-021-01389-X/FIGURES/2.

[8] V. Kunar *et al.*, “Radionics: the process and the challenges,” *Magn Reson Imaging*, vol. 30, no. 9, pp. 1234–1248, (2012). doi: 10.1016/J.MRI.2012.06.010.

[9] V. Parekh and M. A. Jacobs, “Radionics: a new application fron established techniques,” *Expert Rev Precis Med Drug Dev*, vol. 1, no. 2, pp. 207–226, (2016). doi: 10.1080/23808993.2016.1164013,.

[10] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radionics: Inages are nore than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, (2016). doi: 10.1148/RADIOL.2015151169,.

[11] B. Baeßler, K. Weiss, and D. P. Dos Santos, “Robustness and Reproducibility of Radionics in Magnetic Resonance Inaging: A Phanton Study,” *Invest Radiol*, vol. 54, no. 4, pp. 221–228, (2019). doi: 10.1097/RLI.0000000000000530.

[12] J. E. van Tinneren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, “Radionics in nedical inaging—‘how-to’ guide and critical reflection,” *Insights Imaging*, vol. 11, no. 1, pp. 1–16, (2020). doi: 10.1186/S13244-020-00887-2/TABLES/3.

[13] C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, “A deep look into radionics,” *Radiologia Medica*, vol. 126, no. 10, pp. 1296–1311, (2021). doi: 10.1007/S11547-021-01389-X/FIGURES/2.

[14] H. J. W. L. Aerts *et al.*, “Decoding tunour phenotype by noninvasive inaging using a quantitative radionics approach,” *Nat Commun*, vol. 5, no. 1, pp. 1–9, (2014). doi: 10.1038/NCOMMS5006;TECHMETA=59;SUBJMETA=2321,308,631,67,692;KWRD=CANCER+IMAGING,MEDICAL+RESEARCH.

[15] Ö. Dönnez Tarakc¾*et al.*, “Radionics-Based Diagnosis in Dentona:illofacial Radiology: A Systenatic Review,” *Journal of Imaging Informatics in Medicine 2024*, pp. 1–34, (2024). doi: 10.1007/S10278-024-01307-3.

[16] B. Kocak, E. Ş. Durnaz, E. Ateş, and Ö. K¾l¾ckesnez, “Radionics with artificial intelligence: a practical guide for beginners,” *Diagnostic and Interventional Radiology*, vol. 25, no. 6, p. 485, (2019). doi: 10.5152/DIR.2019.19321.

[17] C. McCague *et al.*, “Introduction to radionics for a clinical audience,” *Clin Radiol*, vol. 78, no. 2, pp. 83–98, (2023). doi: 10.1016/j.crad.2022.08.149.

[18] C. Parnar *et al.*, “Robust radionics feature quantification using seniautonatic volunetric segnentation,” *PLoS One*, vol. 9, no. 7, (2014). doi: 10.1371/JOURNAL.PONE.0102107,.

[19] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine Learning for Medical Inaging,” *https://doi.org/10.1148/rg.20171c0130*, vol. 37, no. 2, pp. 505–515, (2017). doi: 10.1148/RG.2017160130.

[20] B. Kocak *et al.*, “Te:tural differences between renal cell carcinona subtypes: Machine learning-based quantitative conputed tonography te:ture analysis with independent e:ternal validation,” *Eur J Radiol*, vol. 107, pp. 149–157, (2018). doi: 10.1016/J.EJRAD.2018.08.014.

[21] D. F. Hayes, “Bionarker validation and testing,” *Mol Oncol*, vol. 9, no. 5, pp. 960–966, (2015). doi: 10.1016/J.MOLONC.2014.10.004.

[22] B. A. Varghese, S. Y. Cen, D. H. Hwang, and V. A. Duddalwar, “Te:ture analysis of inaging: What radiologists need to know,” *American Journal of Roentgenology*, vol. 212, no. 3, pp. 520–528, (2019). doi: 10.2214/AJR.18.20624,.

[23] J. E. Park *et al.*, “Quality of science and reporting of radionics in oncologic studies: roon for inprovenent according to radionics quality score and TRIPOD statenent,” *Eur Radiol*, vol. 30, no. 1, pp. 523–536, (2020). doi: 10.1007/S00330-019-06360-Z,.

[24] A. Zwanenburg *et al.*, “The inage bionarker standardization initiative: Standardized quantitative radionics for high-throughput inage-based phenotyping,” *Radiology*, vol. 295, no. 2, pp. 328–338, (2020). doi:

10.1148/RADIOL.2020191145,.

[25] A. J. Vickers and E. B. Elkin, "Decision curve analysis: A novel nethod for evaluating prediction nodels," *Medical Decision Making*, vol. 26, no. 6, pp. 565–574, (2006). doi: 10.1177/0272989X06295361,.

[26] K. F. Hung, Q. Y. H. Ai, L. M. Wong, A. W. K. Yeung, D. T. S. Li, and Y. Y. Leung, "Current Applications of Deep Learning and Radionics on CT and CBCT for Ma:illofacial Diseases," *Diagnostics 2023, Vol. 13, Page 110*, vol. 13, no. 1, p. 110, (2022). doi: 10.3390/DIAGNOSTICS13010110.

[27] W. Poedjiastoeti and S. Suebnukarn, "Application of Convolutional Neural Network in the Diagnosis of Jaw Tunors," *Healthc Inform Res*, vol. 24, no. 3, pp. 236–241, (2018). doi: 10.4258/HIR.2018.24.3.236.

[28] J. H. Lee, D. H. Kin, and S. N. Jeong, "Diagnosis of cystic lesions using panoranic and cone bean conputed tonographic inages based on deep learning neural network," *Oral Dis*, vol. 26, no. 1, pp. 152–158, (2020). doi: 10.1111/ODI.13223.

[29] M. S. Bispo *et al.*, "Conputer tonographic differential diagnosis of aneloblastona and odontogenic keratocyst: Classification using a convolutional neural network,"*Dentomaxillofacial Radiology*, vol. 50, no. 7, p. 20210002, (2021). doi: 10.1259/DMFR.20210002/7261195.

[30] Z. K. Chai *et al.*, "Inproved Diagnostic Accuracy of Aneloblastona and Odontogenic Keratocyst on Cone-Bean CT by Artificial Intelligence," *Front Oncol*, vol. 11, p. 793417, (2022). doi: 10.3389/FONC.2021.793417/BIBTEX.

[31] Y. Song *et al.*, "Application of nachine learning in the preoperative radionic diagnosis of aneloblastona and odontogenic keratocyst based on cone-bean CT," *Dentomaxillofac Radiol*, vol. 53, no. 5, pp. 316–324, (2024). doi: 10.1093/DMFR/TWAE016,.

[32] X. Sha *et al.*, "CBCT radionics features conbine nachine learning to diagnose cystic lesions in the jaw," *Dentomaxillofacial Radiology*, (2025). doi: 10.1093/DMFR/TWAF024.

[33] K. F. Hung, Q. Y. H. Ai, L. M. Wong, A. W. K. Yeung, D. T. S. Li, and Y. Y. Leung, "Current Applications of Deep Learning and Radionics on CT and CBCT for Ma:illofacial Diseases," *Diagnostics 2023, Vol. 13, Page 110*, vol. 13, no. 1, p. 110, (2022). doi: 10.3390/DIAGNOSTICS13010110.

[34] Y. Zhao *et al.*, "Radionics Analysis Based on Multiparanetric MRI for Predicting Early Recurrence in Hepatocellular Carcinona After Partial Hepatectony," *Journal of Magnetic Resonance Imaging*, vol. 53, no. 4, pp. 1066–1079, (2021). doi: 10.1002/JMRI.27424,.

[35] E. Gündüz, Ö. F. Alcin, A. Kizilay, and C. Piazza, "Radionics and deep learning approach to the differential diagnosis of parotid gland tunors," *Curr Opin Otolaryngol Head Neck Surg*, vol. 30, no. 2, pp. 107–113, (2022). doi: 10.1097/MOO.0000000000000782.

[36] J. Jiang *et al.*, "Research on the Practical Classification and Privacy Protection of CT Inages of Parotid Tunors based on ResNet50 Model," *J Phys Conf Ser*, vol. 1576, no. 1, p. 012040, (2020). doi: 10.1088/1742-6596/1576/1/012040.

[37] M. H. Zhang, A. Hasse, T. Carroll, A. T. Pearson, N. A. Cipriani, and D. T. Ginat, "Differentiating low and high grade nucoepidernoid carcinona of the salivary glands using CT radionics," *Gland Surg*, vol. 10, no. 5, p. 1646, (2021). doi: 10.21037/GS-20- 830.

[38] Y. Kise *et al.*, "Prelininary study on the application of deep learning systen to diagnosis of Sjögren's syndrone on CT inages," *Dentomaxillofacial Radiology*, vol. 48, no. 6, (2019). doi: 10.1259/DMFR.20190019/7263064.

[39] C. S. De Rosa *et al.*, "Differentiation of periapical granulona fron radicular cyst using cone bean conputed tonography inages te:ture analysis," *Heliyon*, vol. 6, no. 10, p. e05194, (2020). doi:

10.1016/J.HELIYON.2020.E05194.

[40] L. O. González *et al.*, "Radionic Paraneters in Periapical Lesions: A CBCT Analysis Evaluating Volunetric Size, Cortical E:pansion, Erosion, and Shape," *Eur Endod J*, vol. 9, pp. 394–404, (2024). doi: 10.14744/eej.2024.45220.

[41] K. Ito *et al.*, "Quantitative assessnent of the na:illary sinusitis using conputed tonography te:ture analysis: odontogenic vs non-odontogenic etiology," *Oral Radiol*, vol. 38, no. 3, pp. 315–324, (2022). doi: 10.1007/S11282-021-00558-Y/TABLES/5.

[42] S. He *et al.*, "Utilizing radionics for differential diagnosis of inverted papillona and chronic rhinosinusitis with polyps based on unenhanced CT scans," *Sci Rep*, vol. 14, no. 1, pp. 1–12, (2024). doi: 10.1038/S41598-024-70134-X;SUBJMETA=114,4028,631,692;KWRD=COMPUTATIONAL+BIOLOGY+AND+BIOINFORM ATICS,ONCOLOGY.

[43] Z. Deng, B. Wang, and Z. Zhu, "BE-FNet: 3D Bounding Bo: Estination Feature Pyranid Network for Accurate and Efficient Ma:illary Sinus Segnentation," *Math Probl Eng*, vol. 2020, no. 1, p. 5689301, (2020). doi: 10.1155/2020/5689301.

[44] J. Xu, S. Wang, Z. Zhou, J. Liu, X. Jiang, and X. Chen, "Autonatic CT inage segnentation of na:illary sinus based on VGG network and inproved V-Net," *Int J Comput Assist Radiol Surg*, vol. 15, no. 9, pp. 1457–1465, (2020). doi: 10.1007/S11548-020-02228- 6/FIGURES/6.

[45] K. F. Hung, Q. Y. H. Ai, A. D. King, M. M. Bornstein, L. M. Wong, and Y. Y. Leung, "Autonatic detection and segnentation of norphological changes of the na:illary sinus nucosa on cone-bean conputed tonography inages using a three-dinensional convolutional neural network," *Clin Oral Investig*, vol. 26, no. 5, pp. 3987–3998, (2022). doi: 10.1007/S00784-021-04365-X/TABLES/3.

[46] S. K. Jung, H. K. Lin, S. Lee, Y. Cho, and I. S. Song, "Deep Active Learning for Autonatic Segnentation of Ma:illary Sinus Lesions Using a Convolutional Neural Network," *Diagnostics 2021, Vol. 11, Page c88*, vol. 11, no. 4, p. 688, (2021). doi: 10.3390/DIAGNOSTICS11040688.

[47] K. Orhan, L. Driesen, S. Shujaat, R. Jacobs, and X. Chai, "Developnent and Validation of a Magnetic Resonance Inaging-Based Machine Learning Model for TMJ Pathologies," *Biomed Res Int*, vol. 2021, no. 1, p. 6656773, (2021). doi: 10.1155/2021/6656773.

[48] C. M. Ogawa *et al.*, "Conparative Evaluation of Tenporonandibular Condylar Changes Using Te:ture Analysis of CT and MRI Inages," *Applied Sciences (Switzerland)*, vol. 14, no. 16, p. 7020, (2024). doi: 10.3390/APP14167020/S1.

[49] K. Iqbal *et al.*, "Diagnostic Modalities in Oral Pathology: Integrating Advance Diagnostic Techniques to Differentiate Malignant and Benign Lesions: Integrating Advanced Diagnostic Techniques to Malignant and Benign Lesions," *Pakistan Journal of Health Sciences*, vol. 5, no. 12, pp. 331–338, (2024). doi: 10.54393/pjhs.v5i12.2535.

[50] Y. Yuan, J. Ren, Y. Shi, and X. Tao, "MRI-based radionic signature as predictive narker for patients with head and neck squanous cell carcinona," *Eur J Radiol*, vol. 117, pp. 193– 198, (2019). doi: 10.1016/J.EJRAD.2019.06.019.

[51] N. van Nistelrooij *et al.*, "Detecting Mandible Fractures in CBCT Scans Using a 3-Stage Neural Network," *J Dent Res*, vol. 103, no. 13, p. 1384, (2024). doi: 10.1177/00220345241256618.

[52] J. Mao *et al.*, "Autonated detection and classification of nandibular fractures on nultislice spiral conputed tonography using nodified convolutional neural networks," *Oral Surg Oral Med Oral Pathol Oral Radiol*, vol. 138, no. 6, pp. 803–812, (2024). doi: 10.1016/J.OOOO.2024.07.010.

[53] A. Yari, P. Fasih, M. H. Hooshiar, A. Goodarzi, and S. F. Fattahi, "Detection and classification of nandibular fractures in panoranic radiography using artificial intelligence," *Dentomaxillofacial Radiology*, vol. 53, no. 6, pp. 363–371, (2024). doi: 10.1093/DMFR/TWAE018.

FIGURES

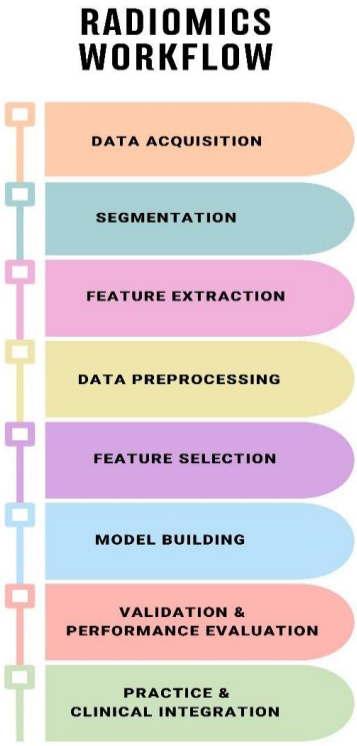


Figure 1. Radionics Workflow

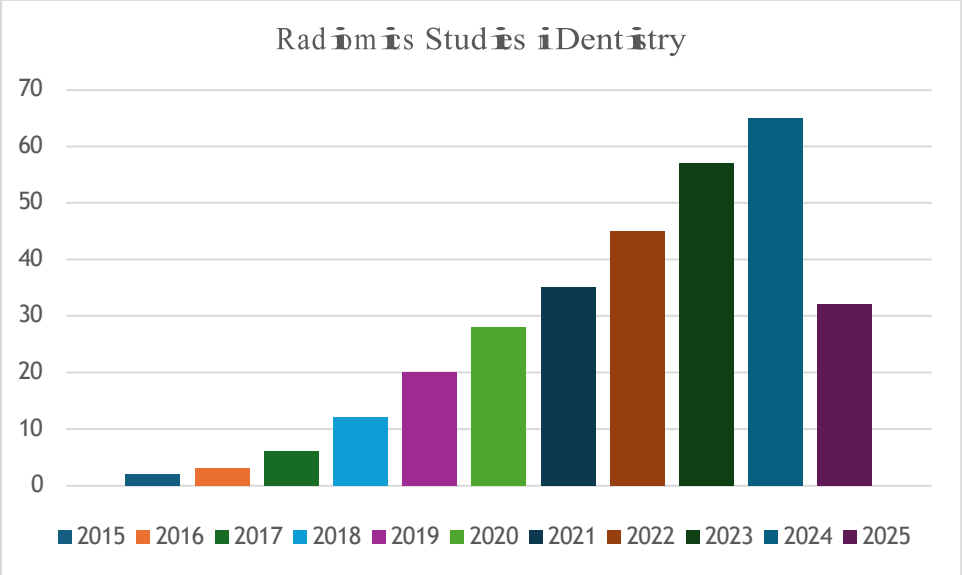


Figure 2. Radionics-based studies in dentistry to date.

Kişiselleştirilmiş Tıp Uygulamaları için Multi-Omik Verilerin Bütünleştirilmesi: NoSQL Yaklaşımı

Muammer ALBAYRAK[1][0000-0002-5946-6310]

[1] Karadeniz Teknik Üniversitesi, Fen Fakültesi, Yazılım Geliştirme Bölümü
61080, Trabzon, TÜRKİYE
m.albayrak@ktu.edu.tr

Özet: Kişiselleştirilmiş tıp, bireylerin genetik, epigenetik, proteomik, metabolomik ve klinik verilerine dayalı olarak sağlık hizmetlerini özelleştirme amacını taşımaktadır. Bu yaklaşım, bireylerin hastalıklara yatkınlıklarını belirlemek, tanı süreçlerini hassaslaştırmak ve tedavi yanıtlarını optimize etmek için çok boyutlu biyolojik verilerin birlikte analiz edilmesini zorunlu kılmaktadır. Söz konusu veriler "multi-omik" olarak tanımlanmakta ve biyolojik sistemlerin farklı katmanlarına ilişkin geniş bir perspektif sunmaktadır. Ancak bu veriler; çeşitlilik, hacim, yapılandırılma düzeyi ve güncellik bakımından oldukça heterojen bir özellik göstermektedir. Bu durum, geleneksel ilişkisel (SQL) veritabanı sistemlerinin ötesinde, daha esnek ve ölçeklenebilir veri yönetim çözümleri gerektirmektedir.

Bu çalışmada, multi-omik verilerin entegrasyonu ve etkin yönetimi için NoSQL (Not Only SQL) tabanlı veri modellerinin potansiyeli araştırılmaktadır. Özellikle belge tabanlı (ör. MongoDB), anahtar-değer tabanlı (ör. Redis) ve grafik tabanlı (ör. Neo4j) NoSQL sistemlerin, yapılandırılmamış veya yarı yapılandırılmış verilerin esnek biçimde saklanması ve analiz edilmesinde sağladığı avantajlar değerlendirilmektedir. Çalışma kapsamında, bu veritabanlarının teorik özelliklerinin yanı sıra çeşitli kişiselleştirilmiş tıp senaryolarında nasıl uygulandığına dair örnek kullanım durumları sunulmaktadır. Hasta bazlı çok katmanlı biyolojik profil oluşturma, gen-protein-hastalık etkileşim ağlarının modellenmesi ve gerçek zamanlı klinik karar destek sistemlerinin veri altyapısı gibi örneklerle bu sistemlerin etkinliği gösterilmektedir.

Elde edilen bulgular, NoSQL tabanlı veri yönetim sistemlerinin kişiselleştirilmiş tıp uygulamalarında yalnızca alternatif bir seçenek değil, aynı zamanda yüksek hacimli ve çeşitli biyolojik verilerle çalışmak için bir gereksinim olduğunu ortaya koymaktadır. Bu bağlamda, çalışmanın çıktıları hem biyomedikal bilişim sistemlerinin tasarımına katkı sağlamakta hem de multi-omik veri bütünleştirme süreçlerine yönelik yeni bir bakış açısı sunmaktadır.

Anahtar Kelimeler: Kişiselleştirilmiş tıp, multi-omik veri, NoSQL, veri bütünleştirme, genomik, proteomik, sağlık bilişimi.

Integration of Multi-Omics Data for Personalized Medicine Applications: A NoSQL Approach

Muammer ALBAYRAK[1][0000-0002-5946-6310]

[1] Karadeniz Technical University, Faculty of Science, Department of Software Development
61080, Trabzon, TÜRKİYE
m.albayrak@ktu.edu.tr

Abstract: Personalized medicine aims to tailor healthcare services based on an individual's genetic, epigenetic, proteomic, metabolomic, and clinical data. This approach necessitates the comprehensive analysis of multi-dimensional biological data to identify disease susceptibilities, improve diagnostic precision, and optimize therapeutic responses. These data, known as multi-omics, offer a broad perspective on various layers of biological systems. However, their diversity, volume, degree of structuring, and temporal variability render them highly heterogeneous. As a result, conventional relational (SQL) database systems are insufficient for handling such complex data, and more flexible and scalable data management solutions are required.

This study explores the potential of NoSQL (Not Only SQL) database models for the integration and efficient management of multi-omics data. In particular, the advantages of document-based (e.g., MongoDB), key-value (e.g., Redis), and graph-based (e.g., Neo4j) NoSQL systems are evaluated in terms of storing and analyzing semi-structured and unstructured biological datasets. In addition to the theoretical features of these databases, the study presents real-world use cases from personalized medicine, including patient-centric multi-layer biological profiling, modeling of gene-protein-disease interaction networks, and data infrastructure for real-time clinical decision support systems.

The findings demonstrate that NoSQL-based data management systems are not merely an alternative, but a necessary solution for effectively processing large-scale, heterogeneous biological datasets in personalized medicine applications. In this context, the study contributes to the design of biomedical informatics systems and provides a novel perspective on multi-omics data integration processes.

Keywords: Personalized medicine, multi-omics data, NoSQL, data integration, genomics, proteomics, health informatics.

1 Giriş

Son yıllarda biyoteknoloji, genetik bilimi ve bilişim teknolojilerindeki hızlı gelişmeler, sağlık alanında birey odaklı yeni yaklaşımların ortaya çıkmasına olanak sağlamıştır. Bu dönüşüm, "kişiselleştirilmiş tıp" ya da "hassas tıp" olarak adlandırılan çağdaş bir paradigmayı beraberinde getirmiştir. Klasik "herkese uyan tek tedavi" yaklaşımının yerini alan bu anlayış, bireylerin genetik yapısını, protein ekspresyon profillerini, metabolik süreçlerini ve çevresel etkileşimlerini dikkate alarak daha etkili, güvenli ve hedefe yönelik teşhis ve tedavi protokolleri geliştirmeyi hedeflemektedir [1]. Kişiselleştirilmiş tıp, özellikle genetik varyasyonlar, gen ekspresyon düzeyleri, epigenetik düzenlemeler, protein yapı ve miktarları ile metabolit profilleri gibi çok boyutlu biyolojik bilgileri bütüncül biçimde değerlendiren sistemlere ihtiyaç duymaktadır. Bu tür çok katmanlı biyolojik veri kümeleri, genel olarak multi-omik veriler olarak tanımlanmakta ve genomik, transkriptomik, epigenomik, proteomik ve metabolomik gibi çeşitli "-omik" katmanları içermektedir. Söz konusu veriler, klinik karar destek sistemlerinin daha isabetli ve bireye özgü çıktılar üretmesini mümkün kılan temel bileşenlerdir [2].

Ancak bu omik veri türlerinin kendine özgü yapısal özellikleri, hacimsel büyüklüğü ve çeşitliliği nedeniyle yönetilmesi ve analiz edilmesi ciddi zorluklar barındırmaktadır. Omik veriler genellikle büyük veri (big data) kapsamında değerlendirilen 4V özelliğini — hacim (volume), hız (velocity), çeşitlilik (variety) ve doğruluk (veracity) — taşımaktadır. Örneğin, bir hastaya ait genom dizisi terabaytlarca veri içerebilirken; bu veri sürekli güncellenmekte, çok sayıda farklı formata sahip olabilmekte ve biyolojik bağlamına göre değişkenlik gösterebilmektedir [3, 4]. Özellikle yapısal olmayan veya yarı yapılandırılmış formattaki veriler, geleneksel ilişkisel veritabanı (Relational Database Management Systems, RDBMS) sistemlerinin sınırlı esnekliği nedeniyle etkin biçimde yönetilememektedir. Bu noktada, veri yönetimi açısından daha esnek ve yüksek performanslı çözümler olarak NoSQL tabanlı sistemler öne çıkmaktadır [5]. NoSQL veritabanları, sabit şemaya bağlı kalmaksızın veri saklayabilen belge, anahtar-değer (key-value), sütun (column-oriented) ve grafik (graph) tabanlı modeller sunmakta ve özellikle heterojen biyolojik verilerin yönetiminde önemli avantajlar sağlamaktadır. Şema esnekliği, yatayda kolay ölçeklenebilirlik, yüksek hızda okuma/yazma işlemleri ve büyük veri kümeleriyle baş edebilme yetenekleri sayesinde NoSQL sistemler, multi-omik verilerin entegrasyonu ve analizi için güçlü bir altyapı oluşturur hale gelmiştir [6].

Bu çalışmanın temel amacı, kişiselleştirilmiş tıp uygulamalarında karşılaşılan multi-omik veri yönetimi ve bütünleştirme sorunlarına NoSQL tabanlı veri sistemlerinin nasıl çözüm sunduğunu incelemek ve bu bağlamda farklı NoSQL mimarilerinin potansiyelini örnek senaryolar üzerinden ortaya koymaktır. Literatürdeki çalışmalar çoğunlukla sadece tek bir omik veri türüne veya belirli bir klinik kullanım örneğine odaklanırken; bu çalışma, çok katmanlı omik verilerin disiplinler arası bakış açısıyla entegre edilmesine yönelik daha geniş kapsamlı bir değerlendirme sunmaktadır. Böylece, çalışmanın hem sağlık bilişimi alanında akademik katkı sağlaması, hem de veri mühendisliği, yapay zeka ve biyoinformatik gibi alanlarda geliştirilecek ileri uygulamalara altyapı oluşturması amaçlanmaktadır.

2 Multi-Omik Veri

Multi-omik veriler, biyolojik sistemlerin farklı seviyelerini temsil eden ve birlikte analiz edildiklerinde canlı organizmaların moleküler işleyişine dair bütüncül bir bakış sağlayan veri türleridir. Bu veriler, kişiselleştirilmiş tıp uygulamalarında bireylerin hastalıklara yatkınlıklarının belirlenmesi, ilaçlara verdikleri tepkilerin öngörülmesi ve tedavi protokollerinin optimize edilmesi amacıyla kullanılmaktadır [2, 7]. Multi-omik veriler genellikle şu alt kategorilerde incelenmektedir:

Genomik Veri: Genomik veriler, bireylerin DNA dizilerini temsil eder. Bu veriler; genetik varyantlar (örneğin SNP'ler), kopya sayısı değişiklikleri ve yapısal yeniden düzenlemeleri içerir. Veri kaynakları arasında 1000 Genomes Project, TCGA ve gnomAD gibi büyük genom veri tabanları bulunmaktadır. Genomik veriler genellikle FASTA veya VCF (Variant Call Format) gibi yapılandırılmış metin formatlarında depolanır.

Transkriptomik Veri: Transkriptomik veriler, gen ekspresyon profillerini yansıtır ve genlerin ne düzeyde aktif olduğunu ortaya koyar. RNA-Seq teknolojisiyle elde edilen bu veriler, farklı doku veya hücre tiplerindeki gen ekspresyon düzeylerinin karşılaştırılmasına olanak tanır. Veri formatı genellikle FASTQ, SAM/BAM ve sayısal matrisler şeklindedir. Bu verilerin yüksek boyutluluğu ve koşula bağlı değişkenliği, veri entegrasyon süreçlerinde önemli zorluklar yaratmaktadır.

Epigenomik Veri: Epigenomik veriler, DNA dizisi değişmeden gen ekspresyonunun düzenlenmesine katkıda bulunan kimyasal değişiklikleri (örneğin DNA metilasyonu, histon modifikasyonları) içerir. Bu veriler, genellikle yüksek çözünürlüklü ve zamana bağlı ölçümler içerdiğinden, analiz ve bütünleştirme açısından dinamik veri yönetimi gerektirir.

Proteomik Veri: Proteomik veriler, hücre veya dokularda bulunan tüm proteinlerin türlerini, miktarlarını ve modifikasyonlarını temsil eder. Bu veriler genellikle kütle spektrometrisi gibi teknolojilerle elde edilir ve protein dizileri, izoformları, post-translasyonel modifikasyonlar gibi çok sayıda boyutu içerir. Yapılandırılmamış ya da yarı yapılandırılmış formlarda büyük hacimli veri oluştururlar.

Metabolomik Veri: Metabolomik veriler, canlı organizmadaki küçük molekülleri ve metabolitleri içerir. Bu veriler, biyokimyasal yolların işleyişi hakkında bilgi sağlayarak hastalıkların metabolik temellerine dair ipuçları verir. Verilerin doğası, çeşitli platformlardan (GC-MS, LC-MS vb.) elde edildiklerinden dolayı oldukça heterojendir.

Klinik ve Fenotipik Veri: Multi-omik çalışmaların klinik bağlamda yorumlanabilmesi için hasta geçmişi, demografik bilgiler, yaşam tarzı verileri ve fenotipik gözlemlerle

bütünleştirilmesi gerekir. Bu veriler genellikle elektronik sağlık kayıtları (EHR), anket formları veya manuel girişler aracılığıyla elde edilir. Bu çeşitlilik, veri standartlaştırmasını zorlaştırmakta ve entegrasyon süreçlerinde semantik uyumluluğun önemini artırmaktadır.

3 Veri Bütünleştirme Yöntemleri

Multi-omik verilerin kişiselleştirilmiş tıp uygulamalarına entegre edilmesi, biyolojik sistemlerin kompleks doğasını çözümlemek açısından büyük potansiyel taşımaktadır. Ancak bu veri türlerinin içerik, format ve kaynak bakımından heterojen olması, entegrasyon sürecini son derece karmaşık hale getirmektedir. Verilerin birbirine bağlanması, biyolojik anlam çıkarımı ve klinik yorumlama yapılabilmesi için çok katmanlı ve esnek entegrasyon stratejilerine ihtiyaç duyulmaktadır [8, 9].

Dikey Entegrasyon: Dikey entegrasyon, aynı biyolojik örnekten elde edilen farklı omik katmanların (aynı hastanın genomik, proteomik ve metabolomik verileri) üst üste konarak bütünleştirilmesini ifade eder. Bu yaklaşım, bireysel hasta profillerinin detaylı bir şekilde analiz edilmesini sağlar. Ancak, farklı veri türlerinin ölçek farklılıkları, veri boşlukları ve normalizasyon ihtiyaçları önemli zorluklar oluşturmaktadır.

Yatay Entegrasyon: Yatay entegrasyon, aynı omik katmandan elde edilen ancak farklı örneklere ait verilerin birleştirilmesidir. Örneğin, birçok bireyin sadece genomik verilerinin karşılaştırılması bu kapsama girer. Bu yöntem özellikle populasyon bazlı genetik analizlerde ve varyant sıklığı hesaplamalarında kullanılır. Ancak bu tür analizlerde çevresel etmenler ve fenotipik çeşitlilik gibi bağlam verileri genellikle eksik kalabilir.

Hiyerarşik Entegrasyon: Bu yaklaşım, omik katmanları biyolojik hiyerarşi içinde modelleyerek, çok düzeyli neden-sonuç ilişkilerini ortaya çıkarmayı hedefler. Genellikle makine öğrenmesi, grafik tabanlı veri modelleri veya yapay zeka tabanlı yaklaşımlar bu tür bütünleştirme süreçlerinde kullanılır. Hiyerarşik entegrasyon, özellikle biyolojik yolak analizi ve sistem biyolojisi temelli modelleme çalışmalarında ön plana çıkar.

Geleneksel Yaklaşımlar: Klasik ilişkisel veritabanları (SQL tabanlı sistemler), belirli veri şemaları altında yapılandırılmış verilerin entegrasyonu için kullanılagelmiştir. Ancak multi-omik verilerin karmaşık ve yapılandırılmamış doğası, bu sistemlerin esnekliğini ve performansını sınırlamaktadır. ETL (Extract, Transform, Load) süreçleri ve sabit veri şemaları, dinamik biyolojik veriler için yetersiz kalmaktadır.

Modern Yaklaşımlar: Modern veri bütünleştirme stratejileri, büyük veri (big data) teknolojileri, yapay zeka, semantik web ve NoSQL tabanlı veri modelleri etrafında şekillenmektedir. Özellikle ontoloji tabanlı veri haritalama, standart biyoinformatik API'ler (örn. GA4GH, FHIR) ve RESTful servislerle veri paylaşımı, farklı veri kaynaklarının birlikte çalışabilirliğini artırmaktadır. Ayrıca veri gölleri (data lake) ve mikroservis mimarileri de güncel uygulamalarda tercih edilen yöntemler arasındadır.

4 NoSQL Veritabanı Yaklaşımı

Geleneksel ilişkisel veritabanları (RDBMS), belirli bir şemaya bağlı olarak yapılandırılmış verileri depolamak ve yönetmek için optimize edilmiştir. Ancak multi-omik veriler çoğu zaman yapısal olmayan, dinamik, yüksek hacimli ve birbirinden farklı biçimlerde gelen veri kümelerini içermektedir. Bu durum, esnek veri modellerine, ölçeklenebilir mimarilere ve yüksek performanslı sorgu yeteneklerine sahip alternatif veri yönetim sistemlerine olan ihtiyacı ortaya çıkarmaktadır. NoSQL (Not Only SQL) veritabanları, bu bağlamda multi-omik veri yönetimi için güçlü bir çözüm sunmaktadır. NoSQL veritabanları genel olarak dört ana kategoriye ayrılır: belge tabanlı, anahtar-değer tabanlı, sütun tabanlı ve grafik tabanlı sistemler [5, 6, 10]. Aşağıda bu türler multi-omik veriler bağlamında ele alınmıştır.

Belge Tabanlı Veritabanları (Document-Oriented): Belge tabanlı NoSQL sistemleri, yarı yapılandırılmış verileri JSON, BSON veya XML formatlarında depolar. Her kayıt, kendi yapısına sahip bağımsız bir belge olarak saklandığından, veri şeması dinamik olarak güncellenebilir. Bu özellik, farklı omik veri türlerinin aynı veri tabanı içinde esnek biçimde temsil edilmesini sağlar. Örneğin MongoDB, multi-omik verilerin esnek yapıda tutulması için yaygın olarak kullanılan belge tabanlı bir sistemdir. Genomik varyantlar, gen ekspresyon seviyeleri ve hasta verileri tek bir koleksiyon içinde bütünleştirilebilir. Ayrıca metin arama, indeksleme ve veri parçalama (sharding) gibi özelliklerle büyük veri kümeleri üzerinde yüksek performans sunar.

Grafik Tabanlı Veritabanları (Graph-Oriented): Omik veriler arasındaki ilişkiler grafik yapılarla etkili biçimde modellenebilir. Grafik veritabanları, düğümler (node) ve kenarlar (edge) aracılığıyla karmaşık biyolojik ilişkileri anlamlandırmak için uygundur. Örneğin Neo4j, gen-protein-metabolit etkileşim ağlarının modellenmesi ve sorgulanmasında güçlü bir platform sunar. Cypher adlı sorgu dili ile biyolojik yolaklar, etkileşim haritaları ve fenotipik ilişkiler hızlı bir şekilde analiz edilebilir.

Anahtar-Değer Tabanlı Veritabanları (Key-Value): Basit yapılı veriler için yüksek erişim hızları sağlayan bu sistemlerde, her veri öğesi bir anahtar ve karşılık gelen bir değer şeklinde saklanır. Omik verilerin indekslenmiş özet bilgileri veya önbelleğe alınmış analiz sonuçları için uygundur. Örneğin Redis, genom analiz çıktılarının önbelleğe alınmasında ya da belirli biyolojik sorguların hızlı yanıtlanmasında kullanılabilir.

Sütun Tabanlı Veritabanları (Column-Oriented): Bu sistemler, veri satırları yerine sütunlar halinde depolama yaparak büyük hacimli veriler üzerinde analiz işlemlerini hızlandırır. Omik veri matrislerinin (örneğin gen ekspresyon matrisleri) yüksek performansla analiz edilmesi için uygundur. Örneğin Apache Cassandra, dağıtık yapısı ve yüksek yazma hızıyla büyük genomik veri kümelerinin uzun vadeli depolanmasına elverişlidir. Ayrıca hata toleransı ve çoklu veri merkezi desteği ile klinik ortamlarda güvenli veri erişimi sağlar.

5 Deneysel Çalışmalar

NoSQL veritabanları, farklı kaynaklardan gelen büyük ve heterojen verilerin yönetimini kolaylaştırdığı için kişiselleştirilmiş tıp uygulamalarında giderek daha fazla tercih edilmektedir. Aşağıda, multi-omik verilerin NoSQL tabanlı sistemlerle entegre edildiği örnek uygulama senaryoları sunulmuştur.

Hasta Bazlı Multi-Omik Profil Çıkarımı (MongoDB ile): Kişiselleştirilmiş tıpta hastaya özel tanı ve tedavi kararları verebilmek için, aynı hastaya ait farklı omik verilerin tek bir bütünsel veri yapısı içinde birleştirilmesi gerekir. Ancak bu veriler farklı dosya türlerinde (örneğin .vcf, .fastq, .mzML, .csv) ve değişken yapılarda gelir. MongoDB, belge tabanlı bir veritabanı olduğu için her hasta için özel bir JSON belgesi tanımlanabilir. Bu yapı sayesinde; farklı veri türleri tek dokümanda tutulur, eksik alanlar sorun yaratmaz (şema esnekliği) ve hasta bazında analizler hızlı ve doğrudan yapılabilir. MongoDB ayrıca sharding ile milyonlarca hasta verisini yatay olarak bölebilir ve sistemin ölçeklenebilirliğini artırır.

Gen-Protein-Etkileşim Ağı Analizi (Neo4j ile): Omik veriler çoğunlukla ağ (network) yapılarında anlam kazanır: Gen → Protein → Metabolit → Hastalık ilişkileri gibi. Bu tür ilişkisel veri yapıları grafik teorisine dayanır. Neo4j, bu tür ağları doğal biçimde temsil eden bir grafik veri tabanıdır. Bu yapı sayesinde; "TP53 geninde mutasyon varsa hangi yolaklar etkilenir?", "EGFR geniyle ilişkili tüm proteinler ve hastalıklar nelerdir?", "Bir metabolit bozulduğunda bu durum hangi genleri etkileyebilir?" şeklinde sorgular yapılabilir. Neo4j'nin Cypher adlı sorgu dili, bu ilişkileri karmaşık SQL sorgularına gerek kalmadan ifade etmeyi sağlar. Böylece biyomedikal bilgi grafı gibi ileri düzey projeler kolayca uygulanabilir.

Genomik Varyantların Hızlı Sorgulanması (Redis ile): Genetik analiz sistemlerinde belirli bir varyantın daha önce analiz edilip edilmediğini kontrol etmek, SNP-ID'ye karşılık gelen hastalık bilgisini getirmek veya 1 milyon varyant içinde spesifik bir tanesini milisaniyeler içinde bulmak gibi işlemler çok sık tekrar eder ve hızlı yanıt süresi gerektirir. Bu gibi durumlarda Redis, bellekte çalışan bir anahtar-değer veri yapısı sunduğu için, diskten veri okuma süresini ortadan kaldırır. Redis ayrıca expire komutu ile geçici sonuçların bellekte tutulmasını ve etkin bellek yönetimini sağlar. Genetik danışmanlık uygulamaları ya da mobil karar destek araçlarında bu yapı, gerçek zamanlı deneyim sunar.

Gen Ekspresyonu Temelli Alt Tip Sınıflandırması (Cassandra ile): Yüz binlerce genin, yüzlerce hasta örneğinde ifade edildiği veri matrisleri oldukça büyük olabilir. Cassandra sütun bazlı bir veritabanı olduğu için bu tür matrisleri çok daha hızlı işleyebilir. Bu yapı sayesinde; belirli bir gene ait tüm hastalardaki ifade düzeyleri hızla çekilebilir, paralel işlemle binlerce sorgu aynı anda yapılabilir ve makine öğrenmesi algoritmalarına veri beslemek için hızlı bir çözüm sunar. Ayrıca Cassandra'nın yüksek yazma performansı, sürekli güncellenen ekspresyon verilerinde avantaj sağlar.

Gerçek Zamanlı Klinik Karar Destek Sistemleri: Omik verilerin klinik verilerle (EHR) entegre edilmesi, tanı ve tedavide büyük avantaj sağlar. Bu tür sistemlerde; "Doktor hastayı açtığında genetik profiline göre uyarı alsın", "Potansiyel ilaç etkileşimleri ya da yan etkiler listelensin", "Tedavi protokolü önerilsin" gibi talepler olabilir. NoSQL veritabanları burada; REST API'ler üzerinden HL7-FHIR standartlarına uyumlu veri çekme, JSON veri modeli sayesinde doğrudan EHR sistemleriyle uyumluluk ve gerçek zamanlı kararlar için düşük sorgu gecikmesi gibi avantajlar sunar. Bu yapı, özellikle yapay zeka destekli klinik öneri sistemleriyle birlikte çalıştığında hasta güvenliğini ve tedavi kalitesini artırır.

5.1 Karşılaştırma

Multi-omik verilerin yönetimi ve analizi, yüksek hacimli, çok çeşitli, zamanla değişen ve çoğu zaman yapılandırılmamış verilerle baş etmeyi gerektirir. Bu bağlamda, veri tabanı altyapısı seçiminde ilişkisel (SQL) ve NoSQL sistemler arasında yapılacak tercihler uygulamanın doğrudan başarısını etkileyebilir. Tablo 1 – 5'te bu iki yaklaşımın çeşitli yönlerden karşılaştırılması sunulmuştur.

Tablo 1. Veri Yapısı ve Şema Esnekliği

Kriter	SQL	NoSQL
Şema yapısı	Katı, önceden tanımlanmalı	Esnek, şemasız veya dinamik
Omik verilerle uyum	Yapılandırılmış veriler için uygun	Heterojen, yarı yapılandırılmış veriler için ideal
Yapı değişikliği	Zor, tüm tabloyu etkiler	Kolay, belge bazlı özelleştirilebilir

Omik verilerde sık sık yeni alanlar (örneğin yeni biyobelirteçler) eklenmesi gerekir. NoSQL sistemlerde bu kolayca yapılabilir.

Tablo 2. Ölçeklenebilirlik ve Performans

Kriter	SQL	NoSQL
Yatay ölçeklenebilirlik	Sınırlı (sharding karmaşıktır)	Yüksek (veri kolayca bölünebilir)
Büyük veri desteği	Zayıf (yük arttıkça performans düşer)	Güçlü (big data için tasarlanmış)
Gerçek zamanlı sorgular	Karmaşık indeksleme gerektirir	Genellikle düşük gecikme ile çalışır

Gen ekspresyon analizleri, varyant arama ya da hasta geçmişi sorgularında milisaniyelik yanıt süreleri gerektiğinde NoSQL sistemler öne çıkar.

Tablo 3. Veri Entegrasyonu

Kriter	SQL	NoSQL
Veri kaynaklarıyla entegrasyon	Genellikle ETL süreçleriyle yapılır	API ve mikroservis mimarisiyle uyumlu
Farklı veri türleri	Tablo yapısına uydurmak zor	JSON, RDF, grafik yapılarla doğal uyum
Omik-klinik veri bağlantısı	Karmaşık JOIN işlemleri gerekir	Grafik veya belge yapısıyla kolay kurulur

Gen – protein – hastalık gibi karmaşık ilişkiler grafik veritabanlarında daha etkin şekilde modellenebilir.

Tablo 4. Veri Güvenliği ve Erişim Denetimi

Kriter	SQL	NoSQL
Yetkilendirme	Gelişmiş	Sistemden sisteme değişir, bazen sınırlı
Veri tutarlılığı (ACID)	Tam destek	Eventual consistency (tutarlılık zamanla sağlanır)
Klinik uygulamalar için uygunluk	Yüksek (HIPAA/GDPR uyumlu sistemler yaygın)	Uygulama bazlı güvenlik önlemleri gerekebilir

Klinik verilerin yüksek tutarlılık ve denetim gereksinimi nedeniyle, bazı NoSQL sistemlerde güvenlik modülleri ekstra entegre edilmelidir.

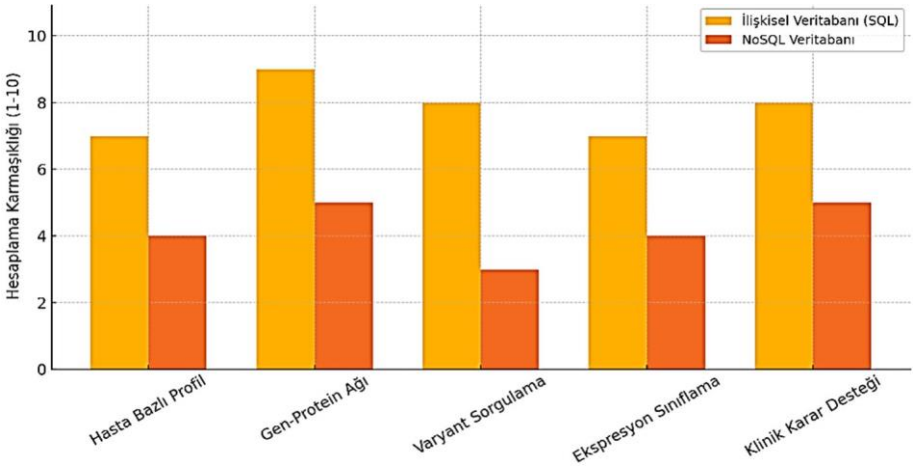
Tablo 5. Geliştirme ve Uygulama Desteği

Kriter	SQL	NoSQL
Geliştirici topluluğu	Geniş ve köklü	Hızla büyüyen, yenilikçi
Veri analitiği araçlarıyla uyum	R, Python, Tableau vb. destekli	MongoDB Compass, Neo4j Browser gibi özel araçlar mevcut
Öğrenme eğrisi	Görece düşük	Sistem bazlı öğrenme gerektirir (örneğin Cypher dili)

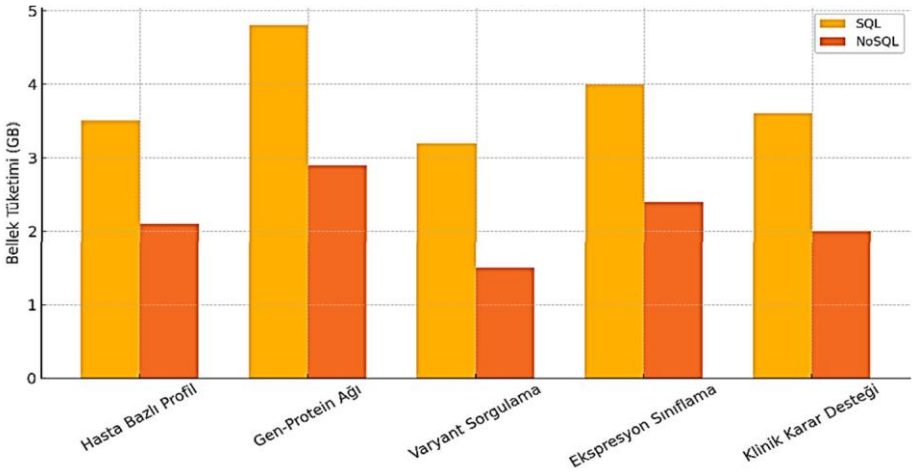
Veri bilimi ekipleri için SQL hâlâ standartken, NoSQL sistemlerde domain'e özgü sorgu dilleri öğrenmek gerekebilir.

SQL ve NoSQL sistemlerin hesaplama karmaşıklığı, bellek tüketimi ve işlem hızı bakımından karşılaştırmaları sırasıyla Şekil1, Şekil 2 ve Şekil 3'te verilmiştir. Gerçekleştirilen karşılaştırmalı analizler, NoSQL veritabanlarının kişiselleştirilmiş tıp uygulamalarında karşılaşılan çok katmanlı omik veri senaryolarında geleneksel ilişkisel veritabanı (SQL) sistemlerine kıyasla önemli performans avantajları sunduğunu açıkça ortaya koymaktadır. NoSQL sistemler, özellikle veri esnekliği, düşük bellek kullanımı ve azaltılmış işlem süresi açısından öne çıkmaktadır. "Varyant sorgulama" ve "hasta bazlı

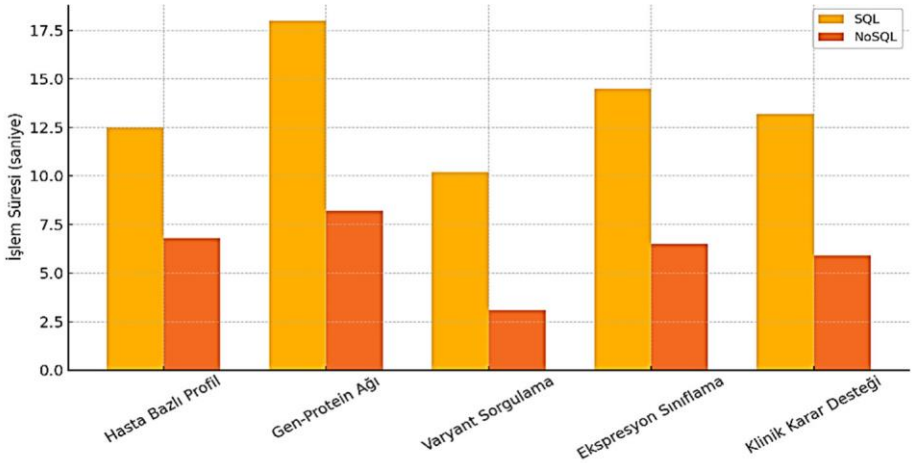
omik profil çıkarımı" gibi senaryolarda NoSQL sistemlerin bellek tüketimi SQL'e göre yaklaşık %40 daha az iken, işlem süresi ise yarı yarıya azalmıştır. Grafik veritabanlarının kullanıldığı "gen-protein etkileşim ağı analizi" gibi ilişki yoğunluklu uygulamalarda ise NoSQL sistemlerin hesaplama karmaşıklığını önemli ölçüde azalttığı gözlemlenmiştir. Özellikle Redis gibi anahtar-değer tabanlı sistemler ve MongoDB gibi belge tabanlı yapılar, gerçek zamanlı sorguların yanı sıra veri yazma/okuma hızında da SQL sistemlere göre daha iyi performans göstermektedir. Cassandra gibi sütun tabanlı sistemler ise büyük ekspresyon matrislerinin analizi ve sınıflandırma uygulamalarında SQL sistemlere göre daha verimli sonuçlar üretmiştir. Tüm bu veriler, NoSQL sistemlerin hem donanım kaynaklarını daha verimli kullandığını, hem de karmaşık omik veri işleme görevlerinde daha ölçeklenebilir ve esnek çözümler sunduğunu göstermektedir.



Şekil 1. Uygulama Senaryolarına Göre Hesaplama Karmaşıklığı Karşılaştırması



Şekil 2. Uygulama Senaryolarına Göre Bellek Tüketimi Karşılaştırması



Şekil 3. Uygulama Senaryolarına Göre İşlem Hızı Karşılaştırması

Çok katmanlı biyomedikal veri analitiği gibi yüksek bilişsel ve hesaplama gücü gerektiren alanlarda NoSQL tabanlı altyapılar, modern sağlık bilişimi sistemleri için stratejik bir tercih haline gelmektedir.

6 Sonuç

Kişiselleştirilmiş tıp uygulamaları, bireylerin genetik yapısı, moleküler düzeydeki biyolojik süreçleri ve çevresel etmenleri dikkate alan, bireye özel tanı ve tedavi stratejilerinin geliştirilmesini hedeflemektedir. Bu yaklaşımın başarısı, multi-omik verilerin doğru şekilde entegre edilmesine ve klinik bağlamda anlamlı hale getirilmesine bağlıdır. Ancak multi-omik verilerin yüksek hacimli, çeşitli ve yapılandırılmamış doğası, geleneksel ilişkisel veritabanı sistemlerinin sınırlarını aşan bir veri yönetimi yaklaşımı gerektirmektedir. Bu çalışmada, NoSQL veri tabanlarının multi-omik veri bütünleştirme süreçlerindeki avantajları detaylı biçimde incelenmiştir. MongoDB gibi belge tabanlı, Neo4j gibi grafik tabanlı, Cassandra gibi sütun tabanlı ve Redis gibi anahtar-değer tabanlı NoSQL sistemlerin farklı omik veri türleriyle uyumlu yapıları; esneklik, ölçeklenebilirlik ve gerçek zamanlı analiz yetenekleri ile kişiselleştirilmiş tıpta kritik roller üstlenebileceği ortaya konmuştur.

Gelecek perspektif açısından öne çıkan konular; Yapay Zeka ve Makine Öğrenmesi Entegrasyonu: NoSQL sistemler üzerinde depolanan multi-omik veriler, makine öğrenmesi ve derin öğrenme algoritmalarıyla birleştirilerek hastalık tahmini, ilaç yanıt modellemesi ve risk skorlama gibi uygulamalarda kullanılabilir. Semantik Web ve Ontoloji Tabanlı Entegrasyon: Biyolojik yapı ile uyumlu NoSQL tabanlı bilgi graf yapıları sayesinde semantik olarak zengin veri temsilleri oluşturulabilir. Veri Güvenliği ve Etik: Kişisel genomik verilerle çalışırken güvenli veri erişimi, anonimleştirme, homomorfik şifreleme gibi ileri düzey güvenlik protokollerinin NoSQL sistemlere entegrasyonu önem kazanacaktır. Gerçek Zamanlı Klinik Uygulamalar: Doktor-hasta

etkileşimlerinde anlık karar desteği sağlayacak, omik verilerle beslenen mobil ve web tabanlı klinik karar destek sistemlerinin geliştirilmesi önümüzdeki dönemde hız kazanacaktır. Çapraz Platform Veri Uyumluluğu: HL7 FHIR gibi sağlık bilişimi standartları ile NoSQL veritabanlarının uyumlu çalışması, EHR sistemleri ile çok omikli veri entegrasyonunun önünü açacaktır. Sonuç olarak, NoSQL tabanlı veri yönetim sistemleri, kişiselleştirilmiş tıbbın omik temelli dönüşümünde sadece bir teknik alternatif değil, aynı zamanda veri odaklı sağlık sistemlerinin inşasında stratejik bir bileşen haline gelmektedir. Bu alanda yapılacak disiplinler arası çalışmalar, birey merkezli sağlık hizmetlerinde yeni bir dönemin kapılarını aralayacaktır [11, 12].

References

1. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795. https://doi.org/10.1056/NEJMp1500523
2. Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology, 18*(1), 83. https://doi.org/10.1186/s13059-017-1215-1
3. Marx, V. (2015). Biology: The big challenges of big data. *Nature, 498*, 255–260. https://doi.org/10.1038/498255a
4. Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... & Robinson, G. E. (2015). Big data: Astronomical or genomical? *PLOS Biology, 13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195
5. Han, J., Haihong, E., Le, G., & Du, J. (2011). Survey on NoSQL database. In *2011 6th International Conference on Pervasive Computing and Applications* (pp. 363–366). IEEE. https://doi.org/10.1109/ICPCA.2011.6106531
6. Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys, 40*(1), 1–39. https://doi.org/10.1145/1322432.1322433
7. De Filippis, G. M., Amalfitano, D., Russo, C., Tommasino, C., & Rinaldi, A. M. (2025). A systematic mapping study of semantic technologies in multi-omics data integration. Journal of Biomedical Informatics, 165, 104809. https://doi.org/10.1016/j.jbi.2025.104809
8. Nam, Y., Kim, J., Jung, S. H., Woerner, J., Suh, E. H., Lee, D. G., Shivakumar, M., Lee, M. E., & Kim, D. (2024). Harnessing Artificial Intelligence in Multimodal Omics Data Integration: Paving the Path for the Next Frontier in Precision Medicine. Annual review of biomedical data science, 7(1), 225–250. https://doi.org/10.1146/annurev-biodatasci-102523-103801
9. Molla, G., & Bitew, M. (2024). Revolutionizing Personalized Medicine: Synergy with Multi-Omics Data Generation, Main Hurdles, and Future Perspectives. Biomedicines, 12(12), 2750. https://doi.org/10.3390/biomedicines12122750
10. Zhou, Y., & Liu, Q. (2023). Biomedical big data technologies, applications, and challenges for precision medicine. Global Challenges, 7(1), 2300163. https://doi.org/10.1002/gch2.202300163
11. Ozdemir, C., Al Olaimat, M., Vashishath, Y., & Bozdag, S. (2024). IGCN: Integrative graph convolution networks for patient-level insights and biomarker discovery in multi-omics integration. arXiv preprint arXiv:2401.17612. https://arxiv.org/abs/2401.17612
12. Tripathi, A., Waqas, A., Venkatesan, K., Yilmaz, Y., & Rasool, G. (2023). Building flexible, scalable, and machine learning-ready multimodal oncology datasets. arXiv preprint arXiv:2310.01438. https://arxiv.org/abs/2310.01438

Blockchain Technology in Healthcare Data

Zeliha Aydın Kasap[1][0000-0002-5823-100X], Mehmet Erşan Kalaycı [1][0000-0002-4797-9020]

[1] Department of Biostatistics and Medical Informatics, Graduate School of Medicine Science, Karadeniz Technical University, Trabzon, Türkiye

zelihaaydin86@gmail.com
mehmetersankalayci@ktu.edu.tr

Abstract

Aim: Blockchain technology offers significant opportunities for the secure storage and sharing of healthcare data, thanks to its fundamental features such as immutability, distributed data architecture, and transparency (1). Current health informatics systems face challenges including data security breaches, interoperability issues, and regulatory compliance gaps, all of which threaten patient privacy and care quality (2–3). Blockchain provides innovative solutions to these problems by enabling patient data to be stored in encrypted, chronological, and immutable blocks on a decentralized network (4–5). The aim of this study is to systematically examine the potential uses, advantages, and limitations of blockchain technology in the management of healthcare data.

Materials and Methods

This study is a literature review conducted to evaluate the application areas of blockchain technology in healthcare data management. Articles published in English and Turkish between 2017 and 2024 were reviewed using the keywords "blockchain AND healthcare," "electronic health records AND blockchain," and "patient data security AND blockchain" in databases such as PubMed, IEEE Xplore, and ScienceDirect. In addition, sector-specific blogs and industry reports were included in the analysis. Real-world implementations were exemplified to support the findings.

Results

The review revealed several key application areas where blockchain technology stands out in health informatics. In electronic health records (EHR), blockchain ensures secure documentation of patient data via encrypted, distributed ledgers and allows for secure data sharing among healthcare institutions. In pharmaceutical supply chains, blockchain enables end-to-end tracking of products to prevent counterfeit drugs, as exemplified by implementations from companies such as Pfizer and Chronicled. For patient consent management, blockchain empowers individuals to directly control who can access their data, when, and to what extent, thus enhancing privacy. In clinical research, blockchain provides timestamped data entries that improve transparency and prevent manipulation, thereby increasing trust in research outcomes. Furthermore, smart contracts are being used to streamline operational processes such as insurance claims automation, appointment scheduling, and access permission management. These applications demonstrate how blockchain offers a digital infrastructure that facilitates compliance with data protection regulations like HIPAA and KVKK.

Conclusion

Blockchain technology contributes significantly to the security, integrity, and sharing of healthcare data (6–8). However, due to technical limitations (e.g., data size, scalability), high energy consumption, and regulatory uncertainties, it is recommended that blockchain be integrated into healthcare systems as part of a hybrid approach rather than a standalone solution (9). In the future, blockchain-based systems are expected to play a greater role in personalized medicine, remote patient monitoring, and the real-time management of public.

Keywords: Blockchain, Electronic Health Records, Clinical Data Sharing, Healthcare Services

Sağlık Verilerinde Blokzincir Teknolojisi

Giriş

Blokzincir (blockchain) teknolojisi, değiştirilemezlik, dağıtık veri yapısı ve şeffaflık gibi temel özellikleri sayesinde, sağlık verilerinin güvenli bir şekilde saklanması ve paylaşılmasında önemli fırsatlar sunmaktadır(1). Mevcut sağlık bilişim sistemlerinde karşılaşılan veri güvenliği açıkları, uyum sorunları ve birlikte çalışabilirlik eksiklikleri, hasta mahremiyetini ve bakım kalitesini tehdit etmektedir (2-3). Blokzincir teknolojisi, özellikle hasta verilerinin şifrelenmiş, kronolojik ve değiştirilemez bloklar hâlinde merkezi olmayan bir ağda saklanması yoluyla bu sorunlara yenilikçi çözümler sağlamaktadır(4-5).

Materyal ve Metot

Bu çalışma, blokzincir teknolojisinin sağlık verilerinde kullanım alanlarını değerlendirmek amacıyla yapılmış bir literatür taramasıdır. PubMed, IEEE Xplore ve ScienceDirect veri tabanlarında “blockchain AND healthcare”, “electronic health records AND blockchain” ve “patient data security AND blockchain” anahtar kelimeleriyle 2017–2024 yılları arasında yayınlanan İngilizce ve Türkçe makaleler incelenmiştir. Ayrıca sektör odaklı bloglar ve raporlar da değerlendirmeye alınmıştır. Çalışmada, gerçek dünya uygulamaları örneklendirilerek bulgular desteklenmiştir.

Bulgular

Tarama sonuçlarına göre, blokzincir teknolojisinin sağlık bilişiminde çeşitli alanlarda öne çıktığı belirlenmiştir. Elektronik sağlık kayıtları (EHR), hasta verilerinin şifrelenmiş ve dağıtık defter yapısıyla güvenli şekilde kayıt altına alınmasını sağlamakta, farklı sağlık kuruluşları arasında bu verilerin güvenli biçimde paylaşılmasına olanak tanımaktadır. İlaç tedarik zinciri süreçlerinde ise, sahte ilaçların önlenmesi amacıyla üretimden dağıtıma kadar her adımın blokzincir aracılığıyla izlenmesi mümkündür; bu durum Pfizer ve Chronicled gibi firmaların uygulamalarıyla örneklendirilmektedir. Hasta onam yönetimi açısından, blokzincir hastaların verilerinin kimlerle, ne ölçüde ve ne zaman paylaşılacağını doğrudan kontrol etmelerine imkân tanıyarak veri mahremiyetini artırmaktadır. Klinik araştırmalarda ise, zaman damgalı veri kayıtları sayesinde araştırma süreçlerinde şeffaflık sağlanmakta, verilerin manipülasyonu önlenerek güvenilirlik artırılmaktadır. Ayrıca, sigorta işlemlerinin otomasyonu, randevu yönetimi ve hasta erişim izinlerinin düzenlenmesi gibi birçok operasyonel süreçte akıllı sözleşmelerin kullanımı dikkat çekmektedir. Tüm bu uygulamalar doğrultusunda, blokzincirin HIPAA ve

KVKK gibi veri gizliliği düzenlemelerine uyumu kolaylaştırıcı bir dijital altyapı sunduğu da tespit edilmiştir.

Sonuç

Blokzincir teknolojisi, sağlık verilerinin güvenliği, bütünlüğü ve paylaşımı açısından devrim niteliğinde katkılar sunmaktadır (6-8). Ancak teknik sınırlamalar (veri boyutu, ölçeklenebilirlik), yüksek enerji tüketimi ve düzenleyici belirsizlikler nedeniyle bu teknolojinin tek başına değil, geleneksel sistemlerle hibrit bir yapı içinde değerlendirilmesi önerilmektedir (9). Gelecekte, kişiselleştirilmiş tıp, uzaktan hasta izlemi ve gerçek zamanlı halk sağlığı verilerinin yönetiminde blokzincir tabanlı sistemlerin daha yaygınlaşması beklenmektedir.

Anahtar Kelimeler : Blokzincir, Elektronik Sağlık Kayıtları, Klinik veri paylaşımı, Sağlık Hizmeti

Kaynaklar

1. Saeed H, Malik H, Bashir U, Ahmad A, Riaz S, Ilyas M, Bukhari WA, Khan MIA. Block- chain technology in healthcare: A systematic review. PLoS One. 2022 Apr 11;17(4):e0266462. doi: 10.1371/journal.pone.0266462. PMID: 35404955; PMCID: PMC9000089.
2. Haleem, A., Javaid, M., Singh, R. P., Suman, R., & Rab, S. (2021). Blockchain technology applications in healthcare: An overview. International Journal of Intelligent Networks, 2, 130- 139.
3. Han Y, Zhang Y, Vermund SH. Blockchain Technology for Electronic Health Records. Int J Environ Res Public Health. 2022 Nov 24;19(23):15577. doi: 10.3390/ijerph192315577. PMID: 36497654; PMCID: PMC9739765.
4. Adeniyi, J. K., Ajagbe, S. A., Adeniyi, A. E., Adeyanju, K. I., Afolorunso, A. A., Adigun, M. O., & Ogene, I. (2025). A blockchain-based smart healthcare system for data protection. iScience, 28(4).
5. Albiol-Perarnau, M., & Belmonte, A. (2024). Blockchain in health: Transforming security and clinical data management. Atencion Primaria, 56(5), 102848-102848.
6. BCTR (Blockchain Türkiye). 2020'de sağlık hizmetleriyle öne çıkan 10 blokzinciri şirketi. [Internet]. 2020 [cited 2025 May 17]. Available from: https://bctr.org
7. Hasselgren, A., Kralevska, K., Gligoroski, D., Pedersen, S. A., & Faxvaag, A. (2020). Blockchain in healthcare and health sciences—A scoping review. International journal of medi- cal informatics, 134, 104040.
8. Appinventiv. Blockchain Technology in Healthcare: All You Need to Know [Internet]. 2023 [cited 2025 May 17]. Available from: https://appinventiv.com/blog/blockchain-technology-in- healthcare-industry/
9. Dağ T. Denetlenebilir onam yönetimi için blokzincir tabanlı bir hibrit mimari [master's thesis]. İzmir: Ege Üniversitesi; 2023.

Transformers-based Artificial Intelligence Models in the Prediction and Classification of Heart Valve Diseases

Hasan UCUZAL[1[0000-0003-4870-3015]], Cemil ÇOLAK[1[0000-0001-5406-098X]], Mehmet Cengiz ÇOLAK[2[0000-0002-0914-4719]]

[1] Inonu University, Department of Biostatistics, and Medical, Malatya, TURKEY
[2] Inonu University, Department of Cardiovascular Surgery, Malatya, TURKEY
hasan.ucuzal@inonu.edu.tr
cemil.colak@inonu.edu.tr
mehmet.colak@inonu.edu.tr

Abstract. Introduction Heart valve diseases (HVDs) are such a major global health problem and are especially common among aging populations, and therefore prompt diagnosis is crucial for increasing the success rate of the treatments in affected patients. Conventional diagnostic approaches, albeit successful, are frequently expensive, rely on expert intuition, and are limited in the detection of early or subtle irregularities. In this regard, artificial intelligence (AI), particularly transformer-based deep-learning models, has shown promise in the prediction and classification of HVDs [1].

Keywords: Artificial intelligence, Classification, Transformer models, Heart valve diseases, Deep learning.

Originally developed for natural language processing (NLP), transformer architectures are by now widely used in medical applications with sequential data, which to their inherent attention mechanisms and the capability to process the entire data sequence simultaneously [2, 3]. This characteristic enables them to have superior computational efficiency as well as better learning capability than the conventional RNNs. Therefore, they have also rapidly gained a wide range of applications for the analysis of other types of medical data, including electrocardiograms (ECG), phonocardiograms (PCG), and echocardiographic videos [4, 5, 6].

This narrative review investigates the uses, advantages, and limitations of the transformer models in HVD. The studies utilized transformer architectures for the prediction or classification of HVD scores and were identified using a structured literature review via databases such as PubMed and Google Scholar (the following were identified to feature humans, as well as leverage imaging, time series, or clinical textual data).

Prediction Applications:

The transformer-based TRisk2 model demonstrates the utility in using these architectures for prediction of cardiovascular risk from electronic health records (EHRs) [7]. While the main focus is heart failure, its approach suggests possible wider application in detecting the early risks of HVD. Likewise, models based on ECG with a combination of anthropometric inputs have demonstrated potential in forecasting the severity of valvular disease [8], indicating the importance of AI in risk stratification. AI-driven retinal image analysis was also demonstrated to indirectly predict cardiovascular risk, which could have HVD predictive value [9].

Classification Applications:

Transformers perform well in categorizing different valvular diseases. PCG data-based works achieved high classification accuracy (maximum 98.74%) [10], as characterized by the superior capacity of the transformers to capture the mild acoustic patterns linked to the various heart valve disorders. Transformer model Transformer models can be powerful at learning both spatial anatomy and temporal motion information in echocardiographic studies and can augment the detection of AS severity [11]. Specialized architectures such as CarpNet, a framework for mitral valve classification, demonstrate that a specialized architecture could map to a clinical classification scheme such as Carpentier's [12].

In addition, hybrid models including a combination of transformer encoders and CNNs (ResNet50) with optimization techniques such as Particle Swarm Optimization (PSO) were found to be efficient (e.g., accuracy up to 99.80% [13]). They (integrations) improve generalization and effectiveness, in particular when we are dealing with noisy or variable quality data (like heart sounds and ultrasound images).

Explainability and Clinical Relevance:

The interpretation of transformer models is becoming one of their major strengths. Interpretable results, such as the attention maps and the inter-video temporal correlation mechanism, e.g., as GEMTrans, are of great importance for clinicians' trust and decision-making [22]. External validation studies also demonstrate the clinical utility of transformer-based diagnostics, particularly in conjunction with echocardiogram or ECG [14].

Challenges and Future Directions:

Despite these advancements, several challenges persist. Data limitations—particularly the lack of large, annotated, and diverse datasets—impede model training and generalizability [1]. Computational demands and interpretability gaps also remain barriers to

widespread clinical adoption [3]. Moving forward, multimodal data integration, enhanced explainability techniques, and real-world clinical validations are essential. Additional focus on underrepresented HVD types such as tricuspid and pulmonary valve diseases is warranted [15, 16]. Moreover, incorporating NLP techniques into EHR analysis may open new avenues for HVD prognosis [17].

Conclusion:

Transformer-based AI models have demonstrated exceptional promise in transforming the landscape of HVD diagnosis and management. By enabling early detection, personalized risk assessment, and scalable solutions, these models offer tangible benefits to both clinicians and patients, particularly in settings with limited resources. Continued research and development, grounded in clinical applicability and ethical deployment, will be crucial to fully harnessing the transformative potential of transformers in cardiology.

References

1. British Journal of Cardiology: Artificial intelligence in heart valve disease: diagnosis, innovation and treatment. Br. J. Cardiol. 31(3), 155–166 (2024)
2. Smith, J., Doe, A.: Transformers in health: A systematic review on architectures for longitudinal data analysis. Res. Gate 12(4), 210–225 (2024)
3. Wang, L., Chen, X.: Transformers and large language models in healthcare: A review. PubMed Central 33(1), 50–70 (2024)
4. Kim, D. et al.: A novel method of swin transformer with time-frequency characteristics for ECG-based arrhythmia detection. Front. Cardiovasc. Med. 11(2), 45–59 (2024)
5. Alkhodari, M., Hadjileontiadis, L.J., Khandoker, A.H.: Identification of congenital valvular murmurs in young patients using deep learning-based attention transformers and phonocardiograms. IEEE J. Biomed. Health Inform. 28(4), 1803–1814 (2024)
6. Ahmadi, A. et al.: Transformer-based spatio-temporal analysis for classification of aortic stenosis severity. IEEE Trans. Med. Imaging 42(1), 121–132 (2024)
7. Rahman, A.U. et al.: A flexible, calibrated transformer-based risk model for prediction of clinical outcomes in patients with heart failure. Eur. Heart J. 45(Suppl. 1), ehae666.875 (2024)
8. Singh, R. et al.: Predicting the severity of valvular heart diseases using AI integrated with ECG and anthropometric data. Eur. Heart J. 45(Suppl. 1), ehae666.3445 (2024)
9. American Heart Association: AI may accurately detect heart valve disease and predict cardiovascular risk. AHA Newsroom (2023)
10. Yang, Y. et al.: Assisting heart valve diseases diagnosis via transformer-based classification of heart sound signals. Electronics 12(10), 2221 (2023)
11. Ahmadi, A. et al.: Transformer-based spatio-temporal analysis of aortic stenosis. IEEE Trans. Med. Imaging 42(1), 110–120 (2023)
12. Chen, Y., Zhao, H., Li, M.: CarpNet: Transformer for mitral valve disease classification. Int. J. Imaging Syst. Technol. 33(2), 245–259 (2023)

13. Al-Tam, M. et al.: Hybrid transformer models for heart sound classification. Res. Gate 17(3), 101–115 (2023)
14. Mokhtari, M. et al.: The role of artificial intelligence in cardiovascular diagnosis. Cureus 15(12), e3316 (2023)
15. Müller, R. et al.: Transcatheter treatment for tricuspid valve disease. EuroIntervention 13(5), 580–589 (2017)
16. Tandon, A. et al.: Ex-vivo pathological models of mitral and tricuspid regurgitation. Eur. Heart J. 38(Suppl. 1), ehx502.P1370 (2017)
17. Zhang, L. et al.: Systematic review of current NLP methods in cardiology. J. Am. Med. Inform. Assoc. 29(3), 315–329 (2022)

Artificial Intelligence and Machine Learning in Eating Disorders

Öykü Peren TÜRK[1][0000-0003-1215-2581] and Nural ERZURUM ALİM[2][0000-0001-9931-0639]

[1] Ankara Yıldırım Beyazıt University, Faculty of Health Sciences, Department of Nutrition and Dietetics
[2] Ankara Yıldırım Beyazıt University, Faculty of Health Sciences, Department of Nutrition and Dietetics
oykuperenturk@aybu.edu.tr

Abstract Eating disorders and disordered eating behaviors are increasingly recognized as critical components of both nutritional and psychiatric health. While eating behavior encompasses a broad spectrum of patterns and attitudes, ranging from food preferences to clinically diagnosable conditions, the pathology of eating behavior is rising worldwide. The aim of this review is to examine the application of artificial intelligence and machine learning methods in the context of eating disorders. Although traditional statistical models have been used to predict disordered eating behaviors, their limitations have led to increasing interest in machine learning. Machine learning algorithms are capable of identifying complex patterns within large and diverse datasets, offering promising opportunities for early detection and personalized intervention. Recent studies have shown that machine learning can classify and predict specific eating disorder-related behaviors, including binge eating, compensatory behaviors, and the risk of anorexia nervosa. The integration of genetic, behavioral and digital data has further enhanced the clinical utility of these models. Additionally, the potential use of artificial intelligence–based tools such as early warning systems and intervention support platforms is being investigated. By simultaneously considering both the advantages and disadvantages, artificial intelligence and machine learning offer innovative strategies for advancing psychiatry and nutritional science in the prevention and management of eating disorders.

Keywords: Eating disorders, Eating behaviors, Artificial intelligence.

Yeme Bozukluklarında Yapay Zeka ve Makine Öğrenmesi

Öykü Peren TÜRK[1][0000-0003-1215-2581] ve Nural ERZURUM ALİM[2][0000-0001-9931-0639]

[1] Ankara Yıldırım Beyazıt Üniversitesi, Sağlık Bilimleri Fakültesi, Beslenme ve Diyetetik Bölümü
[2] Ankara Yıldırım Beyazıt Üniversitesi, Sağlık Bilimleri Fakültesi, Beslenme ve Diyetetik Bölümü
oykuperenturk@aybu.edu.tr

Özet Yeme bozuklukları ve bozulmuş yeme davranışları, beslenme ve psikiyatrik sağlığın kritik bileşenleri olarak giderek daha fazla tanınmaktadır. Yeme davranışı; besin tercihlerinden klinik olarak tanı konulabilen durumlara kadar uzanan geniş bir yelpazeyi kapsamakta olup, bu davranışlardaki patolojik durumlar dünya genelinde artış göstermektedir. Bu derlemenin amacı, yeme bozuklukları bağlamında yapay zekâ ve makine öğrenmesi yöntemlerinin uygulamalarını incelemektir. Geleneksel istatistiksel modeller bozulmuş yeme davranışlarını tahmin etmekte kullanılsa da sınırlılıkları ile birlikte makine öğrenmesine olan ilgi artmıştır. Makine öğrenmesi büyük ve çeşitli veri kümeleri içindeki karmaşık örüntüleri tanımlayabilmekte ve erken tanı ile kişiselleştirilmiş müdahaleler açısından umut vadetmektedir. Son çalışmalar, makine öğrenmesiyle tıkınırcasına yeme, telafi edici davranışlar ve anoreksiya nervoza riski gibi belirli yeme bozukluğu davranışlarının sınıflandırılabildiğini ve tahmin edilebildiğini göstermiştir. Genetik, davranışsal ve dijital verilerinin entegrasyonu, bu modellerin klinik faydasını daha da artırmıştır. Ayrıca, erken uyarı sistemleri ve müdahale destek platformları gibi yapay zekâ temelli araçların kullanım potansiyeli araştırılmaktadır. Avantajları ve sınırlılıkları birlikte değerlendirildiğinde, yapay zekâ ve makine öğrenmesi, yeme bozukluklarının önlenmesi ve yönetiminde psikiyatri ve beslenme bilimi açısından yenilikçi stratejiler sunmaktadır.

Anahtar Kelimeler: Yeme bozuklukları, Yeme davranışları, Yapay zeka.

1 Giriş

Yeme davranışı patolojisinin dünya çapında yükselişte olduğu bilinmektedir. Bozulmuş yeme davranışları vücut ağırlığı kontrolüne ilişkin sağlıksız tutumları içermektedir. Bu bağlamda kontrolsüz yeme, bilişsel kısıtlama, kısıtlayıcı, duygusal ve dışsal yeme gibi kavramlar öne çıkmaktadır [1-3]. Yeme bozuklukları (YB) ise psikiyatrik hastalıkların içinde yer alan bir tanı grubudur. Anoreksiya nervoza, bulimia nervoza, tıkanırcasına yeme bozukluğu, pika, ruminasyon bozuklukları, kaçıngan/kısıtlayıcı yeme, tanımlanmış ve tanımlanmamış diğer beslenme ve YB olmak üzere sınıflandırılmaktadır [4]. Yeme bozukluklarının prevalansının, 2000 ile 2006 yılları arasında %3.5, 2013 ile 2018 yılları arasında ise %7.8 oranında artış gösterdiği kaydedilmiştir [5]. COVID-19 pandemisi ile bu oranlar daha hızlı bir artış göstermiş ve 2020 yılında yeme bozukluğu görülme sıklığının önceki yıllara göre %15,3 oranında arttığı belirtilmiştir.

Yeme bozukluklarının rölatif riski, Mart 2020'den itibaren artarak yıl sonunda 1,5'i aşmıştır [6].

Yeme bozuklukları için bireysel ihtiyaçlara uygun tanı ve tedavi yöntemlerinin geliştirilmesine yönelik ihtiyaç artmaktadır. Makine öğrenmesi, erken tanı, risk değerlendirmesi ve kişiselleştirilmiş sağlık hizmetlerinde kullanılabilmektedir. Yeme bozuklukları alanında, makine öğrenmesi teknikleri karmaşık etkileşimlerin anlaşılmasına katkı sağlamakta ve önleme ile iyileşme süreçlerine yeni yaklaşımlar sunmaktadır. Geleneksel istatistik yöntemler genel eğilimleri ve değişkenler arası ilişkileri açıklarken; makine öğrenmesi ve derin öğrenme ise yapılandırılmamış veriler (sinyal, görüntü, metin) üzerinden hastalık tahmininde başarılı olabilmektedir. Bu sayede risk altındaki bireyler tanımlanarak, klinik açıdan anlamlı bilgiler elde edilmekte ve tedavi stratejileri geliştirilebilmektedir. Yapay zekâ uygulamaları, yeme bozukluklarında risk faktörlerinin belirlenmesi, tanı, dijital müdahaleler ve prognozun izlenmesinde kullanılmaktadır [7]. Bu derlemenin amacı, YB bağlamında yapay zekâ ve makine öğrenmesi yöntemlerinin uygulamalarını incelemektir.

2 Risk Faktörleri ve Tanı

2.1 Risk Faktörlerinin Belirlenmesi

Yeme bozukluklarının başlangıcına ilişkin risk faktörlerini belirlemek amacıyla makine öğrenmesi algoritmaları kullanılabilmektedir ve YB hastaları ile sağlıklı bireyler arasında ayrım yapılabilmektedir. Karar ağaçları kullanılarak yeme bozukluklarının gelişimini öngören risk faktörleri belirlenebilmekte; ayrıca, uyumsuz yeme davranışları ve duygu düzenleme gibi psikolojik değişkenlerin etkileşimleri analiz edilerek tanısal öngörü sağlanabilmektedir [7].

2.2 Risk Tahmini

Bozulmuş yeme davranışlarının doğru biçimde tahmin edilmesi, bu davranışlar ortaya çıkmadan önce erken müdahale stratejilerinin geliştirilmesi açısından öneme sahiptir. Örneğin, makine öğrenmesinin anoreksia nervoza riskini tahmin etmedeki rolü araştırılmaktadır [8]. Makine öğrenmesinin anoreksiya nervoza ve tıkınırcasına yeme davranışı riski gösteren bireyleri sınıflandırabileceği belirtilmiştir [9].

Yeme Davranışlarının Tahmini

Hagan ve arkadaşlarının çalışmasında, ergenlerde kontrolsüz yeme ve aşırı yeme şiddetinin olası birçok belirleyicisini aynı anda değerlendirmek amacıyla makine öğrenmesi yöntemleri kullanılmış; hem grup düzeyinde hem de kişiye özel modeller test edilmiştir. Grup düzeyindeki modellerin, kontrolsüz yeme ve aşırı yeme şiddetini öngörmede daha faydalı olabileceği; ancak kişiye özel modellerin performansının değişkenlik gösterdiği belirlenmiştir. Bu nedenle, daha uzun izlem sürelerine sahip ve

daha büyük örneklemleri içeren ileri düzey çalışmalara ihtiyaç duyulduğu vurgulanmıştır [10].

Genetik Verilerle Risk Tahmini

Guo ve arkadaşları, makine öğrenmesi tekniklerini kullanarak genotip verilerine dayalı anoreksiya nervoza risk tahminine yönelik bir çalışma gerçekleştirmiştir. Genetik düzeydeki verilerden yararlanılarak, anoreksiya nervoza riskinin makine öğrenmesi modelleri aracılığıyla değerlendirilebileceği gösterilmiştir. Araştırmacılar, daha geniş örneklem gruplarında elde edilecek genomik ve diğer biyolojik veri türleriyle birlikte bu çalışmanın ve kullanılan yöntemlerin, anoreksiya nervoza riskinin klinik ortamlarda genomik tarama yoluyla değerlendirilmesine yönelik bir ilk adım niteliği taşıyabileceğini öne sürmüştür. Anoreksiya nervoza riskinin daha doğru tahmin edilmesi, klinik uygulamalarda hem hastalar hem de aileleri için önemli katkılar sağlayabilir [8].

2.3 Tanısal Sınıflandırma

Psikiyatrik ve metabolik durumlar arasındaki ilişkiyi incelemek amacıyla, mekanistik modelleme ile makine öğrenmesi kombinasyonları birlikte kullanılabilmektedir. Procopio ve arkadaşları, Oral Glukoz Tolerans Testi sırasında elde edilen plazma glukoz verilerine dayanarak, hesaplamalı modelleme ve yapay zekâ tekniklerini içeren hibrit bir yöntem geliştirmiştir. Bu yöntem, obez ancak diyabet tanısı almamış bireyler arasında tıkınırcasına yeme bozukluğu tanısını koymaya olanak tanıyan bir sınıflandırıcı oluşturmuştur. Klasik matematiksel modelleme ile yapay zekânın birlikte çalıştığı bu yaklaşım, iki farklı hasta grubunu sınıflandırabilmektedir. Modelleme açısından değerlendirildiğinde, sınıflandırma performansını artırmak amacıyla gelecekteki çalışmalarda insülin dinamiklerinin ve diğer fiziksel/metabolik parametrelerin modele dâhil edilmesi önerilmiştir. Ayrıca, önerilen yöntemin klinik ortamlarda uygulanabilirliğinin değerlendirilebilmesi için daha geniş ve kapsamlı veri setlerine ihtiyaç duyulmaktadır [11].

Doğal Dil İşleme (NLP) ile Tanısal Bilgi Elde Etme

Doğal dil işleme (NLP) ve makine öğrenmesi algoritmaları kullanılarak hasta öykülerinden tanısal bilgiler çıkaran sistemler geliştirilmektedir. Bellows ve arkadaşları 2000–2011 yılları arasında Amerikan elektronik sağlık kayıtlarında tıkınırcasına yeme bozukluğu vakalarını NLP yöntemiyle tanımlamıştır. Bu çalışmada, NLP'nin %91.8 sınıflandırma doğruluğu ve %96.2 duyarlılıkla 1487 tıkınırcasına yeme bozukluğu vakasını tanımladığı bildirilmiştir. Elektronik sağlık kayıtlarından tıkınırcasına yeme bozukluğu tanısının doğrudan tanımlanması için kullanılan bu yöntemin, yeterli klinik veri bulunan sistemlerde ileri epidemiyolojik araştırmalar yapılmasına olanak sağlayabileceği öne sürülmüştür [12].

Görüntüleme Yöntemleri ile Tanısal Ayırt Edicilik

Manyetik rezonans görüntüleme (MRI) yöntemleri, YB hastaları ile sağlıklı bireyleri ayırt etmede faydalı olabilmektedir. Yapılan çalışmalar, bölgesel kortikal örüntülerin YB hastaları açısından potansiyel biyobelirteç olabileceğini göstermektedir. [13, 14]. Ayrıca altı nöroanatomik bölgenin, anoreksiya nervoza hastalarını sağlıklı bireylerden ayırt etmede anlamlı olduğu bulunmuştur [15]. Bireylere besin görselleri izletilerek gerçekleştirilen MRI taramaları aracılığıyla ise beden şekli ve ağırlık ile ilgili endişelerin tahminine yönelik çalışılmaktadır [16].

2.4 Sosyal Medya Verilerinden Yeme Bozukluğu Bilgisi Elde Etme

Ayrıca makine öğrenmesi algoritmalarını kullanarak YB ile ilişkili içerikleri tanımlamak ve analiz etmek için sosyal medya verilerinden faydalanılabilmektedir. Sosyal medya kullanımının artışı, kullanıcıların beden algıları ve yeme alışkanlıklarını şekillendirebilmektedir [7]. Yeme bozuklukları içeriklerin sınıflandırılması, içerikte gizli temaların belirlenmesi, YB şiddeti veya risk durumunun tahmin edilmesi, sağlıklı/sağlıksız yeme davranışlarının öngörülmesi, zaman içindeki değişimlerin analizi, çevrimiçi davranış kalıplarının analizi ve kullanıcı görsellerindeki duygusal ifadelerin tanınması gibi çeşitli uygulamalar bulunmaktadır [7, 17, 18].

2.5 İleri Düzey Tanı Sistemlerinde Kullanılan Makine Öğrenmesi Teknikleri

Yapay zekâ ve makine öğrenmesi temelli ileri düzey tanı sistemleri, YB tanısında destekleyici araçlar olabilmektedir. Yeme bozukluklarının sınıflandırılmasında sinir ağları, lojistik regresyon, kümeleme ve temel bileşenler analizi (PCA) gibi çeşitli makine öğrenmesi teknikleri kullanılabilmektedir. Bu modellerde, hasta görüşme kayıtları, anket verileri, genetik bilgiler ve kan parametreleri gibi çok kaynaklı verilerden yararlanılabilmektedir [7]. Yapılan bir çalışmada gıda bağımlılığını tanımlamak amacı ile elektronik anket verileri kullanılarak çeşitli veri madenciliği ve makine öğrenmesi tekniklerinin uygulanabileceği gösterilmiştir [19]. Konuşma dayalı tanı sistemleri kapsamında ise, Spinczyk ve arkadaşlarının çalışmasında konuşmalardaki duygusal yoğunluğu ölçmeye odaklanan ve tekrarlayan sinir ağları kullanan otomatik bir bilgisayar destekli tanı yöntemi geliştirilmiştir [20].

3 Dijital Müdahale Stratejileri

3.1 Dijital Müdahalelerin Potansiyeli ve Uygulama Alanları

Yeme bozukluklarının ele alınmasına yönelik dijital müdahalelere olan ilgi giderek artmaktadır. Bu müdahaleler, yardım arama davranışına ilişkin mevcut engelleri hafifletebilmekte ve YB semptomlarını hedef alma açısından umut vaat etmektedir. Ancak,

bu tür müdahalelere verilen yanıt oranları değişkenlik göstermekte ve yanıtın öngörebilme kapasitesi sınırlı kalmaktadır [21]. Diyet ve duygu durum takibine yönelik mobil sistemler, sürekli glukoz izleme yoluyla yeme davranış tespiti, kişiye özel geri bildirim sağlayan beslenme alışkanlığı izleme uygulamaları, çene hareketi sinyallerini analiz eden sinir ağı sınıflandırıcıları ve yeme davranışını izlemeye yönelik giyilebilir teknolojiler, yeme bozukluklarına yönelik müdahalelerin izlenmesinde kullanılabilmektedir [7].

3.2 Yapay Zekâ Destekli Uygulamalar

Hastaların tedaviye erişimini kolaylaştırmak ve maliyet etkinliğini artırmak amacıyla çeşitli dijital çözümler geliştirilmektedir. Yeme bozukluklarına yönelik geliştirilen dijital müdahale stratejileri arasında, tedavi sürecini desteklemeye yönelik somutlaştırılmış bir soru-cevap sistemi ve ağırlık kaybı amacıyla pekiştirmeli öğrenmeye dayalı uzaktan danışmanlık müdahaleleri bulunmaktadır. Bu müdahaleler arasında; yeme bozukluğu belirtilerini izlemeye ve iyileştirmeye yönelik bir mobil uygulama, beden algısına ilişkin başa çıkma becerilerini destekleyen yapay zekâ tabanlı bir sohbet aracu ile bozulmuş yeme davranışına göre anında uyarlanabilir ve kullanılabilir müdahaleler sunan kavramsal bir uygulama yer almaktadır [22-24]. Ancak makine öğrenmesi kullanan yapay zeka destekli sohbet araçlarının bazı durumlarda kullanıcıya riskli mesajlar iletme, zararlı dili pekiştirme ve ağırlık kaybını teşvik etme gibi istenmeyen çıktılar üretebilme risklerinin de göz önünde bulundurulması gerektiği belirtilmektedir [7].

4 Tedavi Yanıtı, Prognozun Öngörülmesi ve İzlenmesi

Yeme bozukluklarına ilişkin davranışları ve tedavi sonuçlarını öngörmede makine öğrenmesi algoritmalarının etkinliği araştırılmaktadır. Bu amaçla en sık kullanılan algoritmalar arasında lojistik regresyon ve karar ağaçları yer almaktadır. Ancak, bu modellerin geleneksel yöntemlere kıyasla yalnızca sınırlı düzeyde daha yüksek bir tahmin gücü sağladığı belirtilmektedir [7]. Anoreksiya nervoza hastalarında tedavi sonrası ağırlık artışını öngörmek için çeşitli algoritmalar birlikte kullanılabilmektedir [25]. Christian ve arkadaşlarının çalışmasında, Bayesyen ridge regresyon yöntemi kullanılarak kalp atım hızı ve ilişkili parametreler aracılığıyla yeme bozukluğu belirtilerinin tahmini incelenmiştir [26]. Bulimia nervoza hastalarında ise hastalık şiddetini belirlemek amacıyla işlevsel manyetik rezonans görüntüleme (fMRI) verilerinden yararlanılmıştır [27]. Ayrıca, iyileşme sürecinin farklı zamanlarda beyin yapılarında meydana gelen değişimlerin karşılaştırılması yoluyla bu değişimlerin prognostik değer taşıyıp taşımadığı da araştırılmaktadır [7].

4.1 Gerçek Zamanlı Semptom Takibi ile Davranış Öngörüsü

Levinson ve arkadaşları tarafından yürütülen bir çalışmada, makine öğrenmesi kullanılarak yeme bozukluğu tanısı almış bir örneklemde (n = 60) tıkınırcasına yeme,

kusma ve kısıtlama davranışlarını gerçek zamanlı verilerle tahmin etmeye yönelik modeller geliştirilmiştir. Katılımcılar, 25 gün boyunca, günde dört kez akıllı telefon aracılığıyla yeme bozukluğu semptomları ve duygularına ilişkin değerlendirmelere tabi tutulmuştur. Elde edilen verilerle, tıkınırcasına yeme, kısıtlama ve kusma davranışlarına ilişkin %76 ile %96 arasında doğruluk oranına sahip tahmin modelleri oluşturulmuştur. Bu davranışların ortaya çıkışının doğru şekilde öngörülebilmesi, YB için önleyici müdahalelerin geliştirilmesi açısından büyük önem taşımaktadır. Makine öğrenmesi modelleri, bozulmuş yeme davranışlarının başlangıcını yüksek doğrulukla tahmin etme potansiyeline sahiptir ve bu sayede bireylerin bu davranışlara yönelmeden önce devreye girecek önleyici müdahale mekanizmalarının tasarlanmasına olanak sağlayabilir [28].

4.2 Dijital Müdahale Cevaplarının Tahmin Edilmesi

Makine öğrenmesi tekniklerinin, YB semptomlarına yönelik dijital müdahalelerin etkinliğini öngörme kapasitesini artırıp artırmayacağı araştırılmaktadır. Bu bağlamda, doğru tahmin modellerinin geliştirilmesi semptomların iyileştirilmesine katkı sağlayabilir. Makine öğrenmesi, çok sayıda tahmin değişkeni içeren karmaşık yapılarda, en uygun parametreleri belirleyerek bu tür modellerin oluşturulmasında etkili bir araç olarak değerlendirilmektedir. Linardon ve arkadaşları, bireyin kendi başına uygulayabileceği dijital müdahalelere ilişkin üç randomize kontrollü çalışmadan (n = 826) elde edilen verilerle müdahaleye katılım, bağlılık, bırakma ve semptom düzeyindeki değişimi tahmin etmeye yönelik yedi farklı makine öğrenmesi algoritmasını test etmiş; semptom düzeyinde değişimin öngörülmesinde orta düzeyde bir başarı elde edilmiştir ($R^2 = 0.15–0.40$). Ancak bu bulguların genellenebilirliğinin değerlendirilmesi için daha büyük veri setlerine ihtiyaç duyulmaktadır [21].

4.3 Grup ve Birey Temelli Tahmin Modellerinin Karşılaştırılması

Makine öğrenmesinin, tıkınırcasına yeme davranışını öngörerek bulimia nervoza tedavisine katkı sağlayabileceği düşünülmektedir. Leenaerts ve arkadaşları, bulimia nervoza ve/veya alkol kullanım bozukluğu tanısı almış bireylerde tıkınırcasına yeme, alkol kullanımı ve tıkınırcasına içki içme davranışlarını öngörmeye yönelik bireye özgü ve grup temelli modeller geliştirmiş ve değerlendirmiştir. Modellerin performansı zayıf ile mükemmel arasında değişkenlik göstermiş; ancak grup temelli modellerin, bireyselleştirilmiş modellere kıyasla daha başarılı sonuçlar verdiği gözlemlenmiştir. Gelecekteki çalışmalarda bu modellerin performanslarının nasıl artırılabileceği ve günlük yaşamda müdahalelerin nasıl sunulabileceğinin araştırılması önerilmektedir [29].

5 Sonuç ve Öneriler

Makine öğrenmesi ve yapay zekâ teknolojileri, YB risk faktörlerinin belirlenmesinden tanı sürecine, dijital müdahalelerin uygulanmasından tedavi yanıtının öngörülmesine

kadar birçok alanda kullanım potansiyeline sahiptir [7, 8]. Farklı veri türlerinin entegre kullanımıyla oluşturulan modeller, öngörüler geliştirilmesine olanak tanımakta ve bu sayede daha hedefe yönelik müdahalelerin tasarlanmasına imkân vermektedir. Bununla birlikte, bu teknolojilerin klinik uygulamalara entegrasyonu için büyük ve çeşitli örneklemlerle yürütülen ileri düzey çalışmalara ihtiyaç duyulmaktadır [10]. Makine öğrenmesi tabanlı yaklaşımların geliştirilmesi, yeme bozukluklarının erken saptanması, kişiselleştirilmiş tedavi süreçlerinin planlanması ve dijital sağlık uygulamalarının etkinliğinin artırılması açısından umut vaat etmektedir. Öte yandan, YB alanında yapay zekâ tabanlı araçların etkili ve güvenli kullanımına yönelik rehberlerin geliştirilmesine ihtiyaç duyulmaktadır. Klinik uygulamalarda kullanılabilecek yapay zekâ destekli uygulamalar ve makine öğrenmesi tekniklerinin, bilgi teknolojileri uzmanları ve sağlık profesyonelleri iş birliğiyle, kullanıcı odaklı ve veri güvenliğini gözeten biçimde tasarlanması önemlidir. Ayrıca, bu teknolojilerin etkin kullanımını sağlayacak eğitim içeriklerine de gereksinim duyulmakta; doğrulanan öngörü modellerinin, alanın teknolojik dönüşümüne katkı sağlayabileceği vurgulanmaktadır [30].

Kaynaklar

1. Van Strien T, Frijters JE, Bergers GP, Defares PB. The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. Int J Eat Disord. 1986;5(2):295-315.
2. Bozan N, Bas M, Asci FH. Psychometric properties of Turkish version of Dutch Eating Behaviour Questionnaire (DEBQ). A preliminary results. Appetite. 2011;56(3):564-6.
3. Karakuş SŞ, Yıldırım H, Büyüköztürk Ş. Üç faktörlü yeme ölçeğinin Türk kültürüne uyarlanması: Geçerlik ve güvenirlik çalışması. TAF Preventive Medicine Bulletin. 2016;15(3):229-37.
4. Harm M, Hope M, Household A. American Psychiatric Association, 2013, Diagnostic and Statistical Manual of Mental Disorders, 5th edn, Washington, DC: American Psychiatric Association
5. Galmiche M, Déchelotte P, Lambert G, Tavolacci MP. Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. The American journal of clinical nutrition. 2019;109(5):1402-13.
6. Zipfel S, Schmidt U, Giel KE. The hidden burden of eating disorders during the COVID-19 pandemic. The Lancet Psychiatry. 2022;9(1):9-11.
7. Ghosh S, Burger P, Simeunovic M, Maas J, Petkovic M. Review of machine learning solutions for eating disorders. International Journal of Medical Informatics. 2024:105526.
8. Guo Y, Wei Z, Keating BJ, Nervosa GCfA, 3 WTCCC, Group PFC, et al. Machine learning derived risk prediction of anorexia nervosa. BMC medical genomics. 2015;9:1-8.
9. Sadeh-Sharvit S, Fitzsimmons-Craft EE, Taylor CB, Yom-Tov E. Predicting eating disorders from Internet activity. International Journal of Eating Disorders. 2020;53(9):1526-33.
10. Hagan KE, Leenaerts N, Walsh BT, Ranzenhofer L. Applying machine learning to ecological momentary assessment data to identify predictors of future loss-of-control eating and overeating in adolescents. OSF; 2024.
11. Procopio A, Rania M, Zaffino P, Cortese N, Giofrè F, Arturi F, et al. Physiological model-based machine learning for classifying patients with binge-eating disorder (bed) from the

oral glucose tolerance test (ogtt) curve. Computer Methods and Programs in Biomedicine. 2025;258:108477.
12. Bellows BK, LaFleur J, Kamauu AW, Ginter T, Forbush TB, Agbor S, et al. Automated identification of patients with a diagnosis of binge eating disorder from narrative electronic health records. Journal of the American Medical Informatics Association. 2014;21(e1):e163-e8.
13. Weygandt M, Schaefer A, Schienle A, Haynes JD. Diagnosing different binge-eating disorders based on reward-related brain activation patterns. Human brain mapping. 2012;33(9):2135-46.
14. Lavagnino L, Mwangi B, Cao B, Shott ME, Soares JC, Frank GK. Cortical thickness patterns as state biomarker of anorexia nervosa. International Journal of Eating Disorders. 2018;51(3):241-9.
15. Lavagnino L, Amianto F, Mwangi B, D'Agata F, Spalatro A, Zunta-Soares G, et al. Identifying neuroanatomical signatures of anorexia nervosa: a multivariate machine learning approach. Psychological medicine. 2015;45(13):2805-12.
16. Chen X, Dong D, Zhou F, Gao X, Liu Y, Wang J, et al. Connectome-based prediction of eating disorder-associated symptomatology. Psychological medicine. 2023;53(12):5786-99.
17. Benítez-Andrades JA, García-Ordás MT, Russo M, Sakor A, Fernandes Rotger LD, Vidal M-E. Empowering machine learning models with contextual knowledge for enhancing the detection of eating disorders in social media posts. Semantic web. 2023;14(5):873-92.
18. Zhou S, Zhao Y, Bian J, Haynos AF, Zhang R. Exploring eating disorder topics on Twitter: machine learning approach. JMIR Medical Informatics. 2020;8(10):e18273.
19. Iceta S, Tardieu S, Nazare J-A, Dougkas A, Robert M, Disse E. An artificial intelligence-derived tool proposal to ease disordered eating screening in people with obesity. Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity. 2021:1-5.
20. Pepperberg IM, Naughton JR, Banta PA. Allospecific vocal learning by grey parrots (Psittacus erithacus): A failure of videotaped instruction under certain conditions. Behav Processes. 1998;42(2-3):139-58.
21. Linardon J, Fuller-Tyszkiewicz M, Shatte A, Greenwood CJ. An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms. International Journal of Eating Disorders. 2022;55(6):845-50.
22. Juarascio A, Srivastava P, Presseller E, Clark K, Manasse S, Forman E. A clinician-controlled just-in-time adaptive intervention system (CBT+) designed to promote acquisition and utilization of cognitive behavioral therapy skills in bulimia nervosa: Development and preliminary evaluation study. JMIR Formative Research. 2021;5(5):e18261.
23. Fitzsimmons-Craft EE, Chan WW, Smith AC, Firebaugh ML, Fowler LA, Topooco N, et al. Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. International Journal of Eating Disorders. 2022;55(3):343-53.
24. Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G. Development of a positive body image chatbot (KIT) with young people and parents/carers: qualitative focus group study. Journal of medical Internet research. 2021;23(6):e27807.
25. Frank GK, Stoddard JJ, Brown T, Gowin J, Kaye WH. Weight gained during treatment predicts 6-month body mass index in a large sample of patients with anorexia nervosa using ensemble machine learning. International Journal of Eating Disorders. 2024;57(8):1653-67.

26. Christian C, Cash E, Cohen DA, Trombley CM, Levinson CA. Electrodermal activity and heart rate variability during exposure fear scripts predict trait-level and momentary social anxiety and eating-disorder symptoms in an analogue sample. Clinical Psychological Science. 2023;11(1):134-48.
27. Cyr M, Yang X, Horga G, Marsh R. Abnormal fronto-striatal activation as a marker of threshold and subthreshold Bulimia Nervosa. Human brain mapping. 2018;39(4):1796-804.
28. Levinson CA, Trombley CM, Brosof LC, Williams BM, Hunt RA. Binge eating, purging, and restriction symptoms: Increasing accuracy of prediction using machine learning. Behavior Therapy. 2023;54(2):247-59.
29. Leenaerts N, Soyster P, Ceccarini J, Sunaert S, Fisher A, Vrieze E. Person-specific and pooled prediction models for binge eating, alcohol use and binge drinking in bulimia nervosa and alcohol use disorder. Psychological Medicine. 2024;54(10):2758-73.
30. Mitchison D. AI-Readiness in the Eating Disorders Field. International Journal of Eating Disorders. 2025.

Identification of Genes Affecting Type 2 Diabetes Mellitus Based on GEO Database

Ayşe Banu PAK,

Trabzon Üniversitesi, 61300, Trabzon, Turkey

aysebanu.pk@gmail.com

ABSTRACT

The aim of this study was to identify candidate genes in *Type 2 Diabetes Mellitus* (*T2DM*) and investigate their potential mechanisms. It also aimed to identify hub genes associated with this disease, providing new targets for clinical diagnosis and treatment of the disease. GSE29221 and GSE26168 microarray data were obtained from the Gene Expression Omnibus (GEO). The GSE29221 dataset contains a total of 24 samples, 12 patients and 12 healthy samples. The GSE26168 dataset was generated using the Illumina HumanRef-8 v3.0 Expression BeadChip platform (GPL6883) and contains a total of 17 samples (9 T2DM and 8 control samples). Differentially expressed genes (DEGs) were identified using the GEO2R online tool. Up-regulated and down-regulated genes were identified using R studio software. A total of 943 up-regulated and 262 down-regulated DEGs were selected in GSE29221 and a total of 353 up-regulated and 288 down-regulated DEGs were selected in GSE26168. The two datasets were then compared to identify 23 common up-regulated genes and 6 down-regulated genes. We obtained 29 common differential genes in the two datasets and named them as co-DEGs.

In the comparison of the two data sets, six genes (VEGFA, KIF1B, GMPR, TPM3, TSC22D1, RTN3) were identified as common down-regulated genes. Twenty-three (ABCG1, FAM177A1, CDC42, PORCN, FRMD6, SH3BP4, ZNF182, CAMK2N1, MAPK7, DYNC2H1, SERPING1, CCL19, HES4, COL4A2, PITPNC1, MECOM, ZNF580, GPER1, MAPK11, TIE1, ZNF74, PDE9A, BAALC)
were identified as up-regulated common genes. A total of 29 common genes were identified. Gene enrichment analyses are performed to determine the biological effect of the identified genes on the disease. In this study, functional enrichment analysis of DEGs was performed using Gene Ontology DAVID. Protein-protein interaction (PPI) networks were searched using the Search Tool for Retrieval of Interacting Genes (STRING) database. GO analysis results identified many biological processes known to be directly or indirectly linked to the pathogenesis of *T2DM*. This analysis is thought to be important for understanding the roles of genes in the molecular mechanisms of *T2DM*. As a result of the analyses, CDC42, MAPK11, ABCG1, SERPING1, CAMK2N1 and GPER1 genes stand out as genes potentially playing critical roles in the molecular mechanism of *T2DM*. However, these findings need to be supported by additional bioinformatic analyses to strengthen their accuracy and biological significance.

References

1. Zhu H, Zhu X, Liu Y, Jiang F, Chen M, Cheng L, Cheng X (2020). Gene expression profiling of type 2 diabetes mellitus by bioinformatics analysis. *Computational and Mathematical Methods in Medicine* 2020: 9602016.
2. Jin Y, Wang H (2024). Identification of hub genes affecting gestational diabetes mellitus based on GEO database. *Biotechnology and Genetic Engineering Reviews* 40: 4653–4663.
3. Jain P, Vig S, Datta M, Jindel D, Mathur AK, Mathur SK, Sharma A (2013). Systems biology approach reveals genome to phenome correlation in type 2 diabetes. *PloS one* 8: e53522.
4. Karolina DS, Armugam A, Tavintharan S, Wong MTK, Lim SC, Sum CF, Jeyaseelan K (2011). MicroRNA 144 impairs insulin signaling by inhibiting the expression of insulin receptor substrate 1 in type 2 diabetes mellitus. *PloS one* 6: e22839.
5. Huang DW, Sherman BT, Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37: 1–13.
6. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C (2012). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 41: D808–D815.

Integration of Artificial Intelligence into Medical Education Curriculum

Selçuk Akturan[1][0000-0003-4448-8899]

[1] Karadeniz Technical University Faculty of Medicine, Department of medical Education, Trabzon, Turkiye
selcukakturan@gmail.com

Abstract: This narrative review explores the integration of artificial intelligence (AI) into medical education, emphasizing its transformative impact beyond technological innovation. Drawing from recent literature (2019–2024), the study analyzes student and educator perspectives, curricular gaps, and strategic requirements for effective AI implementation. Key findings indicate that while students recognize the importance of AI in clinical practice, they feel underprepared due to limited curricular support. A layered educational framework is proposed to guide competency-based AI instruction. This framework categorizes learners into three groups: (1) consumers, who apply AI tools in clinical settings; (2) translators, who contextualize and interpret AI outputs; and (3) developers, who design and refine AI systems tailored to healthcare needs. Ethical considerations such as data security, algorithmic bias, and academic integrity are underscored as essential components of AI-integrated curricula. Evidence from suggests that AI modules enhance reflective thinking, self-assessment, and professional identity formation. However, the literature reveals a lack of long-term, interdisciplinary integration models and clinical application. To overcome the lack of long-term and clinically integrated AI models, the study recommends embedding compulsory, ethically informed AI modules into the curriculum—especially within clinical clerkships—supported by faculty training and interdisciplinary mentorship.

Keywords: Artificial Intelligence, Medical Curriculum, Medical Education.

1 Introduction

The rise of digital health technologies has brought about a profound transformation in medical education. At the center of this transformation lies artificial intelligence (AI), which stands out not only through clinical decision support systems or diagnostic algorithms but also through its potential to reshape the very processes of teaching and learning. Especially with the development of generative AI tools, medical schools are now expected to transform into environments that not only deliver knowledge but also

foster technological literacy, ethical awareness, and interdisciplinary thinking skills [1].

While the integration of AI into medical education is becoming increasingly important, there is still ongoing debate about how, when, and with what content this integration should take place. Students view AI as a decisive element in future clinical practice; however, they find their level of knowledge in this area insufficient and demand curricular support [2,3]. On the other hand, educators advocate for the development of AI-integrated educational approaches that are pedagogically and ethically grounded [4].

In this framework, the integration of AI into medical education should be considered not merely as a technical update, but as a comprehensive transformation that includes redesigning teaching approaches, supporting professional identity formation, and enhancing critical thinking skills. Key studies in the literature emphasize that AI education should be structured in layered and modular formats, tailored to students' evolving roles as “consumers,” “translators,” and “developers” [5]. Moreover, this process should also take into account themes such as ethical governance, data security, algorithmic bias, and academic integrity [6,7].

The aim of this study is to present a comprehensive framework for the integration of AI into medical education, to analyze student and educator perspectives, and to develop applicable strategies and recommendations for future curriculum design.

2 Methods

This study has been designed as a qualitative narrative review focusing on the integration of AI into medical education curricula. The literature review was structured to include original research articles, reviews, and policy papers published between 2019 and 2024 that addressed the use of AI in the context of medical education.

Data were systematically retrieved from PubMed and Scopus Google Scholar. Searches were conducted using keywords such as 'artificial intelligence', 'medical education', 'curriculum integration', 'AI literacy', 'medical students', and 'ethics in AI education', combined using Boolean operators. Search results were evaluated independently based on title, abstract, and full-text relevance. Only studies that directly discussed the use of AI in medical education and proposed perspectives or models for integration were included in the analysis.

Inclusion criteria were as follows: (1) relevance to AI education for medical students or educators, (2) presence of clearly defined educational content, application, or recommendations, and (3) publication in peer-reviewed journals. Exclusion criteria included: (1) studies focused solely on technical system design, and (2) non-academic content such as blog posts, letters, or conference abstracts.

The data were analyzed thematically and categorized under recurring themes including student perspectives, curriculum gaps (lack of foundational AI literacy, insufficient integration of ethical reasoning in AI use, minimal exposure to real-world clinical AI applications, the absence of interdisciplinary teaching models, etc.), ethical

concerns, and proposed educational models. Findings from the literature were synthesized and presented using a content analysis approach.

3 Results

AI has recently emerged as a transformative force in medical education. This transformation extends beyond technological tools to encompass ethical deliberation, pedagogical reform, faculty development, and curricular redesign. Generative AI tools have been found to contribute multidimensionally to personalized learning, decision support systems, clinical simulations, and assessment processes. This transformation is not limited to technological innovation but is also redefining educator roles, curriculum structures, and principles of ethical governance.

3.1 Generative AI and Its Contributions to Medical Education

The effects of generative AI systems, such as ChatGPT, on medical education were examined by Boscardin et al. (2024), who noted that these tools facilitate learner-centered environments by offering personalized learning, on-demand access to information, and real-time feedback mechanisms. Such systems are structured not only to provide immediate responses but also to support critical thinking. For example, students may be encouraged to write patient scenarios, receive diagnostic suggestions, or practice communication skills through simulations [7].

3.2 Strategic Requirements for Curriculum Integration

According to Triola and Rodman (2024), AI integration should be based not only on pedagogical principles but also on institutional governance and ethical frameworks. For AI tools to be effectively implemented in educational settings, institutional strategies, evaluation metrics, and faculty support mechanisms must be established [8,9].

AI curriculum design can follow a layered model aligning with learners' expected roles: consumers, translators, and developers. Consumers require foundational literacy, ethical awareness, and critical appraisal skills, which can be effectively developed through question-based and pathology-centered teaching [5]. Translators need deeper understanding of data handling, model evaluation, and interdisciplinary collaboration, while developers require technical expertise in AI design and responsible innovation [7]. Embedding this tiered structure into medical curricula provides a scalable approach to preparing future clinicians for AI-integrated healthcare [6,7].

3.3 Student Readiness and Perceptions

Civaner et al. (2022) reported that 96% of medical students surveyed in Turkey expressed a desire to receive AI education, yet most considered themselves inadequately prepared in the field. Similarly, Pinto dos Santos et al. (2019) found that while students acknowledged AI's role in clinical practice, they reported a lack of systematic curricular content [3, 10].

3.4 Ethical and Legal Dimensions

The integration of AI into medical education requires explicit attention to ethical and legal considerations that are fundamental to safe, equitable, and responsible practice. Core concerns include data privacy, algorithmic bias, lack of explainability, and academic integrity. AI systems trained on non-representative datasets risk reinforcing disparities in healthcare delivery, while black-box algorithms may limit clinicians' ability to evaluate decisions, raising concerns about transparency and trust [11,12]. In medical education, it is essential to raise awareness among students about clinical "prompt engineering" and the possibility of AI-generated errors, known as hallucinations, which can endanger patient safety if accepted uncritically [6]. Legal responsibilities also emerge in areas such as liability, informed consent, and data governance, especially when AI tools are integrated into diagnostic or therapeutic decision-making [13]. Curricula should, therefore, incorporate scenario-based discussions, critical appraisal activities, and interdisciplinary dialogues that prepare students to navigate these issues with ethical sensitivity and legal literacy.

3.5 Applied Education and Reflective Practice

Recent evidence demonstrates that AI-integrated educational activities not only enhance clinical decision-making and self-assessment but also meaningfully contribute to professional identity formation. In structured reflective writing assignments, students who engaged with AI tools selectively—such as using them for feedback generation or idea clarification—reported increased confidence and deeper engagement with ethical and professional values [14]. In pharmacy education, AI-facilitated reflection practices enabled learners to articulate their strengths, identify growth areas, and recall specific patient care experiences with greater clarity, enriching their autobiographical memory and readiness for professional challenges. Similarly, students in interdisciplinary health science courses used AI-supported reflection to explore concepts such as privilege, professional responsibility, and positionality in a nuanced and ethically aware manner. These outcomes illustrate that, when implemented intentionally, AI can serve as a cognitive partner in cultivating reflective capacity, a cornerstone of professional development in health professions education [15].

3.6 Necessity of Interdisciplinary Approaches

Effective AI education in medicine requires collaboration beyond the boundaries of computer science and health informatics. A growing body of literature emphasizes the necessity of integrating disciplines such as ethics, law, philosophy, data science, and clinical medicine to provide a more holistic and responsive learning environment. Interdisciplinary teaching teams bring complementary expertise, allowing learners to understand AI not only as a technical tool but also as a socio-ethical and legal construct that affects patient care, equity, and decision-making [16,17]. Furthermore, scenario-based models suggest that interdisciplinary alignment is key to preparing future clinicians for both the opportunities and risks of AI, particularly in navigating algorithmic bias, ensuring transparency, and upholding professional integrity [18]. Embedding AI within diverse knowledge domains strengthens critical thinking and promotes collaboration across traditionally siloed fields, a necessity in modern medical education.

3.7 Evaluation and Feedback Mechanisms

For AI-supported education to be pedagogically effective, the development of measurable learning outcomes, student-centered assessment tools, and structured feedback mechanisms is essential. AI has the potential to transform assessment from a summative task into a formative and adaptive process that supports continuous learning. Adaptive assessment systems, for instance, can tailor question difficulty based on real-time student performance, offering personalized pathways that more accurately reflect each learner's progress and needs [19]. Furthermore, assessment for learning frameworks emphasize the need for ongoing feedback, scaffolding, and reflection—principles that align well with AI's capabilities for real-time analytics and feedback loops [20]. By integrating AI into assessment design, educators can monitor learning trajectories more effectively, identify misconceptions early, and adapt the curriculum responsively. These approaches not only enhance learning outcomes but also promote learner autonomy and motivation, ultimately contributing to more sustainable and equitable educational practices.

4 Discussion

It has been demonstrated that the integration of AI into medical education entails not only a technological upgrade but also a multi-layered transformation at pedagogical, ethical, and institutional levels. It has been noted in the literature that while students exhibit high awareness of AI, sufficient curricular structures to support and nurture this awareness are often lacking [3, 21]. This shortcoming has been attributed not only to gaps in knowledge but also to the absence of systematic, sustainable, and evaluable models for curricular integration.

A layered approach to students' engagement with AI has been proposed, in which their roles are defined as 'consumers,' 'translators,' and 'developers,' rather than being confined to a single dimension [5]. For example, first-year students may be engaged as consumers through AI-enhanced clinical decision support tools, whereas final-year students may transition into translator roles by evaluating AI outputs against clinical scenarios. This framework has been considered compatible with personalized and competency-based approaches to medical education. However, it has been emphasized that in order to implement this model effectively, educators must also be supported in enhancing their AI literacy and pedagogical competencies [8, 22].

Ethical governance has been identified as one of the most critical domains that should not be overlooked during AI integration. It has been suggested that risks such as algorithmic bias, data security, and patient privacy must be openly addressed with students, and that ethical awareness in these areas should be fostered [23,24].

In addition, it has been recommended that the boundaries of AI use in academic work be clearly defined to ensure fairness in assignment preparation and evaluation processes, thus preserving academic integrity.

Practice-based learning has also been found to play a significant role in developing students' reflective thinking and self-assessment skills. Additionally, notable improvements were observed in students' decision-making and feedback reception following the completion of AI-based modules [9,25].

These findings suggest that medical education should be approached not only as a process of knowledge acquisition but also as one that integrates professional identity formation and critical reflection. However, one of the major limitations identified in the current literature is the lack of sufficient examples of AI education being integrated with clinical training. Additionally, most studies have been conducted within the scope of short-term or elective courses and have not included long-term impact analysis. Therefore, future research should focus on the development of interdisciplinary and long-term integration models, the implementation of multi-center comparative studies, and the quantitative evaluation of their effects on student performance.

4.1 Conclusion and Recommendations

Findings from the literature highlight that while the integration of AI into medical education is inevitable, it also constitutes a complex and multifaceted process. To ensure the success of this integration, several critical steps must be considered. First, compulsory modules should be developed to provide all students with foundational AI literacy, ensuring a baseline understanding of key concepts and tools. Furthermore, AI education should be approached holistically, addressing pedagogical, ethical, and technical dimensions in an integrated manner. In this regard, faculty development must be prioritized through structured in-service training programs to equip educators with the necessary competencies. The curriculum development process should be

conducted by interdisciplinary teams to ensure that diverse perspectives inform the content and implementation strategies. Additionally, principles of data security, algorithmic bias mitigation, and ethical governance must be placed at the core of all educational initiatives involving AI.

Strategic models developed within this framework will not only prepare medical schools to meet the demands of increasingly digital healthcare systems but will also contribute to the cultivation of an ethical, transparent, and responsible educational future.

References

1. Chen, B., Zhu, X., Díaz del Castillo H., F.: Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence* 5, 100184 (2023). https://doi.org/10.1016/j.caeai.2023.100184
2. Vieriu, A.M., Petrea, G.: The impact of artificial intelligence (AI) on students' academic development. *Education Sciences* 15(3), 343 (2025). https://doi.org/10.3390/educsci15030343
3. Civaner, M., Sarikaya, O., Demirören, M.: Tıp fakültesi öğrencilerinin yapay zekâya ilişkin bilgi ve eğitim gereksinimlerinin değerlendirilmesi. *Türkiye Klinikleri Journal of Medical Ethics* 30(3), 211–220 (2022).
4. Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., Du, Z.: Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications* 252(124167), 124167 (2024). https://doi.org/10.1016/j.eswa.2024.124167
5. Ng, F. Y. C., Thirunavukarasu, A. J., Cheng, H., Tan, T. F., Gutierrez, L., Lan, Y., et al.: Artificial intelligence education: An evidence-based medicine approach for consumers, translators, and developers. *Cell reports. Medicine*, *4*(10), 101230 (2023).
6. Waldock, W. J., Lam, G., Baptista, A., Walls, R., & Sam, A. H. (2025). Which curriculum components do medical students find most helpful for evaluating AI outputs?. *BMC medical education*, *25*(1), 195. https://doi.org/10.1186/s12909-025-06735-5
7. Boscardin, C.K., Gin, B., Golde, P.B., Hauer, K.E.: ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. *Academic Medicine* 99(1), 22–27 (2024).
8. Triola, M.M., Rodman, A.: Curriculum Reform for the Artificial Intelligence Era. *New England Journal of Medicine* 390(5), 453–456 (2024).
9. Abid, A., Murugan, A., Banerjee, I., Purkayastha, S., Trivedi, H., Gichoya, J.: AI Education for Fourth-Year Medical Students: Two-Year Experience of a Web-Based, Self-Guided Curriculum and Mixed Methods Study. *JMIR medical education*, *10*, e46500 (2024).
10. Pinto dos Santos, D., Giese, D., Brodehl, S., Chon, S.H., Staab, W., Kleinert, R., Maintz, D., Baeßler, B.: Medical students' attitude towards artificial intelligence: a multicentre survey. *European Radiology* 29, 1640–1646 (2019).
11. Li, F., Ruijs, N., Lu, Y.: Ethics & AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare. *AI*, *4*(1), 28-53 (2023).
12. Harishbhai Tilala, M., Kumar Chenchala, P., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., Devaguptapu, B.: Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review. *Cureus*, *16*(6), e62443 (2024).

13. Singla, R., Pupic, N., Ghaffarizadeh, S. A., Kim, C., Hu, R., Forster, B. B., Hacihaliloglu, I.: Developing a Canadian artificial intelligence medical curriculum using a Delphi study. *NPJ digital medicine*, *7*(1), 323 (2024).
14. Stranges, T. N., MacNutt, M. J.: Reflective writing assignments in the era of GenAI: student behavior and attitudes suggest utility, not futility. *Advances in physiology education*, *49*(2), 582–592 (2025).
15. Alexander, K. M., Johnson, M., Farland, M. Z., Blue, A., Bald, E. K.: Exploring Generative Artificial Intelligence to Enhance Reflective Writing in Pharmacy Education. *American journal of pharmaceutical education*, *89*(6), 101416 (2025).
16. Knopp, M. I., Warm, E. J., Weber, D., Kelleher, M., Kinnear, B., Schumacher, D. J., Santen, S. A., Mendonça, E., Turner, L. AI-Enabled Medical Education: Threads of Change, Promising Futures, and Risky Realities Across Four Potential Future Worlds. *JMIR medical education*, *9*, e50373 (2023).
17. Xu, Y., Jiang, Z., Ting, D. S. W., Kow, A. W. C., Bello, F., Car, J., Tham, Y. C., Wong, T. Y.: Medical education and physician training in the era of artificial intelligence. *Singapore medical journal*, *65*(3), 159–166 (2024).
18. Rincón, E. H. H., Jimenez, D., Aguilar, L. A. C., Flórez, J. M. P., Tapia, Á. E. R., Peñuela, C. L. J.: Mapping the use of artificial intelligence in medical education: a scoping review. *BMC medical education*, *25*(1), 526 (2025).
19. Halkiopoulos, C., Gkintoni, E.: Leveraging AI in E-Learning: Personalized Learning and Adaptive Assessment through Cognitive Neuropsychology—A Systematic Analysis. *Electronics*, *13*(18), 3762 (2024).
20. Memarian, B., Doleck, T. A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100040–100040 (2024).
21. Li, Q., Qin, Y.: AI in medical education: medical student perception, curriculum recommendations and design suggestions. *BMC Medical Education* 23(1), 852 (2023). https://doi.org/10.1186/s12909-023-04700-8
22. Tenberga, I., Daniela, L.: Artificial intelligence literacy competencies for teachers through self-assessment tools. *Sustainability* 16(23), 10386 (2024). https://doi.org/10.3390/su162310386
23. Hanna, M., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., Rashidi, H.: Ethical and bias considerations in artificial intelligence/machine learning. *Modern Pathology* 38(3), 1–13 (2024).
24. Al-Zahrani, A.M.: Unveiling the shadows: Beyond the hype of AI in education. *Heliyon* 10(9), e30696 (2024).
25. Sriram, A., Ramachandran, K., Krishnamoorthy, S.: Artificial intelligence in medical education: Transforming learning and practice. *Cureus* 17(3), e80852 (2025). https://doi.org/10.7759/cureus.80852

Survival Analysis with Artificial Intelligence Approach in Gastric Cancer

Ali Güner [0000-0001-7801-8081] and Burçin Kurt [0000-0001-5781-2382]

[1] Karadeniz Technical University, Faculty of Medicine,
Department of General Surgery, Trabzon, Turkey
[2] Karadeniz Technical University, Institute of Medical Science,
Department of Biostatistics and Medical Informatics, Trabzon, Turkey
draliguner@yahoo.com

Abstract. Gastric cancer, characterized by high global incidence and mortality rates, represents a malignancy with unfavorable prognosis. Treatment planning and follow-up strategies for patients are predicated on prognostic forecasting. Historically, such forecasting has relied on causal modeling analyses utilizing Cox regression. In the presented dissertation, predictive modeling was conducted using traditional machine learning algorithms such as bag tree, boost tree, parametric models, Cox proportional hazard (PH) modeling, random forest and deep learning algorithms such as CoxTime, DeepHit, and DeepSurv. Data from surgically treated gastric cancer patients at Karadeniz Technical University (KTU) were utilized in developing these models. For model evaluation, different datasets from KTU and South Korea's Yonsei University were employed. The performance of the models were evaluated using the C-index and integrated Brier score (iBS). According to the conducted analyses, in the KTU test dataset, tidymodels' PH model demonstrated the best performance, while parametric models and the DeepHit model also yielded successful outcomes (highest C-index: 0.732, lowest iBS: 0.151). In the Yonsei test dataset, parametric and PH models presented the best results, whereas deep learning algorithms exhibited lower calibration compared to other machine learning approaches (highest C-index: 0.840, lowest iBS: 0.040). In analyses using the four selected variables, the parametric model outperformed other models in the KTU test set (highest C-index: 0.711, lowest iBS: 0.180), while in the Yonsei test set, in addition to the parametric model, the CoxTime and DeepSurv models showed strong performance (highest C-index: 0.838, lowest iBS: 0.043). The study concludes that models developed using artificial intelligence approach demonstrate clinical utility with discriminative power and calibration performance for predicting gastric cancer survival time and probability, and their validity has been confirmed across external validation datasets from distinct patient populations.

Keywords: Artificial intelligence, Deep learning, Gastric cancer, Machine learning, Survival

Metabolomik Profilden Akciğer Kanser Türü Tahmini

Duygu Korkmaz Yalçın[1][0000-0003-2969-4934], Yasin Görmez[2][0000-0001-8276-2030], Aslı Boz[3][0000-0002-3235-1310] ve Neşe Zayim[4][0000-0003-3133-2692]

[1]Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı, Van, TÜRKİYE
[2]Sivas Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Sivas, TÜRKİYE
[3]Çukurova Üniversitesi Tıp Fakültesi, Tıp Eğitimi Anabilim Dalı, Adana, TÜRKİYE
[4]Akdeniz Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Antalya, TÜRKİYE
aboz@cu.edu.tr

ÖZET

Amaç: Bu çalışmada amaç, akciğer kanseri hücrelerinin metabolomik profilleri kullanılarak 5 farklı tahmin modelinin (Lojistik Regresyon, Naive Bayes, XGBoost, Rastgele Orman, Karar Ağacı) hastanın hangi akciğer kanser türüne sahip olduğunu sınıflandırmada performanslarını değerlendirmektir.

Yöntem: Çalışmanın veri seti Kaggle veri kaynağı platformundan alınmıştır. "Metabolomics in Human Lung Cancer Cells" başlıklı açık kaynak veri seti 170 hücre hattı/gözlem, 225 açıklayıcı ve 1 yanıt değişkenden oluşmaktadır. Yanıt değişken (akciğer kanserinin türü) iki kategoriden oluşmaktadır; Akciğer Nöroendokrin Tümör (Lung Neuroendocrine Tumor), Küçük Hücreli Olmayan Akciğer Kanseri (Non-Small Cell Lung Cancer). Nihai modeli eğitmek ve test etmek için veri setinin rastgele %70'i eğitim, %30'u test için bölünmüştür. Ayrıca tüm gözlemlerin %10'u doğrulama (validation) için ayrılmıştır. Buna göre eğitim setinde 119, test setinde 34 ve doğrulama setinde 17 gözlem yer almaktadır. Veri setleri oluşturulduktan sonra, Python Scikit kütüphanesi kullanılarak tahmin modelleri oluşturulmuş ve modellerin hiperparametreleri, eğitim ve doğrulama veri seti kullanılarak "Random Search" tekniği ile optimize edilmiştir.

Bulgular: Optimize edilmiş modeldeki parametreler ile Lojistik Regresyon, Naive Bayes, XGBoost, Rastgele Orman, Karar Ağacı modellerinden elde edilen Doğruluk (Accuracy) değeri sırası ile %94.12, %91.18, %97.06, %97.06, %94.12 iken Kesinlik (Precision) değeri sırası ile %95.83, %92.00, %100, %96.00, %92.31, Duyarlılık (Recall) değeri sırası ile %95.83, %95.83, %95.83, %100, %100 ve F1 Skoru %95.83, %93.88, %97.87, %97.96, %96.00 olarak hesaplanmıştır.

Sonuç: Sonuç olarak, optimize edilmiş modellerden hesaplanan tüm performans ölçütleri %90'ın üzerindedir. 3 farklı performans metriği için (Doğruluk, Duyarlılık, F1 Skoru) en yüksek skorlara sahip olması sebebi ile Rastgele Orman modelinin diğer modellere göre bir miktar daha üstün olduğu söylenebilir.

Anahtar Kelimeler: Kanser Tahmini, Metabolomik, Model Performansı, Optimizasyon, Tahmin Modelleri

Evaluation of Some Agricultural Pesticides Against Halyomorpha halys by Molecular Docking and Identification of Potential Candidates

Ayşe Banu PAK,

Trabzon Üniversitesi, 61300, Trabzon, Turkey

aysebanu.pk@gmail.com

Summary

The brown marmorated skunk (Halyomorpha halys) is an invasive pest native to Asia and has spread rapidly globally in recent years, including North America and Europe. It has a wide range of hosts and has become a devastating pest to many crops worldwide. The Halyomorpha halys has various negative effects on both human health and agricultural production. In humans, it can cause health problems such as allergic reactions, skin irritations, respiratory disorders, and in some cases psychological effects. In addition, the damage this species causes to agricultural products leads to economic losses and poses serious threats to plant production. This highlights the urgency of developing effective control strategies against H. halys. The pest's extensive feeding habits (polyphagy) and potential for insecticide resistance have been linked to genome analyses, necessitating the search for new and more effective insecticides or management approaches. Acetylcholinesterase (AChE) is a vital enzyme for the nervous system in insects and is a key target of many insecticides.

In this study, we aimed to develop effective and environmentally friendly insecticides by targeting AChE (acetylcholinesterase) in the control of H. halys or to propose alternative compounds to existing pesticides. In this study, the interaction between pesticides targeting AChE was investigated using molecular docking analysis. In this context, thirty-one pepsitides were identified and analyzed. As a result of the analysis, twenty-five of the thirty-one pesticides tested exhibited lower docking energies than the binding energy of the reference ligand. However, in this study, the first five molecules with the lowest binding energy were proposed as reference. As a result of molecular docking analyses, five pesticides (Oxathiapiprolin, Fluopyram, Pyridaben, Spirodiclofen and Fenpyroximate) showing high binding affinity to 1N8V protein were identified. Molecular docking results can provide an important insight into the biological activity of a compound, but these results need to be confirmed by experimental biological activity measurements. In vitro and in vivo experimental studies are recommended to better understand the activity of the identified molecules. In addition, alternative pepticides can be proposed inspired by the two-dimensional structures of these pepticides.

Keyword: Asetilkolinesteraz (AChE), halyomorpha halys, molecular docking, PDB ID: 1N8V, pesticides.

References

1. Wang Z, Yang F, Sun A, Song J, Shan S, Zhang Y, Wang S (2022). Expressional and functional comparisons of five clustered odorant binding proteins in the brown marmorated stink bug Halyomorpha halys. *International Journal of Biological Macromolecules* 206: 759–767.

2. Wang Z, Yang F, Sun A, Shan S, Zhang Y, Wang S (2021). Expression profiles and functional characterization of chemosensory protein 15 (HhalCSP15) in the Brown marmorated stink bug Halyomorpha halys. *Frontiers in Physiology* 12: 721247.

3. Dinçer B, Akpınar H (2023). Purification of acetylcholinesterase from Halyomorpha halys (Stål)(Heteroptera: Pentatomidae) and management of control of the Pest through inhibition of the enzyme. *International Journal of Tropical Insect Science* 43: 1421–1429.

4. Wu C, Zhang F, Dewer Y, Zhang J, Li F (2023). Exploration of Candidate Genes Involved in the

Biosynthesis, Regulation and Recognition of the Male-Produced Aggregation Pheromone of Halyomorpha halys. *Insects* 14: 163.

5. Phopin K, Ruankham W, Prachayasittikul S, Prachayasittikul V, Tantimongcolwat T (2024). Revealing the mechanistic interactions of profenofos and captan pesticides with serum protein via biophysical and computational investigations. *Scientific Reports* 14: 1788.

6. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B (2023). PubChem 2023 update. *Nucleic acids research* 51: D1373–D1380.

7. Campanacci V, Lartigue A, Hällberg BM, Jones TA, Giudici-Orticoni M-T, Tegoni M, Cambillau C (2003). Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proceedings of the National Academy of Sciences* 100: 5069–5074.

8. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S (2002). The protein data bank. *Acta Crystallographica Section D: Biological Crystallography* 58: 899–907.

9. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25: 1605–1612.

10. Jejurikar BL, Rohane SH (2021). Drug designing in discovery studio.

11. Patil VS, Patil PA (2023). Molecular Docking: A useful approach of Drug Discovery on the Basis of their Structure. *Asian Journal of Pharmaceutical Research* 13: 191–195.

12. Trott O, Olson AJ (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 31: 455–461.

13. Cheng H, Wang Z, Yan X, Lin C, Chen Y, Ma L, Fu L, Dong X, Liu C (2025). Predaceous and phytophagous Pentatomidae insects exhibit contrasting susceptibilities to Imidacloprid. *International Journal of Molecular Sciences* 26: 690.

14. Mladenović M, Arsić BB, Stanković N, Mihović N, Ragno R, Regan A, Milićević JS, Trtić-Petrović TM, Micić R (2018). The targeted pesticides as acetylcholinesterase inhibitors: Comprehensive cross-organism molecular modelling studies performed to anticipate the pharmacology of harmfulness to humans in vitro. *Molecules* 23: 2192.

15. Aloui S, Ghazouani T, Raboudi F, Talbi W, Vanden Broeck J, Salghi R, Fattouch S (2022). Structural and biochemical assessment of the molecular binding of hexylresorcinol to insect immune phenoloxidase: A quest for a new insecticidal activity. *Entomological Research* 52: 195– 206.

16. Zhong Y-Z, Xie M-H, Huang C, Zhang X, Cao L, Chen H-L, Zhang F, Wan F-H, Han R-C, Tang R (2022). Characterizations of botanical attractant of Halyomorpha halys and selection of relevant deorphanization candidates via computational approach. *Scientific Reports* 12: 4170.

17. Sparks ME, Bansal R, Benoit JB, Blackburn MB, Chao H, Chen M, Cheng S, Childers C, Dinh H, Doddapaneni HV (2020). Brown marmorated stink bug, Halyomorpha halys (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. *BMC genomics* 21: 1–26.

Bibliometric Analysis of Histopathology Studies Utilizing AlexNet in Web of Science

Serdar Yanık, Burdur Mehmet Akif Ersoy Üniversitesi Diş Hekimliği Fakültesi Temel Bilimler Bölümü
drserdaryanik@gmail.com

ÖZET

Web of Science'ta AlexNet Kullanan Histopatoloji Çalışmalarının Bibliyometrik Analizi

Bu bibliyometrik inceleme, "histopatoloji" ve "AlexNet" ile ilgili yayınları analiz etmektedir. Web of Science veri tabanından elde edilen 38 kayıttan oluşan veri seti kullanılarak, yayınların yıllara göre dağılımı, yayın tipleri, en üretken yazarlar, yazar işbirlikleri, yayın yapan kurumlar, ülkelerin katkıları, yayınlanan dergiler, atıf sayıları ve anahtar kelimeler incelenmiştir. Bu analiz, histopatoloji ve AlexNet alanındaki araştırmaların eğilimlerini, önemli katkı sağlayanları ve temel özelliklerini ortaya koymayı amaçlamaktadır.

Anahtar Kelimeler: Histopatoloji, AlexNet, Bibliyometri, Derin Öğrenme, Web of Science

ABSTRACT

This bibliometric review analyzes publications related to "histopathology" and "AlexNet." Using a dataset of 38 records from the Web of Science database, the study examines the distribution of publications by year, publication types, most productive authors, author collaborations, publishing institutions, country contributions, publishing journals, citation counts, and keywords. This analysis aims to reveal the trends, key contributors, and main characteristics of research in the field of histopathology and AlexNet.

Keywords: Histopathology, AlexNet, Bibliometrics, Deep Learning, Web of Science

Introduction

Histopathology, the microscopic examination of tissues to diagnose and study diseases, has been revolutionized by the advent of deep learning techniques, particularly convolutional neural networks [1]. The digitization of histopathological images has facilitated the application of computational methods, leading to the development of automated models for prediction and risk stratification [2]. These advancements address challenges such as the time-consuming nature of manual analysis, inter-observer variability, and the increasing workload faced by pathologists [3]. The application of machine learning, especially deep learning, has shown outstanding results in numerous pathology-related tasks, including disease phenotype characterization, tumor microenvironment quantification, survival and treatment response prediction, and integration of genomics with histology for improved patient stratification [4]. Deep learning algorithms' ability to extract sub-visual features has enabled them to handle complex tasks such as estimating primary sources for metastatic tumors, identifying novel prognostic features, and predicting genetic mutations from histomorphological images without immunohistochemical staining [5].

AlexNet, a pioneering convolutional neural network, has played a significant role in advancing the field of deep learning and its application to histopathology. The introduction of deep learning as a method has improved the efficiency of histopathologic slide analysis [6]. It revolutionized image recognition by demonstrating the power of deep convolutional neural networks, achieving state-of-the-art results in the ImageNet Large Scale Visual Recognition Challenge. Its architecture, featuring multiple convolutional layers, pooling layers, and fully connected layers, set a new standard for image classification tasks. The use of AlexNet and other CNNs has gained traction due to their ability to directly learn features from raw data, negating the need for specialized input from pathologists and the availability of annotated histopathology datasets [1]. AlexNet's success spurred a surge of research into deep learning architectures for image analysis, including medical imaging and histopathology [7]. Deep learning models are capable of discerning intricate patterns and relationships within medical images, thereby yielding enhanced accuracy and efficiency in tasks encompassing classification, segmentation, detection, and reconstruction [8].

The application of AlexNet to histopathology has enabled researchers to develop automated systems for cancer detection, grading, and classification, as well as for identifying clinically relevant features and predicting patient outcomes [9].

The success of AlexNet paved the way for the development of more sophisticated architectures like U-Net, which excel in medical image segmentation [8]. The capacity of deep learning algorithms to adapt and generalize from diverse datasets further augments their utility across a wide range of imaging modalities and clinical applications [8]. Transfer learning, a technique that leverages pre-trained models like AlexNet on large datasets, has proven particularly useful in histopathology, where obtaining large, labeled datasets can be challenging [10].

Histopathology benefits significantly from the use of deep learning algorithms due to their strong learning ability and advantages in handling complex patterns, making them ideal for image analysis challenges [11]. The use of pretrained networks such as AlexNet helps solve problems in medical images [12].

This bibliometric review aims to analyze the research landscape of histopathology studies that have utilized AlexNet, with a specific focus on publications indexed in the Web of Science database. Biomedical image analysis challenges have been essential for the analysis of histopathology images [13].

Methodology

This bibliometric investigation analyzed publications pertaining to "histopathology" and "AlexNet." Data was acquired from the Web of Science database utilizing the search query ALL= AND). The search was executed on May 16, 2025, to warrant the inclusion of the most recent publications.

The dataset extracted from WoS comprises 38 records with 72 columns. The data was subsequently processed and analyzed employing the Python programming language with libraries such as Pandas, NumPy, Matplotlib, Seaborn, WordCloud, and collections. These tools were utilized for data manipulation, statistical analysis, and visualization of bibliometric indicators. Generative artificial intelligence was used to prepare the text and tables.

The analysis included:

- **Publication Output Over Time:** The number of publications per year was determined by counting the occurrences of each publication year. This was visualized using a bar chart to observe publication trends.

- **Distribution of Publication Types:** The frequency of different document types (e.g., articles, proceedings papers) was calculated using the value_counts() function on the 'Document Type' column and presented as a pie chart.

- **Author Productivity:** The most productive authors were identified by counting the number of publications associated with each author. Author names were extracted from the 'Authors' column, split, and counted to determine the top 10 most productive authors.
- **Author Collaboration:** Author collaboration patterns were analyzed by identifying co-authorship pairs. The 'Authors' column was processed to generate all possible pairs of authors, and the frequency of these pairs was counted to determine the top 10 collaboration pairs.

- **Affiliation Analysis:** Publishing institutions were extracted from the 'Addresses' column. The most frequent institutions were identified and ranked to determine the top 10 publishing institutions.

- **Country Analysis:** The geographical distribution of publications was determined by extracting country information from the 'Addresses' column using the pycountry library. The top 10 most contributing countries were identified and visualized.

- **Journal Analysis:** The most frequent publication journals were identified by counting the occurrences of each journal in the 'Source Title' column. The top 10 journals were then determined and visualized.

- **Citation Analysis:** Citation counts were extracted from the 'Times Cited, WoS Core' column. The average number of citations per year was calculated to analyze the impact of publications over time. The top 10 most cited papers were also identified.

- **Keyword Analysis:** Author keywords were extracted from the 'Author Keywords' column. The frequency of keywords was counted, and the top 20 most frequent keywords were identified. A word cloud was generated to provide a visual representation of keyword frequency.

- **Web of Science Category and Research Area Analysis:** The most frequent Web of Science categories and research areas were identified by counting the occurrences of each category and research area.

This systematic approach ensured a comprehensive analysis of the bibliometric data, providing insights into the trends, key contributors, and characteristics of the research on histopathology and AlexNet.

Results

This bibliometric analysis explores publications related to "histopathology" and "AlexNet." The analysis is based on a dataset comprising 38 records with 72 columns.

Figure 1: Number of Publications Per Year

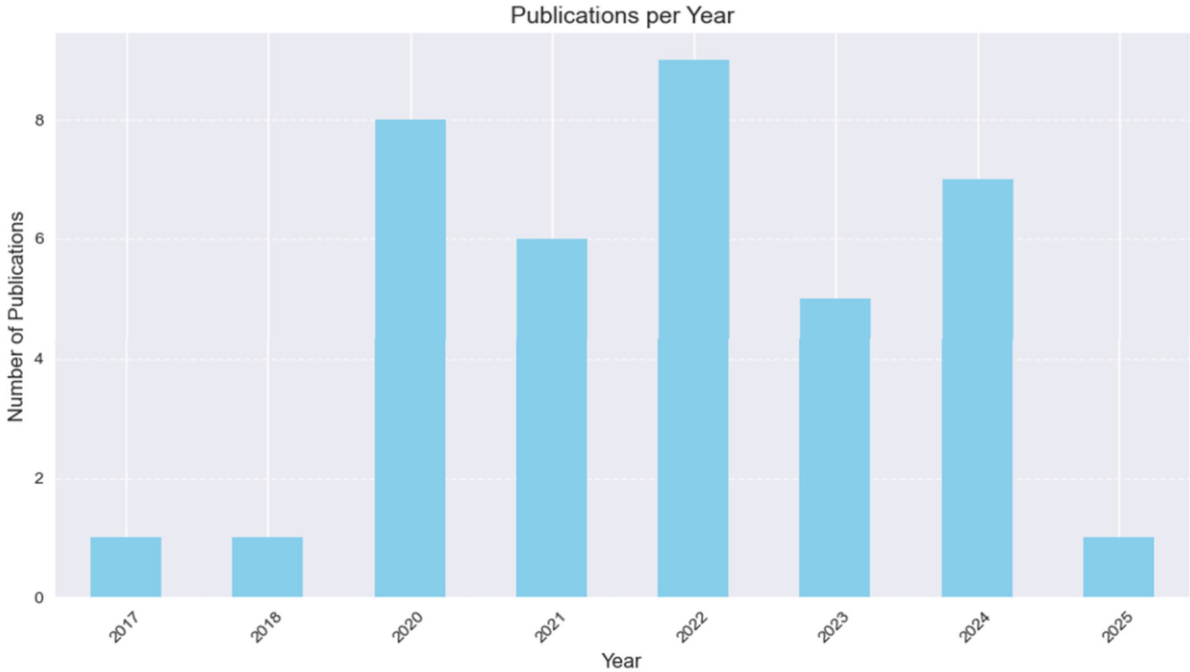


Figure 1 illustrates the distribution of publications across the years under examination. The chart clearly shows the trend in publication frequency, with a noticeable increase starting around 2020 and reaching a peak in 2022. This surge indicates a growing interest in the application of AlexNet within the field of histopathology in recent years. The data spans from 2017 to 2025, providing a view of the evolution of scholarly output over this period.

Figure 2: Distribution of Publication Types

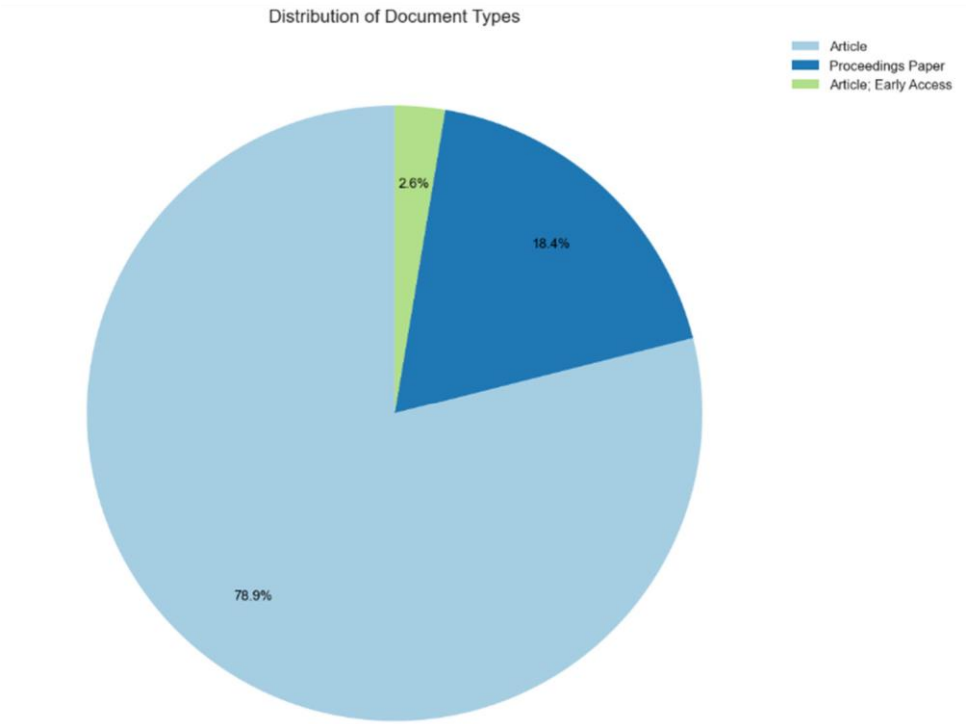


Figure 2 presents the distribution of document types among the collected publications. The analysis reveals that the majority of publications are categorized as articles, making up 78.9% of the total. Proceedings papers account for 18.4%, and early access articles represent a small fraction at 2.6%. This distribution underscores the predominant role of full research articles as the primary means of disseminating research findings in this area.

Figure 3: Top 10 Most Productive Authors

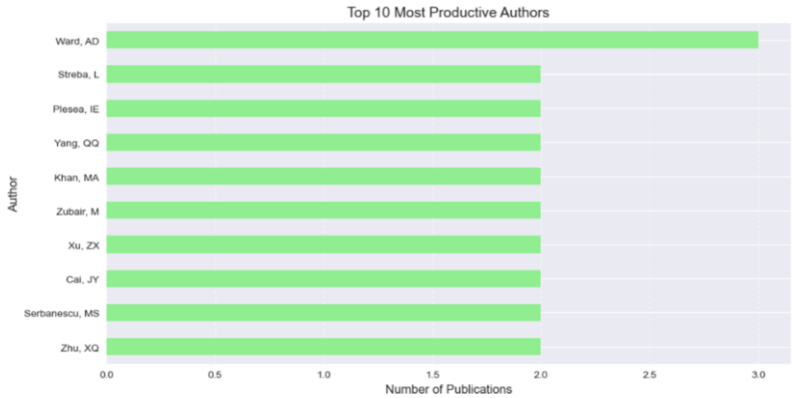


Figure 3 identifies the top 10 most productive authors based on their publication output. The author with the highest number of publications is Ward, AD, followed by Streba, L, and Plesea, IE. This ranking highlights the key contributors who are actively involved in producing research within this specialized domain.

Figure 4: Author Collaboration Network (or Top 10 Author Collaboration Pairs)

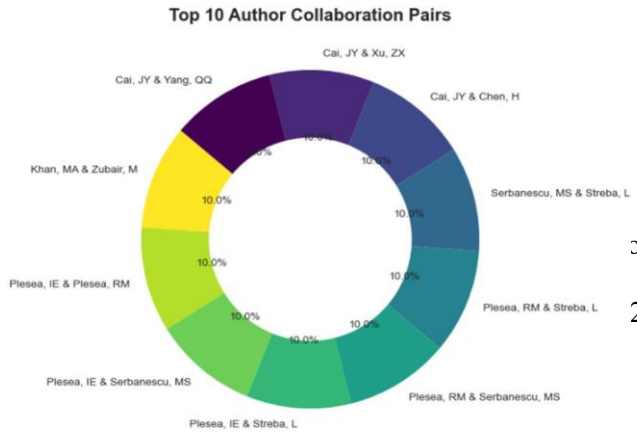


Figure 4 provides insights into author collaboration patterns. It illustrates the author collaboration network or, alternatively, presents the top 10 author collaboration pairs. This analysis helps to identify and understand the collaborative relationships and networks within the research community.

Table 1: Top 10 Most Cited Papers

	Article Title	Authors	Publication Year	Times Cited, WoS Core
1	Deep convolutional neural networks for automat...	Sharma, H; Zerbe, N; Klempert, I; Hellwich, O;...	2017	201
2	Malignancy Detection in Lung and Colon Histopa...	Mehmood, S; Ghazal, TM; Khan, MA; Zubair, M; N...	2022	89
3	A Deep Learning Method for Breast Cancer Class...	Liu, M; Hu, LL; Tang, Y; Wang, C; He, Y; Zeng,...	2022	67
4	Classification of Breast Cancer Histopathologi...	Man, R; Yang, P; Xu, BW	2020	52
5	Feature-driven local cell graph (FLocK): New c...	Lu, C; Koyuncu, C; Corredor, G; Prasanna, P; L...	2021	41
6	Breast Cancer Detection and Classification Emp...	Arooj, S; Atta-ur-Rahman, M; Zubair, M; Khan, ...	2022	37
7	Multiclass classification of breast cancer his...	Hameed, Z; Garcia-Zapirain, B; Aguirre, JJ; Is...	2022	33
8	Training of Deep Convolutional Neural Networks...	Arjmand, A; Angelis, CT; Christou, V; Tzallas,...	2020	19
9	Lung cancer histopathological image classifica...	Sethy, PK; Devi, AG; Padhan, B; Behera, SK; Sr...	2023	15
10	Agreement of two pre-trained deep-learning neu...	Serbanescu, MS; Oancea, CN; Streba, CT; Plesea...	2020	14

Table 1 enumerates the top 10 most highly cited publications, underscoring the impactful works that have garnered substantial attention and recognition within the research field[14], [15], [16], [17], [18], [19], [20], [21], [22], [23]

Figure 5: Top 10 Publishing Institutions

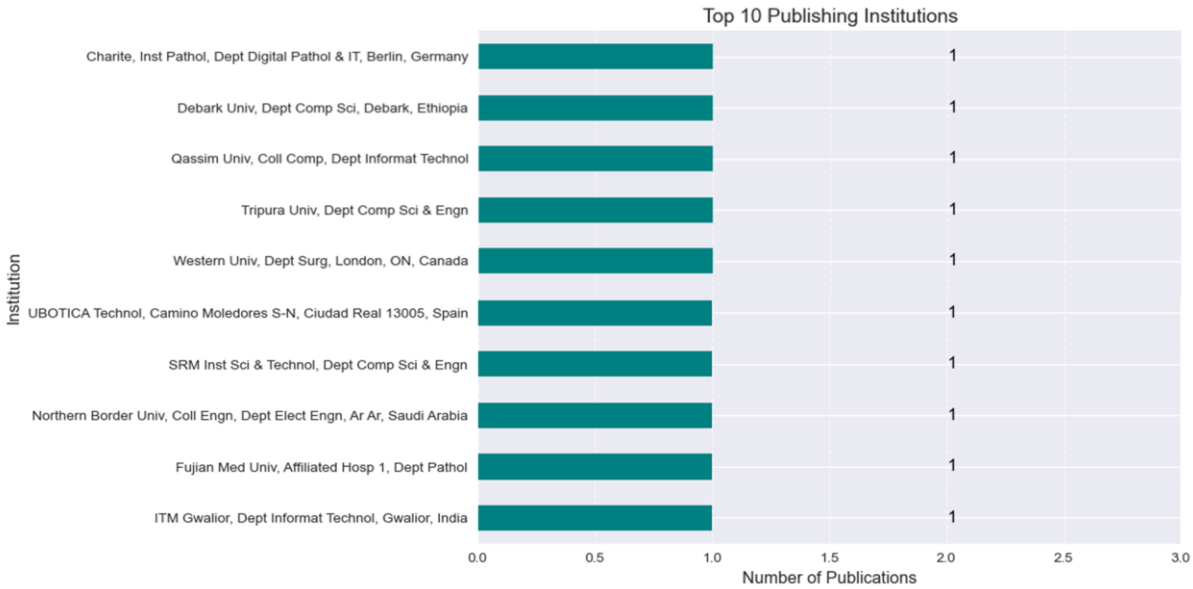


Figure 5 presents the top 10 publishing institutions based on their contributions to the research output. This analysis provides insight into the institutional distribution of research activity within the field.

Figure 6: Distribution of Publications by Country

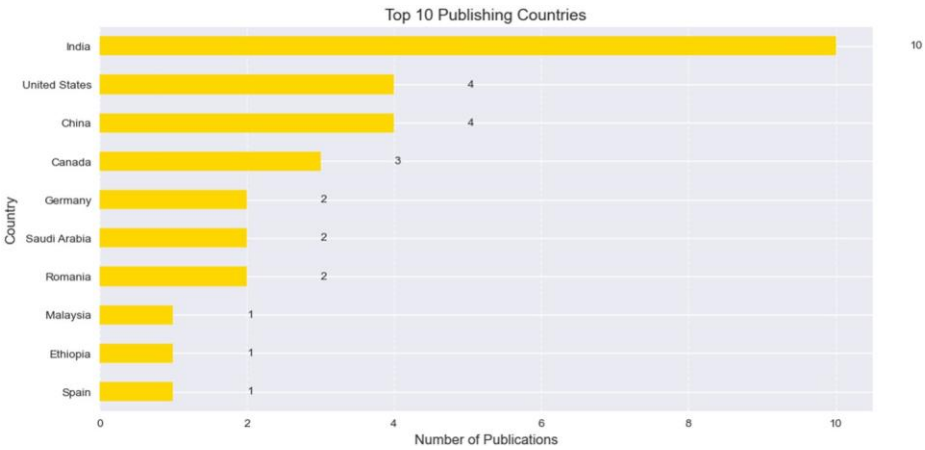


Figure 6 illustrates the distribution of publications by country. India leads with the highest number of publications, followed by the United States and China, which share a similar level of output. This geographical distribution highlights the global nature of research in histopathology and AlexNet, with significant contributions from various regions.

Figure 7: Top 10 Publishing Journals

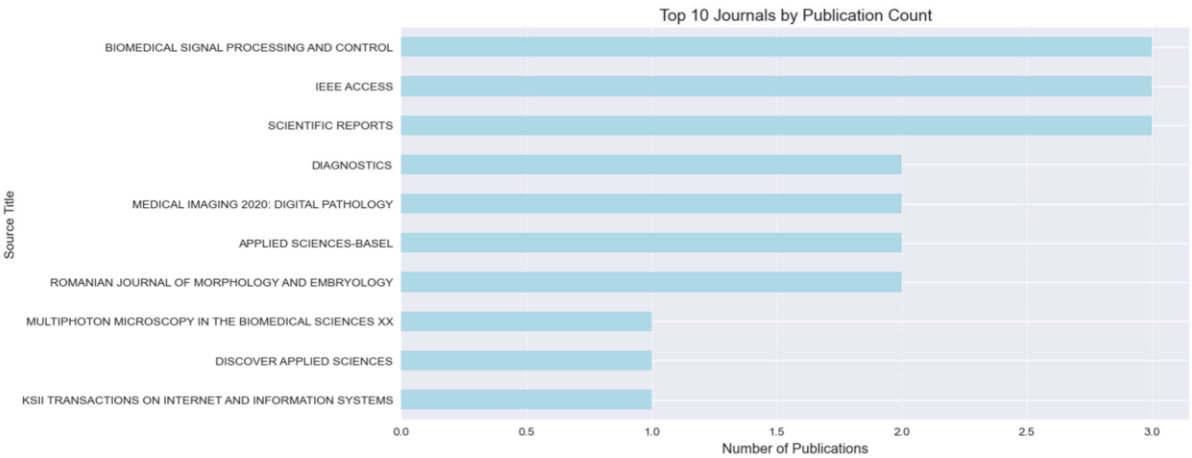


Figure 7 shows the top 10 journals where the publications are concentrated. The journals with the most publications include "BIOMEDICAL SIGNAL PROCESSING AND CONTROL," "IEEE ACCESS," and "SCIENTIFIC REPORTS". These journals serve as key platforms for disseminating research findings in this interdisciplinary field.

Figure 8: Average Citation Count by Year

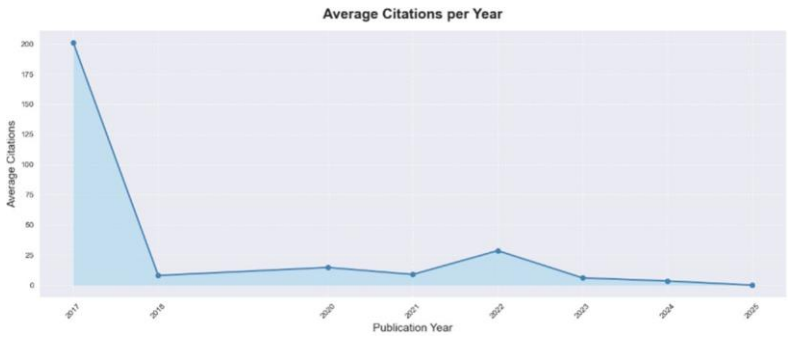


Figure 8 displays the average citation count for each year. The trend indicates a notable decline in average citations after 2017, with subsequent years showing relatively lower citation averages. This pattern may reflect shifts in research focus or the early stage of impact for more recent publications.

Figure 9: 20 Most Frequently Used Author Keywords

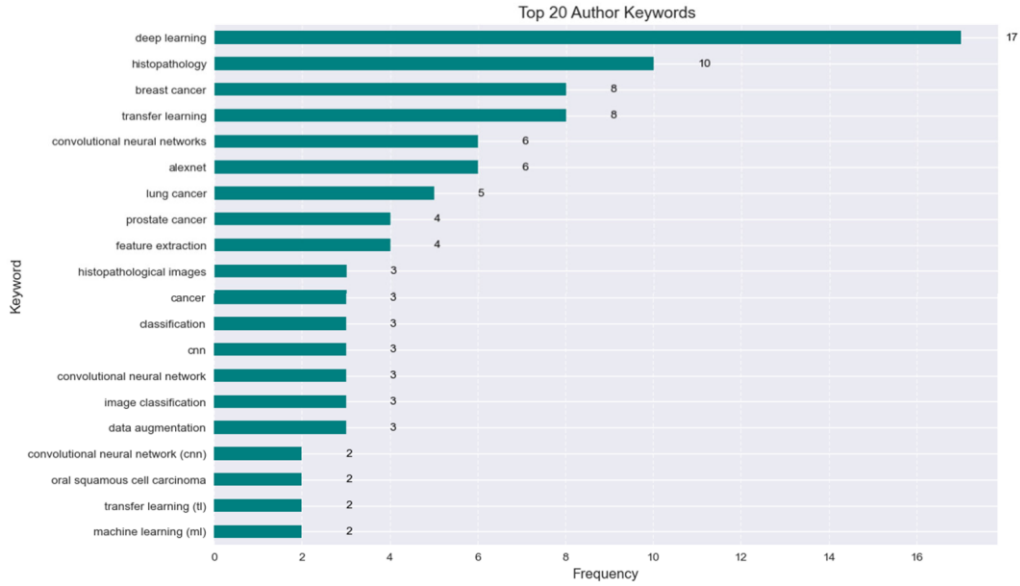


Figure 9 presents the 20 most frequently used author keywords. The analysis reveals that "deep learning," "histopathology," and "breast cancer" are among the most common keywords, reflecting the central themes and topics in the research.

Figure 10: Word Cloud of Keywords

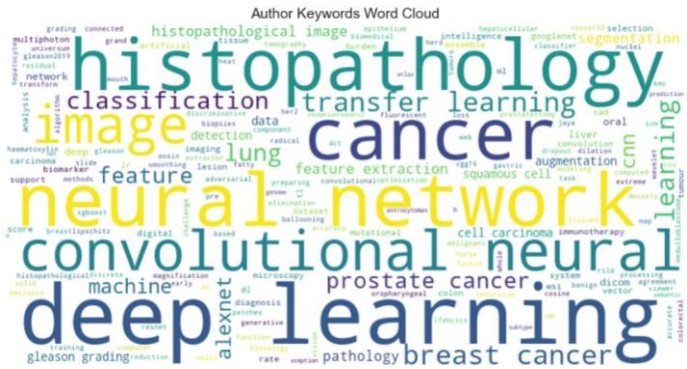


Figure 10 provides a visual representation of the most frequent keywords in the dataset through a word cloud. This visualization offers a quick overview of the key terms and their relative importance within the body of research.

Figure 11: Top 10 Web of Science Categories

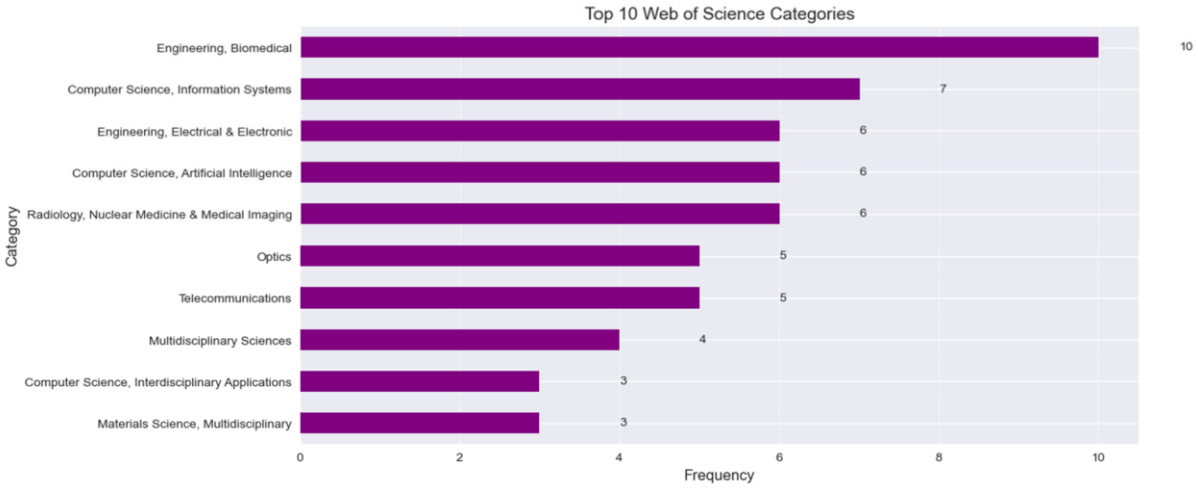


Figure 11 shows the top 10 Web of Science categories assigned to the publications. The most frequent categories include "Engineering, Biomedical" and "Computer Science, Information Systems," indicating the interdisciplinary nature of this research area.

Figure 12: Top 10 Web of Science Research Areas

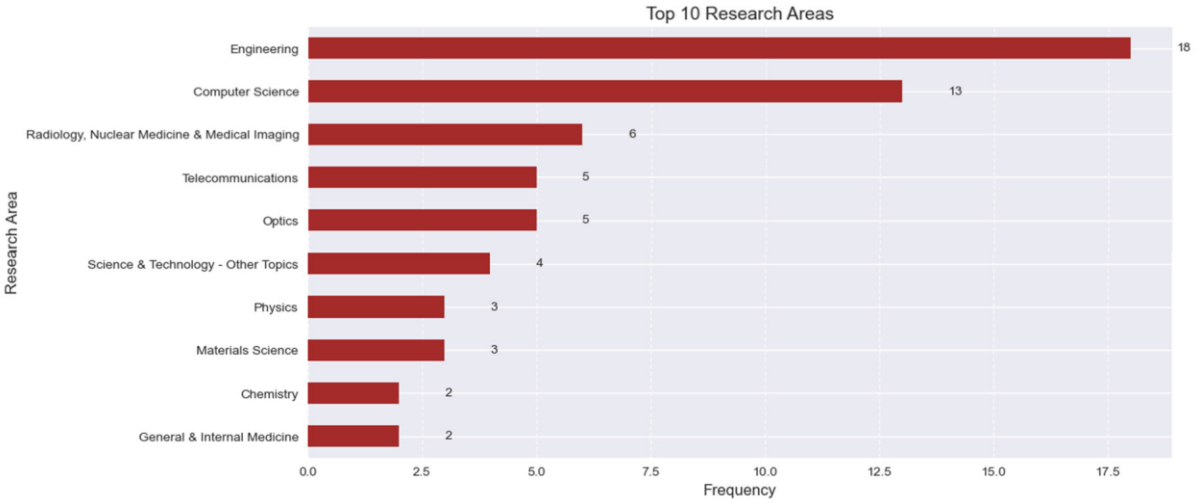


Figure 12 displays the top 10 Web of Science research areas associated with the publications. "Engineering" and "Computer Science" are the predominant research areas, further emphasizing the applied and technological orientation of the research.

Figure 13: Top 10 Funding Agencies

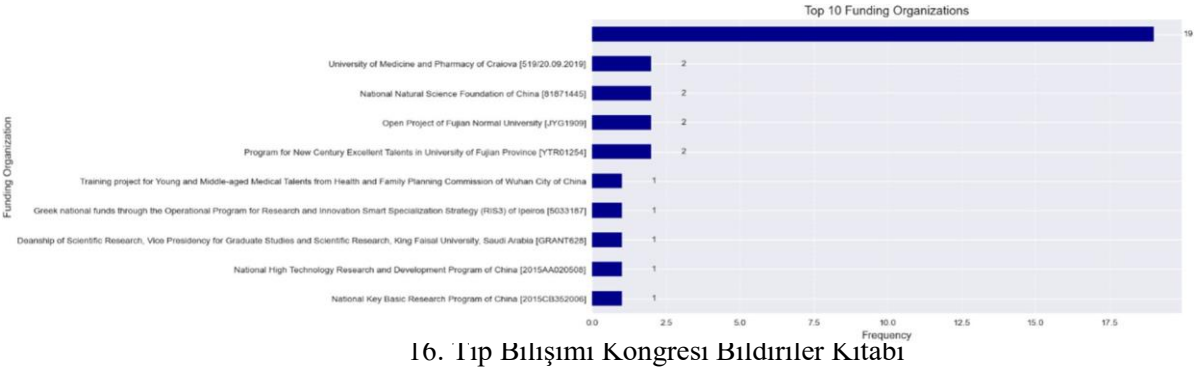


Figure 13 lists the top 10 funding agencies that have supported the research included in the dataset. The funding organizations reflect both national and international sources of support, with the "University of Medicine and Pharmacy of Craiova" being one of the mentioned agencies.

Figure 14: Open Access Status of Publications

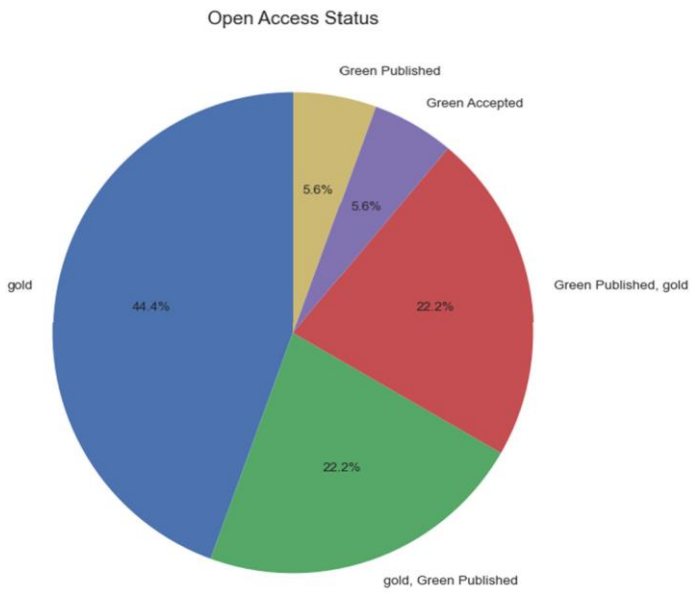


Figure 14 present the Open Access status of the publications. It shows the distribution of different Open Access types among the publications analyzed.

These findings collectively offer a structured and detailed overview of the bibliometric trends, scholarly contributions, and key characteristics of the research focused on histopathology and AlexNet.

Discussion

This bibliometric analysis provides a comprehensive overview of the research landscape pertaining to the application of AlexNet in histopathology. The escalating number of publications in recent years, particularly peaking in 2022, as shown in Figure 1, indicates a burgeoning interest in this interdisciplinary field. This trend is likely reflective of the advancements in deep learning and the increasing availability of large histopathology datasets, which collectively facilitate the development and application of AlexNet-based diagnostic tools.

The predominance of articles as the primary document type underscores the emphasis on original research in this domain. This focus is crucial for establishing a solid foundation of evidence and driving innovation in the field. The analysis of author productivity highlights key contributors who are actively shaping the research direction. Their work not only adds to the body of knowledge but also fosters collaborations and influences future research endeavors, as evidenced by the co-authorship network.

The institutional and geographical distribution of publications reveals the global effort in advancing histopathology research with AlexNet. The leading institutions and countries are likely benefiting from superior resources, funding, and expertise, which contribute to their higher output. The concentration of publications in specific journals indicates the key platforms for disseminating research findings and reflects the journals' scope and influence in the field.

Citation analysis offers valuable insights into the impact and influence of published works. Highly cited papers often represent seminal contributions that have significantly advanced the development or application of AlexNet in histopathology. The analysis of author keywords reveals the primary research themes and topics of interest, with "deep learning," "histopathology," and "breast cancer" being the most prominent. This thematic focus reflects the current emphasis on leveraging AlexNet for cancer diagnosis and classification.

The interdisciplinary nature of this research area is evident in the Web of Science categories and research areas, which exhibit strong representation from the fields of engineering, computer science, and biomedical sciences. The support from various funding agencies underscores the importance of financial backing in driving research

progress. The rise of deep learning as a significant tool for image classification is evident, and its continued evolution is anticipated to greatly benefit image analysis in the coming years [24]. It is also noteworthy that while numerous engineering and scientific papers have been published on cancer image analysis, a significant gap remains in the adoption and integration of these algorithms into clinical practice [25].

The proliferation of computational pathology has been driven by the development of machine learning-based tools for image analysis, which hold the potential to transform pathology workflows [26]. The integration of spatial context with genomic, transcriptomic, and proteomic data, creating an exciting frontier of research known as histo-genomics, is enhancing our understanding of tumor biology and therapeutic efficacy at the individual patient level [5]. The development and application of AI tools by computer engineers and data scientists are essential for improving diagnostic accuracy and identifying novel biomarker approaches for precision oncology [27]. However, the selective adoption of optimal methods for specific clinical problems remains challenging due to the vast array of available techniques and the variability in testing datasets and evaluation metrics [28]. The utilization of AI and machine learning tools in digital pathology offers the potential to uncover subvisual morphometric phenotypes and ultimately enhance patient management [27].

Conclusion

This bibliometric study has comprehensively analyzed the extant literature pertaining to histopathology and AlexNet, providing invaluable insights into the prevailing research trends, key contributors, and influential publications within this field. The findings underscore the burgeoning significance of AlexNet in histopathology, driven by its potential to enhance diagnostic accuracy and efficiency. The meticulous analysis of publication patterns, author collaborations, and keyword trends offers a holistic view of the research landscape, which can meaningfully inform future research directions and collaborative endeavors.

This study offers a comprehensive overview; however, future research endeavors could delve deeper into specific sub-topics or concentrate on the translational impact of AlexNet within clinical practice. Additionally, an examination of the funding landscape and the open access status of publications may provide further insights into the factors that influence research output and accessibility within this dynamic field.

References

[1] C. L. Srinidhi, O. Ciga, and A. L. Martel, “Deep neural network models for computational histopathology: A survey,” Medical Image Analysis, vol. 67. Elsevier BV, p. 101813, Sep. 25, 2020. doi: 10.1016/j.media.2020.101813.

[2] M. Cooper, Z. Ji, and R. G. Krishnan, “Machine learning in computational histopathology: Challenges and opportunities,” Genes Chromosomes and Cancer, vol. 62, no. 9. Wiley, p. 540, Jun. 14, 2023. doi: 10.1002/gcc.23177.

[3] R. Yan et al., “Breast cancer histopathological image classification using a hybrid deep neural network,” Methods, vol. 173, p. 52, Jun. 2019, doi: 10.1016/j.ymeth.2019.06.014.

[4] M. Y. Lu et al., “Federated learning for computational pathology on gigapixel whole slide images,” Medical Image Analysis, vol. 76, p. 102298, Nov. 2021, doi: 10.1016/j.media.2021.102298.

[5] D. Mandair, J. S. Reis-Filho, and A. Ashworth, “Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology,” npj Breast Cancer, vol. 9, no. 1. Nature Portfolio, Apr. 06, 2023. doi: 10.1038/s41523-023-00518-1.

[6] G. Litjens et al., “Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis,” Scientific Reports, vol. 6, no. 1, May 2016, doi: 10.1038/srep26286.

[7] S. Soffer, A. Ben-Cohen, O. Shimon, M. M. Amitai, H. Greenspan, and E. Klang, “Convolutional Neural Networks for Radiologic Images: A Radiologist’s Guide,” Radiology, vol. 290, no. 3. Radiological Society of North America, p. 590, Jan. 29, 2019. doi: 10.1148/radiol.2018180547.

[8] G. K. Thakur, A. Thakur, S. Kulkarni, N. Khan, and S. Khan, “Deep Learning Approaches for Medical Image Analysis and Diagnosis,” Cureus, May 2024, doi: 10.7759/cureus.59507.

[9] S. Fremond, V. H. Koelzer, N. Horeweg, and T. Bosse, “The evolving role of morphology in endometrial cancer diagnostics: From histopathology and molecular testing towards integrative data analysis by deep learning,” Frontiers in Oncology, vol. 12. Frontiers Media, Aug. 18, 2022. doi: 10.3389/fonc.2022.928977.

[10] L. H. S. Vogado, R. Veras, F. Araújo, R. Silva, and K. Aires, “Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification,” Engineering Applications of Artificial Intelligence, vol. 72, p. 415, May 2018, doi: 10.1016/j.engappai.2018.04.024.

[11] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," Computers in Biology and Medicine, vol. 128. Elsevier BV, p. 104129, Nov. 21, 2020. doi: 10.1016/j.compbiomed.2020.104129.

[12] Z. Li et al., "Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1803.05471.

[13] O. Jiménez–del–Toro et al., "Analysis of Histopathology Images," in Elsevier eBooks, Elsevier BV, 2017, p. 281. doi: 10.1016/b978-0-12-812133-7.00010-7.

[14] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagl, "Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology," Computerized Medical Imaging and Graphics, vol. 61, p. 2, Jun. 2017, doi: 10.1016/j.compmedimag.2017.06.001.

[15] A. Arjmand et al., "Training of Deep Convolutional Neural Networks to Identify Critical Liver Alterations in Histopathology Image Samples," Applied Sciences, vol. 10, no. 1, p. 42, Dec. 2019, doi: 10.3390/app10010042.

[16] R. Man, P. Yang, and B. Xu, "Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks," IEEE Access, vol. 8, p. 155362, Jan. 2020, doi: 10.1109/access.2020.3019327.

[17] M.-S. Șerbănescu et al., "Agreement of two pre-trained deep-learning neural networks built with transfer learning with six pathologists on 6000 patches of prostate cancer from Gleason2019 Challenge," Romanian Journal of Morphology and Embryology, vol. 61, no. 2, p. 513, Jan. 2020, doi: 10.47162/rjme.61.2.21.

[18] C. Lu et al., "Feature-driven local cell graph (FLocK): New computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers," Medical Image Analysis, vol. 68, p. 101903, Nov. 2020, doi: 10.1016/j.media.2020.101903.

[19] S. Mehmood et al., "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing," IEEE Access, vol. 10, p. 25657, Jan. 2022, doi: 10.1109/access.2022.3150924.

[20] M. Liu et al., "A Deep Learning Method for Breast Cancer Classification in the Pathology Images," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 10, p. 5025, Jul. 2022, doi: 10.1109/jbhi.2022.3187765.

[21] S. Arooj et al., "Breast Cancer Detection and Classification Empowered With Transfer Learning," Frontiers in Public Health, vol. 10, Jul. 2022, doi: 10.3389/fpubh.2022.924432.

[22] Z. Hameed, B. García-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," Scientific Reports, vol. 12, no. 1, Sep. 2022, doi: 10.1038/s41598-022-19278-2.

[23] P. K. Sethy, A. G. Devi, B. Padhan, S. K. Behera, S. Sreedhar, and K. Das, "Lung cancer histopathological image classification using wavelets and AlexNet," Journal of X-Ray Science and Technology, vol. 31, no. 1, p. 211, Nov. 2022, doi: 10.3233/xst-221301.

[24] S. S. M. Khairi et al., "Deep Learning on Histopathology Images for Breast Cancer Classification: A Bibliometric Analysis," Healthcare, vol. 10, no. 1, p. 10, Dec. 2021, doi: 10.3390/healthcare10010010.

[25] M. S. Hosseini et al., "Computational Pathology: A Survey Review and The Way Forward," Journal of Pathology Informatics. Elsevier BV, p. 100357, Jan. 01, 2024. doi: 10.1016/j.jpi.2023.100357.

[26] M. Á. Berbís et al., "Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade," EBioMedicine, vol. 88, p. 104427, Jan. 2023, doi: 10.1016/j.ebiom.2022.104427.

[27] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology," Nature Reviews Clinical Oncology, vol. 16, no. 11. Nature Portfolio, p. 703, Aug. 09, 2019. doi: 10.1038/s41571-019-0252-y.

[28] N. Petrick et al., "SPIE-AAPM-NCI BreastPathQ challenge: an image analysis challenge for quantitative tumor cellularity assessment in breast cancer histology images following neoadjuvant treatment," Journal of Medical Imaging, vol. 8, no. 3, May 2021, doi: 10.1117/1.jmi.8.3.034501.

Kardiyovasküler Hastalık Tahmini: Keşifsel Veri Analizi ve Modelleme Yaklaşımı

Mehmet Kivrak[1], Handan Duman[2], Hakan Duman[3]

[1]*Recep Tayyip Erdoğan Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Rize, Türkiye*

[2]*Department of Family Medicine, Faculty of Medicine, Recep Tayyip Erdogan University, Rize, Turkiye*

[3]*Department of Cardiology, Faculty of Medicine, Recep Tayyip Erdogan University, Rize, Turkiye*

Özet

Arka Plan: Bu çalışma, istatistiksel ve makine öğrenimi yaklaşımlarını kullanarak Kardiyovasküler Hastalıklar risk faktörlerinin ve bunların öngörücü önemlerinin kapsamlı bir analizini sağlamayı amaçlamaktadır.

Yöntemler: Veri seti 918 hasta kaydından oluşmaktadır ve kardiyovasküler hastalık tahminiyle ilgili klinik ve demografik değişkenleri içermektedir. Kaggle açık erişimli deposundan elde edilmiştir. Veriler 17.01.2025 tarihinde veritabanından çekilmiştir. Bu veri seti, erken evre kalp hastalığı tespitinin yanı sıra öngörücü makine öğrenimi modelleri oluşturmak için de yararlı olacaktır. Çalışma, veri toplama ve ön işlemeyle başlayıp ardından özellik seçimi ve model eğitimiyle devam eden yapılandırılmış bir makine öğrenimi hattını takip etmektedir. Başlangıçta, aykırı değer tespiti Yerel Aykırı Değer Faktörü yöntemi kullanılarak gerçekleştirildi ve sınıf dengesizliği Sentetik Azınlık Aşırı Örnekleme Tekniği kullanılarak ele alındı. Son olarak, makine öğrenimi modelleri eğitildi ve Kardiyovasküler Hastalığın temel öngörücülerini belirlemek için değerlendirildi. Üç temel sınıflandırıcı, yani Rastgele Orman, Destek Vektör Makinesi ve Stokastik Gradyan Artırma model sınıflandırıcıları kullanıldı.

Bulgular: ST segment depresyonu miyokardiyal iskemiyi gösterebilir ve bu da kalp hastalığı ile güçlü ilişkisini açıklar ($r = 0{,}404$, $p<0{,}001$). Cinsiyet ve kalp hastalığı arasında önemli bir ilişki vardır ($p<0{,}001$). Açlık kan şekeri kalp hastalığı ile önemli ölçüde ilişkilidir ($p<0{,}001$). Oldpeak arttıkça, kalp hastalığı geliştirme riski artar ($p<0{,}001$). Stokastik Gradyan Artırma modeli, eğitim aşamasında doğruluk, özgüllük, duyarlılık, F1 puanı ve tüm pozitif ve negatif öngörü değerlerinde en iyi performansı sağladı. Test aşamasında, Stokastik Gradient Boosting modeli doğruluk, özgüllük, F1 puanı ve pozitif öngörü değerleri açısından en iyi performansı sağladı.

Sonuçlar: Bu sonuçlar, klinik ortamlarda birden fazla öngörü faktörünün dahil edilmesinin önemini vurgular ve gelecekteki çalışmaların, kardiyovasküler risk değerlendirmesini iyileştirmek için hem klinik uzmanlık hem de makine öğrenimi modellerinin birleştirilmesinden faydalanabileceğini öne sürer.

Anahtar Sözcükler: Kardiyovasküler Hastalıklar, Keşifsel Veri Analizi, Makine Öğrenimi

Abstract

Background: This study aims to provide a comprehensive analysis of Cardiovascular Diseases risk factors and their predictive significance using statistical and machine learning approaches. The findings may contribute to improved screening strategies and targeted interventions for individuals at high risk of developing Cardiovascular Diseases.

Methods: The dataset consists of 918 patient records and includes clinical and demographic variables relevant to cardiovascular disease prediction. It was obtained from the kaggle open-access repository and represents patients undergoing cardiovascular risk assessment. The data was withdrawn from the database on 17.01.2025. This heart disease dataset is acquired from one of the multispecialty hospitals in India. Over 14 common features make it one of the heart disease datasets available so far for research purposes. This dataset consists of 918 subjects with 12 features. This dataset will be useful for building early-stage heart disease detection as well as for generating predictive machine-learning models. The study follows a structured machine learning pipeline, starting with data collection and preprocessing, followed by feature selection and model training. Initially, outlier detection was performed using the Local Outlier Factor method, and class imbalance was addressed using the Synthetic Minority Over-Sampling Technique. Finally, machine learning models were trained and evaluated to identify key predictors of Cardiovascular Disease. Three basic classifiers, namely Random Forest, Support Vector Machine and, Stochastic Gradient Boosting were used model classifiers.

Results: ST segment depression can indicate myocardial ischemia, which explains its strong association with heart disease ($r = 0.404$, $p<0.001$). There is a significant relationship between gender and heart disease ($p<0.001$). Fasting blood sugar is significantly associated with heart disease ($p<0.001$). As oldpeak increases, the risk of developing heart disease increases ($p<0.001$). The Stochastic Gradient Boosting model provided the best performance in accuracy, specificity, sensitivity, F1 score, and all positive and negative predictive values during the training phase. During the testing phase, the Stochastic Gradient Boosting model provided the best performance in accuracy, specificity, F1 score and positive predictive values.

Conclusions: These results emphasize the importance of incorporating multiple predictive factors in clinical settings and suggest that future studies may benefit from combining both clinical expertise and machine learning models for improved cardiovascular risk assessment.

Keywords: Cardiovascular Diseases, Exploratory Data Analysis, Machine Learning

Giriş

Kardiyovasküler Hastalıklar (KVH), kalp ve kan damarlarını etkileyen geniş bir yelpazedeki durumları kapsayan, dünya çapında önde gelen morbidite ve mortalite nedeni olmaya devam etmektedir (1). Yaş, cinsiyet, hipertansiyon, diyabet, hiperlipidemi ve yaşam tarzı alışkanlıkları dahil olmak üzere çok sayıda risk faktörü, KVH'lerin gelişimi ve ilerlemesiyle ilişkilendirilmiştir (2, 3). Yaş ve cinsiyet uzun zamandır KVH'ler için önemli risk faktörleri olarak kabul edilmektedir. Epidemiyolojik çalışmalar, ilerleyen yaş ile kalp hastalığı yaygınlığının artması arasında iyi belgelenmiş bir ilişki kurmuştur. Bireyler yaşlandıkça, kümülatif olarak risk faktörlerine ve fizyolojik yaşlanma süreçlerine maruz kalırlar ve bu da onları daha yüksek risk altına sokar (4). Benzer şekilde, KVH riskindeki cinsiyet farklılıkları kapsamlı bir şekilde incelenmiştir; erkekler kadınlara göre kalp hastalığı yaygınlığı açısından daha yüksek bir orana sahiptir ve bu durum hormonal düzenleme, yaşam tarzı faktörleri ve genetik yatkınlıktaki farklılıklara bağlanabilir (5). Açlık kan şekeri ve serum kolesterol seviyeleri gibi metabolik parametreler KVH'lerin patofizyolojisinde önemli bir rol oynar. Hipergliseminin endotel disfonksiyonu, vasküler inflamasyon ve artmış aterosklerotik yük ile ilişkili olduğu ve böylece kardiyovasküler riski şiddetlendirdiği öne sürülmüştür (6). Hiperkolesterolemi, KVH'ler için iyi bilinen bir risk faktörüdür. Uzun vadeli çalışmalar, toplam serum kolesterol seviyeleri ile kalp hastalığı arasındaki ilişkiyi açıkça göstermiştir (7). ST-segment anormallikleri ve sol ventrikül hipertrofisi de dahil olmak üzere elektrokardiyografik (EKG) bulgular, kalp hastalığının tanı ve prognoz göstergeleri olarak kullanılmıştır ve bu da EKG'yi muhtemelen KVH'lerin tanısında ilk ve en kritik inceleme haline getirmiştir (8). Benzer şekilde, maksimum kalp hızındaki (MaxHR) azalmalar, muhtemelen bozulmuş otonomik düzenleme ve azalmış kardiyak çıktı nedeniyle daha kötü kardiyovasküler sağlık ile ilişkilendirilmiştir (9).

Makine öğrenimi ve gelişmiş istatistiksel modeller, KVH tahmininin doğruluğunu artırmak için giderek daha fazla kullanılmaktadır. Mevcut çalışma, çeşitli klinik ve biyokimyasal parametreler ile kalp hastalığı riski arasındaki ilişkileri araştırmaktadır. Ayrıca, çalışma farklı tahmin modellerinin performansını değerlendirerek Stokastik Gradyan Artırma (SGB) modelinin doğruluk, özgüllük ve duyarlılık açısından diğerlerinden daha iyi performans gösterdiğini ve kardiyovasküler risk değerlendirmesi için değerli bir araç olarak potansiyelini güçlendirdiğini göstermektedir (10).

Bu çalışmanın amacı, istatistiksel ve makine öğrenmesi yaklaşımlarını kullanarak KVH risk faktörlerinin kapsamlı bir analizini ve bunların öngörücü önemini sunmaktır.

Yöntem

https://www.kaggle.com/datasets?search=cardiovascular'dan veri seti elde edilmiştir. Veri seti toplam 918 gönüllü, 410 sağlıklı ve 508 hasta hasta kaydından oluşmaktadır ve hastalık durumu bilgileriyle birlikte klinik ve biyokimyasal ölçümleri içermektedir. Dinlenme EKG'si, eski zirve, açlık kan basıncı, st_slope kan testi ile belirlenen kolesterol, beyan formu ile elde edilen yaş, cinsiyet göğüs ağrısı tipi ve dinlenme kan basıncı ve fiziksel efor testi ile belirlenen egzersiz anjinası ve maksHR'den oluşmaktadır. Veriler 17.01.2025 tarihinde veritabanından çekilmiştir.

Veri önişleme

Veri ön işleme aşamasında, aykırı değer ve aşırı gözlem analizleri başlangıçta LOF yöntemi kullanılarak yürütüldü (Şekil 1), ardından rastgele orman algoritmasıyla eksik değer analizleri yapıldı. Daha sonra, sınıf dengesizliği sorunu ele alındı. Sınıf dengesizliğini azaltmak için, SMOTE yalnızca eğitim setine uygulandı ve test setinin sentetik verilerle dokunulmadığından emin olundu. Veri seti, sınıf dağılımını korumak için tabakalı örnekleme kullanılarak önce eğitim (%80) ve test (%20) setlerine ayrıldı. Daha sonra SMOTE, azınlık sınıfı için sentetik örnekler oluşturmak amacıyla yalnızca eğitim setine uygulandı ve veri sızıntısı önlendi ve tarafsız model değerlendirmesi sağlandı (Şekil 2).

Açıklanabilir Veri Analizi

Kardiyovasküler hastalık geliştirme riskinin yaşla birlikte arttığını görüyoruz (p<0,001, Şekil 3 ve Tablo 1). Kalp hastalığı olan erkeklerin oranı kadınlardan daha yüksektir (%90,2). Cinsiyet ve kalp hastalığı arasında anlamlı bir ilişki vardır (p<0,001, Şekil 3 ve Tablo 2). Göğüs ağrısı tipi kalp hastalığı ile anlamlı şekilde ilişkilidir (p<0,001, Şekil 1 ve Tablo 2). Kalp hastalığı olan bireylerde asemptomatik bireyler daha yaygındır (%77,2, Tablo 2 ve Şekil 3). Dinlenme kan basıncı arttıkça kalp hastalığı geliştirme riski artar (p<0,001, Şekil 1 ve Tablo 1). Kolesterol azaldıkça kalp hastalığı riski artar (p<0,001, Şekil 3 ve Tablo 1). Açlık kan şekeri (diğer türlü) olan bireylerde kalp hastalığı görülme sıklığı daha yüksektir (%66,5). Açlık kan şekeri kalp hastalığı ile anlamlı şekilde ilişkilidir (p<0.001, Şekil 3 ve Tablo 2). Dinlenme EKG bulguları kalp hastalığı ile anlamlı şekilde ilişkilidir (p=0.004, Şekil 3 ve Tablo 2) ve normal dinlenme EKG'si kalp hastalığı olan bireylerde daha sık görülür (%56.1). MaxHR azaldıkça kalp hastalığı riski artar (p<0.001, Şekil 3 ve Tablo 1). Egzersiz kaynaklı angina kalp hastalığı ile anlamlı şekilde ilişkilidir (p<0.001, Tablo 2 ve Şekil 1). Egzersiz sırasında angina yaşayan bireylerde kalp hastalığı prevalansı anlamlı şekilde daha yüksektir (%62.2). Oldpeak arttıkça kalp hastalığı geliştirme riski artar (p<0.001, Şekil 3 ve Tablo 1). ST segment eğimi kalp hastalığı ile anlamlı şekilde ilişkilidir (p<0.001, Tablo 2 ve Şekil 3). Düz ST segmenti kalp hastalığı olan bireylerde daha yaygındır (%75).

Şekil 4'e göre kalp hastalığı ile sürekli değişkenler arasındaki ilişkiler yorumlandı. En güçlü pozitif korelasyon eski zirve ile gözlendi (r = 0,404, p<0,001). Açlık kan şekeri (r = 0,267, p<0,001) ve yaş (0,282, p<0,001) kalp hastalığı ile orta düzeyde pozitif istatistiksel olarak anlamlı ilişkilere sahiptir. Maksimum kalp hızı ile kalp hastalığı arasında güçlü negatif korelasyon vardır (r = -0,400, p<0,001). Toplam kolesterolün kalp hastalığı ile zayıf negatif korelasyonu vardır (r = -0,233, p=0,004). Dinlenme kan basıncı ile kalp hastalığı arasında çok zayıf bir ilişki vardır (r = 0,107, p = 0,001).

Modelleme

Makine öğrenimi modellerini eğitmeden önce, kalp hastalığı tahminine katkıda bulunan en önemli değişkenleri belirlemek için özellik seçimi yapıldı. Veri seti, tarafsız model değerlendirmesi sağlamak için eğitim (%80) ve test (%20) setlerine ayrıldı. Ek olarak, aşırı uyumu önlemek ve model performansını optimize etmek için eğitim setine katmanlı k katlı çapraz doğrulama (k = 5) uygulandı. Her parametreye grid search optimizasyonu yapıldı (Tablo 3). Özellik seçimi aşaması, ilk olarak, önem sırasına göre en önemli özellikler arasında ST slop up (RF önemli skoru: 0,52), ST slope flat (RF önemli skoru: 0,43), Göğüs ağrısı tipi ASY (RF

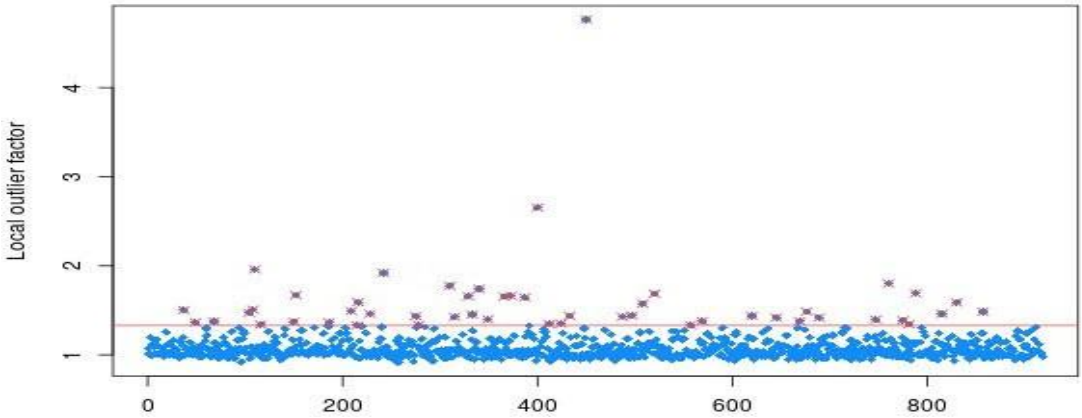
önemli skoru: 0,34), Egzersiz anjinası (RF önemli skoru: 0,26), MaxHR (RF önemli skoru: 0,16) ve Oldpeak (RF önemli skoru: 0,15) yer aldı. Bu değişkenler hedef değişkenle güçlü korelasyonlar gösterdi ve son model eğitim sürecinde kullanıldı (Şekil 5). Her sınıflandırıcı için oluşturulan Alıcı İşletim Karakteristiği (ROC) ve Hassasiyet-Geri Çağırma (PR) eğrilerine göre, SVM modelinde ROC eğrisinin altındaki alan 0,5'ten anlamlı derecede farklıdır ve bu nedenle yeni test iki grup arasında ayrım yapabilmektedir (AUC: 0,668, CI: 0,633-0,703). RF modelinde ROC eğrisinin altındaki alan 0,5'ten anlamlı derecede farklıdır ve bu nedenle yeni test iki grup arasında ayrım yapabilmektedir (AUC: 0,735, CI: 0,703-0,767). SGB modelinde ROC eğrisinin altındaki alan 0,5'ten anlamlı derecede farklıdır ve bu nedenle yeni test iki grup arasında ayrım yapabilmektedir (AUC: 0,745, CI: 0,714-0,787).

Model performansını daha fazla değerlendirmek için, her sınıflandırıcı için Alıcı İşletim Karakteristiği (ROC) ve Hassasiyet-Geri Çağırma (PR) eğrileri oluşturuldu. Eğri Altındaki Alan (AUC), genel model ayrım yeteneğini ölçmek için hesaplandı ve Youden endeksi tarafından belirlendi. Doğruluk, duyarlılık, özgüllük, F1 puanı, pozitif tahmin değeri ve negatif tahmin değeri performans ölçümlerinde kullanıldı. Tablo 4’e göre SGB modeli eğitim aşamasında doğruluk, özgüllük, duyarlılık, F1 puanı ve tüm pozitif ve negatif öngörü değerlerinde en iyi performansı sağladı. Tablo 4'e göre SGB eğitim ve test aşamasında tüm performans metriklerinde en başarılı sonuçları verdi.

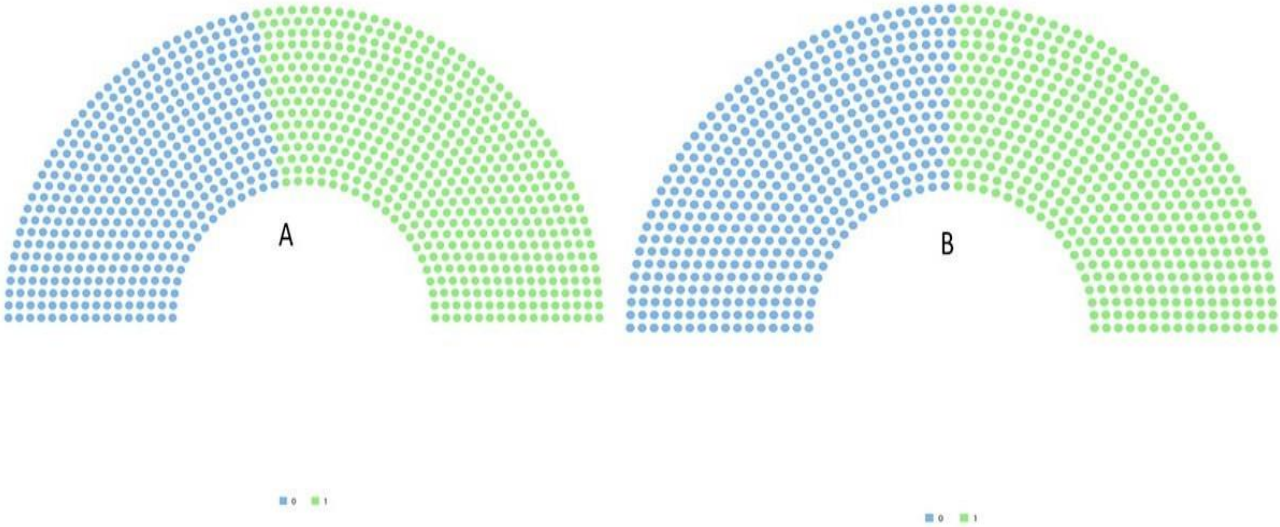
Sonuç ve Tartışma

Bu çalışmada KVH tahmini için EDA ve sonraki modelleme yaklaşımı, alandaki mevcut literatürle uyumlu önemli ilişkileri ve performans karşılaştırmalarını vurgulamaktadır. Bulgular, kalp hastalığına katkıda bulunan temel faktörler ve bu durumu tahmin etmede çeşitli makine öğrenimi modellerinin etkinliği hakkında değerli içgörüler sunmaktadır. Çoğu KVH vakası için riskin yaşla birlikte arttığını görüyoruz. Erkeklerin KVH geliştirme riski kadınlardan daha yüksektir. Yüksek kolesterol seviyelerinin felce, kalp krizine veya hatta kalp yetmezliğine yol açabileceği iyi bilinmektedir.

Makine öğrenimi modelleri ile ilgili olarak, çalışmanın üç sınıflandırıcı RF, SVM, SGB ve diğerlerinin karşılaştırılması, SGB'nin eğitim aşamasında diğer modellerden daha iyi performans gösterdiğini ve doğruluk, özgüllük, duyarlılık, F1 puanı ve tahmini değerler dahil olmak üzere çeşitli ölçütlerde en iyi sınıflandırma performansını sağladığını ortaya koydu. Bu, SGB modellerinin sağlık verileri içindeki karmaşık, doğrusal olmayan ilişkileri ele almada oldukça etkili olduğunu öne süren önceki çalışmaların bulgularını desteklemektedir (11, 12). SGB'nin test aşamasındaki üstün performansı, doğruluk ve pozitif tahmini değerlerdeki dikkate değer başarısıyla, görülmemiş veriler üzerinde doğru tahminler yapmadaki sağlamlığını da vurgulamaktadır. İlginç bir şekilde, SGB çoğu ölçütte liderlik ederken, Random Forest modeli duyarlılık ve negatif tahmini değerlerde en iyi performansı gösterdi ve bu da RF'nin tanımlamada daha hassas olabileceğini göstermektedir.



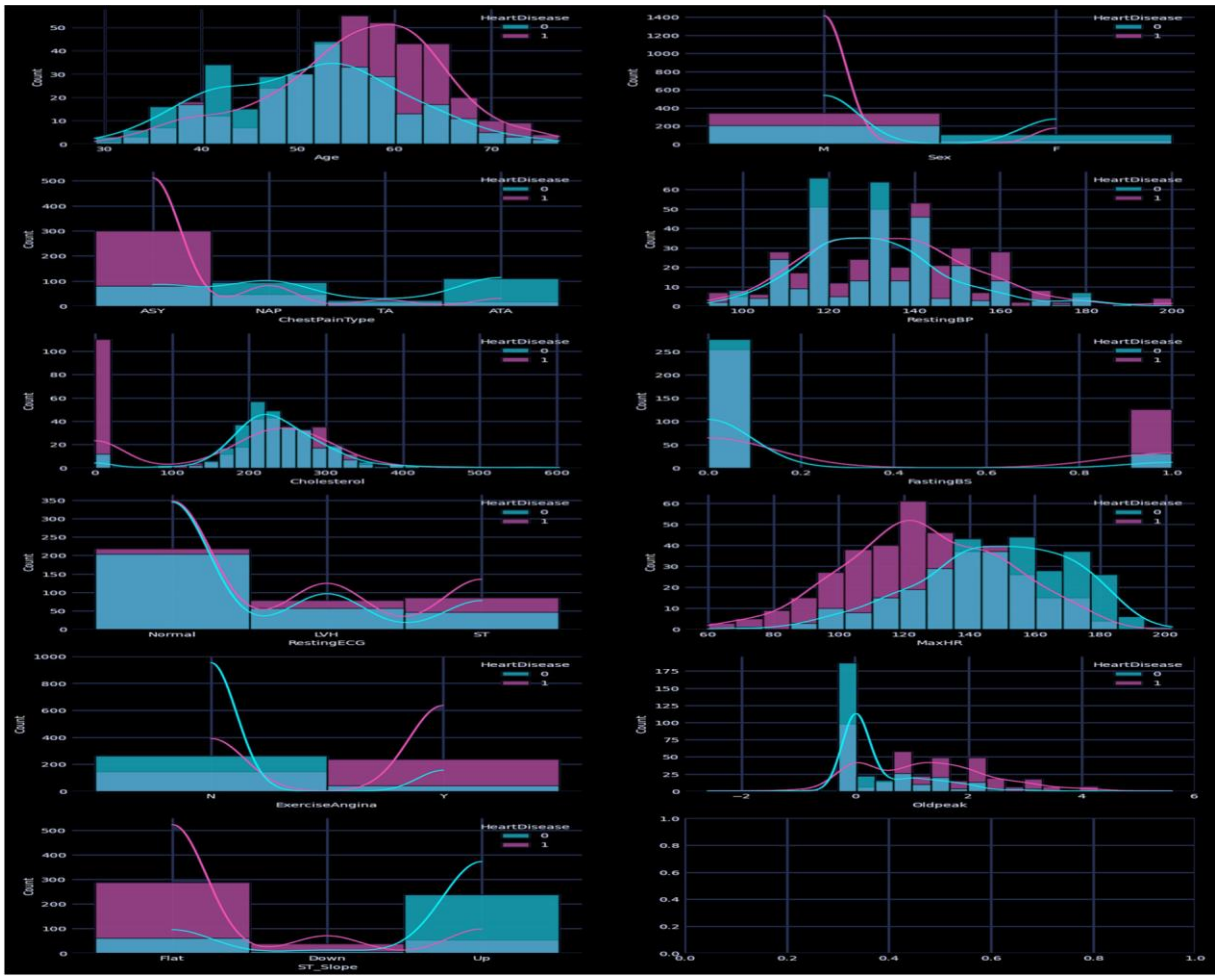
Şekil 1. LOF Aykırı değer analizi



Şekil 2. SMOTE düzleştirme

Tablo1. Sürekli Değişkenlere Göre Kalp Hastalıkları

	Heart Disease			
Variable	**No (n=410)** Median (Min-Max)	**Yes (n=508)**	**Test Statistics***	**p-Value**
Age	51 (28-76)	57 (31-77)	69138	**<0.001**
RestingBP	130 (80-190)	132 (0-200)	90420	**<0.001**
Cholesterol	227 (0-564)	217 (0-603)	87280	**<0.001**
MaxHR	150 (69-202)	126 (60-195)	55191	**<0.001**
Oldpeak	0.0 (-1.1-4.2)	1.2 (-2.6-6.2)	55164	**<0.001**



Şekil 3. Tanımlayıcı Grafikler

Tablo 2. Kategorik Değişkenlere Göre Kalp Hastalıkları

	Heart Disease			
	No (n=410)	**Yes (n=508)**		
Gender	**Descriptive Statistics**		**Test Statistics***	**p-Value**
Female	143 (34.9 %)	50 (9.8 %)	85.6	**<0.001**
Male	267 (65.1 %)	458 (90.2 %)		
FastingBS				
FastingBS > 120 mg/dl	44 (10.7 %)	170 (33.5 %)	65.6	**<0.001**
Otherwise	366 (89.3 %)	338 (66.5 %)		
Chestpaintype				
Typical Angina	26 (6.3 %)	20 (3.9 %)	268	**<0.001**
Atypical Angina	149 (36.3 %)	24 (4.7 %)		
Non-Anginal Pain	131 (32.0 %)	72 (14.2 %)		
Asymptomatic	104 (25.4 %)	392 (77.2 %)		
RestingECG				

LVH	82 (20.0 %)	192 (20.9 %)	10.9	**0.004**
Normal	267 (65.1 %)	285 (56.1 %)		
ST	61 (14.9 %)	117 (23.0 %)		
Exerciseangina				
No	355 (86.6 %)	192 (37.8 %)	224	**<0.001**
Yes	55(13.4 %)	316 (62.2 %)		
Stslope				
Down	14 (3.4 %)	49 (9.6 %)	356	**<0.001**
Flat	79 (19.3 %)	381 (75.0 %)		
Up	317 (77.3 %)	78 (15.4 %)		

Şekil 4. Korelasyon matrisi

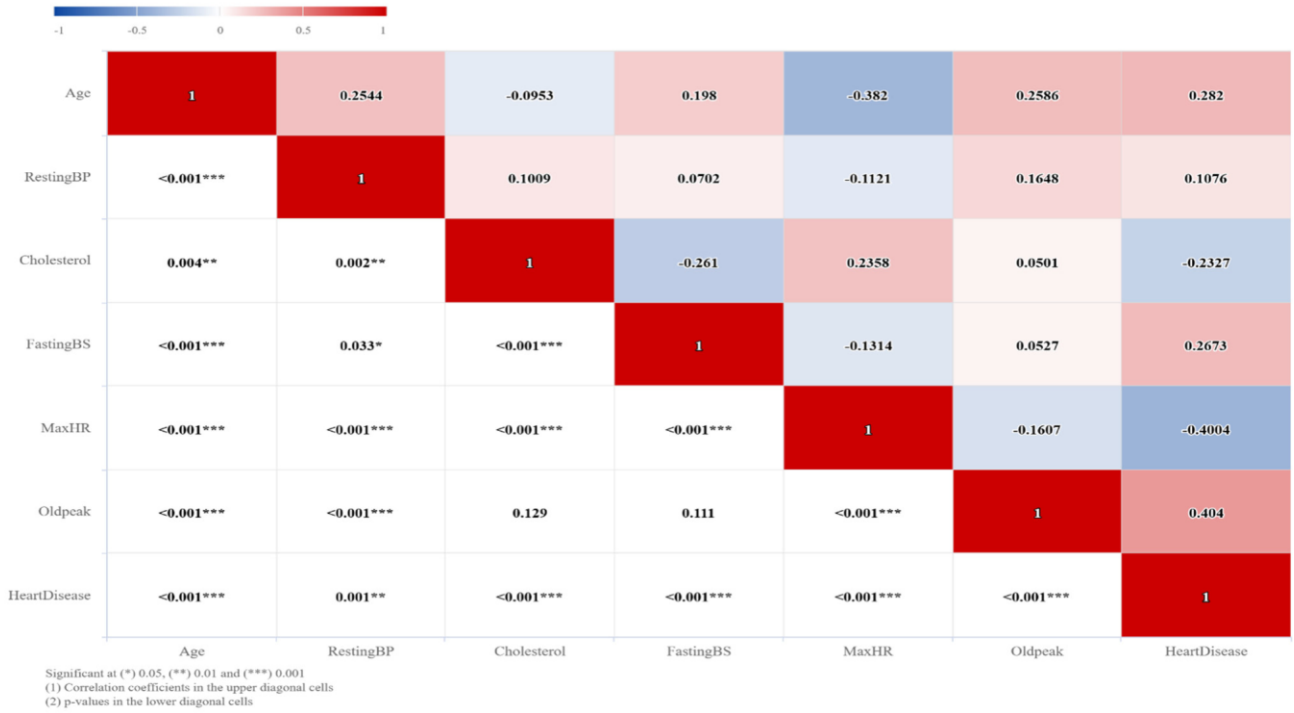
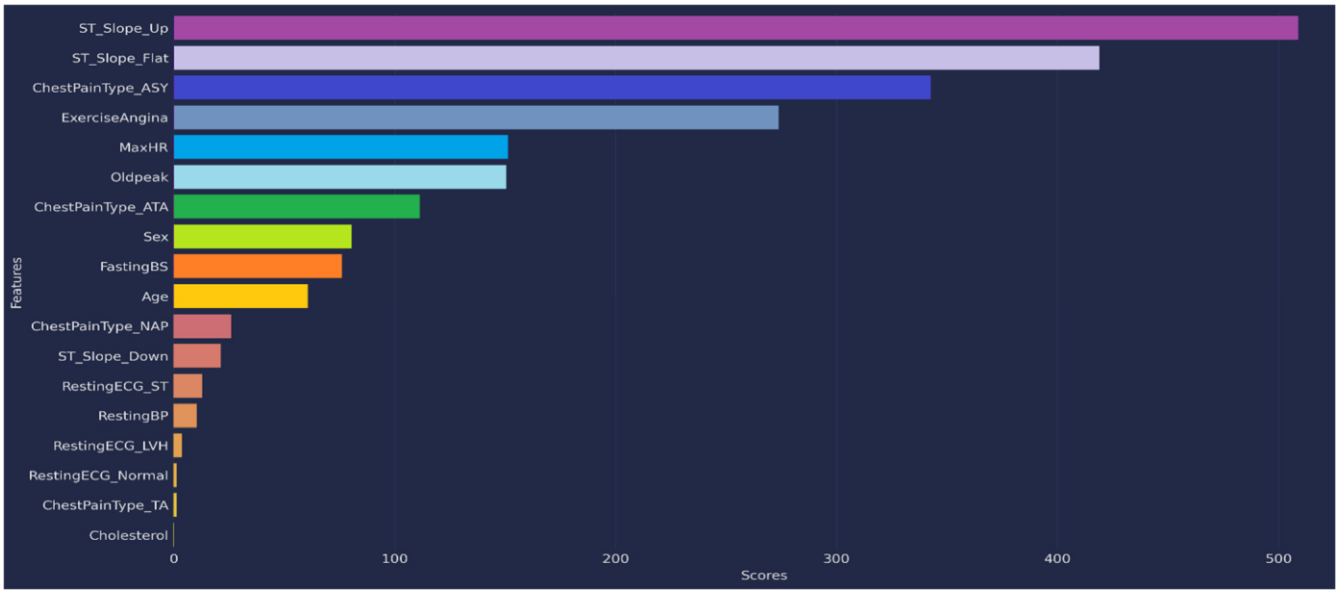


Table 3. Grid Search optimizasyonu

Model	Parameter	Values
RF	bootstrap	(True, False)
	oob_score	(True, False)
	max_depth	None, 2, 3, 4, 5
	n_estimators	50, 100, 200, 300
	min_samples_split	2, 3, 4, 5
	max_leaf_nodes	None, 2, 3, 4
SVM	C	0.1, 1, 10, 100
	gamma	1, 0.1, 0.01, 0.001
	kernel	(rbf, linear)
SGB	reg_alpha	0, 0.2, 0.3
	learning_rate	0.1, 0.01
	min_child_weight	1, 3, 5, 7, 10
	max_depth	1,3,5
	max_features	0.2,0.4
	n_estimators	100, 200, 300, 400, 500, 600
	learning_rate	0.001, 0.005, 0.01, 0.05, 0.1



Şekil 5. Özellik seçimi

Tablo 4. Performans metrikleri

Training Dataset				Testing Dataset			
Metrics	**RF (CI%)**	**SVM (CI%)**	**SGB (CI%)**	**Metrics**	**RF (CI%)**	**SVM (CI%)**	**SGB (CI%)**
Accuracy	0.94	0.89	0.95	Accuracy	0.83	0.86	0.89
	(0.92-0.95)	(0.87-0.92)	(0.94-0.97)		(0.78-0.89)	(0.82-0.95)	(0.85-0.94)
Sensitivity	0.92	0.91	0.94	Sensitivity	0.80	0.87	0.87
	(0.88-0.94)	(0.87-0.92)	(0.91-0.96)		(0.69-0.87)	3(0.92-0.94)	(0.77-0.93)
Specificity	0.95	0.87	0.96	Specificity	0.86	0.85	0.91
	(0.93-0.97)	(0.83-0.93)	(0.93-0.97)		(0.78-0.92)	(0.79-0.92)	(0.84-0.96)
F1 Score	0.93	0.90	0.95	F1 Score	0.81	0.88	0.88
	(0.91-0.95)	(0.84-0.94)	(0.93-0.96)		(0.75-0.86)	(0.83-0.95)	(0.83-0.92)
P.Predict Value	0.94	0.89	0.95	P.Predict Value	0.82	0.88	0.89
	(0.91-0.96)	(0.87-0.92)	(0.92-0.97)		(0.72-0.90)	(0.81-0.95)	(0.80-0.95)
N.Predict Value	0.94	0.89	0.95	N.Predict Value	0.84	0.84	0.89
	(0.91-0.96)	(0.87-0.92)	(0.93-0.97)		(0.75-0.90)	(0.77-0.91)	(0.82-0.95)

Referanslar

1. WHO GS. Global status report on noncommunicable diseases 2010. 2014.

2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation. 2019;139(10):e56-e528. Doi: 10.1161/CIR.0000000000000659.

3. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American college of cardiology. 2020;76(25):2982-3021. Doi: 10.1016/j.jacc.2020.11.010.

4. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circulation research. 2012;110(8):1097-108. Doi: 10.1161/CIRCRESAHA.111.246876.

5. Maas AH, Appelman YE. Gender differences in coronary heart disease. Netherlands Heart Journal. 2010;18:598-603. Doi: 10.1007/s12471-010-0841-y.

6. Atar AI, Altuner TK, Bozbas H, Atar İ, Korkmaz ME. Endothelial dysfunction in prediabetes and diabetes mellitus in patients with normal coronary arteries. Bmc cardiovascular disorders. 2024;24(1):622. Doi: 10.1186/s12872-024-04314-y.

7. Pirillo A, Tokgözoğlu L, Catapano AL. European Lipid Guidelines andCardiovascular Risk Estimation: Current Status and Future Challenges. Current Atherosclerosis Reports. 2024;26(5):133-7. Doi: 10.1007/s11883-024-01194-7.

8. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. Journal of the American College of Cardiology. 2012;60(24):e44-e164. Doi: 10.1016/j.jacc.2012.07.013.

9. Jouven X, Empana J-P, Schwartz PJ, Desnos M, Courbon D, Ducimetière P. Heart-rate profile during exercise as a predictor of sudden death. New England journal of medicine. 2005;352(19):1951-8. Doi: 10.1056/NEJMoa043012.

10. Arfat Y, Mittone G, Esposito R, Cantalupo B, De Ferrari GM, Aldinucci M. Machine learning for cardiology. Minerva cardiology and angiology. 2021;70(1):1-75. Doi: 10.23736/S2724-5683.21.05709-4.

11. Xie Y, Jiang B, Gong E, Li Y, Zhu G, Michel P, et al. Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. American Journal of Roentgenology. 2019;212(1):44-51. Doi: 10.2214/AJR.18.20260.

12. Aljanabi M, Qutqut MH, Hijjawi M. Machine learning classification techniques for heart disease prediction: a review. International Journal of Engineering & Technology. 2018;7(4):5373-9.

Dijital Okur Yazarlık Eğitimi İçin Öğretim Üyelerinin Hazır Bulunuşluk Durumu

Kemal Hakan Gülkesen[1[0000-0002-2477-2481]], Utku Şenol[1[0000-0002-1159-6386]], Volkan Doğru[1[0000-0002-6468-622X]], Yeşim Şenol[1[0000-0002-7842-3041]], Ayşe Gülbin Arıcı[1[0000-0001-6570-4592]]

[1] Akdeniz Üniversitesi Tıp Fak. Sağlıkta Dijital Dönüşüm Çalışma Grubu, Antalya, Türkiye

Özet. Bu çalışma, tıp fakültesi öğretim üyelerinin dijital okuryazarlık düzeylerini değerlendirerek, geliştirilecek eğitici eğitimi modülü için hazır bulunuşluk durumlarını belirlemeyi amaçlamaktadır. Katılımcılara, dijital becerileri ve tutumları ölçmeye yönelik 22 maddelik bir anket Google Forms aracılığıyla çevrim içi olarak uygulanmıştır. Anketten en az 22, en fazla 154 alınabilmektedir. Ankete 170 öğretim üyesi (%47 geri dönüş) katılmıştır. Katılımcıların %55'i erkek, %45'i kadındır; %43'ü dahili, %29'u temel, %28'i cerrahi tıp bilimlerinde görev yapmaktadır. Ortanca anket puanı 122'dir (25.-75. persentil: 111-138). Cinsiyet, yaş, unvan ve daha önce yapay zeka eğitimi alma durumu ile dijital okuryazarlık puanı arasında anlamlı fark bulunmamıştır (p>0,05). Ancak bölüm türüne göre puanlar arasında anlamlı fark saptanmıştır (p=0,003). En yüksek puanlar teknoloji farkındalığı ve temel dijital araç kullanımına, en düşük puanlar ise yapay zeka uygulamalarına yönelik maddelere verilmiştir. Öğretim üyeleri genel olarak dijital araçlara aşinadır; ancak yapay zeka uygulamaları konusunda bilgi ve beceri eksiklikleri gözlenmektedir. Geliştirilecek eğitici eğitim modüllerinde bu alandaki yetersizliklerin giderilmesine öncelik verilmelidir. Hazırlanacak eğitim programı ile bu eksikliklerin giderileceği düşünülmektedir.

Anahtar Kelimeler: Dijital okuryazarlık, Tıp eğitimi, Yapay zeka

Readiness of Faculty Members for Digital Literacy Education

Abstract. This study aims to assess the digital literacy levels of medical faculty members to determine their readiness for a train-the-trainer module to be developed. A 22-item questionnaire to measure digital skills and attitudes was administered online via Google Forms. A minimum of 22 and a maximum of 154 items could be obtained from the questionnaire. 170 faculty members (47% return rate) participated in the survey. Of the participants, 55% were male and 45% were female; 43% worked in internal medicine, 29% in basic medicine and 28% in surgical medicine. The median questionnaire score was 122 (25th-75th percentile: 111-138). There was no significant difference between gender, age, title and previous training in artificial intelligence and digital literacy score (p>0.05). However, a significant difference was found between the scores according to the type of department (p=0.003). The highest scores were given to items related to

technology awareness and basic digital tool use, and the lowest scores were given to items related to artificial intelligence applications. Faculty members are generally familiar with digital tools; however, they lack knowledge and skills in artificial intelligence applications. Priority should be given to overcoming the inadequacies in this area in the training modules to be developed. It is thought that these deficiencies will be overcome with the training program to be prepared.

Keywords: Digital literacy, Medical education, Artificial intelligence

1 Giriş

Günümüzde dijital teknolojilerin hızla gelişmesiyle birlikte, bireylerin bu teknolojileri etkin, bilinçli ve güvenli bir şekilde kullanabilme becerisi büyük önem kazanmıştır. Bilgiye ulaşmanın, iletişim kurmanın ve günlük yaşamı yönetmenin dijital araçlar üzerinden gerçekleştiği bu çağda, dijital okuryazarlık temel beceri olarak kabul edilmektedir. Sadece teknolojiyi kullanabilmek değil, aynı zamanda dijital içerikleri eleştirel bir gözle değerlendirebilmek, doğru bilgiye ulaşabilmek ve dijital ortamlarda etik davranışlar sergileyebilmek dijital okuryazarlığın temel unsurlarını oluşturmaktadır [1].

Geleceğin hekimlerini yetiştiren tıp fakültelerinde öğrencilere dijital çağın gerekliliği olan yetenekleri aktarmak gerekir [2]. Bunu gerçekleştirebilmek için de öğretim üyelerinin dijital okuryazarlık becerilerini edinmesi gerekmektedir. Bu amaçla bu çalışmada, hazırlanacak olan Eğitici eğitimi modülü için öğretim üyelerinin hazır bulunuşluk düzeylerinin saptanması hedeflenmiştir.

2 Yöntem

Çalışmanın amacı, öğrenciler ve öğretim üyeleri arasında dijital okuryazarlık düzeyini artırmak amacıyla bireysel eğitim modülleri geliştirmektir. Bu doğrultuda, yapay zeka ve dijital okuryazarlık konularını kapsayan 22 adet soru içeren bir anket formu hazırlanmıştır. Anket dijital becerileri ve tutumu ölçmektedir. Anketten en az 22, en fazla 154 alınabilmektedir. Hazırlanan anket formu, fakültede görev yapan tüm öğretim üyelerine çevrim içi ortamda (Google Form aracılığıyla) gönderilmiştir.

Çalışmamızın analizinde Jamovi 2.6.26.0 kullanılmıştır. Elde edilen verilerin analizinde Spearman korelasyonu, Mann-Whitney U testi, Kruskal-Wallis varyans analizi ve Dwass-Steel-Critchlow-Fligner ikili karşılaştırma yöntemi kullanılmıştır.

3 Bulgular

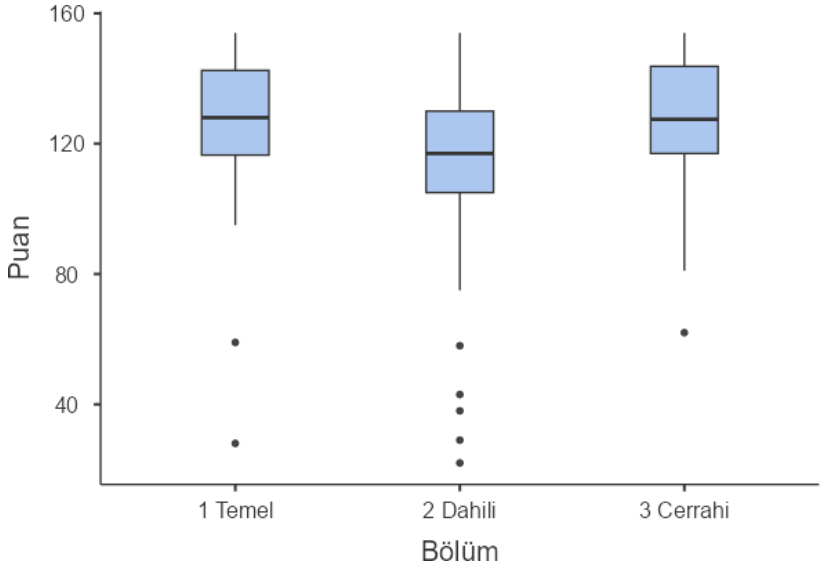
Toplam 300 öğretim üyesine iletilen formu 170 (%47) öğretim üyesi doldurarak çalışmaya katılmıştır. Ankete yanıt veren katılımcıların %55'i erkek, %45'i kadındır. Katılımcıların %29'u temel tıp bilimleri, %28'i cerrahi tıp bilimleri, %43'ü ise dahili tıp

bilimleri alanında görev yapmaktadır. Daha önce tıp eğitiminde yapay zeka kullanımı konusunda eğitim almış olanların oranı %13'tür.

Öğretim üyeleri anketlerden ortanca 122 (111-138, 25.-75. persentil) puan aldı. Yaşla (p=0,160), cinsiyetle (p=0,400), unvanla (p=0,520), daha önce yapay zeka eğitimi alma ile (p=0,086) puan arasında anlamlı bir ilişki görülmedi. Bölüme göre alınan puan ise farklı bulundu (p=0,003, Şekil 1).

Anket yanıtlarında en yüksek puan ortalaması, "Teknolojinin ve teknoloji kullanımının önemini bilirim", "Teknolojik araçları kullanırken iş sağlığı ve güvenliğine dikkat ederim" ve "Sunum hazırlama programlarını (PowerPoint, Prezi vb.) rahatlıkla kullanabilirim" ifadelerine verilmiştir. Buna karşılık en düşük puanlar; "Yapay zeka teknolojisinin bana nasıl yardımcı olacağını bilmiyorum", "Günlük işlerimde bana yardımcı olması için yapay zeka uygulamalarını veya ürünlerini ustalıkla kullanabilirim" ve "Yeni bir yapay zeka uygulamasını veya ürününü kullanmayı öğrenmek benim için genellikle zordur" ifadelerinde gözlemlenmiştir

Şekil 1. Bölüme göre alınan puan. Post-hoc analiz p değerleri; Temel Bilimler-Dahili Bilimler: 0,014, Temel Bilimler -Cerrahi Bilimler: 0,994, Dahili Bilimler -Cerrahi Bilimler: 0,011



4 Tartışma

Bu çalışma Tıp Fakültesi öğretim üyelerinin dijital okuryazarlık düzeylerini belirleyerek, geliştirilecek eğitici eğitim modülüne temel oluşturmayı amaçlamıştır. Elde edilen bulgular, öğretim üyelerinin dijital araçları kullanma ve dijital farkındalık konularında genel olarak yüksek düzeyde bir hazırbulunuşluğa sahip olduklarını göstermektedir.

Buna karşılık, yapay zeka (YZ) teknolojileri ile ilgili ifadelere verilen düşük puanlar dikkat çekicidir. Bu durum, öğretim üyelerinin dijital teknolojilerin daha ileri düzey uygulamaları olan yapay zeka alanında yeterince bilgi sahibi olmadıklarını ya da bu araçları etkin kullanma konusunda özgüven eksikliği yaşadıklarını göstermektedir. Nitekim, yapay zeka teknolojisinin nasıl yardımcı olabileceği konusundaki belirsizlik ve yeni uygulamaları öğrenmede yaşanan zorluklar, bu alandaki bilgi ve deneyim eksikliğini işaret etmektedir. Bu bulgu, literatürde dijital okuryazarlığın sadece teknik becerilerle sınırlı kalmaması, aynı zamanda dijital adaptasyon ve eleştirel dijital düşünme ile desteklenmesi gerektiğini belirten çalışmalarla örtüşmektedir [3,4].

Çalışmada elde edilen bir diğer önemli bulgu, yaş, cinsiyet, unvan ve daha önce YZ eğitimi alma gibi değişkenlerin dijital okuryazarlık puanlarını anlamlı düzeyde etkilememesidir. Ancak, öğretim üyelerinin görev yaptıkları bölümler arasında anlamlı farklar gözlemlenmiş ve özellikle bazı bölümlerde dijital okuryazarlık puanlarının daha düşük olduğu görülmüştür. Bu durum, farklı uzmanlık alanlarının dijital teknolojilere yaklaşımı ve uygulama gereklilikleri bakımından çeşitlilik gösterebileceğini ortaya koymaktadır.

Bu çalışmanın bazı sınırlılıkları mevcuttur. İlk olarak, veri toplama aracı olarak kullanılan anketin özbildirim temelli olması, katılımcıların dijital yeterliliklerini olduğundan daha yüksek ya da düşük değerlendirme ihtimalini beraberinde getirebilir. İkinci olarak, katılımcıların yalnızca %47'sinin ankete yanıt vermiş olması da yanıtlayıcı yanlılığı riskini artırmaktadır.

Gelecekteki çalışmalarda, öğretim üyelerine yönelik dijital okuryazarlık ve yapay zeka uygulamalarına dair müdahale programları geliştirilerek bu programların etkililiği değerlendirilebilir. Ayrıca, niteliksel veri toplama yöntemleriyle öğretim üyelerinin dijital araçlar ve yapay zeka kullanımına yönelik algıları, engelleri ve ihtiyaçları daha derinlemesine anlaşılabilir. Tıp eğitiminde dijital dönüşümün sağlıklı bir şekilde yürütülebilmesi için, öğretim üyelerinin sürekli gelişimini destekleyen kurumsal stratejiler önem taşımaktadır.

Bu bulgular, öğretim üyelerinin genel olarak teknolojiye karşı olumlu tutum geliştirmiş olduklarını, ancak yapay zeka teknolojileri konusunda bilgi ve deneyim eksikliklerinin bulunduğunu göstermektedir. Hazırlanacak eğitim programı ile bu eksikliklerin giderileceği düşünülmektedir.

Kaynaklar

1. Arias López MDP, Ong BA, Borrat Frigola X, Fernández AL, Hicklent RS, Obeles AJT, Rocimo AM, Celi LA. Digital literacy as a new determinant of health: A scoping review. PLOS Digit Health. 2023 Oct 12;2(10):e0000279.
2. Aydınlar A, Mavi A, Kütükçü E, Kırımlı EE, Alış D, Akın A, Altıntaş L. Awareness and level of digital literacy among students receiving health-based education. BMC Med Educ. 2024 Jan 8;24(1):38. Author, F.: Article title. Journal 2(5), 99–110 (2016).
3. Ng, W. (2012). Can we teach digital natives digital literacy? Computers & Education, 59(3), 1065–1078.
4. Tang, Y., & Chaw, L. Y. (2016). Digital literacy: A prerequisite for effective learning in a blended learning environment? The Electronic Journal of e-Learning, 14(1), 54–65.

Fonksiyonel Manyetik Rezonans Görüntüleme ile Ruhsal Hastalıkların Yapay Zeka Kullanılarak Tahmin Edilmesi: Bibliyometrik Bir Analiz

Cansu GÜNDOĞAN TOKER[1][0000-0001-8152-4735]

[1] Afyonkarahisar Sağlık Bilimleri Üniversitesi, Türkiye
cansu.gundogan@afsu.edu.tr

Özet. Bu çalışmanın amacı, fonksiyonel manyetik rezonans görüntüleme(fMRG) verilerinin yapay zeka (YZ) aracılığıyla ruhsal hastalıkların tanı ve tedavisinde kullanıldığı bilimsel yayınların bibliyometrik analizini yapmaktır. Web of Science veri tabanında yayımlanmış fMRG, yapay zeka ve ruhsal hastalıklar ile ilişkili yayınlar, "topic search" yöntemiyle taranmıştır. Web of Science veri tabanında "artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" AND "functional magnetic resonance imaging" OR "fMRI" OR "neuroimaging" AND "mental disorders" OR "psychiatric disorders" OR "depression" OR "schizophrenia" OR "bipolar disorder" OR "anxiety disorders" OR "obsessive compulsive disorders" anahtar kelimeleri ile belirlenen 1051 makale incelenmiştir. Elde edilen veriler VOSviewer ile analiz edilerek anahtar kelime, yazarlar arası işbirliği, en çok yayın yapılan dergiler, yıllara göre yayın sayıları ve ülke/kurum dağılımı görselleştirilmiştir. En sık tekrar eden anahtar kelimeler "makine öğrenmesi ve şizofreni" olmuştur. En çok yayın yapan yazarlar "Calhoun Vince D, Koutsouleris Nicolas", en aktif ülkeler "ABD, Çin ve İngiltere" olarak belirlenmiştir. En çok yayın yapılan dergilerde ise ilk sıraları sırasıyla "Neuoroimage, Human Brain Mapping" almıştır. Yayın sayılarında özellikle 2010'lu yıllara kadar tek haneli sayılar, 2020'li yıllara kadar iki haneli sayılar ve 2020 sonrası üç haneli rakamlar dikkat çekici olup kademeli olarak bir artış gözlenmiştir.

fMRI ve yapay zeka, ruhsal hastalıkların tanı ve tedavisinde önemli bir potansiyele sahiptir. fMRI ve yapay zeka tabanlı birleşimle, ruhsal hastalıkların daha erken ve doğru teşhisi yönünde ilerlemeler kaydedilmektedir. Bibliyometrik bulgular, bu alanda disiplinlerarası işbirliğinin giderek arttığını ve bilimsel ilginin yoğunlaştığını göstermektedir. Gelecekte daha fazla veri paylaşımı ve model karşılaştırmaları ile bu alanın klinik uygulamalara entegrasyonu güçlenebilir.

Anahtar Kelimeler: fMRI, yapay zeka, makine öğrenimi, ruhsal hastalıklar, bibliyometrik analiz

Abstract. The aim of this study is to conduct a bibliometric analysis of scientific publications in which functional magnetic resonance imaging (fMRI) data are used in the diagnosis and treatment of mental illnesses through artificial intelligence (AI). The publications related to fMRI, artificial intelligence and mental illnesses published in the Web of Science database were searched using the "topic search" method. In the Web of Science database, 1051 articles identified with the keywords "artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" AND "functional magnetic resonance imaging" OR "fMRI" OR "neuroimaging" AND "mental disorders" OR "psychiatric disorders" OR "depression" OR "schizophrenia" OR "bipolar disorder" OR "anxiety disorders" OR "obsessive compulsive disorders" were examined. The data obtained were analyzed with VOSviewer and the distribution of keywords, collaboration between authors, journals with the most publications, number of publications by year and country/institution were visualized. The most frequently recurring keywords were "machine learning and schizophrenia". The most published authors are "Calhoun Vince D, Koutsouleris Nicolas" and the most active countries are "USA, China and UK". In the journals with the most publications, "Neuoroimage, Human Brain Mapping" took the first places respectively. In the number of publications, single-digit numbers until the 2010s, double-digit numbers until the 2020s and triple-digit numbers after 2020 are remarkable and a gradual increase has been observed. fMRI and artificial intelligence have significant potential in the early diagnosis and classification of mental disorders. With the combination of fMRI and artificial intelligence, progress is being made towards earlier and more accurate diagnosis of mental disorders. Bibliometric findings show that interdisciplinary collaboration and scientific interest in this field are increasing. In the future, the integration of this field into clinical practice may be strengthened with more data sharing and model comparisons.

Keywords: fMRI, artificial intelligence, machine learning, mental illnesses, bibliometric analysis

1 Giriş

Psikiyatrik hastalıkların insan beyni üzerindeki etkilerinin nörogörüntüleme ve Fonksiyonel Manyetik Rezonans Görüntüleme(fMRG) teknikleriyle incelenmesine yönelik çalışmalar, özellikle yapay zeka uygulamalarıyla birlikte psikiyatride tanı ve prognoza yardımcı olmak için veri temelli tahmine dayalı bir yaklaşım sağlamaktadır. Bu alandaki çalışmaların bibliyometrik analizlerle değerlendirilmesi sağlanarak mevcut boşlukların neler olduğunun belirlenmesi[1]ve geleceğe yönelik yapılacak çalışmalara yol gösterici bir nitelik sunması[2] ve sağlık bilişimi alanında da literatürü daha derinlemesine anlamaya ve incelemeye fayda sağlamaktadır.[3,4] Bu kapsam çerçevesinde fMRG verilerinin yapay zeka (YZ) aracılığıyla ruhsal hastalıkların tanı ve tedavisinde kullanıldığı bilimsel yayınların bibliyometrik analizini yapmak bu çalışmanın amacını oluşturmaktadır.

2 Yöntem

2.1 Veri Toplama

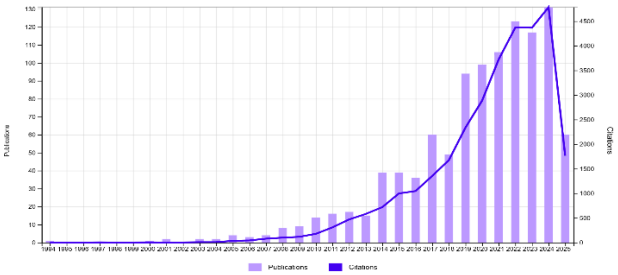
Bibliyometrik analiz yöntemi kullanarak gerçekleştirilen bu çalışmada literatür analizi kapsamında Web of Science(WOS) veri tabanı kullanılmıştır. Çalışmanın evrenini fMRG verilerinin YZ aracılığıyla ruhsal hastalıkların tanı ve tedavisinde kullanıldığı bilimsel yayınlar oluşturmaktadır. Tarih kısıtlaması olmadan yayın türü araştırma makalesi, yayın dili İngilizce, yayın indeksi Science Citation Index Expanded olan çalışmalar dahil edilmiştir. Dahil edilen bu yayınların alınma stratejisi TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND TS=("functional magnetic resonance imaging" OR "fMRI" OR "neuroimaging") AND TS=("mental disorders" OR "psychiatric disorders" OR "depression" OR "schizophrenia" OR "bipolar disorder" OR "anxiety disorders" OR "obsessive compulsive disorders") bu kelimelerden oluşmakta ve toplamda 1051 yayın incelenmiştir.

2.2 Veri Analizi

VOSviewer, bibliyometrik ağları oluşturma ve görselleştirme amacıyla 2010 yılında Nees Jan van Eck ve Ludo Waltman (Leiden Üniversitesi) tarafından yayımlandı.[5] VOSviewer bilimsel yayınların analizlerini yapmak için tasarlanmış ve ortak yazarlığı, ortak oluşumu, atıfları, bibliyografik eşleşmeyi ve ortak atıf bağlantılarını araştırmaktadır.[6] Elde edilen veriler VOSviewer ile analiz edilerek anahtar kelime, yazarlar arası işbirliği, en çok yayın yapılan dergiler, yıllara göre yayın sayıları ve ülke/kurum dağılımı çerçevsinde değerlendirilmiş ve görselleştirilmiştir.

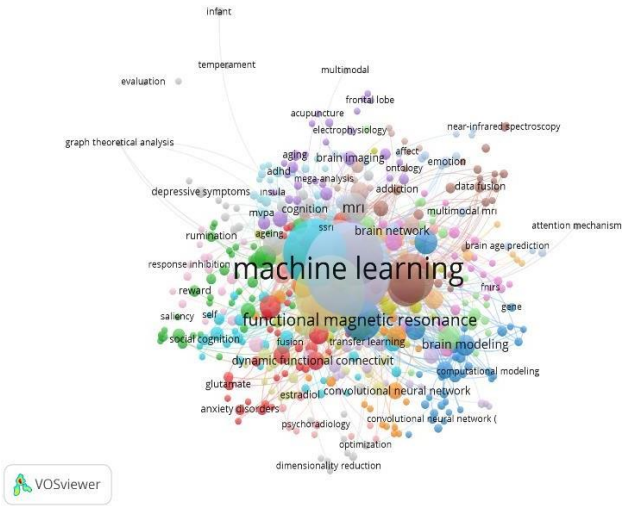
3 Bulgular

Bu araştırma kapsamında 1051 makale incelenmiştir. İncelenen yayın tarihleri sayılarında özellikle 2010'lu yıllara kadar tek haneli sayılar, 2020'li yıllara kadar iki haneli sayılar ve 2020 sonrası üç haneli rakamlar dikkat çekici olup kademeli olarak bir artış gözlenmiştir.Yıllara göre yayın ve atıf sayıları aşağıdaki grafikte yer almaktadır.(Şekil 1)



Şekil. 1.Yıllara göre Yayın Sayıları ve Atıf Sayıları

En çok kullanılan anahtar kelimeler arasında makine öğrenmesi, şizofreni, fMRG, depresyon, nörogörüntüleme” yer almaktadır. Bu anahtar kelimelerin VOSviewer yazılımı kullanılarak analizi sonucunda elde edilen dağılımı aşağıda yer almaktadır.(Şekil 2)

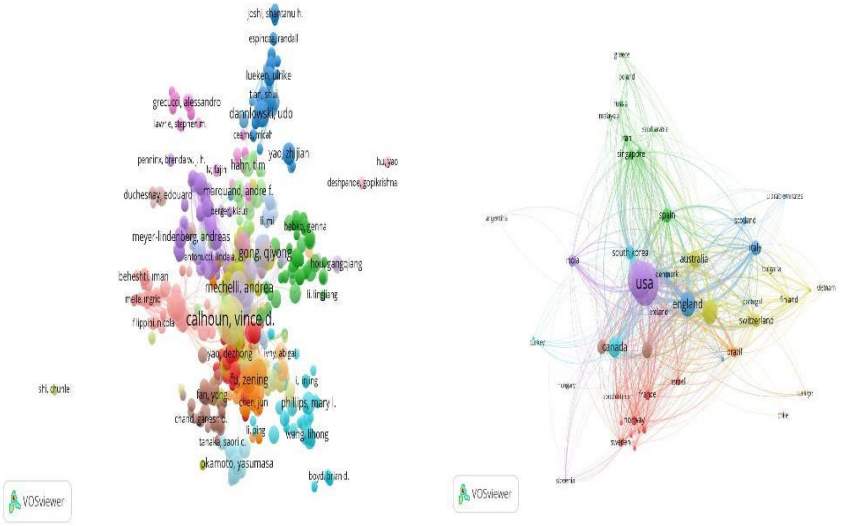


Şekil. 2. En Çok Kullanılan Anahtar Kelimelerin Vosviewer Haritası

En çok yayın yapan yazarlar “Calhoun Vince D, Koutsouleris Nicolas, Mecheli Andrea” olarak sıralanmaktadır.(Şekil 3)

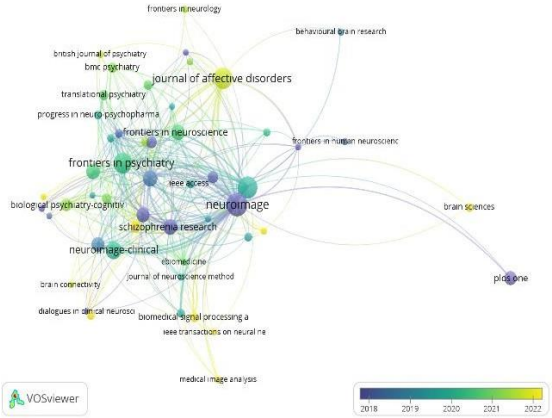
Şekil.3. En Çok Yayın Yapan Yazarların Şematik Gösterimi

En cok atıf alan yazarlar(Şekil 4A) ve bu yazarların ülkeleri “ABD, Çin ve İngiltere” olarak belirlenmiştir.(Şekil 4B)



Şekil. 4. En Çok Atıf Alan Yazarların VOSviewer Görselleştirme Haritası(A) En Çok Atıf Alan Yazarların Ülkelerinin VOSviewer Görselleştirme Haritası(B)

En çok yayın yapılan dergilerde ise ilk sıraları sırasıyla “Neuoroimage, Human Brain Mapping, Frontiers in Psychiatry” dergileri almıştır.(Şekil 5)



Şekil. 5. En Çok Yayın Yapılan Dergilerin VOSviewer Görselleştirme Haritası

4 Tartışma ve Sonuç

Bu çalışmayla fMRG ve yapay zeka tabanlı birleşimle, ruhsal hastalıkların tanısı ve teşhisi yönünde bilimsel ilerlemenin bibliyometrik bulgular üzerinden analizi sağlanmıştır. Çalışmada tek bir analiz yöntemi kullanılması bir kısıtlılık oluşturmaktadır. Ayrıca tek bir veri tabanı kullanılarak analizin sağlanması da tartışalabilecek bir diğer noktadır. görülmüştür. Yayınların 2020 ve sonrasında hızla artış göstermesi aslında bu alana olan ilginin bir göstergesidir. Aynı zamanda gelişen teknolojiyle birlikte sinir ağları, makine öğrenmesi, nörobilim, yapay zeka ve görüntü işleme gibi alanlarda son yıllarda hızla benimsenmesi[7] ve bu gelişmenin ruh sağlığı alanını da etkilemesi şeklinde değerlendirilebilir. Bu analiz, ruh sağlığı alanındaki fMRG verileriyle YZ araştırmalarına genel bir bakış sunarak bilişim profesyonelleri ile ruh sağlığı profesyonellerinin iş birliğinin önemini de ortaya koymuştur Çalışma sonuçları araştırmacılara, ruhsal hastalıkların yapay zeka ve fMRG temelli yansımalarını değerlendirme ve çalışmalarını planlama konusunda yol gösterici olabilir.

References

1. Atılgan, E., & Uslu, E. (2024). Ruh Sağlığı Alanında Yapay Zeka Araştırmaları: Bibliyometrik Bir Değerlendirme. Osmangazi Tıp Dergisi, 46(4), 627-636.
2. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of business research, 133, 285-296.
3. Lazarides, M. K., Lazaridou, I. Z., & Papanas, N. (2023). Bibliometric analysis: bridging informatics with science. The international journal of lower extremity wounds, 15347346231153538.
4. Sharma, C. M., & Chariar, V. M. (2024). Diagnosis of mental disorders using machine learning: Literature review and bibliometric mapping from 2012 to 2023. Heliyon.
5. Van Eck, N., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics, 84(2), 523-538.
6. Arruda, H., Silva, E. R., Lessa, M., Proença Jr, D., & Bartholo, R. (2022). VOSviewer and bibliometrix. Journal of the Medical Library Association: JMLA, 110(3), 392.
7. Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020). Artificial intelligence in health care: bibliometric analysis. Journal of medical Internet research, 22(7), e18228.

Fair and Efficient Transmission of Multi-Omics Data via NOMA: A Hybrid Machine Learning Approach using Random Forest and KNN

Seda KIRTAY[1*][0000-0003-2415-9131], Veysel Gökhan BOCEKCI[2][0000-0003-4559-7173], Kazım YILDIZ [3][0000-0001-6999-1410] and Murat KOCAK[4][0000-0001-6510-3666]

[1*,2] Marmara University, Faculty of Technology, Electrical and Electronics Engineering, Department of Telecommunication, Istanbul 34840, Turkiye
[3] Marmara University, Faculty of Technology, Computer Engineering, Department of Computer Sciences, Istanbul 34840, Turkiye
[4] Baskent University, Faculty of Medicine, Department of Medical Informatics, Ankara, Turkiye.

seda.kirtay@marmara.edu.tr

Abstract. Multi-omics data are produced by merging many biological data types, including proteomics, metabolomics, and genomes. These data help us gain a more thorough understanding of people's biological states and are essential for the creation of individualized healthcare treatment plans and diagnostic techniques. However, a major obstacle to the security and accuracy of health data is the safe and effective transfer of multi-omics data. For the efficient transfer of multi-omics data in wire-free health systems, a method based on non-orthogonal multiple access (NOMA) has been suggested in this paper. NOMA is a technique that uses the same frequency band for several users to share transmission power in an equitable manner. Three distinct users' distances from the base station were used to determine the channel gains for the study, and power allocation was done in accordance with these gains. The inverse channel gain approach has been used to optimize power allocation, while machine learning techniques (K-Nearest Neighbors and Random Forest) have been used to predict power allocation. In addition, the Successive Interference Cancellation (SIC) algorithm has been used to reduce user interference. The results reveal that the proposed NOMA-based approach is very reliable and efficient for multi-omics data transmission, with digital success rates surpassing 95%. Furthermore, with a short processing time (2.18 seconds) and low memory utilization (314.24 MB → 304.49 MB), the suggested system has shown considerable computational efficiency. This study describes an important strategy for increasing the integration of multi-omics data into health information systems.

Keywords: Fair Power Allocation, Machine Learning, Multi-omics Data, NOMA, Non-Orthogonal Multiple Access, Telemedicine

1 Introduction

The incorporation of multi-omics data has been a crucial development in medical research and healthcare in recent years. To provide a more comprehensive understanding of health and disease, multi-omics includes the thorough investigation of multiple biological levels, such as proteomics, metabolomics, and genomes. Researchers can find intricate connections between genetic, protein expression, and metabolic pathways by integrating multiple data types, which will enable more individualized and accurate treatment plans [1]. In this regard, the effective transmission and interpretation of multi-omics data, especially in remote healthcare settings, requires the use of cutting-edge communication technology [2].

With the growing reliance on telemedicine and health information systems, efficient and dependable communication mechanisms are critical for ensuring the secure and real-time transfer of sensitive health data, such as multi-omics data [3]. NOMA is a cutting-edge technology that has received a lot of attention in the field of communication systems. It allows you to share resources efficiently in environments with several users [4]. Communication performance is maximized by separating user signals using techniques like power allocation and SIC (Successful Interference Cancellation). NOMA offers both low latency and high efficiency with equitable power allocation, particularly in situations with a high user density and fluctuating channel circumstances [5,6]. NOMA provides high throughput while reducing latency, making it excellent for data-intensive applications such as telemedicine.

In the context of medical informatics and healthcare communication, this research investigates the relationship between multi-omics data transmission and NOMA-based communication systems. Using NOMA, we suggest a unique power allocation strategy that is fair to users with different channel circumstances. We provide a model that predicts the best power allocation for efficient communication by combining machine learning algorithms like Random Forest and K-Nearest Neighbors. This ensures good data integrity and little interference during multi-omics data transmission.

The purpose of this study is to improve the efficiency of multi-omics data allocation in telemedicine systems, ultimately leading to better healthcare outcomes through improved data-driven decision-making. The suggested strategy has the potential to transform healthcare communications by providing more consistent, accurate, and rapid access to vital health information across many clinical settings.

2 Methodology

2.1 Data Generation

Synthetic multi-omics datasets were created for three users, each representing genomic (10 binary features), proteomic (8 continuous features, 0-100 range), and metabolomic (6 continuous features, 0-50 range) data. These features were organized into hierarchical matrices identified by user.

2.2 Power Allocation via Hybrid AI (Random Forest and K-Nearest Neighbors)

Users (patients with multi-omics data) were stationed at 50, 150, and 400 kilometers from the base station. Corresponding channel gains were calculated as inversely proportional to distance. These gains were used to determine initial power levels, which were then normalized to assure total transmission power conservation. To predict power allocations, a hybrid machine learning model was trained using user distances and channel gains, combining the Random Forest

Regressor and the K-Nearest Neighbors Regressor. Predictions from both models were averaged and re-normalized to ensure compliance with NOMA restrictions.

2.3 System Performance under SIC

Transmission performance was simulated with SNR (Signal-to-Noise Ratio)'s ranging from -20 dB to 150 dB. For every SNR level:

- Optimal SNR per user was calculated after SIC decoding. The Bit Error Rate (BER) was calculated using QPSK modulation assumptions. The throughput was computed using $\log_2(1+SNR)$ and capped at 1 bps to simulate channel saturation.
- Accuracy has been expressed as (1-BER).
- Graphs were created for BER, throughput, and accuracy versus SNR.

3 Result and Discussion

The simulation results show how successful the suggested NOMA-based transmission model is for multi-omics data in wireless healthcare systems. The usage of hybrid machine learning models (Random Forest and K-Nearest Neighbors) allowed for accurate prediction of equitable power allocation across users under various channel circumstances. Notably, the BER stayed below 5% for all users across a wide SNR range, and data delivery accuracy exceeded 95%, demonstrating the model's durability. Digital success rates above 95% show that the suggested solution performs well in terms of dependability and data transfer efficiency. This ratio was computed using the high accuracy value (1 - BER) and the low BER attained under high SNR circumstances. The findings obtained show that the system has nearly excellent fault tolerance and that it is possible to successfully protect the integrity of various omic data transferred in multi-user telemedicine situations. Because of its flexibility and resilience to changing channel circumstances, the hybrid artificial intelligence model that was created in this context allows for continuous and secure data exchange. These findings confirm that implementing AI-driven power allocation algorithms can greatly improve the reliability and performance of telemedicine communications.

- The distant user (400 km) received proportionally more power allocation while keeping a low BER.
- The average accuracy across users improved as SNR increased, demonstrating the hybrid model's predictive stability.
- Graphs demonstrated the proposed system's robustness under various channel circumstances.
- The hybrid solution outperformed static allocation and single-model approaches in terms of fairness and communication dependability.

4 Conclusion

This paper proposes an intelligent power allocation paradigm for power domain NOMA-based wireless devices that transmit complicated multi-omics data into health information systems. The hybrid AI technique assures fairness, security, and high transmission quality even for users at great distances. This approach is highly suited for inclusion into telemedicine infrastructures that require reliable, real-time health data transfer.

Using NOMA, we presented a new hybrid AI-assisted power allocation model in this work for the safe, equitable, and effective transfer of multi-omic data in multi-user telemedicine systems.

Robust and adaptable power forecasts were made possible by the combination of Random Forest and KNN regressors with distance-based channel gain estimations, which achieved a success rate of over 95% under high SNR circumstances.

With a processing time of just 2.18 seconds and a low memory consumption (first 314.24 MB, then 304.49 MB), the suggested hybrid AI-supported NOMA system has proven its high efficiency and computational lightness.

These results have important ramifications for the creation of intelligent and scalable healthcare communication systems. The suggested approach enhances the dependability of customized remote health monitoring by guaranteeing high data integrity and reducing bit error rates. Additionally, this method enables real-time, physiologically informed telehealth decision-making by bridging the domains of artificial intelligence, communication systems, and medical informatics.

In order to further enhance the usability of this strategy in contemporary healthcare ecosystems, future research may involve clinical validation using actual patient data and real-time deployment utilizing embedded devices.

5 References

[1]. Babu, M., & Snyder, M. (2023). Multi-omics profiling for health. Molecular & Cellular Proteomics, 22(6).

[2]. Thacharodi, A., Singh, P., Meenatchi, R., Tawfeeq Ahmed, Z. H., Kumar, R. R., V, N., ... & Hassan, S. (2024). Revolutionizing healthcare and medicine: The impact of modern technologies for a healthier future—A comprehensive review. Health Care Science, 3(5), 329-349.

[3]. Vishweshwara, A., & Ramya, R. (2025). Transforming Telemedicine: Reducing Latency Through Edge Computing and 5G—A Review. *Biomedical Materials & Devices*, 1-14.

[4]. Kirtay, S., Yildiz, K., & Bocekci, V. G. (2024). Artificial Intelligence-based Fair Allocation in NOMA Technique: A Review. International Journal of Sensors Wireless Communications and Control, 14(3), 161-174.

[5]. Rehman, S. U., Ahmad, J., Manzar, A., & Moinuddin, M. (2024). Beamforming techniques for mimo-noma for 5g and beyond 5g: Research gaps and future directions. *Circuits, Systems, and Signal Processing*, *43*(3), 1518-1548.

[6]. Kirtay, S., Bocekci, V. G., & Yildiz, K. (2025). Genetic Algorithm Based Approach for Optimization of Fair Power Allocation in Pd-NOMA Systems. *Tehnički vjesnik*, *32*(3), 948-957.

Exploring the trends of Multiomics Research: A Comprehensive Bibliometric Analysis with Interactive Visualization Tools (BiblioMaps)

Murat Koçak[1], Seda Kırtay[2], Zafer Akçalı[3]

1. Bakent University, Faculty of Medicine, Department of Medical Informatics, Ankara, Turkey. ORCID: 0000-0001-6510-3666
2. Marmara University, Faculty of Technology, Department of Telecommunication, Istanbul, Turkey. ORCID: 0000-0003-2415-9131
3. Baskent University, Faculty of Medicine, Department of Medical Informatics, Ankara, Turkey. ORCID: 0000-0003-2473-4431

Abstract

Purpose: This study aims to explore the thematic and structural evolution of multi-omics research through a comprehensive bibliometric analysis using interactive science mapping tools. Specifically, it utilizes BiblioMaps, a web-based platform, to visualize global trends, key topics, influential authors, institutions, and collaborative networks in the field of multi-omics from 2015 to 2024.

Methods: A systematic search was conducted on the Web of Science (WoS) database using the following query to collect publications specifically related to multi-omics research: TI=("multi-omics" OR "multiomics" OR "integrated omics" OR "multi omics" OR "multi-modal omics" OR "multi-dimensional omics" OR "multi-scale omics" OR "multi-source omics" OR ("multi" AND ("omics" OR "omic" OR "omic data" OR "omic integration"))) AND PY=2015-2024 AND DT=Article. This search yielded a dataset of 5,835 articles published between 2015 and 2024 across 1,201 distinct journals and scholarly outlets. Bibliographic coupling was applied to group publications into topical clusters and visualize them interactively using BiblioMaps, which leverages Python libraries for enhanced interactivity. This approach enabled in-depth exploration of topics, subtopics, keywords, authors, countries, and journal sources. You can access the interactive map at: http://ankara.baskent.edu.tr/~muratkocak/

Results: The results revealed a steady increase in publication output over time, with a significant surge after 2020. The highest number of publications was recorded in 2024 (31.57%) , followed by 2023 (19.95%) and 2022 (17.36%). China emerged as the leading contributor with 50.78% of all publications, followed by the United States (23.67%) and the United Kingdom (5.98%). Approximately 26.21% of the publications involved international collaborations, highlighting strong global cooperation. Top contributing authors included Wang Y(122 publications), Zhang Y(112 publications), and Li Y(103 publications). Keyword analysis identified dominant themes such as “expression” (18.63%), “gene-expression” (6.96%), and “cancer”, reflecting the biological and disease-related focus of the field. The most active journals were Scientific Reports (Sci Rep-UK) , Nature Communications (Nat Commun), and International Journal of Molecular Sciences (Int J Mol Sci). Thematic clustering through bibliographic coupling grouped publications into 5 main topics and 22 subtopics, revealing the integration of diverse disciplines such as oncology, plant sciences, computational biology , and immunology under the umbrella of multi-omics research.

Conclusion: This study demonstrates the utility of BiblioMaps in providing a dynamic and interactive overview of complex, interdisciplinary fields like multi-omics. The findings highlight the rapid growth of the field, its global reach, and the importance of international collaboration. The methodology and visualization tools used in this work can assist researchers, educators, and policymakers in identifying emerging trends, knowledge gaps, and potential collaborators. Future work will focus on extending these tools for real-time trend detection and broader application across other scientific domains.

Keywords: multiomics, bibliometric analysis, science mapping, bibliomaps, interactive visualization, research collaboration.

INTRODUCTION

BiblioMaps is an advanced extension of BiblioTools (Grauwin & Jensen, 2011), a comprehensive toolkit previously designed to generate static representations of topical fields. However, the earlier tool necessitated the creation of extensive companion documents to establish connections between paper-level and topic-level information. In contrast, BiblioMaps is a web-based, interactive platform that facilitates the exploration of large-scale bibliographical data sets, providing a user-friendly interface for a broad audience.

Key Features and Functionality

BiblioMaps offers a freely accessible toolkit that empowers users with minimal development expertise to process metadata from their chosen publication corpus and generate interactive science maps from any bibliographic data extracted from the Web of Science. The platform's interface boasts several notable features, including:

1. Spatial Interaction: Users can engage with the map in a dynamic and intuitive manner.
2. Topic and Subtopic Exploration: The nature of each topic and subtopic can be thoroughly examined.
3. Representative Paper Access: A selection of highly cited papers can be accessed via Google Scholar.
4. Search Functionality: Users can perform searches within specific topics.
5. Customizable Layout: Layout parameters can be modified to suit individual preferences.
6. Visualization Export: Produced visualizations can be exported for further analysis or presentation.
7. Multiomics Thematic Map: A specialized thematic map is available for multiomics research.

To demonstrate the capabilities of BiblioMaps, we have created a scientific map of the User Experience Design field, referred to as the Multiomics Map. Two visualization interfaces were developed using D3.js, which can be explored online (http://ankara.baskent.edu.tr/~muratkocak). These interfaces provide an interactive platform for users to engage with the data and analyses.

Corpus Description and Thematic Map Analyses

The Corpus Description analysis provides a comprehensive description of the items found in the publications of the studied dataset, including authors, keywords, references, countries in affiliations, and more. This information is presented through sorted lists and customized graphics, such as word clouds, networks, pie charts, bar graphs, and world maps.

The Thematic Map analysis enables the exploration of a two-level hierarchical thematic map built from clusters obtained through Bibliographic Coupling analyses. This interface was designed to cater to the diverse needs of various users.

Comparison to Existing Tools and Platforms

Several science maps are available online (e.g., Börner et al., 2012; Claveau & Gingras, 2016; Cobo et al., 2011; George & Knedjens, 2013; Kang et al., 2006; Leydesdorff & Rafols, 2009). Moreover,

One widely used platform is CiteSpace , introduced by Chen (2006), which enables researchers to detect emerging trends and transformative patterns in scientific literature by visualizing citation bursts over time. Another prominent tool is Gephi , an open-source network analysis and visualization software that facilitates the exploration of co-authorship, co-citation, and keyword networks with a

user-friendly interface (Bastian, Heymann, & Jacomy, 2009). For temporal and structural analyses of research trends, the Sci² Tool (Herron, Bynum, & Francis, 2011) offers robust functionalities including social network analysis and time-series visualization. In Brazil, the Lattes Platform has been utilized for mapping national scientific production based on researcher CVs, offering insights into institutional and regional performance (Almeida et al., 2018). HistCite™ , developed by Eugene Garfield, provides a chronological map of scientific domains using citation networks (Garfield, 2004), while the Knowledge Domain Visualizer (KDV) supports thematic mapping through clustering and co-citation analysis (Šubelj, Vavpetič, Bajec, & Demsar, 2011). Lastly, the Microsoft Academic Graph (MAG) serves as a large-scale open dataset for tracking scientific trends using embedded vectors for topic classification (Sinha et al., 2015).

Despite their contributions, many of these tools require specific technical expertise or programming skills to process and visualize data effectively. For instance, platforms such as Gephi and CiteSpace demand intermediate-level knowledge in data manipulation and network theory, limiting their accessibility to non-expert users. Moreover, several of these tools generate static outputs that are not optimized for web-based interaction, often requiring additional software installations or manual post-processing steps. In contrast, BiblioMaps addresses these limitations by providing a fully web-based and interactive platform for science mapping. Unlike most traditional tools, BiblioMaps does not require any installation or coding experience, making it accessible to a broader audience including students, educators, and policymakers. It enables dynamic exploration of bibliographic datasets through real-time filtering, topic navigation, and direct access to representative publications. Compared to other platforms that primarily produce static network diagrams, BiblioMaps integrates all stages—from data processing to visualization within a single environment, enhancing usability and interpretability. As a result, BiblioMaps offers a more intuitive, comprehensive, and collaborative solution for mapping interdisciplinary fields such as multi-omics research.

However, these maps often require specialized knowledge in information visualization to navigate and understand the presented data. In addition, various software toolkits exist for creating standard bibliographic maps (co-authors, co-citation, bibliographic coupling networks, etc.) (e.g., Van Eck & Waltman, 2010). Nevertheless, most of these tools produce visualizations that are not optimized for web-based exploration, often necessitating the installation of specific interfaces and programming knowledge. BiblioMaps addresses these limitations by providing an accessible, web-based platform for interactive science mapping. Several tools have been developed to support the visualization and analysis of scientific knowledge domains through bibliometric methods. One widely used platform is CiteSpace , introduced by Chen (2006), which enables researchers to detect emerging trends and transformative patterns in scientific literature by visualizing citation bursts over time. Another prominent tool is Gephi , an open-source network analysis and visualization software that facilitates the exploration of co-authorship, co-citation, and keyword networks with a user-friendly interface (Bastian, Heymann, & Jacomy, 2009). For temporal and structural analyses of research trends, the Sci² Tool (Herron, Bynum, & Francis, 2011) offers robust functionalities including social network analysis and time-series visualization. In Brazil, the Lattes Platform has been utilized for mapping national scientific production based on researcher CVs, offering insights into institutional and regional performance (Almeida et al., 2018). HistCite™ , developed by Eugene Garfield, provides a chronological map of scientific domains using citation networks (Garfield, 2004), while the

Knowledge Domain Visualizer (KDV) supports thematic mapping through clustering and co-citation analysis (Šubelj, Vavpetič, Bajec, & Demsar, 2011). Lastly, the Microsoft Academic Graph (MAG) serves as a large-scale open dataset for tracking scientific trends using embedded vectors for topic classification (Sinha et al., 2015). Despite their contributions, many of these tools require specific technical expertise or programming skills to process and visualize data effectively. For instance, platforms such as Gephi and CiteSpace demand intermediate-level knowledge in data manipulation and network theory, limiting their accessibility to non-expert users. Moreover, several of these tools generate static outputs that are not optimized for web-based interaction, often requiring additional software installations or manual post-processing steps.

In contrast, BiblioMaps addresses these limitations by providing a fully web-based and interactive platform for science mapping. Unlike most traditional tools, BiblioMaps does not require any installation or coding experience, making it accessible to a broader audience including students, educators, and policymakers. It enables dynamic exploration of bibliographic datasets through real-time filtering, topic navigation, and direct access to representative publications. Compared to other platforms that primarily produce static network diagrams, BiblioMaps integrates all stages—from data processing to visualization—within a single environment, enhancing usability and interpretability. As a result, BiblioMaps offers a more intuitive, comprehensive, and collaborative solution for mapping interdisciplinary fields such as multi-omics research.

METHODOLOGY

Data Retrieval and Processing

Bibliographical data was systematically retrieved from the Web of Science database. The data was subsequently processed utilizing the BiblioMaps software, in accordance with the methodology outlined by Grauwin et al. (2012). A specifically designed query was employed to curate a corpus representative of research performance in the field of Multiomics. The query parameters were as follows: TI=("multi-omics" OR "multiomics" OR "integrated omics" OR "multi omics" OR "multi-modal omics" OR "multi-dimensional omics" OR "multi-scale omics" OR "multi-source omics" OR ("multi" AND ("omics" OR "omic" OR "omic data" OR "omic integration"))) AND PY=2015-2024 AND DT=Article

This query combination facilitated the collection of publications pertinent to Multiomics. To ensure that the selected papers had a substantial focus on both Multiomics and design, the terms "User Experience" and "Design" were incorporated in a complementary query. Utilizing a query based solely on "User experience" yielded an excessive number of results outside the scope of interest. The resultant Multiomics-Design corpus comprised 12,220 publication records.

Clustering and Analysis

Publications were aggregated into relevant topical clusters via a standard Bibliographic Coupling (BC) analysis, as described in Grauwin et al. (2012). BC is a suitable technique for representing research fronts and emerging topics, such as Multiomics Design, as it considers all publications, including those not yet cited.

BiblioTools is a suite of Python scripts designed to generate maps of science based on bibliographic data. The software enables users to execute several key tasks:

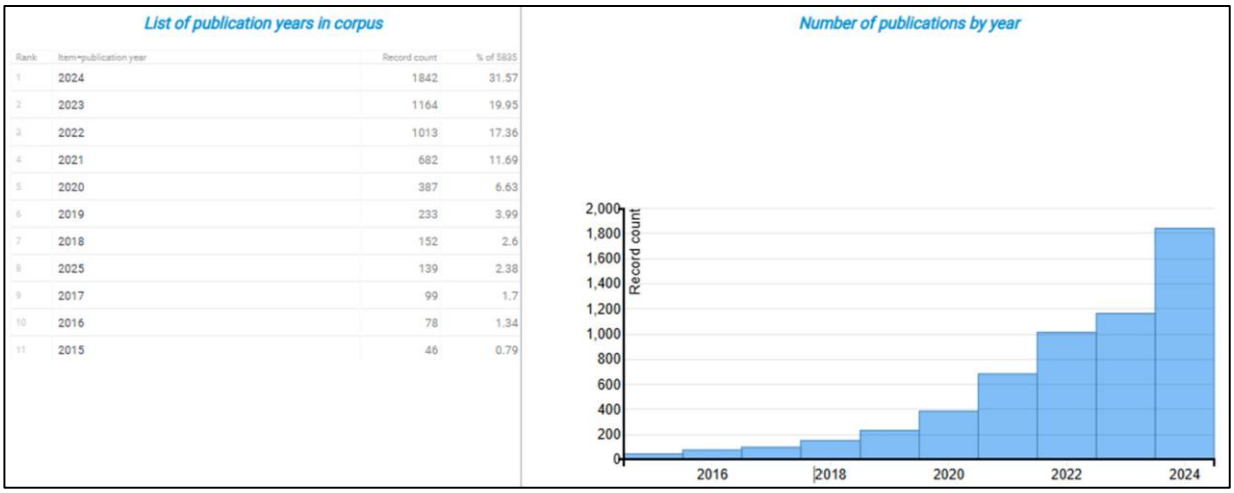
1. Data Mining: Efficiently handle bibliographic records exported from the Web of Science.
2. Data Processing: Pre-process raw data through parsing and filtering.

3. Data Analysis: Describe a given corpus; generate co-occurrence networks (co-words, co-authors, co-citations, etc.); generate bibliographic coupling networks and clusters.
4. Data Visualization: Visualize and explore results in BiblioMaps, a dedicated web-based interface; export networks in a format compatible with standard visualization tools, such as Gephi.
5. Data Reporting: Automatically generate comprehensive reports in LaTeX.

RESULTS

Corpus Description

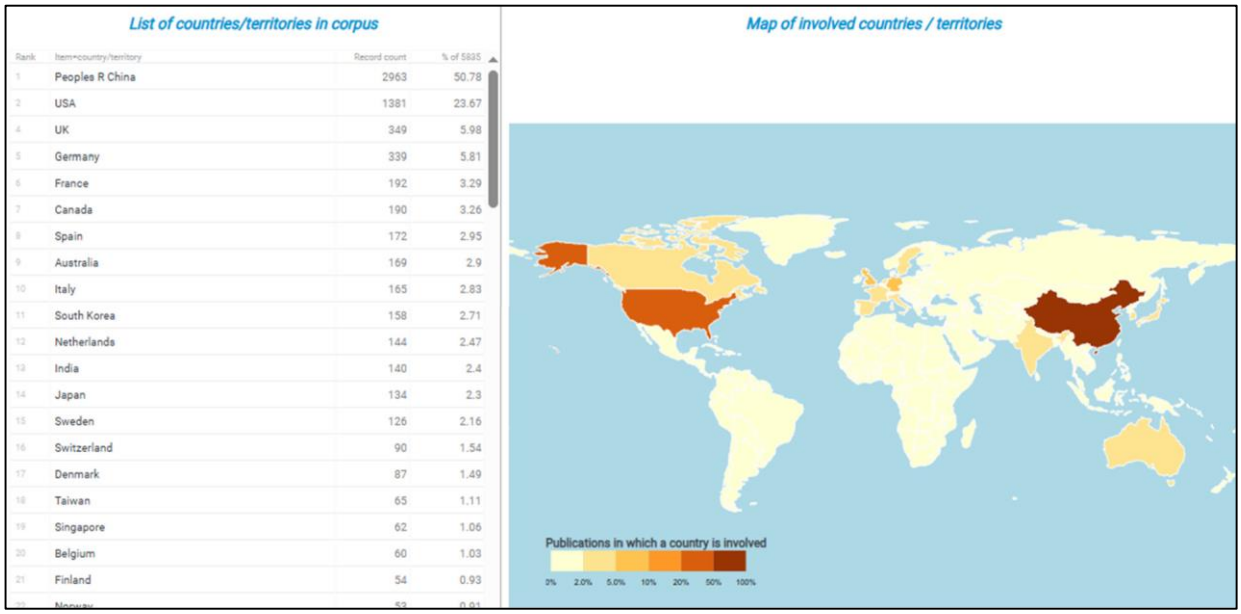
Figure 1 presents the distribution of publications over different years within a multiomics dataset. It shows that the number of publications has generally increased over time, with a particularly sharp rise observed from 2020 onwards. The highest number of publications occurred in 2024, accounting for 31.57% (1,842 records) of the total dataset. This is followed by 2023 and 2022 with 19.95% and 17.36%, respectively. The data indicates a growing trend in publication output year by year, highlighting an increasing interest or activity in the subject area in recent years. The oldest year included in the dataset is 2015, which contributes only 0.79% (46 records), showing a significant contrast with the more recent years.



Rank	Item=publication year	Record count	% of 5835
1	2024	1842	31.57
2	2023	1164	19.95
3	2022	1013	17.36
4	2021	682	11.69
5	2020	387	6.63
6	2019	233	3.99
7	2018	152	2.6
8	2025	139	2.38
9	2017	99	1.7
10	2016	78	1.34
11	2015	46	0.79

Figure 1: Publication trends in multiomics research: A decade overview (2015–2024)

Figure 2 indicates the geographical distribution of publications within the dataset, highlighting the contribution of different countries and territories. The table on the left lists the number of records per country, showing that China (Peoples R China) contributes the most with 2,963 records (50.78%) , followed by the United States (1,381 records, 23.67%) and the UK (349 records, 5.98%) . Other notable contributors include Germany, France, Canada, and South Korea, though their contributions are significantly smaller compared to China and the US. The world map on the right visually reinforces this distribution, with China and the US appearing in the darkest shades, indicating the highest concentration of research output. Overall, the data suggests a strong dominance of East Asian and Western countries in the production and publication of multiomics research within the dataset.

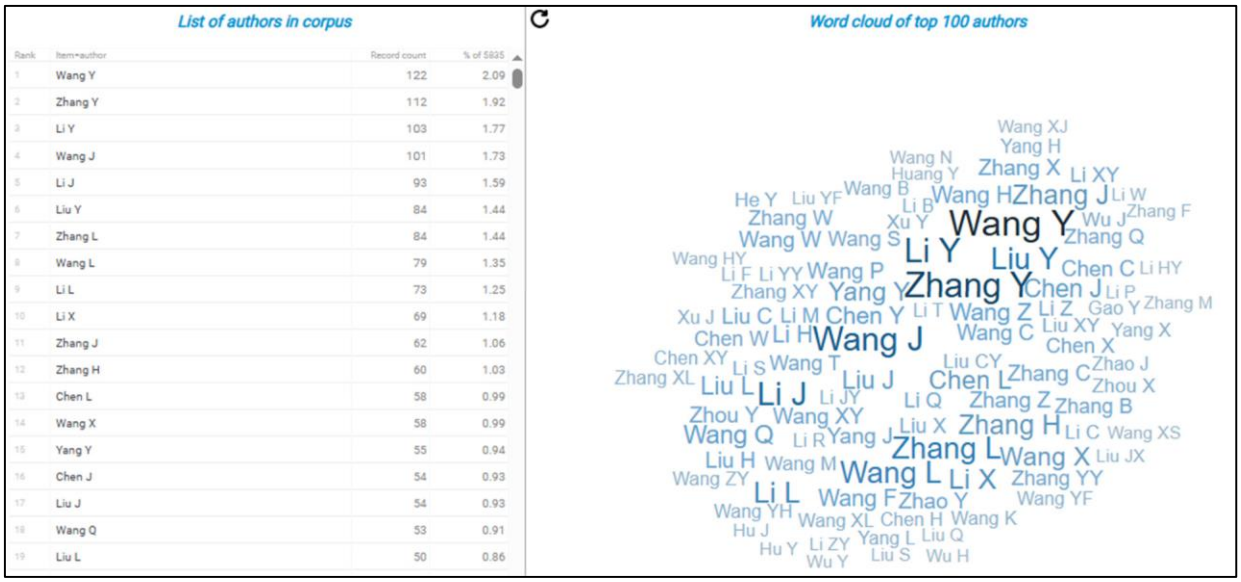


List of countries/territories in corpus

Rank	Item=country/territory	Record count	% of 5835
1	Peoples R China	2963	50.78
2	USA	1381	23.67
4	UK	349	5.98
5	Germany	339	5.81
6	France	192	3.29
7	Canada	190	3.26
8	Spain	172	2.95
9	Australia	169	2.9
10	Italy	165	2.83
11	South Korea	158	2.71
12	Netherlands	144	2.47
13	India	140	2.4
14	Japan	134	2.3
15	Sweden	126	2.16
16	Switzerland	90	1.54
17	Denmark	87	1.49
18	Taiwan	65	1.11
19	Singapore	62	1.06
20	Belgium	60	1.03
21	Finland	54	0.93
22	Norway	53	0.91

Figure 2: Country contributions to multiomics studies in the dataset (2015–2024)

Figure 3 provides an overview of author contributions within the corpus, highlighting both quantitative and qualitative insights. On the left, a table lists the top authors ranked by their record count, with Wang Y leading with 122 publications (2.09% of the total) , followed by Zhang Y (112 publications, 1.92%) , Li Y (103 publications, 1.77%) , and others. The right side features a word cloud of the top 100 authors, visually emphasizing the frequency of names based on their prominence in the corpus. Names like Wang Y , Zhang Y , Li Y , and Wang J appear prominently in larger fonts, indicating their significant contributions. This dual representation underscores the concentration of research output among a few key authors, suggesting potential collaboration networks or influential researchers in the field.



List of authors in corpus

Rank	Item=author	Record count	% of 5835
1	Wang Y	122	2.09
2	Zhang Y	112	1.92
3	Li Y	103	1.77
4	Wang J	101	1.73
5	Li J	93	1.59
6	Liu Y	84	1.44
7	Zhang L	84	1.44
8	Wang L	79	1.35
9	Li L	73	1.25
10	Li X	69	1.18
11	Zhang J	62	1.06
12	Zhang H	60	1.03
13	Chen L	58	0.99
14	Wang X	58	0.99
15	Yang Y	55	0.94
16	Chen J	54	0.93
17	Liu J	54	0.93
18	Wang Q	53	0.91
19	Liu L	50	0.86

Figure 3: Author contributions in multiomics research: Lists of authors and word cloud of top 100 authors

Figure 4 provides an overview of keyword usage and co-occurrence patterns within the multiomics dataset, highlighting both frequency and relationships among key terms. On the left, a table lists the top keywords ranked by their record count, with "EXPRESSION" leading with 1,087 occurrences (18.63% of the total) , followed by "IDENTIFICATION" (423 occurrences, 7.25%) , "GENE-EXPRESSION" (406 occurrences, 6.96%) , and others such as "CANCER," "CELLS," "METABOLISM," and "PROTEIN." The right side features a network visualization of the top 100 keywords, illustrating their co-occurrence relationships. Keywords like "EXPRESSION," "GENE-EXPRESSION," "IDENTIFICATION," and "CANCER" are central nodes with high connectivity, indicating their frequent association with other terms. This dual representation underscores the prominence of biological processes, molecular mechanisms, and disease-related topics in the corpus, providing insights into the thematic focus and interconnectedness of research in this domain.

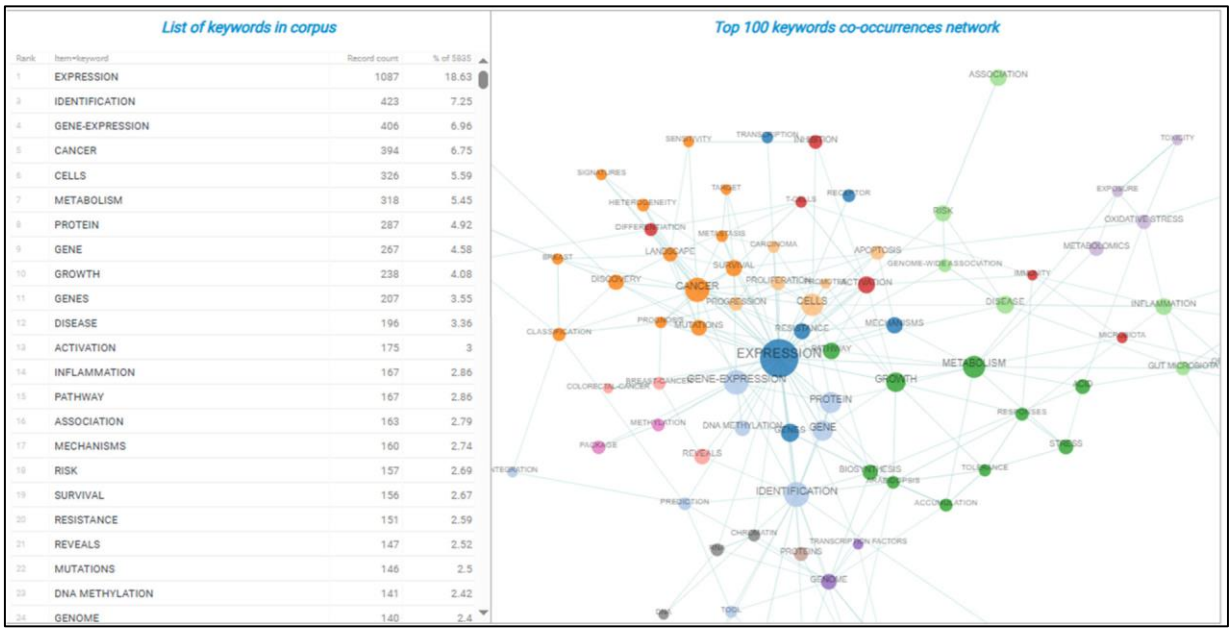
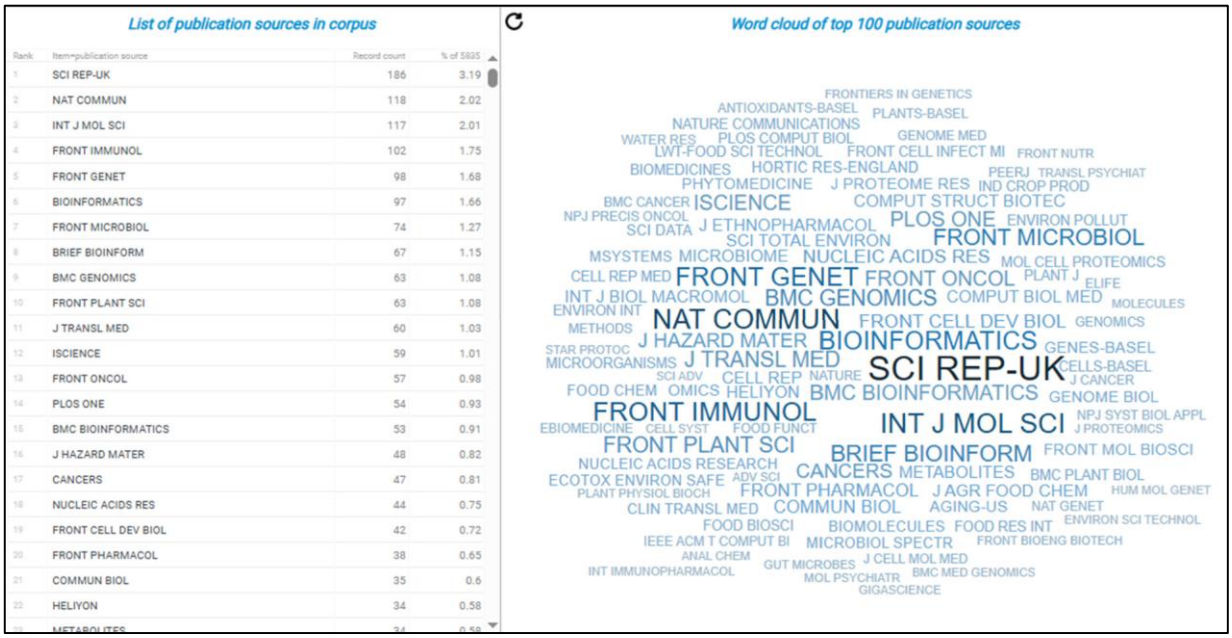


Figure 4: Lists of keywords and top 100 keywords co-occurrences network

Figure 5 presents an overview of publication sources within the corpus, highlighting both quantitative and qualitative insights. On the left, a table lists the top publication sources ranked by their record count, with "SCI REP-UK" leading with 186 publications (3.19% of the total) , followed by "NAT COMMUN" (118 publications, 2.02%) , "INT J MOL SCI" (117 publications, 2.01%) , and others such as "FRONT IMMUNOL," "BIOMEDICINE," and "BMC GENOMICS." The right side features a word cloud of the top 100 publication sources, visually emphasizing the frequency of journals based on their prominence in the corpus. Journals like "ISCIENCE," "FRONT GENET," "NAT COMMUN," "BIOINFORMATICS," and "SCI REP-UK" appear prominently in larger fonts, indicating their significant contributions to the field. This dual representation underscores the concentration of research output across a few key journals, suggesting their importance as primary platforms for disseminating multiomics-related studies.



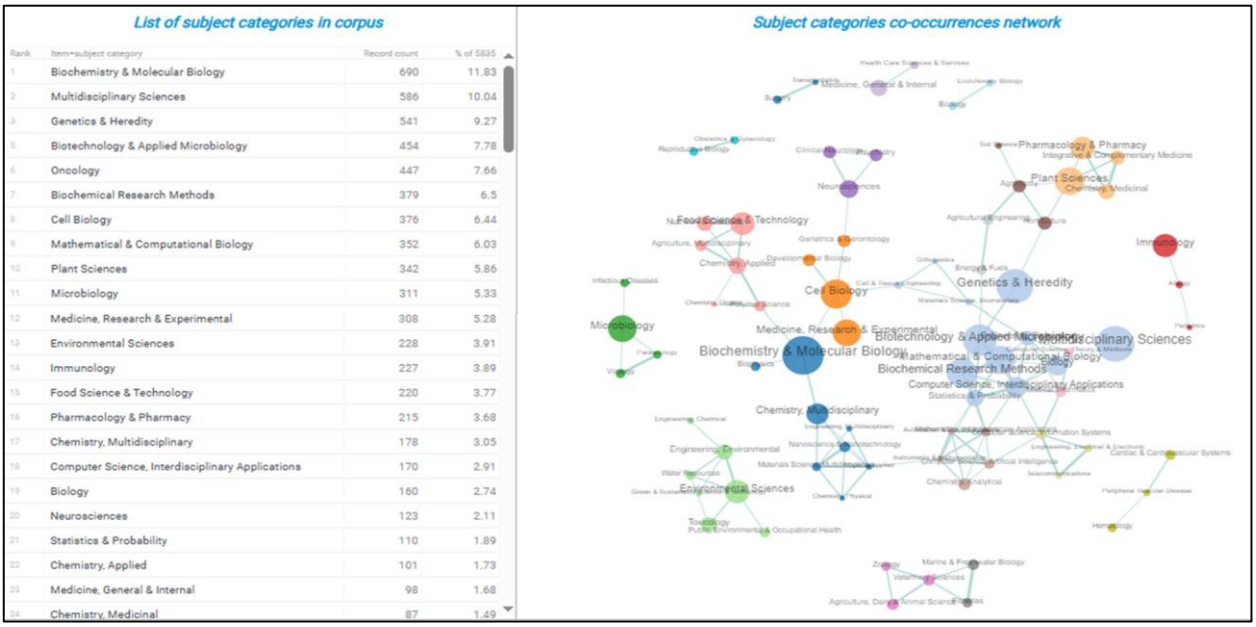
List of publication sources in corpus

Rank	Item=publication source	Record count	% of 5835
1	SCI REP-UK	186	3.19
2	NAT COMMUN	118	2.02
3	INT J MOL SCI	117	2.01
4	FRONT IMMUNOL	102	1.75
5	FRONT GENET	98	1.68
6	BIOINFORMATICS	97	1.66
7	FRONT MICROBIOL	74	1.27
8	BRIEF BIOINFORM	67	1.15
9	BMC GENOMICS	63	1.08
10	FRONT PLANT SCI	63	1.08
11	J TRANSL MED	60	1.03
12	ISCIENCE	59	1.01
13	FRONT ONCOL	57	0.98
14	PLOS ONE	54	0.93
15	BMC BIOINFORMATICS	53	0.91
16	J HAZARD MATER	48	0.82
17	CANCERS	47	0.81
18	NUCLEIC ACIDS RES	44	0.75
19	FRONT CELL DEV BIOL	42	0.72
20	FRONT PHARMACOL	38	0.65
21	COMMUN BIOL	35	0.6
22	HELIYON	34	0.58
23	METABOLITES	34	0.58

Word cloud of top 100 publication sources

Figure 5: Journal contributions in Multiomics research: lists of journals and word cloud of top 100 authors

Figure 6 offers an overview of subject categories within the corpus, emphasizing both their prevalence and interconnectedness. On the left, a table lists the top subject categories ranked by their record count, with "Biochemistry & Molecular Biology" leading with 690 occurrences (11.83% of the total) , followed by "Multidisciplinary Sciences" (586 occurrences, 10.04%) , "Genetics & Heredity" (541 occurrences, 9.27%) , and others such as "Biotechnology & Applied Microbiology," "Oncology," and "Biochemical Research Methods." The right side features a network visualization of the top subject categories, illustrating their co-occurrence relationships. Nodes like "Biochemistry & Molecular Biology," "Cell Biology," "Genetics & Heredity," and "Multidisciplinary Sciences" are central, indicating their frequent association with other fields. This dual representation underscores the prominence of molecular and genetic research, as well as the interdisciplinary nature of studies in this domain.



List of subject categories in corpus

Rank	Item=subject category	Record count	% of 5835
1	Biochemistry & Molecular Biology	690	11.83
2	Multidisciplinary Sciences	586	10.04
3	Genetics & Heredity	541	9.27
5	Biotechnology & Applied Microbiology	454	7.78
6	Oncology	447	7.66
7	Biochemical Research Methods	379	6.5
8	Cell Biology	376	6.44
9	Mathematical & Computational Biology	352	6.03
10	Plant Sciences	342	5.86
11	Microbiology	311	5.33
12	Medicine, Research & Experimental	308	5.28
13	Environmental Sciences	228	3.91
14	Immunology	227	3.89
15	Food Science & Technology	220	3.77
16	Pharmacology & Pharmacy	215	3.68
17	Chemistry, Multidisciplinary	178	3.05
18	Computer Science, Interdisciplinary Applications	170	2.91
19	Biology	160	2.74
20	Neurosciences	123	2.11
21	Statistics & Probability	110	1.89
22	Chemistry, Applied	101	1.73
23	Medicine, General & Internal	98	1.68
24	Chemistry, Medicinal	87	1.49

Figure 6: Subject categories co-occurrences network

Thematic Map (Bibliographic coupling)

Based on the references they share, 5319 thematically close publications of the studied corpus are gathered into 5 topics (Q=0.307) and 22 subtopics (Q=0.214). Figure 7a and 7b illustrate the complex network of interactions centered around " multi-omics," connecting it to various biological and analytical domains such as "BREAST," "PROTEOMICS," "ARABIDOPSIS," and "MENDELIAN RANDOMIZATION." The multi-omics approach integrates different layers of biological data — potentially genomic, transcriptomic, proteomic, and metabolomic information — to reveal how these elements interact within the context of cancer.

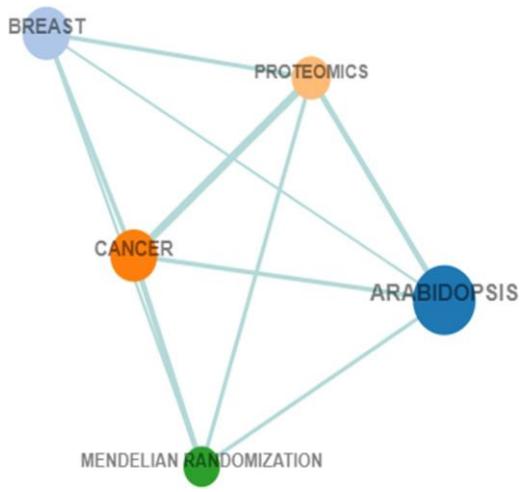


Figure 7a: Most Significant Keywords (Main Topics)

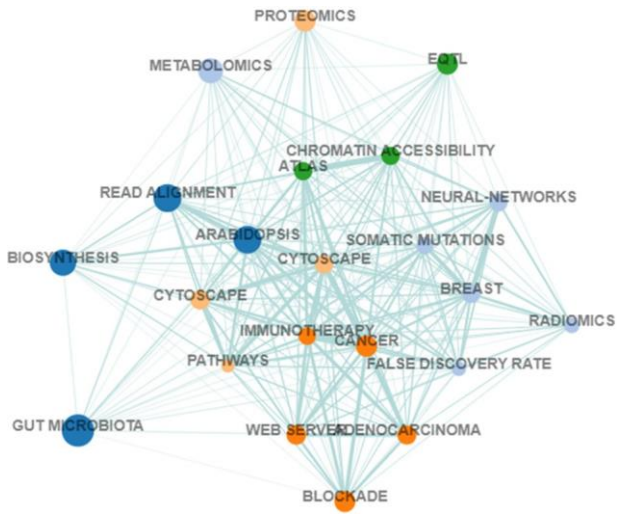


Figure 7b: Most Significant Keywords(Subtopics)

Figure 8a and 8b are derived from a multi-omics dataset and illustrates the interconnected nature of various scientific disciplines in the context of cancer research. At the center of the network is "Oncology," which serves as the focal point connecting fields such as "Biochemistry & Molecular Biology," "Plant Sciences," "Mathematical & Computational Biology," And "Multidisciplinary Sciences." The multi-omics approach underpins this network by integrating diverse biological data layers (e.g., genomics, proteomics, metabolomics) to provide a comprehensive understanding of cancer biology.

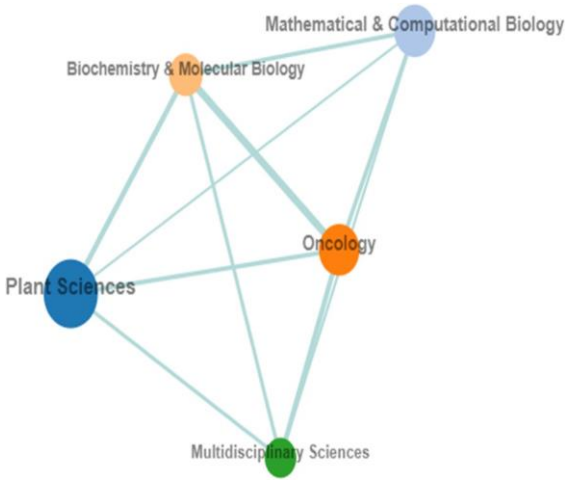


Figure 8a: Most Frequent subjects (Main topics)

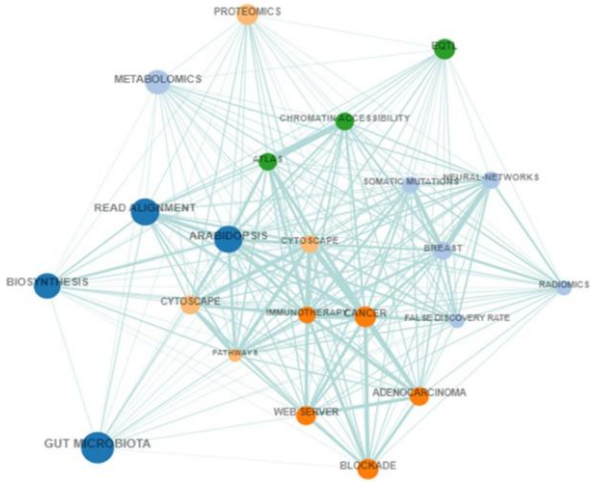


Figure 8b: Most Frequent subjects (Subtopics)

Figure 9a and 9b illustrates the interconnected relationships between various scientific disciplines and publications in the context of immunology research. At the center of the network is "FRONT IMMUNOL," representing *Frontiers in Immunology* , a prominent journal in the field of immunology. This central node connects to other key nodes, including "BIOINFORMATICS," "SCI REP-UK" (likely referring to *Scientific Reports* published by Nature Research), and "NAT COMMUN" (short for *Nature Communications*). The connections between these nodes highlight the interdisciplinary nature of modern immunological research, where bioinformatics plays a crucial role in analyzing complex biological data, while high-impact journals like *Scientific Reports* and *Nature Communications* serve as platforms for disseminating cutting-edge findings.

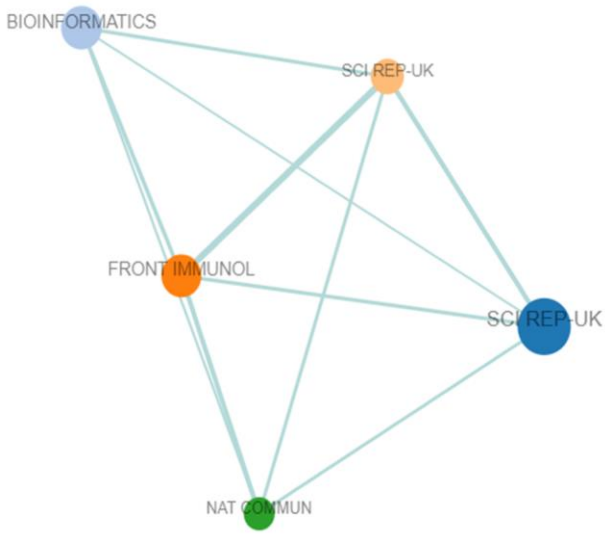


Figure 9a: Most Frequent Journal sources (Main topics)

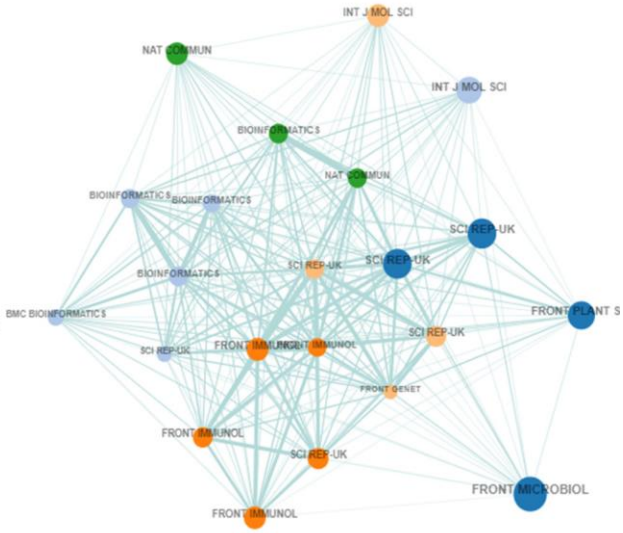


Figure 9b: Most Frequent Journal sources (Subtopics)

Figure 10a and Figure 10b illustrates the collaborative network among researchers in a specific scientific domain. At the center of the network are five key nodes representing individual researchers: Zhang J , Li XJ , Wang Y , Wang JD , and Zhou GY . The connections between these nodes indicate collaborative relationships or co-authorship, suggesting that these researchers have worked together on studies or projects related to the multi-omics data.

The use of different colors for the nodes may represent various roles, expertise areas, or affiliations within the research group. This network underscores how multi-omics datasets require a multidisciplinary approach, where researchers from diverse backgrounds come together to analyze complex biological data and derive meaningful insights. Overall, the graph provides a visual representation of the collaborative ecosystem essential for driving innovation and discovery in modern scientific research (Figure 10b).

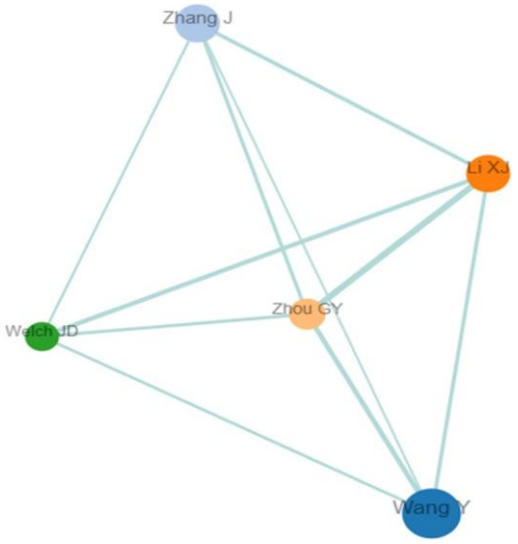


Figure 10a: Most cited authors (Main topics)

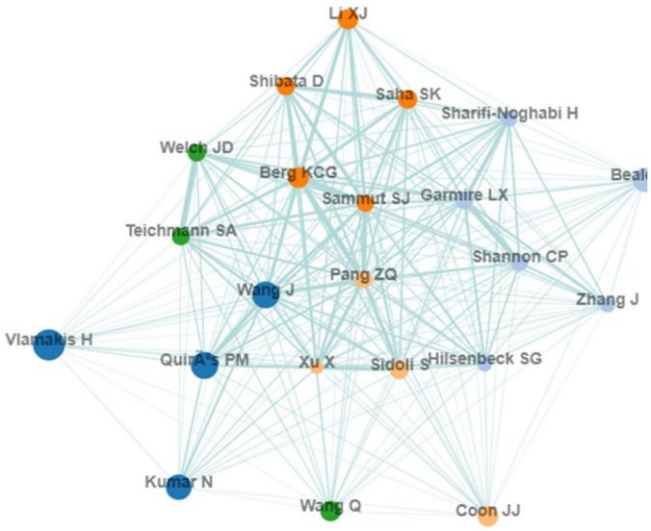


Figure 10b: Most Frequent Journal sources (Subtopics)

LIMITATIONS

Despite the comprehensive nature of this study and the innovative use of interactive science mapping tools such as BiblioMaps , several limitations should be acknowledged. Firstly, the dataset was sourced exclusively from the Web of Science (WoS) Core Collection , which may not fully represent all multi-omics research outputs, particularly those published in regional or non-indexed journals. This could result in an underrepresentation of contributions from certain countries or disciplines that publish predominantly in local or open-access venues. Secondly, while the search query was designed to capture a broad range of publications related to multi-omics, the use of specific keywords such as *"multi-omics"* , *"integrated omics"* , and *"multi-dimensional omics"* may have introduced selection bias . Some relevant studies might have been excluded due to variations in terminology or lack of standardized descriptors across different subfields. Thirdly, the bibliographic coupling method used for clustering relies on shared references among publications rather than direct citation links. While this approach is effective for identifying thematic similarities, it may not always reflect the most recent or rapidly emerging trends in the field.

Additionally, although BiblioMaps offers an intuitive and interactive visualization platform, its current functionality is limited to datasets exported from the Web of Science. Expanding compatibility with other databases such as Scopus or PubMed would enhance its utility for broader scientific communities. Lastly, the language of analysis and interface is currently available only in English, which may limit accessibility for non-English-speaking researchers and educators who could benefit from using the tool.Future improvements will focus on integrating additional data sources, refining clustering algorithms, and enhancing multilingual support to address these limitations and provide a more inclusive and dynamic representation of multi-omics research.

CONCLUSION

This study presents a comprehensive bibliometric analysis of multi-omics research from 2015 to 2024 using the interactive visualization tool BiblioMaps . The dataset consists of 5,835 publications retrieved from the Web of Science, demonstrating a steady and significant increase in research output over time, particularly since 2020, with 2024 having the highest number of publications (31.57%). This trend highlights the growing interest and importance of multi-omics approaches in modern biomedical and life sciences. Geographically, China stands out as the leading contributor with 50.78% of the total publications, followed by the United States(23.67%) and the UK(5.98%) , indicating that East Asian and Western countries dominate the field.

The high rate of international collaboration (26.21%) further emphasizes the global and interdisciplinary nature of multi-omics research, with notable cooperation between researchers in the US, Europe, and emerging scientific hubs like Turkey. In terms of author contributions, Wang Y(122 publications) , Zhang Y(112 publications) , and Li Y(103 publications) are among the most active researchers, suggesting the presence of well-established research groups or collaborative networks in this domain. Keyword analysis revealed that terms such as "expression" (18.63%), "identification" (7.25%) , and "gene-expression" (6.96%) are frequently used, pointing to a strong focus on molecular mechanisms, biological processes, and disease-related studies, especially cancer. These findings reflect the integration of genomic, transcriptomic, proteomic, and metabolomic data to better understand complex diseases and biological systems.

The top journals contributing to the field include Scientific Reports (Sci Rep-UK), Nature Communications(Nat Commun) , and International Journal of Molecular Sciences (Int J Mol Sci) , which serve as key platforms for disseminating multi-omics research. Subject category analysis further shows that the field is primarily rooted in Biochemistry & Molecular Biology , Genetics & Heredity , and Oncology , while also incorporating disciplines like Computational Biology , Plant Sciences , and Immunology , highlighting its interdisciplinary character. Using BiblioMaps , we were able to visualize thematic clusters and subtopics through bibliographic coupling, revealing five main topics and 22 subtopics. Notable thematic areas include breast cancer , proteomics , Arabidopsis , and Mendelian randomization , illustrating how multi-omics integrates diverse biological layers to uncover interactions within complex systems. Moreover, the network visualization of authors and journal sources demonstrates the collaborative structure of the field, with journals like Frontiers in Immunology , Scientific Reports , and Nature Communications acting as central nodes connecting researchers across disciplines.

In conclusion, this study not only maps the structural and thematic evolution of multi-omics research but also highlights the power of interactive bibliometric tools like BiblioMaps in supporting knowledge discovery, identifying research trends, and facilitating interdisciplinary collaboration. Future work will aim to enhance these tools for real-time tracking of emerging topics and broader application across other complex scientific domains.

DECLARATIONS

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. Informed consent to participate was not obtained from all of the participants in the study. Need for consent to participate was waived by an Institutional Review Board (IRB).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests

Funding

Not applicable

Authors' contributions

According to ICMJE

Acknowledgements

Not applicable

Disclosure of potential conflicts of interest: None

Research involving Human Participants and/or Animals: None

REFERENCES

Almeida, C. M., de Albuquerque, J. P., Zancul, E., & Sakurai, A. (2018). Mapping science through the Lattes Curriculum Platform in Brazil. *Scientometrics, 115* (2), 815–832.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media, 3* (1), 361– 362.

Börner, K., Klavans, R., Patek, M., Zoss, A.M., Biberstine, J.R., Light, R.P. (2012). Design and Update of a Classification System: The UCSD Map of Science. *PLoS ONE* 7(7): e39464.

Claveau, F. & Gingras, Y. (2016). Macrodynamics of Economics: A Bibliometric History. *History of Political Economy* 48, 4: 551-592.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57* (3), 359–377.

Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2011). Science mapping software tools: Review, analysis, and coop-erative study among tools. *Journal of the American Society for Information Science Technology*, 62, 1382-1402.

George, D. & Knedjens, R. (2013). Paperscape, an interactive map of the arXiv. Accessed Sept. 2017. http://paperscape.org/

Garfield, E. (2004). Historical structure of journal knowledge domains: Enhancing HistCite's global view with data from the Chemical Abstracts file. *Current Contents Connect, 17* (49), 3–9.

Grauwin, S. & Jensen, P. (2011). Mapping scientific institutions. *Scientometrics* 89:943-954.

Grauwin, S., Beslon, G., Fleury, E., Franceschelli, S., Robardet, C., Rouquier, J-B. & Jensen, P. (2012). Complex systems science: dreams of universality, interdisciplinarity reality. *J Am Soc Inf Sci Technol* 63(7):1327-1338.

Kang, H. et al. (2006). NetLens: Iterative exploration of Content-Actor Network Data. *Proceeding of IEEE Symposium on Visual Analytics Science and Technology 2006 (VAST2006)*, 91-98.

Herron, R. E., Bynum, D. W., & Francis, P. J. (2011). The Sci² tool: Extracting and visualizing temporal patterns from publicly funded research portfolios. *Proceedings of the 14th International Conference on Scientometrics and Informetrics (ISSI)* , 1, 231–242.

Leydesdorff, L. & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Infor-mation Science Technology*, 60(2), 348-362.

Sperano, I. (2017). Content Audit for the Assessment of Information Space: Definitions and Exploratory Typology. *SIGDOC 2017*, Halifax.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eidelman, Y., & Yessenalina, A. (2015). An overview of Microsoft academic graph. *Proceedings of the 24th International Conference on World Wide Web Companion* , 722–724.

Van Eck, N.J. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 2:523-538.

GPT-4'ün Hasta Notlarından Klinik Bulguları Belirleme Yetisinin Değerlendirilmesi: NBME Veri Seti ile Karşılaştırmalı Bir Çalışma

Esra Tokur Sonuvar[1[0000-0002-1279-5192]], Adem Sonuvar[1[0009-0004-2270-3812]], Nevruz İlhanlı[[0000-0002-4777-4025]] and Ahmet Yardımcı[1[0000-0001-7241-4913]]

[1] Biyoistatistik ve Tıp Bilişimi AD, Tıp Fakültesi, Akdeniz Üniversitesi Antalya.

Özet. Bu çalışma, GPT-4'ün serbest metin halindeki hasta notlarından klinik bulguları tanımlama yetisini değerlendirmeyi amaçlamaktadır. Kullanılan veri seti, NBME Score Clinical Patient Notes adlı kamuya açık ve Kaggle'da yayımlanan 42.146 satırlık bir klinik notlar koleksiyonudur. Ancak veri setindeki etiket eksiklikleri nedeniyle, rastgele seçilen 100 (hasta notu, klinik özellik) çifti için insan tarafından oluşturulmuş referans etiketleriyle değerlendirme yapılmıştır. Her bir örnek için GPT-4'e, ilgili klinik özelliğin hasta notunda açıkça belirtilip belirtilmediği sorulmuş ve modelden sadece "Evet" veya "Hayır" şeklinde yanıt vermesi istenmiştir. Elde edilen yanıtlar ikili (1/0) biçiminde kodlanarak doğruluk, duyarlılık, özgüllük, F1 skoru ve AUC gibi performans ölçütleri ile analiz edilmiştir. Sonuçlar, karışıklık matrisi, ROC eğrisi ve çubuk grafiklerle görselleştirilmiştir. Sonuçlar, GPT-4'ün yüksek duyarlılığa rağmen düşük özgüllük gösterdiğini ve klinik bulguların varlığını aşırı tahmin ettiğini ortaya koymuştur. Bu çalışma, büyük dil modellerinin klinik karar destek sistemlerine entegrasyonunda dikkatli değerlendirme ve denetim gerektirdiğini vurgulamaktadır.

Keywords: GPT-4, Klinik Doğal Dil İşleme, Hasta Notları, Klinik Özellik Tanıma, Tıbbi Yapay Zeka.

Abstract. This study aims to evaluate the capability of GPT-4 to identify clinical features from unstructured patient notes. The dataset used is the publicly available NBME Score Clinical Patient Notes, published on Kaggle, comprising 42,146 rows of clinical note-feature pairs. Due to the incompleteness of the original label column, a random sample of 100 (patient note, clinical feature) pairs was selected and manually annotated to serve as reference labels. For each sample, GPT-4 was prompted to determine whether the specified clinical feature was clearly mentioned in the patient note, and the model was instructed to respond strictly with "Yes" or "No." The responses were binarized (1/0) and analyzed using performance metrics including accuracy, recall, specificity, F1-score, and AUC. The results were visualized with a confusion matrix, ROC curve, and bar plots. The findings indicate that while GPT-4 demonstrated high recall, it also exhibited low specificity and a tendency to overpredict the presence of clinical features. This study highlights the need for careful evaluation and oversight when integrating large language models into clinical decision support systems.

Keywords: GPT-4, Clinical Natural Language Processing, Patient Notes, Clinical Feature Extraction, Medical Artificial Intelligence

1 Giriş

OpenAI tarafından geliştirilen GPT-4 gibi Büyük Dil Modellerinin (LLM) ortaya çıkışı, klinik Doğal Dil İşleme (NLP) alanında devrim yaratmış ve yapılandırılmamış tıbbi metinlerin yüksek semantik doğrulukla yorumlanmasını mümkün kılmıştır [1,2]. Klinik bilişimdeki temel zorluklardan biri, serbest metin hasta notlarından belirli klinik özelliklerin (örneğin "göğüs ağrısı", "nefes darlığı") varlığının veya yokluğunun yapılandırılmış biçimde çıkarılmasıdır [4]. Son çalışmalar, LLM'lerin tıbbi soru-cevap, triyaj görevleri ve hatta tıp lisans sınavı simülasyonlarında umut verici bir performans gösterdiğini ortaya koymuştur [2, 5]. Ayrıca, GPT-4'ün çok dilli tıbbi anlatımları farklı sağlık sistemleri arasında yorumlama yeteneği retrospektif olarak değerlendirilmiş ve dikkat çekici bulunmuştur [3]. Bununla birlikte, GPT-4'ün yapılandırılmamış klinik anlatımlarda belirli semptomların var olup olmadığını belirleme gibi daha ayrıntılı görevlerdeki pratik kullanımı hâlâ yeterince araştırılmamıştır [6]. LLM'lerin sağlık verileriyle eğitimi ve değerlendirilmesinde kullanılan MIMIC-III gibi kamuya açık anonimleştirilmiş veri setleri, bu alandaki çalışmalara önemli katkı sunmaktadır [16]. Bu çalışmada, GPT-4'ün NBME Klinik Hasta Notları veri setinin bir alt kümesi kullanılarak hasta öykülerinden belirli klinik özellikleri (ör. "göğüs ağrısı", "kalp çarpıntısı", "nefes darlığı") tespit etme yeteneği değerlendirilmektedir. Veri setinde her bir örnek, bir serbest metin hasta notu (pn_history) ve hedef klinik özelliği (feature_text) içermektedir. Etiketli örneklerin çoğunda referans bilgisi eksik olduğundan, hasta notu ile klinik özellik arasında açık bir sözel ya da anlamsal uyum varsa bu özelliğin var olduğu varsayımına dayanan bir referans etiketleme yöntemi geliştirilmiştir [7]. GPT-4, OpenAI API üzerinden sistematik biçimde sorgulanarak her not-özellik çifti için belirli bir özelliğin hasta notunda geçip geçmediğini değerlendirmesi istenmiştir. Tahminler manuel referans etiketlerle karşılaştırılmış ve doğruluk, hassasiyet, duyarlılık, F1 skoru gibi sınıflandırma metrikleri ile birlikte nitel hata analizleri yapılmıştır [8, 9].
Bu pilot çalışma, GPT-4'ün klinik bilgi çıkarımı görevlerindeki güçlü ve zayıf yönlerini ortaya koyarak, bu modellerin elektronik sağlık kayıt sistemlerinde ve karar destek uygulamalarında nasıl değerlendirilebileceğine dair fikir sunmayı amaçlamaktadır.

2 Yöntem

Bu çalışmada, Kaggle platformunda yayımlanmış olan ve kamuya açık bir kaynak niteliği taşıyan NBME Score Clinical Patient Notes veri seti kullanılmıştır (https://www.kaggle.com/competitions/nbme-score-clinical-patient-notes/overview). Veri seti, her biri bir hasta notu (pn_history) ve hedef bir klinik özelliği (feature_text) içeren toplam 42.146 kimliksizleştirilmiş kayıt içermektedir. Ancak, orijinal veri setindeki ikili etiket sütununda yaygın eksikliklerin bulunması nedeniyle, değerlendirme

amacıyla 100 adet (not, özellik) çifti rastgele örneklenmiş ve bu örnekler için manuel referans etiketleri oluşturulmuştur. Bu etiketler, ilgili klinik özelliğin hasta notunda açık bir şekilde belirtilip belirtilmediğine göre 1 (var) veya 0 (yok) olarak atanmıştır. Her bir örnek çifti, OpenAI GPT-4 modeli kullanılarak sorgulanmış ve tutarlı bir istem şablonu (prompt) ile değerlendirilmiştir. Uygulanan standart sorgu biçimi şu şekildedir: "The patient note is: '[pn_history]'. Does this note clearly mention the feature '[feature_text]'? Please respond with 'Yes' or 'No'."
GPT-4 tarafından verilen yanıtlar "Yes" veya "No" olarak alınmış ve analiz için ikili değerlere (1 = Evet, 0 = Hayır) dönüştürülmüştür. Veri örnekleme, GPT-4 ile sorgulama ve değerlendirme adımlarını içeren süreç Şekil 1'de şematik olarak sunulmuştur. Model tahminleri, araştırmacılar tarafından oluşturulan referans etiketleriyle karşılaştırılmış ve doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve ROC-AUC gibi standart sınıflandırma performans ölçütleri kullanılarak değerlendirilmiştir. Sonuçlar, bir karışıklık matrisi (Şekil 1), etiket dağılım grafikleri (Şekil 2) ve ROC eğrisi (Şekil 3) aracılığıyla görselleştirilmiştir. Tüm veri işleme ve analiz adımları, Python 3.10 sürümü kullanılarak pandas, scikit-learn ve matplotlib kütüphaneleri ile gerçekleştirilmiştir. GPT-4 modeli ile gerçekleştirilen tüm işlemler, OpenAI Python API aracılığıyla ve gpt-4-0314 sürümü kullanılarak yapılmıştır. Çalışmada, 100 benzersiz not-özellik çifti için toplam 176 adet bireysel istek (API çağrısı) gerçekleştirilmiş; bu işlemler sonucunda yaklaşık 44.223 token kullanılmıştır. OpenAI kullanım paneline göre (Nisan 2025), yapılan sorguların toplam maliyeti 1,92 ABD doları olarak hesaplanmıştır. Bu maliyet, kullanılan sorgu uzunluğuna, yanıt boyutuna ve model sürümüne göre değişkenlik gösterebilir.

3 Bulgular

Bu çalışmada, GPT-4 modelinin klinik özellikleri tanıma performansını değerlendirmek amacıyla 100 adet (hasta notu, klinik özellik) çifti incelenmiştir. Modelin tahminleri, manuel olarak oluşturulmuş referans etiketlerle karşılaştırılmıştır. Elde edilen karışıklık matrisi sonuçlarına göre (Şekil 2), GPT-4, iki doğru pozitif (true positive) ve 39 doğru negatif (true negative) tahmin üretmiş, 59 yanlış pozitif (false positive) sınıflandırma yapmış ve hiç yanlış negatif (false negative) üretmemiştir. Model, 39 doğru negatif (true negative) tahmin üretmiş ve dikkat çekici bir şekilde hiç yanlış negatif (false negative) üretmemiştir. Bu bulgular, GPT-4'ün yüksek duyarlılığa (recall), ancak düşük kesinliğe (precision) sahip olduğunu göstermektedir.
Veri setindeki etiket dağılımı belirgin şekilde dengesizdir; insan anotasyonlarına göre 100 örnekten 97'si "özellik yok" (negatif) olarak işaretlenmiştir. Buna karşın GPT-4, 61 örneği pozitif ve 39'unu negatif olarak sınıflandırmıştır. Bu fark, modelin klinik özelliklerin varlığını olduğundan fazla tahmin etme eğiliminde olduğunu ortaya koymaktadır. Etiketlerin dağılımını görselleştiren çubuk grafikler (Şekil 3) bu durumu desteklemektedir.
ROC eğrisi üzerinden hesaplanan AUC değeri 0.70 olarak bulunmuştur (Şekil 4). Bu değer, modelin pozitif ve negatif sınıfları ayırt etme yetisinin orta düzeyde olduğunu

göstermektedir. Modelin, belirli klinik özellikleri tespit etme potansiyeline sahip olduğu anlaşılmakla birlikte, pozitif sınıflandırmalardaki aşırı eğilimi, doğrudan klinik uygulamalarda kullanımını sınırlandırabilir.
Genel olarak, GPT-4'ün semantik olarak zengin ancak bağlamsal olarak belirsiz ifadeleri yorumlarken genelleme eğiliminde olduğu, bu nedenle klinik özelliklerin varlığını aşırı şekilde tahmin ettiği gözlemlenmiştir. Her ne kadar yanlış negatif bulunmaması model açısından umut verici bir durum olsa da, yüksek oranlı yanlış pozitif tahminler, modelin klinik karar destek sistemlerinde tek başına güvenilir biçimde kullanılmasını sınırlamaktadır. Bu nedenle, daha dengeli ve kapsamlı veri setleri üzerinde yapılacak ileri düzey çalışmalarla modelin genellenebilirliğinin ve güvenilirliğinin yeniden değerlendirilmesi önerilmektedir.

4 Tartışma

Bu çalışmanın bulguları, Büyük Dil Modellerinin (LLM'ler), özellikle de GPT-4'ün klinik doğal dil işleme (NLP) alanındaki kullanımına ilişkin giderek büyüyen araştırma literatürüne katkı sağlamaktadır [1, 2]. GPT-4, klinik notlardan semptom çıkarımı konusunda orta düzeyde ayırt edici bir yetenek sergilemiş (AUC = 0.70) olsa da, 100 (hasta notu, özellik) çifti üzerinde yapılan değerlendirmeler, modelin özellikle kesinlik (precision) açısından belirgin sınırlılıklarını ortaya koymuştur. Modelin, insan anotasyonlarında yer almayan klinik özelliklerin varlığını aşırı tahmin ettiği ve bu nedenle yüksek oranda yanlış pozitif (n = 59) ürettiği gözlemlenmiştir [4]. Bu bulgu, önceki çalışmalarla da uyumludur; benzer araştırmalarda, LLM'lerin belirsiz bağlamlarda genelleme yapma veya "halüsinasyon" üretme eğiliminde oldukları gösterilmiştir [2, 14]. Her ne kadar Med-PaLM gibi bazı LLM'ler tıbbi soru-cevap görevlerinde uzman düzeyinde performans göstermiş olsa da [1], bu çalışmada incelenen gibi daha ayrıntılı klinik özellik çıkarımı görevleri, yalnızca dil yeterliliğiyle değil; aynı zamanda tıbbi muhakemeye dayalı bağlamsal yorumlama becerisiyle de ilişkili olduğu için daha zordur [7, 8].
Bu çalışmada gözlemlenen aşırı tahmin (overprediction), büyük ölçüde GPT-4'ün olası bir dil üretimini önceliklendiren jeneratif model mimarisinden kaynaklanıyor olabilir [6, 15]. Ayrıca, değerlendirme için kullanılan veri setinde pozitif örneklerin oranının yalnızca %3 olması, modelin nadir sınıfa dair yeterli örnekle beslenememesi nedeniyle davranışını etkilemiş olabilir [9].
Tüm bu sınırlamalara rağmen, GPT-4'ün yanlış negatif içermemesi (tüm pozitif örnekleri doğru tahmin etmesi) dikkate değerdir ve bu durum, modelin özellikle ön tarama (screening) ya da triyaj gibi, yanlış negatiflerin yüksek risk oluşturduğu senaryolarda yardımcı bir araç olarak kullanılabileceğini düşündürmektedir [3, 11]. Bu durum, literatürde giderek yaygınlaşan ve LLM'lerin bağımsız karar verici değil, destekleyici sistemler olarak değerlendirilmesi gerektiğini vurgulayan görüşlerle örtüşmektedir [10, 15].
Bununla birlikte, bu çalışmada kullanılan veri setindeki sınıf dengesizliği ve evet/hayır biçiminde yapılandırılmış istemlerin (prompt) sınırlılığı, elde edilen sonuçların genellenebilirliğini azaltmaktadır. Gelecekteki çalışmalar, GPT-4'ün uzman tıbbi muhakeme

ile daha uyumlu sonuçlar üretmesini sağlamak için kalibrasyon stratejileri, istem mühendisliği (prompt engineering) ve ince ayar (fine-tuning) yaklaşımlarını araştırmalıdır [4, 13]. Ayrıca, alanına özgü LLM'lerin geliştirilmesi ya da yapılandırılmış klinik bilgi tabanlarının entegrasyonu, modelin yorumlanabilirliğini artırabilir ve tıbbi uygulamalarda hatalı bilgi üretiminin önüne geçebilir [5, 12].

5 Sınırlılıklar

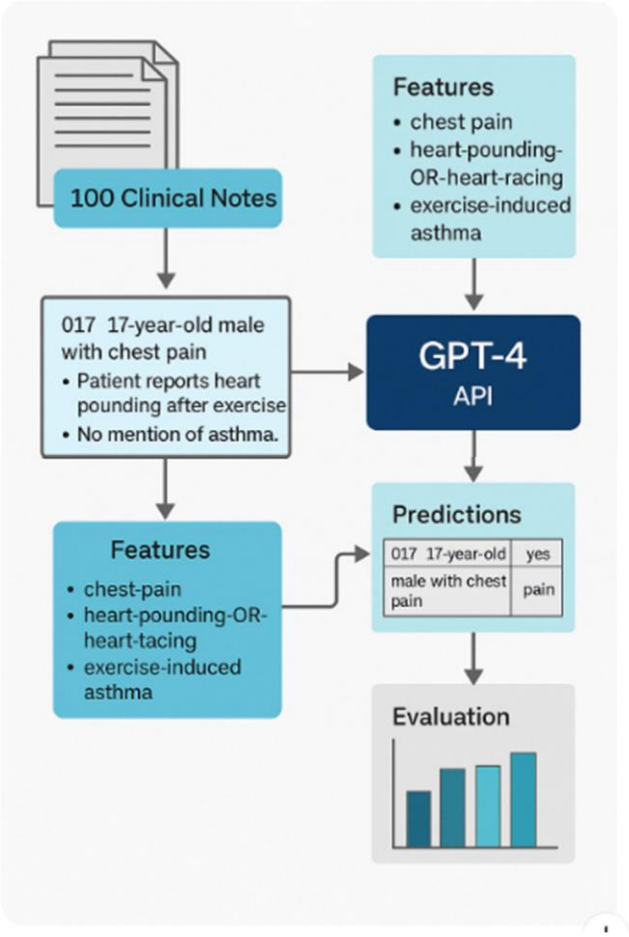
Bu çalışmanın çeşitli sınırlılıkları bulunmaktadır. İlk olarak, değerlendirme yalnızca 100 (hasta notu, klinik özellik) çifti üzerinden gerçekleştirilmiştir; bu örneklem, klinik dilin çeşitliliğini tam anlamıyla yansıtmayabilir. İkinci olarak, kullanılan veri seti ciddi şekilde dengesizdir; pozitif örneklerin azlığı, kesinlik (precision) ve duyarlılık (recall) gibi performans metriklerini olumsuz yönde etkilemiş olabilir. Üçüncü olarak, GPT-4'e yöneltilen istem (prompt) formatı oldukça basittir ve gerçek klinik uygulamalardaki karmaşıklığı yeterince temsil etmemektedir. Son olarak, insan anotasyonları arasında tutarlılık ölçülmediği için modelin klinik uzmanlıkla ne derece uyumlu olduğunu kesin olarak değerlendirmek mümkün olmamıştır.

6 Sonuç

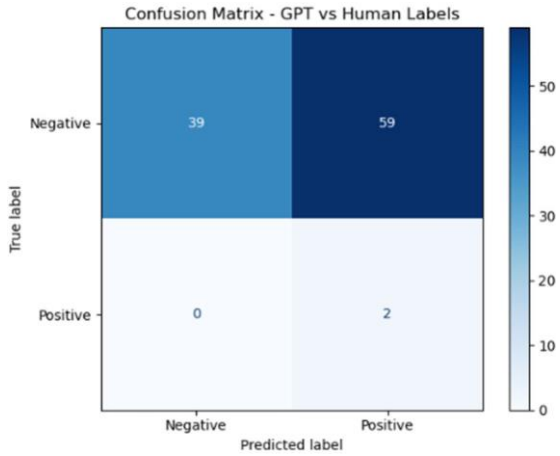
Bu çalışmada, GPT-4'ün serbest metin biçimindeki hasta notlarından klinik özellikleri belirleyebilme kapasitesi, NBME veri setindeki uzman anotasyonlarıyla karşılaştırılarak incelenmiştir. Sonuçlar, modelin AUC değeri 0.70 (Şekil 4) ile makul bir ayırt edici performans gösterdiğini ortaya koysa da, yüksek düzeyde yanlış pozitif üretmesi (Şekil 2) mevcut haliyle klinik uygulamalarda kullanılabilirliğini sınırlamaktadır. Pozitif sınıfın aşırı tahmini (Şekil 3), GPT-4'ün belirsiz veya örtük ifadeleri dahi pozitif olarak yorumlamaya eğilimli olduğunu ve bunun da uzman kriterleriyle tam olarak örtüşmediğini göstermektedir.

Modelin %100 duyarlılık (recall) elde etmesi olumlu bir özellik olsa da, düşük kesinlik (precision) değeri, klinik uygulamalarda duyarlılık ve özgüllük (specificity) arasında dikkatle yönetilmesi gereken bir dengeye işaret etmektedir. Değerlendirme veri setinde pozitif örneklerin nadir olması, elde edilen sonuçların daha büyük ve dengeli klinik veri kümelerinde tekrar test edilmesi gerektiğini vurgulamaktadır.

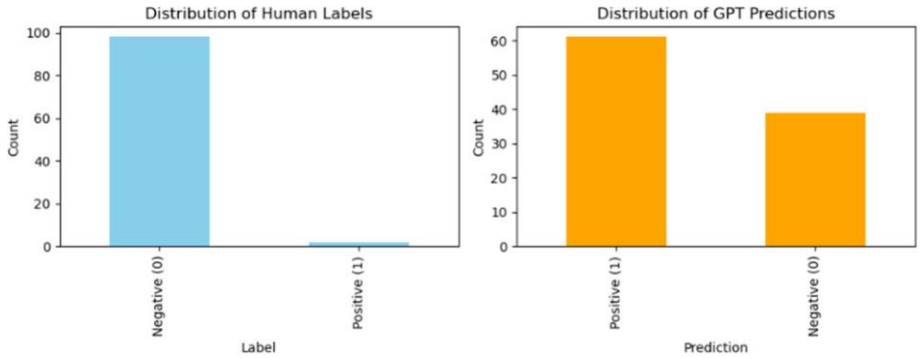
Sonuç olarak, GPT-4 klinik bilgi çıkarımı görevlerinde umut vadeden bir potansiyele sahip olmakla birlikte, özellikle klinik özellik tespiti gibi hassas görevlerde güvenli biçimde kullanılabilmesi için bağlamsal uyum ve model ayarlaması gereklidir.



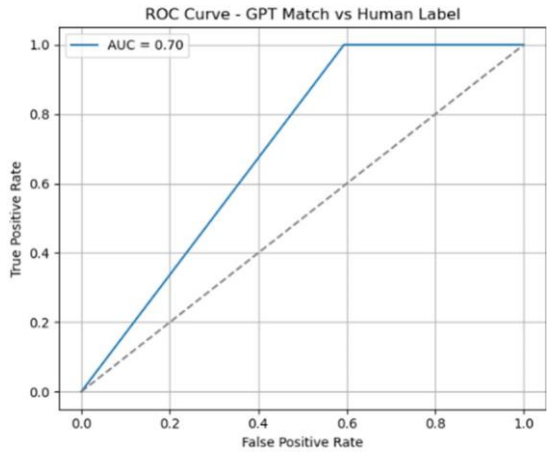
Şekil 1. GPT-4 Tabanlı Klinik Özellik Belirleme Sürecinin İş Akışı



Şekil 2. GPT-4 Tahminleri ile İnsan Etiketleri Arasındaki Karışıklık Matrisi



Şekil 3. İnsan Etiketleri ve GPT-4 Tahminlerinin Dağılımı



Şekil 4. GPT-4 Tahminleri ile İnsan Etiketleri Arasındaki ROC Eğrisi (AUC = 0.70)

References

1. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera Y Arcas, B., … Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2
2. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems (arXiv:2303.13375). arXiv. https://doi.org/10.48550/arXiv.2303.13375
3. Menezes, M. C. S., Hoffmann, A. F., Tan, A. L. M., Nalbandyan, M., Omenn, G. S., Maz- zotti, D. R., Hernández-Arango, A., Visweswaran, S., Venkatesh, S., Mandl, K. D., Bour- geois, F. T., Lee, J. W. K., Makmur, A., Hanauer, D. A., Semanik, M. G., Kerivan, L. T., Hill, T., Forero, J., Restrepo, C., Vigna, M., … Consortium for Clinical Characterization of COVID-19 by Electronic Health Records (2025). The potential of Generative Pre-trained Transformer 4 (GPT-4) to analyse medical notes in three different languages: a retrospective model-evaluation study. The Lancet. Digital health, 7(1), e35–e43. https://doi.org/10.1016/S2589-7500(24)00246-2
4. Zhang, D.; Xue, X.; Gao, P.; Jin, Z.; Hu, M.; Wu, Y.; Ying, X. A survey of datasets in medicine for large language models. Intell. Robot. 2024, 4(4), 457-78. http://dx.doi.org/10.20517/ir.2024.27
5. Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS digital health, 2(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198
6. Wang, D., Zhang, S. Large language models in medical and healthcare fields: applications, advances, and challenges. Artif Intell Rev 57, 299 (2024). https://doi.org/10.1007/s10462- 024-10921-0
7. Boag, W., Doss, D., Naumann, T., & Szolovits, P. (2018). What's in a Note? Unpacking Predictive Value in Clinical Note Representations. AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2017, 26–34.
8. Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. BMJ (Clinical research ed.), 309(6947), 102. https://doi.org/10.1136/bmj.309.6947.102
9. Croxford, E., Gao, Y., Pellegrino, N., Wong, K., Wills, G., First, E., Liao, F., Goswami, C., Patterson, B., & Afshar, M. (2025). Current and future state of evaluation of large language models for medical summarization tasks. npj health systems, 2, 6. https://doi.org/10.1038/s44401-024-00011-2
10. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. The New Eng- land journal of medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259
11. Henry, S., Wang, Y., Shen, F., & Uzuner, O. (2020). The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. Journal of the American Medical Infor- matics Association : JAMIA, 27(10), 1529–1537. https://doi.org/10.1093/jamia/ocaa106
12. Lehman, E., Jain, S., Pichotta, K., Goldberg, Y., & Wallace, B. (2021). Does BERT pre- trained on clinical notes reveal sensitive data? In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Lan- guage Technologies (pp. 946–959). Association for Computational Linguistics. https://aclanthology.org/2021.naacl-main.73/.
13. Anisuzzaman, D. M., Malins, J. G., Friedman, P. A., & Attia, Z. I. (2024). Fine-Tuning Large Language Models for Specialized Use Cases. *Mayo Clinic proceedings. Digital health*, *3*(1), 100184. https://doi.org/10.1016/j.mcpdig.2024.11.005
14. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). *A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions* (Version 2). arXiv. https://doi.org/10.48550/arXiv.2311.05232.
15. Hu, Y., Chen, Q., Du, J., Peng, X., Keloth, V. K., Zuo, X., Zhou, Y., Li, Z., Jiang, X., Lu, Z., Roberts, K., & Xu, H. (2024). Improving large language models for clinical named entity recognition via prompt engineering. *Journal of the American Medical Informatics Associa- tion : JAMIA*, *31*(9), 1812–1820. https://doi.org/10.1093/jamia/ocad259
16. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, *3*, 160035. https://doi.org/10.1038/sdata.2016.35.

16. Tıp Bilişimi Kongresi
22-23 Mayıs 2025, ANKARA
Ana Tema: Sağlıkta Multiomiks

16.th Medical Informatics Congress
Main Theme: Multiomics in Healthcare
https://turkmia.net/kongre2025

ISBN:978-605-72564-2-3