



KARADENİZ
TEKNİK ÜNİVERSİTESİ
KARADENİZ TECHNICAL UNIVERSITY
1955

15. Tıp Bilişimi Kongresi

30-31 MAYIS 2024, TRABZON

Ana Tema: Sağlıkta Yapay Zeka ve Büyük Dil Modelleri

BİLDİRİLER KİTABI

15th Medical Informatics Congress

Main Theme: Artificial Intelligence and Large Language Models in Healthcare

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



TÜBİTAK

<https://turkmia.net/kongre2024>

TIP BİLİŞİMİ DERNEĞİ YAYINLARI

Sertifika No: 7163

Yayın Tarihi: 05/09/2024 Yayın No:2

15. Tıp Bilişimi Kongresi Bildiriler Kitabı

Proceedings of 15th 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

Dijital olarak basılmış ve çevrimiçi olarak dağıtılmıştır.
348 sayfadır.

Ücretsizdir.

© 2024 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 978-605-72564-1-6

ISBN 978-605-72564-1-6



9 786057 256416 >

15. TIP BİLİŞİMİ KONGRESİ

15th Turkish Congress of Medical Informatics Association

30-31 Mayıs 2024, Trabzon, Türkiye

Karadeniz Teknik Üniversitesi
Osman Turan Kültür ve Kongre Merkezi

BİLDİRİLER KİTABI

Proceedings

Editörler *Editors*

Kemal Turhan, Ahmet Yardımcı, Adil Alpkoçak, Abdullah Utku Şenol, Aslı Suner Karakulah, Kemal Hakan Gülkesen, Uğur Bilge, Oğuz Dicle, Filiz İşleyen, Muammer Albayrak, Mehmet Erşan Kalaycı

Organizasyon *Organized by*



TÜBİTAK



Destekleyen Kuruluşlar *In Cooperation with*



Ana Sponsor *Main Sponsored by*



Sponsorlar *Sponsored by*



KONGRE BAŐKANI Congress Chair

Kemal TURHAN, Karadeniz Teknik Üniversitesi

DÜZENLEME KURULU İ Editorial Board

Kemal TURHAN, Karadeniz Teknik Üniversitesi

Ahmet Yardımcı, Akdeniz Üniversitesi

Filiz İşleyen, Sağlık Bakanlığı

Kemal Hakan Gülkesen, Akdeniz Üniversitesi

Adil Alpkoçak, İzmir Bakırçay Üniversitesi

Abdullah Utku Şenol, Akdeniz Üniversitesi

Aslı Suner Karakülah, Ege Üniversitesi

Oğuz Dicle, Dokuz Eylül Üniversitesi

Neşe Kaklıkaya, Karadeniz Teknik Üniversitesi

Ali Güner, Karadeniz Teknik Üniversitesi

Burçin Kurt, Karadeniz Teknik Üniversitesi

Tolga Berber, Karadeniz Teknik Üniversitesi

Cavit Boz, Karadeniz Teknik Üniversitesi

Sedat Bostan, Karadeniz Teknik Üniversitesi

Cemil Çolak, İnönü Üniversitesi

Mehmet Kıvrak, Recep Tayyip Erdoğan Üniversitesi

Nurhan Gürsel Özmen, Karadeniz Teknik Üniversitesi

Sibel Kul, Karadeniz Teknik Üniversitesi

Uğur Şevik, Karadeniz Teknik Üniversitesi

Engin Yenilmez, Karadeniz Teknik Üniversitesi

Mine Duman, Karadeniz Teknik Üniversitesi

KONGRE SEKRETERİ Congress Secretary

Muammer Albayrak, Karadeniz Teknik Üniversitesi

Mehmet Erşan Kalaycı, Karadeniz Teknik Üniversitesi

Sinem Özdemir, Karadeniz Teknik Üniversitesi

Zeliha Aydın Kasap, Karadeniz Teknik Üniversitesi

BİLİMSEL KURUL Scientific Committee

Ahmet Yardımcı, Akdeniz Üniversitesi (Başkan)

Kemal Turhan, Karadeniz Teknik 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
Barış Emre Dayanç, İzmir Ekonomi Üniversitesi
Burçin Kurt, Karadeniz Teknik Üniversitesi
Ebru Akçapınar Sezer, Hacettepe Üniversitesi
Esin Öztürk Işık, Boğaziçi Üniversitesi
Filiz İşleyen, Sağlık Bakanlığı
Gizem Tatar Yılmaz, Karadeniz Teknik 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
Leman Tomak, Ondokuz Mayıs Üniversitesi
Mehmet Serdar Güzel, Ankara Üniversitesi
Mehmet Ünlütürk, Yaşar Üniversitesi
Menderes Tarcan, Osmangazi Üniversitesi
Murat Osman Ünalır, Ege Üniversitesi
Mustafa Özmen, Hacettepe Üniversitesi
Nese Zayim, Akdeniz Üniversitesi
Oğuz Dicle, Dokuz Eylül Üniversitesi
Orhan Er, İzmir Bakırçay Üniversitesi
Öznur Özkasap, Koç Üniversitesi
Pınar Özbay, Boğaziçi Üniversitesi
Pembe Keskinoglu, Dokuz Eylül Üniversitesi
Semih Utku, Dokuz Eylül Üniversitesi
Suat Özdemir, Hacettepe Üniversitesi
Sibel Kul, Karadeniz Teknik Üniversitesi
Timur Köse, Ege 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
Cemil Çolak, İnönü Üniversitesi
Mustafa Necmi İlhan, Gazi Üniversitesi
Engin Yenilmez, Karadeniz Teknik Üniversitesi
Nurhan Gürsel Özmen, Karadeniz Teknik Üniversitesi
Cengizhan Öztürk, Boğaziçi Üniversitesi
Şükrü Özen, Akdeniz Üniversitesi
Murat Canpolat, Akdeniz Üniversitesi
Tuncay Aydoğan, Isparta Uygulamalı Bilimler Üniversitesi
Gökhan Ertay, Yeditepe Üniversitesi
Azim Çelik, GE Healthcare
Yiğit Ali Üncü, Akdeniz Üniversitesi
Kılıçhan Gürleyik, University of Pittsburgh
Selen Bozkurt, Emory Üniversitesi
Yılmaz Kemal Yüce, Alanya Alaaddin Keykubat Üniversitesi
Mustafa Mahir Ülgü, Sağlık Bakanlığı
Mehmet Kıvrak, Recep Tayyip Erdoğan Üniversitesi
Fatih Gürcan, Karadeniz Teknik Üniversitesi
Tolga Berber, Karadeniz Teknik Üniversitesi
Ali Güner, Karadeniz Teknik Üniversitesi
Neşe Zaim, Akdeniz Üniversitesi
Uğur Şevik, Karadeniz Teknik Üniversitesi
Engin Yenilmez, Karadeniz Teknik Üniversitesi
Mine Duman, Karadeniz Teknik Üniversitesi
Cavit Boz, Karadeniz Teknik Üniversitesi
Tunca Doğan, Hacettepe Üniversitesi
Emel Timuçin, Acıbadem Mehmet Ali Aydınlar Üniversitesi
Osman Uğur Sezerman, Acıbadem Mehmet Ali Aydınlar Üniversitesi

YEREL DÜZENLEME KOMİTESİ
Local Organizing Committee

İlknur Buçan Kırkbir, Karadeniz Teknik Üniversitesi

Gülbahar Merve Çakmak Şilbir, Karadeniz Teknik Üniversitesi

Ayşe Banu Pak, Karadeniz Teknik Üniversitesi

Nihat Burak Zihni, Karadeniz Teknik Üniversitesi

Işık Çakmak, Karadeniz Teknik Üniversitesi

Tuğba Kurt, Karadeniz Teknik Üniversitesi

YEREL ÖĞRENCİ DÜZENLEME KOMİTESİ
Local Student Organizing Committee

İlknur Buçan Kırkbir, Karadeniz Teknik Üniversitesi

Ahmat Safa Çolak, Karadeniz Teknik Üniversitesi

Berk Bayraktar, Karadeniz Teknik Üniversitesi

Beyza Naz Hakyemez, Karadeniz Teknik Üniversitesi

Buket Gürler, Karadeniz Teknik Üniversitesi

Elif Aydın, Karadeniz Teknik Üniversitesi

Emir Bahadır Kadioğlu, Karadeniz Teknik Üniversitesi

İclal Oruç, Karadeniz Teknik Üniversitesi

Melisa Kavalcı, Karadeniz Teknik Üniversitesi

Mert Can Alpaydın, Karadeniz Teknik Üniversitesi

Rabia Kalyon, Karadeniz Teknik Üniversitesi

Rabia Tomruk, Karadeniz Teknik Üniversitesi

Yiğit Emre Kokoç, Karadeniz Teknik Üniversitesi

Zehra Şirin, Karadeniz Teknik Üniversitesi

GRAFİK TASARIM
Graphic Design

Belgin İpek, Karadeniz Teknik Üniversitesi

İLETİŞİM ADRESİ E-mail Address
kongre2024@turkmia.net

KONGRE WEB Congress Web
<https://turkmia.net/kongre2024/>

İçindekiler <i>Contents</i>	Sayfa No <i>Page #</i>
Önsöz <i>Foreword</i>	<i>xi</i>
LSTM Tabanlı Derin Öğrenme Yaklaşımı ile Bayesian Hiper-Parametre Optimizasyonu Kullanılarak Hasta Şikayetlerinden Hastalık Tahmini Aslı Boz*, Duygu Korkmaz Yalçın, Yasin Görmez, Neşe Zayim	<i>1</i>
DrugAutoML: İlaç Keşfi için Açık Kaynaklı, Uçtan Uca Otomatikleştirilmiş Bir Makine Öğrenimi ve İstatistiksel Değerlendirme Aracı Ayça BEYHAN*, Aslı SUNER	<i>6</i>
In Silico Screening of Momordica Charantia Compounds for the Treatment of Type 2 Diabetes Mellitus: Molecular Docking Ayşe Banu PAK*, Işık ÇAKMAK, Mustafa Emre ERCİN	<i>8</i>
Interpreting Chest X-ray with ChatGPT: Can It Serve as a Tool for Justifying Computed Tomography? Nur Hürsoy*, Hafsa Kolluk, Merve Solak, Kubilay Kağan Budak, Esat Kaba	<i>22</i>
Sağlık Alanındaki Çalışmalarda Yardımcı Bir Araç Olarak Yapay Zeka'nın Kullanımının Değerlendirilmesi: ChatGPT örneği Yasemin Demir Avcı*, Esra Tokur Sonuvar, Dilek Yapar	<i>32</i>
A Comparative Study of Parkinson's Disease Detection using different Deep Learning Architectures Beyza GÜL*	<i>44</i>
Prediction of retinopathy through machine learning in diabetes mellitus Tarık Keçeli*, Nevruz İlhanlı, Kemal Hakan Gülkesen	<i>56</i>
Epilepside Kopya Sayısı Varyasyonlarının Tüm Ekzom Verileri Kullanılarak Araştırılması Ezgi Özyurt*, Alper Han Çebi, Ali Cansu, Ayberk Türkyılmaz, Kemal Turhan, Pınar Özkan Kart	<i>65</i>
B Cell Epitope Analysis of the SARS-CoV-2 Spike Protein: Accessibility, Polymorphisms, and Implications for Vaccine Design Mehmet Erşan KALAYCI*, Güven ŞAHİN, Muammer ALBAYRAK, Kemal TURHAN	<i>74</i>
Parkinson's Disease Related Alterations in Common Spatial Patterns of EEG Signals Fatih Onay, Bilge Karaçalı	<i>76</i>
TheraPose: A Large Video Dataset for Physiotherapy Exercises Hamdi Yalın Yalçın*, Ahmet Emre Usta, Ümit Atıla, Alaettin Uçan, Adem Ali Yılmaz, Ali Yaşar Yiğit	<i>88</i>
Gamification in Health: A Bibliometric Perspective Esra Tokur Sonuvar*, Kemal Hakan Gülkesen	<i>98</i>
DMFT ASSESSMENT USING DEEP LEARNING BASED CONVOLUTIONAL NEURAL NETWORK ALGORITHM WITH PANORAMIC IMAGES Onur MUTLU*, Elif ASLAN, Canberk ULUSOY, Ali MERT	<i>105</i>
Identifying Malignancy Risk Factors in Thyroid Nodules: A Machine Learning Approach in Trabzon, Türkiye Zeliha Aydın Kasap*, Burçin Kurt, Ali Güner, Elif Özsağır, Mustafa Emre Ercin	<i>110</i>
Klinik Metinlerden İlişki Çıkarma için Bağlam İçi Öğrenme Aidana Aidynkyzy*, Şebnem Bora, Oğuz Dikenelli	<i>111</i>

Bilgi Güvenliği Yönetim Sistemi Kurulumu: Sağlık Bakanlığı Deneyimi Filiz İşleyen*, Mustafa Mahir Ülgü, Dilek Şen Karakaya, Betül Yılmaz	120
Tıp Alanında Yapay Zekâ Kullanılarak Yapılan Klinik Çalışmaların Bibliyometrik Analizi Hülya Kılbahri*, Pınar Deniz Tosun	132
Predictors of Neuropathy in Diabetes Mellitus Nevruz İlhanlı*, Kemal Hakan Gülkesen	141
Fine-Tuning GPT-2 for Contextually Relevant Text Generation to Discover New Immune Checkpoint Research Mehmet Erşan KALAYCI*, Muammer ALBAYRAK, Kemal TURHAN	147
Integration of Non-Negative Matrix Factorization to Identification of Immune System Checkpoints Beyzanur SİYAH*, Mehmet Erşan KALAYCI, Sinem ÖZDEMİR, Tolga BERBER, Kemal TURHAN	158
Siroz Hastalarında Hepatik Ensefalopati için Semptom Tespit ve İzleme Aracı Prototipinin Arayüz Tanıtımı Dilek YAPAR*, Sıdıka Zeynep GÖK SARGIN, Ahmet YARDIMCI	171
Cancer Prediction via Binary Tuna Swarm Optimization Muharrem Korkmaz*, Eda Özkul, Özge Tezel	180
Improvement of Encoder Layers in Segmentation Network Models for Automatic Polyp Detection Nurhan DEĞİRMENÇİ BİNGÖL, Özkan BİNGÖL, Kemal TURHAN	185
Using Meta-Heuristic Techniques to Estimate Logistic Regression Parameters and Applying to Health Data Orhan KESEMEN, Buket Nur ALTUN*	186
Diagnosing Retinal Disorders with Artificial Intelligence: The Role of Large Language Models in Interpreting Pattern Electroretinography Data Aslan Aykut *, Büşra Akgün, Almıla Sarıgül Sezenöz , Mehmet Orkun Sevik, Özlem Şahin	191
Investigation of the Inhibition Mechanism of Acetylcholinesterase Enzyme for Alzheimer's Treatment by Molecular Docking Method Merve Baldan*, Gizem Tatar Yılmaz	201
CAN LARGE LANGUAGE MODELS SURPASS HUMAN PERFORMANCE IN THE MEDICAL SPECIALIZATION EXAM? Yeşim Aygül*, Müge Oluçoğlu, Adil Alpkoçak	202
Detection of psychological disorders from structural MRI with deep learning methods Osman Tayfun Bişkin*, Cemre Candemir, Mustafa Alper Selver, Pembe Keskinöglü	206
Ses Dalgalarından Öfke Belirleme İçin Evrişimli Sinir Ağı Modeli Mehmet YİĞİT*, Mehmet Erşan KALAYCI, Kemal TURHAN	212
An Example Application For Machine Learning-Assisted Retrosynthesis Güven ŞAHİN*, Mehmet Erşan KALAYCI, Kemal TURHAN	223
Pulmonary Nodule Decision Support Model Using Convolutional Neural Network-Based Transfer Learning Approach Hilal TİRYAKİ*, Burçin KURT, Polat KOŞUCU, Selçuk AKKAYA	225
Radyoloji Uzmanlık Eğitiminde Yapay Zekanın Yeri: Ulusal Anket Çalışması Emre Emekli*, Özlem Coşkun, Işıl İrem Budakoğlu	226
Tümör Tedavi Alanları (Ttfields) Terapisinin Glioblastomada (Gbm) Hayatta Kalma ile İlişkisi: Sistematik Bir İnceleme ve Meta-Analiz	232

Mehmet KIVRAK , Hamit YILMAZ	
Analysis of Findings in Radiology Reports Using Non- Negative Matrix Factorization Emir KARAYAĞIZ*, Tolga BERBER	235
In Silico Screening of the Interaction of Factor Xa and Melilotus Officinalis Phytochemicals Işık ÇAKMAK*, Ayşe Banu PAK, Mustafa Emre ERCİN	244
Parkinson Hastalığının Teşhisinde EEG-F1 Kanal Aktivitesinin Rolü Neslihan BAKİ, Nurhan GÜRSEL ÖZMEN	252
A Bibliometric Analysis of the Use of Machine Learning Methods in Variant Effect Prediction Gülbahar Merve ŞILBİR*, Burçin KURT	260
Transforming Human-Computer Interaction: Gesture Recognition Technologies for Accessible Smart Home Ufuk Beşenk*, Ahmet Özkurt, Ömür Öcal, Tuğçe Toprak, Pembe Keskinoglu, M. Alper Selver	261
Türkiye’de Tıp Bilişimi Alanında Üretilen Tezlerin Analizi Hazal TAŞ*, Neşe ZAYİM	273
The Inclusion of Medical Informatics Topics in the Curricula of Medical Faculties in Turkey: A Document Analysis Study Ayşenur Duman Dilbaz*, Yasemin Güner, Melek Şahin, Barış Sezer	280
An Ethical Examination of ChatGPT Use in the Field of Medicine Belkıs Birden*	284
Statistical Analysis of Physiological Stress Signals and Exam Performance Ayşegül K. Kasap*, Burçin KURT	290
OFTALMOLOJİDE YAPAY GÖRME (RETİNA İMPLANTI) ÜZERİNE YAPILAN ÇALIŞMALAR Evren Sezgin*, Ahmet Yardımcı	291
Derin Öğrenme Öznelik Çıkarım Algoritmalarının Makine Öğrenmesi Modelleri ile Kullanılması: Diyabetik Retinopati Örneği Fatih Serdar AYDEMİR*, Uğur ŞEVİK	299
Utilizing Pretrained GPT-2 Model for Analyzing Drug Reviews Sinem ÖZDEMİR*, Kemal TURHAN	306
Automatic Article Reader and Article Section Classifier Application with Web Scraping Nihat Burak Zihni*	307
Nesnelerin İnterneti ve Giyilebilir Sensörlerle Yaşlı Bireylerin Sağlık Durum Değerlendirilmesi Mustafa Yazıcı*, Sedat Görmüş, Mustafa Ulutaş, Mukadder Okuyan	318
Gamified Facial Exercises Software for Facial Paralysis Patients: A Fun and Effective Method Hüsnü Canbolat*, Sabiha Kaya	322
Alterations in Gene Expression Related to Neuroendocrine Functions in Long-Term Endurance and Resistance Training Elderly Arif Kamil Salihoglu*	325
Hastane Bilgi Yönetim Sistemleri Merkezinin Yeniden Yapılandırılması; Hasta, Hasta Yakını ve Çalışan Memnuniyetinin Değerlendirilmesi İsmail YILDIZ*	326

Convolutional Neural Networks Utilizing Image Representation of Amino Acid Sequence for Variant Effect Prediction Gülbahar Merve ŞILBİR*, Burçin KURT	328
Makine Öğrenimi Algoritmalarıyla Meme Kanseri Sağkalım Tahmini: Performans Karşılaştırması ve Klinik Uygulamaları Yunus Emre KAYAOĞLU*, Uğur ŞEVİK	329
The Impact of ChatGPT in Medical Informatics Education Nur Hürsoy*, Merve Solak, Ayşenur Topçu, Yusuf Çubukçu, Enver Turan, Esat Kaba	335
Apandisit Tanısı Alan Çocuklarda Perforasyon Risk Faktörlerinin İncelenmesi Tuğba KURT, İlknur BUÇAN KIRKBİR*	339
Uzmanlık ve Yüksek Lisans Programlarında Görüntüleme Bilişimi Eğitimi Nur Hürsoy*	346

ÖNSÖZ Foreword

Sayın Tıp Bilişimi Araştırmacıları,

Tıp Bilişimi Derneği olarak, tıp bilişimi alanındaki en yeni bilimsel ve teknolojik gelişmeleri ülkemize kazandırma ve bu alandaki akademik yapının güçlenmesine katkı sağlama misyonunu üstlenmiş bulunmaktayız. Özellikle günümüzde giderek daha fazla önem kazanan sağlıkta yapay zeka ve büyük dil modelleri konularında yenilikçi fikirlerin paylaşılması, derinlemesine tartışmaların yapılması ve bu teknolojilere dair daha gerçekçi perspektifler kazanılması, toplumun geniş kesimlerine ışık tutacak bir etki yaratmaktadır.

Bilimsel kongreler, yapay zekanın sağlık bilimlerindeki dönüştürücü potansiyelini keşfetmek ve küresel ölçekte rekabet edebilir uygulamalar geliştirmek için kritik bir platform sunmaktadır. Bu bağlamda, 15. Tıp Bilişimi Kongresi'nde sunulan bilimsel çalışmaları sizlerle paylaşmaktan büyük bir memnuniyet duyuyoruz. Bu etkinlik, sağlık bilişimi disiplininin akademik yapısını daha da güçlendirmekle kalmayıp, sektördeki bilimsel iş birliğini teşvik etmek adına önemli bir adım olarak değerlendirilmektedir. Sizler de bu çabanın bir parçası olarak, araştırmalarınızla kongremize büyük bir değer kattınız.

Bu kitabın, siz değerli okuyucularımıza ilham vermesini ve yeni fikirler ile projeler geliştirme konusunda motivasyon sağlamasını diliyoruz. Bu eserin ortaya çıkmasında katkı sağlayan herkese teşekkürlerimizi sunarız.

Dear Medical Informatics Researchers,

As the Medical Informatics Association, we have undertaken the mission of bringing the latest scientific and technological advancements in medical informatics to our country and contributing to the strengthening of the academic structure in this field. In particular, the increasing significance of topics such as artificial intelligence in healthcare and large language models has led to the sharing of innovative ideas, in-depth discussions, and the development of more realistic perspectives on these technologies, creating a far-reaching impact on broader society.

Scientific conferences provide a critical platform for exploring the transformative potential of artificial intelligence in health sciences and for developing globally competitive applications in this field. In this context, we are pleased to share with you the scientific works presented at the 15th Medical Informatics Congress. This event not only strengthens the academic structure of the medical informatics discipline but also represents a significant step in fostering scientific collaboration within the sector. By contributing your research, you have added great value to our congress as part of this collective effort.

We hope that this book will inspire you, our esteemed readers, and motivate you to develop new ideas and projects. We extend our sincere thanks to everyone who contributed to the creation of this work.

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

LSTM Tabanlı Derin Öğrenme Yaklaşımı ile Bayesian Hiper-Parametre Optimizasyonu Kullanılarak Hasta Şikayetlerinden Hastalık Tahmini

Aslı Boz¹[0000-0002-3235-1310], Duygu Korkmaz Yalçın²[0000-0003-2969-4934], Yasin Görmez³[0000-0001-8276-2030] ve Neşe Zayım⁴[0000-0003-3133-2692]

¹Çukurova Üniversitesi Tıp Fakültesi, Tıp Eğitimi Anabilim Dalı, Adana, TÜRKİYE

²Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı, Van, TÜRKİYE

³Sivas Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Sivas, TÜRKİYE

⁴Akdeniz Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Antalya, TÜRKİYE
aboz@cu.edu.tr

Özet: Tıp alanında teşhis ve tedavide nicel ve nitel verilerinden elde edilen sonuçlar kullanılmaktadır. Bu çalışmada ise uzun kısa vadeli ağlar (long short-term memories-LSTM) tabanlı derin öğrenme ile Bayesian Hiper-Parametre Optimizasyonu kullanılarak hastanın sözel olarak ifade ettiği şikayetlerden hareketle hastalık tahmini yapmak amaçlanmıştır. Çalışmanın materyali, hasta şikayetlerinin kısa açıklamasını (semptomlar), 24 hastalık türünü içeren ve 1200 gözlemden oluşan Kaggle'dan alınan Symptom2Disease veri setidir. Çalışmada semptomların kısa açıklamasını içeren doğal dilde metinlerden, hastalıkların tahmin edilebilmesi için LSTM tabanlı derin öğrenme modeli geliştirilmiş ve modelde gömme, sönümlenme, tam bağlı katman yapılarından yararlanılmıştır. Modellerin optimizasyonu için ise Bayesian optimizasyon yöntemi kullanılmıştır. Eğitim, test ve doğrulama olarak 3 kısıma ayrılan veri seti kullanılarak farklı katman sayıları (1 LSTM katmanlı, 2 LSTM katmanlı, 3 LSTM katmanlı) ile 3 tahmin modeli oluşturulmuştur. Modelin performans ölçütleri hesaplandığında ise (Başarı Oranı, Kesinlik, F1, Fbeta) en yüksek değerler 2 LSTM katmanlı modelden elde edilmiştir. LSTM tabanlı derin öğrenme ve Bayesian Hyper-Parameter Optimization kullanılarak hastalık tahmini yapılabilmesinin amaçlandığı bu çalışmada oluşturulan tahmin modellerinden özellikle de yüz yüze görüşmenin zor olduğu veya tercih edilmediği durumlarda hastanın şikayetlerinden, hastalığı ilgilendiren birime yönlendirilmesi konusunda yardımcı sistemlerin yanı sıra hekimlerin tanı koyma süreçlerinde veya klinik karar destek sistemlerinin geliştirilmesinde yararlanılabileceği sonucuna ulaşılmıştır.

Anahtar Kelimeler: LSTM, Bayesian Hiper-Parametre Optimizasyonu, Hastalık Tahmini.

1. Giriş

Tıp alanında tarama, teşhis ve tedavide; demografik bilgiler, anamnez ve fizik muayene, laboratuvar tetkikleri, elektronik sağlık kayıtları, görüntüleme vb. verilerinden elde edilen sonuçlar kullanılmaktadır [1].

Bu sonuçların sınıflandırılması ve modellemesinde yapay zekanın dallarından biri olan Yapay Sinir Ağları'ndan (YSA) yararlanılmaktadır.

Yinelemeli Sinir Ağları (RNN) sıralı veriler ve zaman serisi verileri için kullanılmaktadır. Önceki bilgileri saklayan ve sonraki adımlarda bu bilgiyi kullanan katmanlara sahip olan yapay sinir ağıdır. RNN'nin dil işleme ve konuşma tanıma gibi alanlarda kullanımı yaygın olup RNN bir önceki adımın çıktısını bir sonraki adımın girdisi olarak kullanmakta ve bilgiyi kısa bir süre saklamaktadır [2-4]. Uzun süre saklayamaması ise önemli bilgilerin ağın başında yitirilmesine sebep olabilmektedir. RNN'de yaşanan bu problemin üstesinden gelmek için Hochreiter ve Schmidhuber tarafından uzun kısa süreli bellek algoritması (LSTM) geliştirilmiştir [3].

RNN ağı ile birbirine çok benzeyen LSTM sinir ağının kullanım alanı da RNN ile aynıdır.

Unutma, girdi ve çıktı olmak üzere 3 katmandan oluşan RNN'den farklı olarak uzun süreli gerçekleşen işlemleri öğrenebilen LSTM'nin unutma katmanı isminden de anlaşılacağı gibi gelen bilginin tutulmasına veya unutulmasına, girdi katmanı bellekte saklanıp saklanmayacak bilgiye, çıktı katmanı ise bilginin çıktı olup olmama durumuna karar verir [5-7].

Snoek ve ark. (2012) tarafından literatüre kazandırılan Bayesian Hiper-Parametre Optimizasyonu literatürde sıkça kullanılmaktadır [8].

Bu yöntemde istatistikte sıkça kullanılan Bayes mantığına dayanarak ön bilgiden hareketle olasılık hesaplarına dayalı tahmin yapmaktadır. Bu yöntem ile hiper parametrelerinin optimum değerleri hesaplanmaktadır ve daha sonra hesaplanan değerler diğer hiper parametrelerin seçiminde kullanılmaktadır [9].

Sağlık alanında ise LSTM tıbbi notlar, semptom tanımları gibi nitel verilerin analizinde kullanılmaktadır [10,11]. Sağlık araştırmalarında nitel verilerden biri olan hastalık şikayetleri; hastalığın tanı, teşhis ve tedavi planı için kullanılmaktadır. Ayrıca hasta şikayetleri ve anamnez gibi hastanın sözel olarak ifade etmesiyle toplanan veriler hastanın tanı ve bakım süreçlerinde temel bir kaynak olarak kullanılır ve neticesinde hastaların sağlık sonuçlarında önemli bir iyileşme gözlenebilir.

Bu çalışmada ise LSTM tabanlı derin öğrenme ile Bayesian Hiper-Parametre Optimizasyonu kullanılarak hastanın sözel olarak ifade ettiği şikayetlerden hareketle hastalık tahmini yapmak amaçlanmıştır.

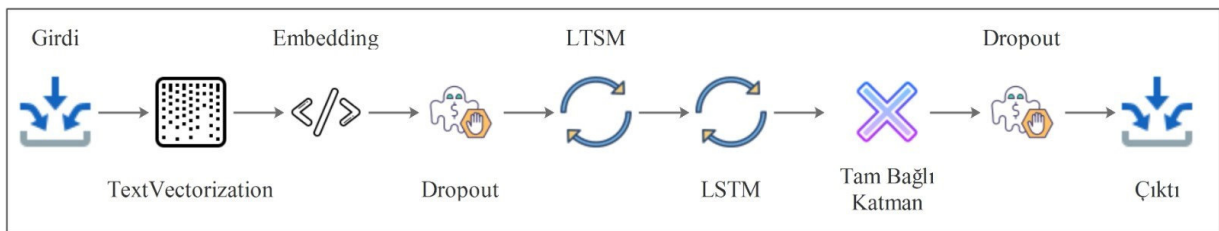
2. Materyal ve Metot

2.1. Materyal

Bu çalışmada açık kaynak platformundan (Kaggle) elde edilen Symptom2Disease veri seti kullanılmıştır [12]. Symptom2Disease veri seti 1200 gözlemden, etiket ve metin olmak üzere 2 sütundan oluşmaktadır. Etiket sütunu hastalık türünü, metin sütunu ise doğal dilde hasta şikayetlerinin kısa açıklamasını içermektedir. Veri setinde toplam 24 farklı hastalık türü (Psoriasis, Varicose Veins, Typhoid, Chicken pox, Impetigo, Dengue, Fungal infection, Common Cold, Pneumonia, Dimorphic Hemorrhoids, Arthritis, Acne, Bronchial Asthma, Hypertension, Migraine, Cervical spondylosis, Jaundice, Malaria, urinary tract infection, allergy, gastroesophageal reflux disease, drug reaction, peptic ulcer disease, diabetes) ve her bir hastalık için 50 farklı şikayet mevcuttur.

2.2. Metot

Çalışmamızda semptomların kısa bir tanımının verildiği metinlerden hastalıkların tahmin edilebilmesi için uzun kısa vadeli ağlar (long short-term memories-LSTM) tabanlı derin öğrenme modeli geliştirilmiştir. Önerilen modelde LSTM katmanlarının yanı sıra gömme (Embedding), sönümlenme (Dropout) ve tam bağlı katman (fully connected layer-FLC) yapıları kullanılmıştır. Bu bağlamda LSTM katman sayısı 1 ile 3 arasında değiştirilerek farklı derinliklerde modeller tasarlanmıştır. Şekil 1'de iki katmana sahip derin öğrenme modelinin mimari yapısı gösterilmektedir. Diğer iki model için tek farklılık LSTM katman sayılarının farklı olmasıdır.



Şekil 1. İki adet seri bağlı LSTM katmanına sahip önerilen derin öğrenme modeli mimari yapısı.

Şekil 1'de gösterilen model Python dilindeki Keras kütüphanesi kullanılarak geliştirilmiştir [13]. Önerilen modelin ilk aşamasında metin verilerinin okunduğu girdi katmanı bulunmaktadır. Girdi katmanında, bir hastalık için verilen kısa tanımın okunduğu tek nöronlu bir yapı bulunmaktadır. Girdi katmanından alınan verilerin, temsil vektörlerine dönüştürülmesi için ise gömme katmanından faydalanılmıştır. Bu aşamada girdi katmanı ile gömme katmanı arasına yine Keras kütüphanesinde var olan TextVectorization yapısı eklenmiştir.

TextVectorization katmanı için kelime dağarcığının maksimum boyut (`max_token`) parametresi 1000, çıktı boyutu 500 olarak ayarlanmıştır. Gömme katmanının çıktı boyutu ise 128 olarak belirlenmiş bu sayede tek boyutlu alınan bir metin 128 uzunluğunda sabit bir vektöre dönüştürülmüştür. Bu aşamadan sonra öğrenme kabiliyetini artırmak için modele söndürme katmanı eklenmiş bu sayede regülarizasyon yapılması amaçlanmıştır. Söndürme katmanını LSTM katmanları takip etmiştir. Bu aşamada LSTM katman sayısı bir (`LSTM_Deep_1`), iki (`LSTM_Deep_2`) ve üç (`LSTM_Deep_3`) olan üç farklı model elde edilmiştir (Şekil 1, Deep modelini göstermektedir). Her bir modeldeki son LSTM katmanı hariç tüm LSTM katmanlarının `return_sequences` parametresi `True` olarak ayarlanmış, aktivasyon fonksiyonları ise `relu` yapılmıştır. LSTM katmanları için çıktı boyutunun uzunluğu (`units`) parametresi optimize edilmiştir. Önerilen modelde LSTM katmanını, FLC ve söndürme katmanları takip etmiştir. Her iki söndürme katmanı içinde düşürülecek giriş birim oranları 0.5 olarak belirlenmiştir. FLC katmanındaki nöron sayısı optimize edilmiş ve aktivasyon fonksiyonu ise `relu` olarak belirlenmiştir. Çıktı katmanında ise nöron sayısı sınıf sayısına eşit, aktivasyon fonksiyonu `softmax` olan bir FLC katmanı kullanılmıştır. Modellerimizde optimizasyon olarak Adam fonksiyonu ve kayıp fonksiyonu olarak `categorical_crossentropy` tercih edilmiştir. Modelin epok sayısı ve öğrenme adımı gibi diğer iki önemli parametresi ise optimize edilmiştir. Bunun yanı sıra modelimize erken durdurma (`early_stopping`) ve öğrenme adımını düşürme (`lr_callback`) olmak üzere iki adet geri çağırma fonksiyonu eklenmiştir. Bu fonksiyonlara göre validasyon veri setinde art arda iki adımda iyileşme olmaması durumunda öğrenme adımı iki ile bölünmüş, art arda altı adımda iyileşme olmaması durumunda ise eğitim sonlandırılmıştır.

Bu çalışmanın ilk aşamasında veri seti, eğitim, test ve validasyon olmak üzere üçe bölünmüştür. Bu kapsamda `sklearn` kütüphanesinde yer alan `train_test_split` fonksiyonu kullanılarak öncelikle tüm veri setinin %20'si test verisi olarak seçilmiştir [14]. Daha sonra kalan %80'lik kesimin %10'u ise aynı fonksiyon kullanılarak validasyon veri setini oluşturmak üzere seçilmiştir. Validasyon ve test veri setleri oluşturulduktan sonra kalan örnekler kullanılarak ise eğitim veri seti oluşturulmuştur. Bu aşamada `train_test_split` fonksiyonunda bulunan `stratify` parametresi kullanılmış bu sayede eğitim, test ve validasyon veri setlerindeki her bir sınıfa ait örneklerin dengeli dağılımı sağlanmıştır. Bölütleme aşaması sonunda eğitim veri setinde 864, test veri setinde 240 ve validasyon veri setinde 96 örnek olduğu gözlemlenmiştir.

Veri seti bölütleme aşamasından sonra modellerimize hiper parametre optimizasyonu Bayesian optimizasyon yaklaşımı kullanılarak oluşturulmuştur. Bayesian optimizasyon yönteminin özellikle derin öğrenme modellerinde diğer hiper-parametre optimizasyonu yöntemlerine üstünlük sağladığı için bu yöntem çalışmamızda kullanılmıştır [8,15]. Bayesian optimizasyon yöntemi Python dilinde var olan `scikit-optimize` [16] kütüphanesi kullanılarak geliştirilmiştir. Bu kütüphanede var olan `gp_minimize` fonksiyonunun, `acq_func` parametresi "`EI`" ve `n_calls` parametresi 25 olarak ayarlanmış ve hiper-parametre optimizasyonu yapılmıştır. Parametre optimizasyonu aşamasında her bir model için parametreler ayrı ayrı optimize edilmiş ve modeller eğitim veri seti ile eğitilerek validasyon veri setindeki başarı oranını maksimum elde etmek hedeflenmiştir.

3. Bulgular

Tablo 1'de her bir model için parametre değerinin türü, parametrenin adı, parametre uzayındaki en küçük değer, parametre uzayındaki en büyük değer ve her bir model için elde edilen optimum değer gösterilmektedir.

Tablo 1. Önerilen her bir model için parametre optimizasyonu detayları.

Hiper-Parametre adı	Hiper-Parametre değeri türü	En küçük değer	En büyük değer	LSTM_Deep_1 modeli optimum değeri	LSTM_Deep_2 modeli optimum değeri	LSTM_Deep_3 modeli optimum değeri
Öğrenme Adımı	Gerçek Değer	10 ⁻⁴	10 ⁻¹	0.0010962	0.0473750	0.0015790
Döngü Sayısı	Tam Sayı	5	250	132	120	143
Batch sayısı	Tam Sayı	1	16	8	4	4
LSTM birim sayısı	Tam Sayı	32	256	122	231, 236	143, 156, 105
FLC Nöron sayısı	Tam Sayı	128	1024	449	846	711

Bu tabloda LSTM birim sayısı için LSTM_Deep_2 ve LSTM_Deep_3 modellerinde var olan virgülle belirtilmiş çoklu değerler sırası ile bağlı olan LSTM katmanları için birim sayılarını temsil etmektedir. Optimizasyon aşamasında elde edilen en uygun parametrelerin atandığı önerilen modeller eğitim veri seti kullanılarak eğitilmiş ve modellerin performans değerleri test veri seti kullanılarak hesaplanmıştır. Tablo 2'de, her bir model için hesaplanmış olan başarı oranı, kesinlik, f1 ve fbeta değerleri gösterilmektedir.

Tablo 2. Optimum parametre ile eğitilmiş önerilen modellerin performans skorları.

Model Adı	Başarı Oranı	Kesinlik	F1	Fbeta
LSTM_Deep_1	95.83%	92.65%	92.16%	92.42%
LSTM_Deep_2	96.66%	96.99%	96.65%	96.81%
LSTM_Deep_3	95.00%	95.35%	94.98%	95.15%

4. Sonuç

Kullanılan 3 modelden (LSTM_Deep_1, LSTM_Deep_2 ve LSTM_Deep_3) hesaplanan performans ölçütleri incelendiğinde LSTM_Deep_2 modelinin en yüksek değerlere sahip olduğu görülmektedir. Katman sayısı arttıkça modelin performansının artması beklenirken bu çalışmada kullanılan veri setindeki gözlem sayısının düşük olması sebebiyle 2 katmanlı model en iyi sonuçlara sahiptir. Bu nedenle gözlem sayısı artırıldığında LSTM_Deep_3 modelinin daha iyi performansa sahip olabileceği düşünülmektedir. Bu veri setinde gözlem sayısının düşük olmasının yanı sıra veri seti 24 hastalık türü ile sınırlıdır. Bu nedenle modelin tüm hastalık türlerini tahminleme performansı hakkında çıkarımda bulunmak güçtür. Ancak modellerin performans ölçütleri daha sonraki çalışmalar için umut vaat etmektedir. Bu yüz yüze görüşmenin zor olduğu veya tercih edilmediği durumlarda hasta şikayetlerinden olası hastalık tahmini yaparak hastaları yönlendirmek için örneğin hastaların semptomlarına göre hangi branşa başvuracağı konusunda yardımcı sistemlerin, hekimlerin tanı koyma süreçlerinde veya eğitim amaçlı kullanılan klinik karar destek sistemlerinin geliştirilmesinde yararlanılabilir.

Kaynaklar

1. FDA, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/electronic-source-data-clinical-investigations>, Erişim Tarihi 06/05/2024.
2. Graves, A., Graves, A.: Long short-term memory. Supervised sequence labelling with recurrent neural networks, 37-45, Springer, Heidelberg (2012).
3. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735-1780, (1997).
4. Talathi, S. S., Vartak, A.: Improving performance of recurrent neural network with relu nonlinearity. arXiv preprint arXiv:1511.03771, (2015).
5. Kara, A.: uzun-kısa süreli bellek ağı kullanarak global güneş ışınımı zaman serileri tahmini. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 7(4), 882-892, (2019).
6. Yılmaz, M. C., Orman, Z.: LSTM derin öğrenme yaklaşımı ile Covid-19 pandemi sürecinde twitter verilerinden duygu analizi. Acta Infologica, 5(2), 359-372, (2021).
7. Tanışman, S., Karcıoğlu, A. A., Çur, A., Bulut, H.: LSTM sinir ağı ve ARMA zaman serisi modelleri kullanılarak bitcoin fiyatının tahminlenmesi ve yöntemlerin karşılaştırılması. Avrupa Bilim ve Teknoloji Dergisi, (32), 514-520, (2021).
8. J. Snoek, H. Larochelle and R. P. Adams.: "Practical bayesian optimization of machine learning algorithms", Advances in neural information processing systems, pp. 2951-2959, (2012).
9. Tanyıldızı, E., Demirtaş, F.: Hiper parametre optimizasyonu hyper parameter optimization. In 2019 1st International Informatics and Software Engineering Conference (IBMYK), pp. 1-5., (2019).
10. Edara, D. C., Vanukuri, L. P., Sistla, V., Kolli, V. K. K.: Sentiment analysis and text categorization of cancer medical records with LSTM. Journal of Ambient Intelligence and Humanized Computing, 14(5), 5309-5325, (2023).
11. Liu, J., Zhang, Z., Razavian, N.: Deep ehr: Chronic disease prediction using medical notes. In Machine Learning for Healthcare Conference, pp. 440-464., PMLR, (2018)
12. Kaggle, <https://www.kaggle.com/datasets/niyarbarman/symptom2disease>, Erişim Tarihi 01/05/2024.
13. Keras, <https://keras.io/>, Erişim Tarihi 01/05/2024.
14. <https://scikit-learn.org/stable/>, Erişim Tarihi 01/05/2024.
15. Salama, A. Y. A., Hassanien, A. E., Fahmy, A.: Sheep identification using a hybrid deep learning and bayesian optimization approach. IEEE Access, 7, 31681-31687, (2019).
16. <https://pypi.org/project/scikit-optimize/>, Erişim Tarihi 01/05/2024.

DrugAutoML: İlaç Keşfi için Açık Kaynaklı, Uçtan Uca Otomatikleştirilmiş Bir Makine Öğrenimi ve İstatistiksel Değerlendirme Aracı

Ayça BEYHAN^[1,2] ve Aslı SUNER^[1]

¹ Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Tıp Fakültesi, Ege Üniversitesi, İzmir, Türkiye

² İşletme Yönetimi Bölümü, Milli savunma Üniversitesi, Balıkesir, Türkiye
aycapamukcu9@gmail.com

Özet: "DrugAutoML: İlaç Keşfi için Açık Kaynaklı, Uçtan Uca Otomatikleştirilmiş Bir Makine Öğrenimi ve İstatistiksel Değerlendirme Aracı" başlıklı makale, ilaç keşfi sürecindeki karmaşık aşamaları otomatikleştirmek için geliştirilen yenilikçi bir AutoML aracını tanıtmaktadır. DrugAutoML, veri toplama, ön işleme, model seçimi, hiperparametre optimizasyonu ve performans değerlendirme gibi temel adımları otomatik olarak yönetir. Açık kaynaklı bu platform, bilim insanlarına ve sağlık profesyonellerine karmaşık kodlama bilgisi gerektirmeyen erişilebilir bir arayüz sunar. Ayrıca, ilaç keşif süreçlerini hızlandırarak maliyetleri düşürmeyi ve yeni tedavi edici ilaçların piyasaya sürülme süresini kısaltmayı hedefler. Platformun performansı, DUD-E ve MUV veri setleri üzerinde değerlendirilmiş ve yüksek doğruluk ve F1 skorları ile klasik sınıflandırma algoritmalarına üstünlük sağlamıştır. Bu çalışma, AutoML'nin ilaç keşfi alanında kullanım potansiyelini etkili bir şekilde göstermektedir.

Anahtar Kelimeler: İlaç Keşfi, AutoML, Makine Öğrenmesi

Kaynaklar

1. Beck, D.A., Carothers, J.M., Subramanian, V.R., Pfaendtner, J.: Data science: Accelerating innovation and discovery in chemical engineering. *AIChE J.* 62(5), 1402–1416 (2016)
2. Beck, H., Härter, M., Haß, B., Schmeck, C., Baerfacker, L.: Small molecules and their impact in drug discovery: A perspective on the occasion of the 125th anniversary of the Bayer Chemical Research Laboratory. *Drug Discov. Today* 27(6), 1560–1574 (2022).
3. Hutter, F., Kotthoff, L., Vanschoren, J.: Automated machine learning: methods, systems, challenges. Springer Nature, p. 219 (2019)

DrugAutoML: An Open-Source, End-to-End Automated Machine Learning and Statistical Evaluation Tool for Drug Discovery

Ayça BEYHAN^[1,2] ve Aslı SUNER^[1]

¹ Department of Biostatistics and Medical Informatics, Faculty of Medicine, Ege University, İzmir, Türkiye

² Department of Business Administration, National Defense University, Balıkesir, Türkiye
aycapamukcu9@gmail.com

Abstract: The article titled "DrugAutoML: An Open Source, End-to-End Automated Machine Learning and Statistical Evaluation Tool for Drug Discovery" introduces an innovative AutoML tool developed to automate the complex stages of the drug discovery process. DrugAutoML automatically manages key steps such as data accusation, preprocessing, model selection, hyperparameter optimization, and performance evaluation. This open-source platform provides scientists and healthcare professionals with an accessible interface that does not require complex coding knowledge. Additionally, it aims to reduce costs and shorten the time to market for new therapeutic drugs by accelerating drug discovery processes. The performance of the platform has been evaluated on the DUD-E and MUV datasets, where it demonstrated superiority over traditional classification algorithms with high accuracy and F1 scores. This study effectively showcases the potential use of AutoML in the field of drug discovery.

Key Words: Drug Discovery, AutoML, Machine Learning

In Silico Screening of Momordica Charantia Compounds for the Treatment of Type 2 Diabetes Mellitus: Molecular Docking

Ayşe Banu PAK^{1,2}[0000-0002-9490-3504], Işık CAKMAK¹[0000-0003-4758-4077] and Mustafa Emre ERCİN³[0000-0002-7340-8045]

¹Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey

²Distance Education Application and Research Center, Trabzon University, Trabzon, Turkey

³Ankara Atatürk Sanatorium Training and Research Hospital, Department of Medical Pathology, 06290, Ankara, Turkey
aysebanu.pk@gmail.com

Abstract. *Type 2 Diabetes Mellitus (T2DM)* is the most prevalent form of diabetes, affecting millions of individuals worldwide. Antidiabetic agents are employed in the treatment of this disease. The significance of *T2DM* lies in the fact that various complications resulting from elevated blood sugar levels have the potential to impact the patient's quality of life and even lead to premature mortality. The primary objective in the management of *T2DM* is to reduce hyperglycaemia.

Momordica charantia has the potential to be effective against *T2DM* disease. However, the mechanism by which it exerts this effect is not fully understood. This study aimed to identify novel plant-derived antidiabetic compounds from *Momordica charantia* in order to gain insight into the molecular basis of antidiabetic activity. To this end, the phytochemicals of *Momordica charantia* were identified using the IMPPAT and NEI-MPDB databases. The phytochemicals were downloaded in the structure data format (sdf) with the help of the PubChem database. Glucagon-like peptide-1 (GLP-1) was identified as the target. The protein was obtained from the RCSB PDB (3C5T). The protein and phytochemicals were prepared for the identified target, and docking analysis was performed. The binding energy of the ligand in the protein was used in the evaluation of the analysis results. It was observed that 68 of the 171 phytochemicals screened had binding energies less than that of the reference ligand. Protein-ligand interactions of phytochemicals with binding energies inferior to that of the reference ligand were analysed using Discovery Studio. These interactions may explain the antidiabetic potential of plants. In our study, 6 phytochemicals were identified as suitable molecules. These compounds may interact with the GLP-1 receptor residue, thus activating the inhibition of glucagon release. It is thought that this will result in a reduction in blood glucose levels. In conclusion, Quercetin, Luteolin, Riboflavin, cis-Zeatin riboside, Momordicine I, Zeatin riboside can be suggested as potential antidiabetic compounds. However, we recommend this study as a

reference for further investigation of these compounds as antidiabetic by in vitro and in vivo methods.

Keywords: Autodock Vina, admetSAR, IMPPAT, NEI-MPDB, Molecular docking, *Momordica charantia*, swissADME and T2DM,

1 Introduction

Type 2 Diabetes Mellitus (T2DM), the most common type of *Diabetes Mellitus*, is considered to be a progressive disease in which pancreatic function decreases over time [1]. A variety of agents are employed in the treatment of this disease, with differing effects on glycaemic control. However, these agents have different side effects. The high cost of existing antidiabetic agents and the side effects associated with their use have prompted a search for natural products that could be used as effective anti-diabetic drugs with fewer side effects [2]. It is known that many plants have been used traditionally for the management and control of diabetes as an alternative to antidiabetic drugs [3]. *Momordica charantia* (*M. charantia*), a member of the Cucurbitaceae family, has attracted considerable attention due to its antidiabetic properties [1] [2] [4] [5] [6]. The scientific evidence currently available supports the traditional use of this plant as one of the most promising herbs for the treatment of diabetes (3). However, its molecular mechanism has not been fully elucidated. *Momordica charantia* (bitter lemon) is a member of the Cucurbitaceae family, which is a source of traditional medicine used to treat a range of conditions, including cancer and diabetes.

Momordica charantia has the potential to be effective against *T2DM* disease. However, its mechanism on the disease is not fully known. GLP-1 receptor agonists are used for the treatment of the disease. Glucagon-like peptide-1 receptor (GLP-1 R) is among the well-known targets for diabetes drugs due to its clinical values in the regulation of glucose metabolism. Although the antidiabetic mechanism of action of *M. Charantia* has been reported, the molecular mechanism of compounds derived from the plant has not been fully elucidated. In this context, it is thought that it is worth elucidating the mechanism underlying the glycemic control of *M. charantia*. The aim of this study was to analyze potential therapeutic targets for glycemic control. It has the potential to be antidiabetic through its binding mechanism to GLP-1 receptor residues. These compounds can interact with the GLP-1 receptor residue, thereby activating the inhibition of glucagon release. It may reduce blood sugar levels.

2 Material and Method

2.1 Protein Preparation for Molecular Docking

In our study, the protein with PDB ID: 3C5T for the target (GLP-1) was selected as the receptor. The three-dimensional crystallographic structure of GLP-1, whose PDB ID: 3C5T, was downloaded from the Protein Data Bank (www.rcsb.org). In order to make this protein suitable for docking analysis, cofactors, ions and water molecules in the

protein were removed using Chimera 1.16, missing atoms were checked, if there were missing atoms, they were completed, hydrogen and charges were added and prepared (protein.pdbqt) for docking analysis.

2.2 Preparation of Ligands for Molecular Docking

Using Indian Medicinal Plants, Phytochemistry And TherapeuticsI (IMPPAT) and An integrated North East India-Medicinal Plant Database (NEI-MPDB), 171 phytochemicals of *Momordica charanti* were identified. These phytochemicals were downloaded in structure data format (sdf) using PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). These downloaded phytochemicals were first optimised to have the most stable structure using General Amber Force Field (GAFF). They were then converted to pdbqt using OpenBabel and prepared for docking analysis.

2.3 Determination of Active Site

It is necessary to determine important binding site residues for ligand interaction. In this context, the binding site was determined using the 10M ligand in the protein. Active region the grid parameters were occurred as 16x14x16 Å x, y, z dimensions, 1.0 Å space and -30.304, 30.337, 3.417 x, y, z centers.

2.4 Molecular Docking

Molecular docking studies represent a pivotal tool in the field of drug design [7]. Molecular docking is a method that assists in determining the suitability of a ligand for a specific target and facilitates the development of new drugs [8]. The theoretical basis of molecular docking is thought to be the study of protein-ligand interactions and binding conformations. Molecular docking is employed to ascertain the most stable conformation between protein and ligand. By examining the interactions between ligand(s) and receptor protein, the success rate of drug discovery can be increased by focusing on the most appropriate molecules. A molecular docking analysis was conducted using the prepared protein (protein.pdbqt) and ligand/(s) (ligand.pdbqt). The analysis was conducted using Autodock Vina [9].

2.5 Drug Likeness and In Silico ADME Prediction

The determination of drug similarity properties and Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties of screened phytochemicals prior to clinical trials has been demonstrated to increase the success rate of clinical trials. The pharmacokinetic properties (Absorption, Distribution, Metabolism and Excretion (ADME)) and drug similarity of the screened phytochemicals were evaluated using the swissADME database (<http://www.swissadme.ch/index.php>). AdmetSAR (<http://lmmd.ecust.edu.cn/admetSar2/>) was employed to predict toxicity behaviour. The drug similarity of the phytochemicals was estimated based on Lipinski's Rule of 5

(Ro5), which stipulates that a compound should exhibit a binding energy better than that of the reference ligand. According to this rule, "drug-like" molecules should have $\text{Log } P \leq 5$, molecular weight (M.W.) ≤ 500 , number of hydrogen bond acceptors (HBA) ≤ 10 , and number of hydrogen bond donors (HBD) ≤ 5 . Lipinski's rule of five was applied to select probable ligands [10]. It is not possible for molecules outside this range to be bioavailable orally as drugs [11-14].

3 Results

3.1 Molekular Docking

Molecular docking analysis was performed using Autodock Vina [9] to investigate the antidiabetic properties of the phytochemicals. All 171 phytochemicals prepared were analysed by molecular docking with the protein prepared PDB ID: 3c5t protein was subjected to molecular docking analysis. When the analysis results were analysed, the binding energies of the identified phytochemicals against GLP-1 R agonist ranged from -7.4 kcal/mol to -2.4 kcal/mol. It was compared with the reference ligand binding energy (-4.7) and 67 phytochemicals exhibited lower binding energy, i.e. better binding, than the binding energy of the reference ligand. The highest binding affinities between two molecules occur when their binding energies are the most negative and show the lowest values [15]. Below are the ligand molecules with better binding energies than the binding energy of the reference ligand as a result of the analysis (Table 1).

Table 1. Kenetlenme analiz sonuçları

S. NO	Phytochemical name	PubChem ID	BE (kcal/mol)
M0	10M	10345813	-4.7
M1	beta-Amyrin	73145	-7.4
M2	Oleanolic acid	10494	-6.9
M3	Taraxerol	92097	-6.6
M4	Glutinol	9932254	-6.6
M5	beta-Sitosterol-beta-D-glucoside	12309055	-6.6
M6	24-Methylenecycloartanol	94204	-6.1
M7	Stigmasta-7,22,25-trien-3beta-ol	157009953	-6.1
M8	Cycloartenol	92110	-6
M9	Diosgenin	99474	-6
M10	(+)-Tirucallol	101257	-6
M11	β -Sitosterol- β -D-glucoside	5742590	-5.9
M12	Simiarenol	12442794	-5.9
M13	Momordicine II	101019708	-5.9
M14	4alpha-Methylzymosterol	22212495	-5.8
M15	beta-Ylangene	25244198	-5.8

M16	(+)-delta-Cadinene	441005	-5.7
M17	Quercetin	5280343	-5.7
M18	Luteolin	5280445	-5.7
M19	Stigmasterol	5280794	-5.7
M20	alpha-Spinasterol	5281331	-5.7
M21	Obtusifoliol	65252	-5.6
M22	Momorcharaside B	131828	-5.6
M23	Lophenol	160482	-5.6
M24	beta-Sitosterol	222284	-5.6
M25	beta-Caryophyllene	5281515	-5.6
M26	(-)-alpha-Cadinol	10398656	-5.6
M27	Momordicoside K	57330180	-5.6
M28	Momordicoside L	101743788	-5.6
M29	Cycloeucalenol	101690	-5.5
M30	Retinol	445354	-5.5
M31	Riboflavin	493570	-5.5
M32	Kaempferol	5280863	-5.5
M33	Bicyclogermacrene	13894537	-5.5
M34	Momordicoside F1	44445566	-5.5
M35	Cholesterol	5997	-5.4
M36	Lanosterol	246983	-5.4
M37	Phytosterols	12303662	-5.4
M38	Cucurbitadienol	14543446	-5.4
M39	Farnesylacetone	1711945	-5.3
M40	Humulene	5281520	-5.3
M41	Epi-Cubenol	12046149	-5.3
M42	momordicoside F2	44445567	-5.3
M43	Momordicoside I	71717036	-5.3
M44	Momordicoside G	91895422	-5.3
M45	Stigmasta-7,25-dien-3beta-ol	6321372	-5.2
M46	Isogermacrene D	91723653	-5.2
M47	Epi-Cubebol	91753433	-5.2
M48	p-Cymene	7463	-5.1
M49	15-Demethyl plumieride	453214	-5.1
M50	Humulene epoxide II	10704181	-5.1
M51	cis-Zeatin riboside	13935024	-5.1
M52	Momordicine I	101293615	-5.1
M53	Sabinene	18818	-5

M54	Zeatin riboside	6440982	-5
M55	(3R,6E)-nerolidol	11241545	-5
M56	DL-Tryptophan	1148	-4.9
M57	Geraniol	637566	-4.9
M58	beta-Ionone	638014	-4.9
M59	Indole	798	-4.8
M60	Alpha-Pinene	6654	-4.8
M61	4'-Methylacetophenone	8500	-4.8
M62	alpha-Terpineol	17100	-4.8
M63	Nerolidol	5284507	-4.8
M64	(3S,4S)-4-ethenyl-4-methyl-3-prop-1-en-2-ylcyclohexene	10607083	-4.8
M65	Isophorone	6544	-4.7
M66	Nerol	643820	-4.7
M67	cis-3-Hexenyl salicylate	5371102	-4.7

The phytochemicals with the highest binding energy against GLP-1 R agonist are presented in the table above. A comparison of 171 phytochemicals with the reference ligand revealed that 67 of them exhibited a superior binding energy. It is hypothesised that these phytochemicals may be identified as anti-diabetic chemical compounds in *M. charantia*.

3.2 Investigation of Protein-Ligand Interaction

The interactions of 18 phytochemicals identified with 3iol targets were examined. Protein-ligand interactions were examined using Discovery.

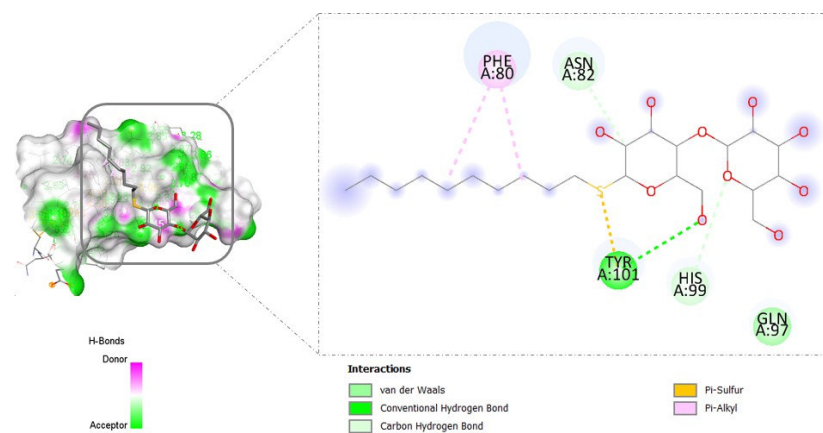
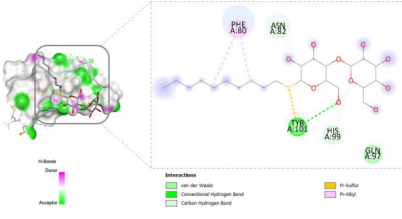
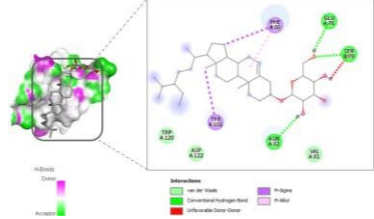
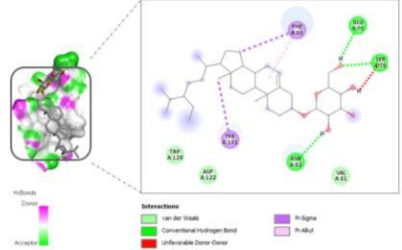
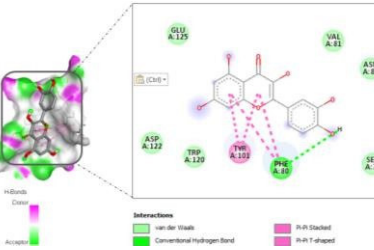
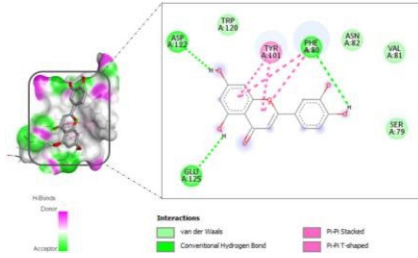


Fig. 1. 3D and 2D representation of reference ligand (10M) interaction with 3C5T protein

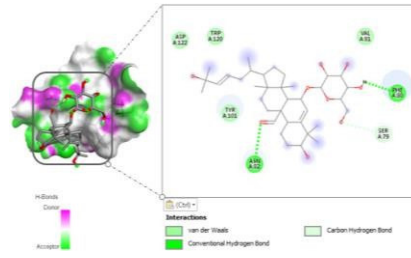
There are hydrogen bond, hydrophobic and other interactions between the 3C5T protein and the reference ligand (10M). In this interaction area, amino acids TYR101, HIS99, ASN82, TYR101, PHE80 are located. The reference ligand was found to engage in hydrogen bonding with TYR101, HIS99, ASN82, pi-alkyl interactions with PHE80 and pi-sulfur interactions with TYR101. The ligand's stability in the binding pocket of proteins is contingent upon the presence of favourable hydrophobic, hydrogen bonding and pi-interaction properties. In the course of our investigation, 18 molecules were identified that exhibited binding energies lower than that of the reference ligand. These molecules were found to engage in similar interactions with the reference ligand. The interactions of all the phytochemicals identified in Table 1 were analysed. It was observed that these phytochemicals predominantly engage in hydrophobic interactions with the target protein. Furthermore, it has both hydrogen bond and hydrophobic interactions with the molecules presented in the table (Table 2) below.

Table 2. 3D and 2D views of selected molecules after interaction analysis

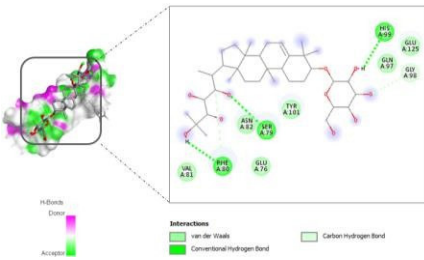
Ligand ID - PubChem ID- Name			Ligand ID - PubChem ID - Name				
	M0	10345813	10M		M13	101019708	Momordicine II
	M11	5742590	β -Sitosterol- β -D-glucoside		M11	5742590	β -Sitosterol- β -D-glucoside



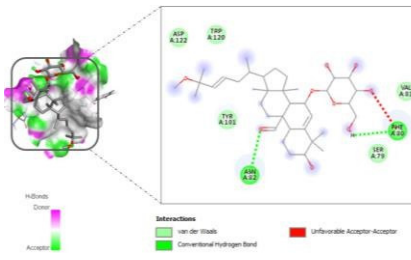
M18 5280445 Luteolin



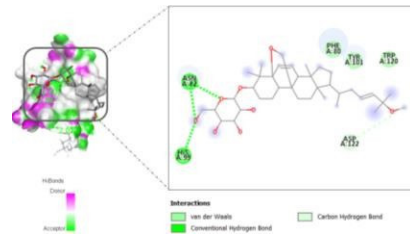
M28 10174378
8 Momordicoside L



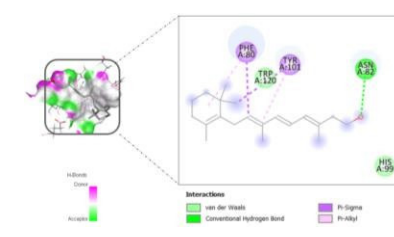
M22 131828 Momorcharaside B



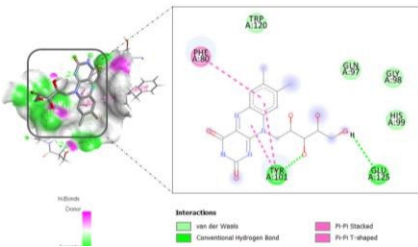
M27 57330180 Momordicoside K



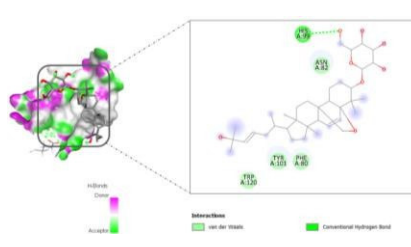
M34 44445566 momordicoside F1



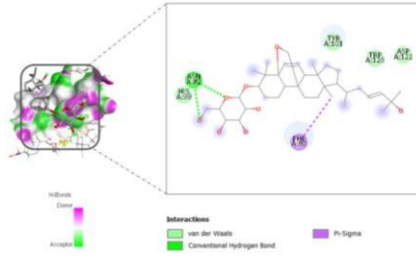
M30 445354 Retinol



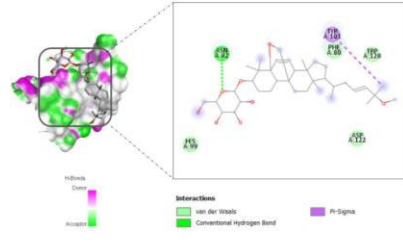
M31 493570 Riboflavin



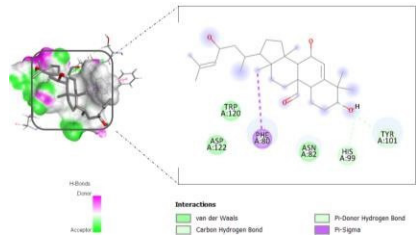
M42 44445567 momordicoside F2



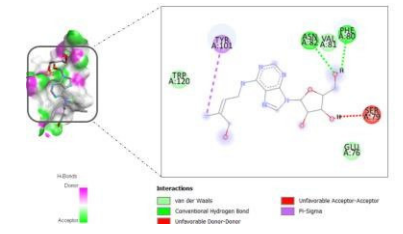
M43 71717036 Momordicoside I



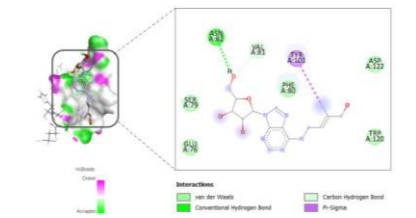
M44 91895422 Momordicoside G



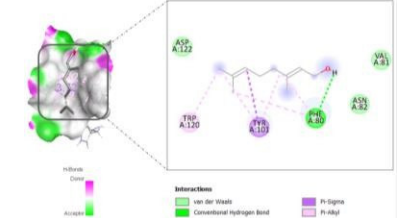
M52 10129361
5 Momordicine I



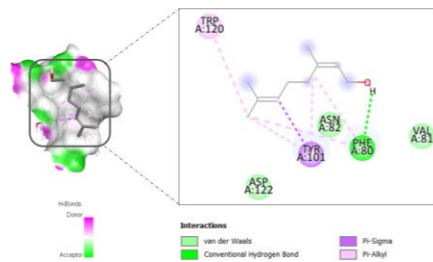
M51 13935024 cis-Zeatin riboside



M54 6440982 Zeatin riboside



M57 637566 Geraniol



M66 643820 Nerol

When the interaction sites are examined, it is suggested that ligands that interact at the same site as the reference ligand have the potential to be anti-diabetic by binding to GLP-1 receptor residues. This is a promising approach for the future development of GLP-1 receptor agonists with improved efficacy for the treatment of diabetes and weight management.

3.3 Drug Likeness and In Silico ADME Prediction

Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) analysis is used to evaluate the drug candidate similarity and pharmacokinetic properties of drug candidate molecules. Compounds with the lowest binding energy and similar interaction in the protein target were selected and swissADME was used to determine the drug-like properties of these phytochemicals. When the drug similarity properties and ADMET properties were analysed considering the reference ligand properties, it can be said that the following molecules have drug-like properties. All of these ligands met the rules of Lipinski's. However, our reference ligand violated only one rule. In drug discovery and development, it is important to evaluate the drug similarity property of phytochemicals through Lipinski's Rule of Five [15]. The following table (Table 3) shows the results of drug similarity and ADMET prediction properties of phytochemicals.

Table 3. Table captions should be placed above the tables.

L.No	MW	HBA	HDB	TPSA	iLOGP	Lipins ki	RB	GI	BBB	LogS	LD50, mol/kg	Carcino- gen
M0	498.63	10	7	194.6	3.17	1	14	Low	No	-2.6632	2.0531	Non- car- cino- gens
M17	302.24	7	5	131.36	1.63	0	1	High	No	-2.9994	3.02	Non- car- cino- gens
M18	286.24	6	4	111.13	1.86	0	1	High	No	-2.9994	3.02	Non- car- cino- gens
M31	376.36	8	5	161.56	0.97	0	5	Low	No	-3.6043	1.6067	Non- car- cino- gens
M51	351.36	8	5	145.78	2.08	0	6	Low	No	-2.2537	2.7694	Non- car- cino- gens

M52	472.7	4	3	77.76	4.25	0	5	High	No	-4.0296	2.9633	Non-carcinogens
M54	351.36	8	5	145.78	2.08	0	6	Low	No	-2.2537	2.7694	Non-carcinogens

RB: Rotatable bonds, HBA: H-bond acceptors, HDB: H-bond donors, GI: GI absorption, BBB: Blood–Brain Barrier permeant

The swissADME analysis of phytochemicals is very important for drug design and drug potential. From the results of Table x, we can conclude that our identified compounds fulfil the criteria defined in Lipinski's rules and that there are no violations. The results of the drug similarity estimates of the phytochemicals, shown in Table 4, indicate that the identified compounds comply with the Ro5 rule and thus have good permeability or absorption properties. All of these phytochemicals screened conform to the drug similarity rule and LogS, as well as the lethality factor concentration (LD50). The values of ADME properties were all within the permissible ranges. Promising pharmacological drug candidates should have pharmacokinetic ADMET properties acceptable for the therapeutic target. Toxicity predictions ensure safety by guiding the selection of ligands for molecular docking.

4 Discussion and Conclusion

In the present era, as a consequence of scientific advances, a considerable number of herbal plants have been identified as possessing a variety of medicinal properties. Molecular docking analysis was conducted to assess the impact of GLP1 receptor agonists, which play a pivotal role in the treatment of diabetes, on type 2 diabetes. It was observed that seven of the 171 phytochemicals of *M. charantia* which were considered within the scope of the study, demonstrated similar and enhanced interactions with GLP-1 RAs in the treatment of the disease. The interactions included several hydrogen bonds and other hydrophobic interactions. The docking energies exhibited a range from -7.4 to -4.7, which was lower than the binding energy of the reference ligand. The interaction domains of these phytochemicals suggest that six of them may be suitable candidates for drug development, given their similarity to existing drugs and their predicted ADMET properties. These molecules may be less toxic, cost-effective, and superior drug candidates for the treatment of type 2 diabetes. In order to overcome the challenges posed by *T2DM*, it is necessary to adopt innovative approaches. Molecular docking is being employed in studies aimed at identifying enzyme inhibitors and receptor agonists for the treatment of the disease.

The results indicate that the phytochemicals of *Momordica Charantia* include quercetin, luteolin, riboflavin, cis-zeatin riboside, momordicine I, and zeatin riboside. These

compounds are regarded as potential antidiabetic candidates. The screened phytochemicals were found to comply with the drug similarity rule and LogS, as well as the lethal factor concentration (LD50). The binding energy scores, hydrogen bond lengths, and interactions between amino acids and ligands indicate that all phytochemicals exhibit good binding energy with the T2D receptor. These phytochemicals are therefore considered to be better ligands for the related target (GLP-1). In comparison to the reference compounds, six phytochemicals were identified as potential antidiabetic drug candidates, namely quercetin (-5.7 kcal/mol), luteolin (-5.7 kcal/mol), riboflavin (-5.5 kcal/mol), cis-zeatin riboside (-5.1 kcal/mol), momordicine I (-5.1 kcal/mol) and zeatin riboside (-5.0 kcal/mol). The results of the drug similarity analysis indicate that our novel molecules exhibit a favourable physicochemical profile. These findings are anticipated to inform future studies on the development of antidiabetic drugs. Furthermore, this study may serve as a reference for further investigation of these compounds as antidiabetics, employing in vitro and in vivo methods.

These recommended phytochemicals are also found in various plants and are recommended for the treatment of various diseases. Quercetin is known to have anti-cancer, antioxidant [16], anti-inflammatory [17], anti-carcinogenic [17], anti-Alzheimer [18] etc. effects. Their therapeutic potential has been evaluated in the treatment of various diseases [18]. Molecular docking studies are being carried out to understand the molecular mechanisms of these compounds.

In conclusion, the antidiabetic activity of phytochemicals identified from the databases of *Momordica charantia* was investigated by molecular docking analysis against the therapeutic target of *T2DM*. Initially, 171 phytochemicals were screened on the basis of high binding energy against the receptor of *T2DM*. The principal analyses indicated that six compounds may be suitable candidates. The identified compounds included quercetin, luteolin, riboflavin, cis-zeatin riboside, momordicine I, and zeatin riboside.

References

- [1] C. Ning *et al.*, “Recent advances in the managements of type 2 diabetes mellitus and natural hypoglycemic substances,” *Food Sci. Hum. Wellness*, vol. 11, no. 5, pp. 1121–1133, 2022.
- [2] S. O. Famuyiwa *et al.*, “Investigating the multitargeted anti-diabetic potential of cucurbitane-type triterpenoid from *Momordica charantia*: an LC-MS, docking-based MM\GBSA and MD simulation study,” *J. Biomol. Struct. Dyn.*, pp. 1–12, 2023.
- [3] O. O. Elekofehinti *et al.*, “Potential use of bitter melon (*Momordica charantia*) derived compounds as antidiabetics: in silico and in vivo studies,” *Pathophysiology*, vol. 25, no. 4, pp. 327–333, 2018.
- [4] B. Joseph and D. Jini, “Antidiabetic effects of *Momordica charantia* (bitter melon) and its medicinal potency,” *Asian pacific J. Trop. Dis.*, vol. 3, no. 2, pp. 93–102, 2013.
- [5] M. A. S. Hairani, F. A. Abdul Majid, N. H. Zakaria, D. Hudiyanti, A. Fadhlina,

- and H. I. Sheikh, "Anti-diabetic properties of traditional herbal concoction containing *Eleutherine palmifolia* (L.) Merr., *Momordica charantia* L., and *Syzygium polyanthum* (Wight.): a bibliometric analysis," *Food Prod. Process. Nutr.*, vol. 5, no. 1, p. 60, 2023.
- [6] K. R. Anilakumar, G. P. Kumar, and N. Ilaiyaraja, "Nutritional, pharmacological and medicinal properties of *Momordica charantia*," *Int. J. Nutr. Food Sci.*, vol. 4, no. 1, pp. 75–83, 2015.
- [7] İ. Çapan, S. Servi, İ. Yıldırım, and Y. Sert, "Synthesis, DFT study, molecular docking and drug-likeness analysis of the new Hydrazine-1-carbothioamide, triazole and thiadiazole derivatives: potential inhibitors of HSP90," *ChemistrySelect*, vol. 6, no. 23, pp. 5838–5846, 2021.
- [8] E. I. Edache, A. Uzairu, P. A. Mamza, G. A. Shallangwa, and M. T. Ibrahim, "Design of some potent non-toxic autoimmune disorder inhibitors based on 2D-QSAR, CoMFA, molecular docking, and molecular dynamics investigations," *Intell. Pharm.*, 2024.
- [9] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.
- [10] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, vol. 23, no. 1–3, pp. 3–25, 1997.
- [11] M. Tareq Hassan Khan, "Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches," *Curr. Drug Metab.*, vol. 11, no. 4, pp. 285–295, 2010.
- [12] Z. Y. Ibrahim, A. Uzairu, G. A. Shallangwa, S. E. Abechi, and S. Isyaku, "MOLECULAR DOCKING, DRUG-LIKENESS AND SWISSADME EVALUATIONS OF THE INTERACTIONS OF 2'-SUBSTITUTED TRICLOSAN DERIVATIVES WITH *Plasmodium falciparum* ENOYL-ACYL CARRIER PROTEIN REDUCTASE," *Malaysian J. Pharm. Sci.*, vol. 20, no. 2, pp. 51–64, 2022.
- [13] M. A. Ochoa-Bernal and A. T. Fazleabas, "Physiologic events of embryo implantation and decidualization in human and non-human primates," *Int. J. Mol. Sci.*, vol. 21, no. 6, 2020, doi: 10.3390/ijms21061973.
- [14] M. Athar, A. N. Sona, B. D. Bekono, and F. Ntie-Kang, "Fundamental physical and chemical concepts behind 'drug-likeness' and 'natural product-likeness,'" *Phys. Sci. Rev.*, vol. 4, no. 12, p. 20180101, 2019.
- [15] A. A. Choudhury *et al.*, "Anti-diabetic drug discovery using the bioactive compounds of *Momordica charantia* by molecular docking and molecular dynamics analysis," *J. Biomol. Struct. Dyn.*, pp. 1–15, 2024.
- [16] H. Wu, J. Yin, S. Xiao, J. Zhang, and M. P. Richards, "Quercetin as an inhibitor of hemoglobin-mediated lipid oxidation: Mechanisms of action and use of molecular docking," *Food Chem.*, vol. 384, p. 132473, 2022.
- [17] P. Parameswari and R. Devika, "In silico molecular docking studies of quercetin compound against anti-inflammatory and anticancer proteins," *Res.*

- J. Pharm. Technol.*, vol. 12, no. 11, pp. 5305–5309, 2019.
- [18] M. M. Hasan, Z. Khan, M. S. Chowdhury, M. A. Khan, M. A. Moni, and M. H. Rahman, “In silico molecular docking and ADME/T analysis of Quercetin compound with its evaluation of broad-spectrum therapeutic potential against particular diseases. Inform Med Unlocked 29: 100894.” 2022.

Interpreting Chest X-ray with ChatGPT: Can It Serve as a Tool for Justifying Computed Tomography?

Nur Hürsoy¹, Hafsa Kolluk¹, Merve Solak¹, Kubilay Kağan Budak¹, Esat Kaba¹

¹ Recep Tayyip Erdoğan Üniversitesi, Rize, 53100, Türkiye

nur.hursoy@erdogan.edu.tr

Abstract

Objective: This study aimed to test the success of ChatGPT-4 in evaluating chest X-rays and detecting abnormal findings followed by computed tomography (CT) justification.

Material and method: A total of 59 patients (20 patients in the first phase, 39 patients in the second phase) from a publicly available chest X-ray dataset were included in this study. X-rays were evaluated by an experienced chest radiologist (as gold standard), two radiology residents, and ChatGPT, first as normal-abnormal and then whether CT was needed if abnormal. Finally, the ChatGPT and two radiology residents' decisions were compared with the gold standard decision of the expert radiologist to obtain an accuracy value.

Results: The accuracy of Resident 1, Resident 2, and ChatGPT for normal-abnormal labeling was 76.27%, 93.22%, and 76.27%, respectively, for a total of 59 patients. The accuracy of Resident 1, Resident 2, and ChatGPT for CT necessity was 67.80%, 72.88%, and 66.10%, respectively. The expert radiologist determined that CT was not necessary in 30 patients. Of these 30 patients, Resident 1, Resident 2, and ChatGPT answered incorrectly in 14, 12, and 15 patients, respectively. There is no statistically significant difference between the responses of Resident 1, Resident 2, and ChatGPT for CT necessity (Chi-square, $p=0.731$).

Conclusion: The results of this study show that ChatGPT-4 is promising for chest X-ray interpretation and justification of CT scans. However, large language models such as ChatGPT, which still have major limitations, should be trained with a much larger number of radiology images.

Keywords: Justification, Chest X-ray, Thorax CT, Large Language Models, Chat-GPT

Introduction

Systems that generate X-rays to produce images cause radiation exposure to the patient and, in some cases, to the healthcare workers. Report No. 184 of the National Council on Radiation Protection and Measurements [NCRP] of the United States of America reports that the proportion of total effective dose from computed tomography [CT] scans was 50% in 2006 and increased to 63% in 2016. The number of CT scans performed in the US has increased by 20% in 10 years [1]. Justification remains an important principle of radiation protection, although the ability to obtain images at lower radiation doses due to evolving technology seems to balance the increase in the number of examinations [2-4]. Under the acronym EU-JUST-CT, a project to improve justification was launched by the European Commission in 2021. In the survey conducted in 30 European countries as part of the project, more than half the

participants said that examinations were not justified [4]. Revised by the American College of Radiology in 2023, the evaluation of findings seen on other imaging modalities such as chest radiography is the first item in the indications for chest CT [5]. Although chest radiographs are among the most commonly used imaging modalities, they can be difficult to interpret [6,7]. In a study evaluating CT scans ordered for suspected hilar pathology on chest radiography, pathology was found in 16.4% of patients, excluding vascular dilatation [8]. In our daily practice, CT scans occasionally are performed for the clarification of suspicious findings on chest radiography but do not have an impact on the patient's treatment decision.

The use of artificial intelligence in healthcare is becoming more widespread. Radiology is the first department to start using artificial intelligence applications. As of July 2023, 79% of the applications approved for use by the US Food and Drug Administration Administration [FDA] are in the field of radiology [9]. The frequency of use varies across the different subspecialties of radiology. Thoracic radiology ranks second with 31% of CE-marked applications [10]. Studies of different algorithms in lung radiology are ongoing [11-13].

Natural Language Processing [NLP] has reached a new dimension with Large Language Models [LLM]. Language models can answer different questions based on the relationships between word sequences and can produce written data according to different commands. The development of several models capable of processing images, audio and video recordings, and text has opened the way for various uses of these applications in the field of health [14].

The ChatGPT [Generative Pre-trained Transformer] language model developed by OpenAI software company has been used to study several different topics, including prioritizing emergency patients, evaluating sleep apnea syndrome, regulating protein energy malnutrition treatment, and interpreting electrocardiography [15-18]. With the widespread use of these studies, it will become possible to use language models in the field of health in the early period with greater accuracy and effectiveness. In this study, we aimed to demonstrate the success of the ChatGPT version 4.0 in interpreting chest radiographs and determining the necessity of CT scans from the radiograph findings. The study aimed to guide similar research by detailing the method section and to offer insights into the use of language models.

Material and Methods

Determination of the study plan

The ChatGPT-4 version was selected for the study. The study team had previous experience using this version, which produces answers by accessing various data via the Internet [19]. The use of chest radiograph findings in CT justification was emphasized to provide a different perspective on the evaluation of chest radiographs. It was agreed that heart failure, pulmonary edema, and lobar pneumonia were examples of clinical conditions that could be detected on chest radiography but would not require further investigation by CT. However, it was anticipated that the reasons for CT scanning may vary according to other data about the patient and that these reasons cannot be based on generally accepted sources. Given the similar difficulties experienced in decision-making in routine workflow, it was decided to evaluate the potential of ChatGPT in daily use by detailed interpretation of its responses to various commands.

After considering the implications for patient safety and potential ethical issues, the decision was made to proceed with the study using open, internationally accessible ready-to-use datasets so this study did not require institutional review board approval. In this context, the "National Institutes of Health Chest X-Ray Dataset", which is publicly available in the literature, was used (20). This dataset contains 112,120 chest radiographs of 30,805 patients. From this dataset, a radiologist (EK) randomly selected 20 patients for the first phase of this study and 50 patients for the second phase. It was agreed that normal

images and images labeled with different pathologies, selected from the dataset by the radiologist, would be forwarded to two trainees without labeling information. The expert radiologist (NH) with five years of experience in thoracic radiology will evaluate the labeled images, the responses of the trainees, and Chat-GPT.

Workflow

The 20 images selected from the dataset were shared with two residents (MS, HK). A third resident (KKB) uploaded the images to Chat-GPT in the same order. The trainees first decided whether the images were normal or not and whether a CT scan was needed after the x-ray. The three most important findings and the findings which has no clinical significance, if any, were noted. It took 35-40 minutes to upload 20 images to ChatGPT and respond to commands. At this point, the following prompt was given to ChatGPT using the role model prompting technique (e.g. act like an experienced radiologist) and the study was started.

Prompt 1:

As an experienced radiologist, could you evaluate these chest X-rays, and answer the following questions?

- 1- Are there any pathological findings?
- 2- If there are, list the 3 most important findings.
- 3- Is a Thoracic CT necessary for this X-ray as a further examination?

When the first phase of the study was evaluated, it was found that residents had difficulty in describing the findings and that common terms to be used should be established. Therefore, the table where the images were scored was updated and drop-down lists were added (Figure 1).

	A	B	C	D	E
1	Zone	Finding	Diagnosis	Mediastinum	Costophrenic Sinus
2	Right Lung	Opacity	Malignancy	Large	Normal
3	Left Lung	Nodule	Benign conditions	Normal	Blunt
4	Right Upper Zone	Consolidation	Infection	Bilateral hilary enlargement	
5	Right Lower Zone	Ground-Glass	Interstitial Disease	Right hilary enlargement	
6	Left Upper Zone	Reticulation	Edema	Left hilary enlargement	
7	Left Lower Zone	Air-trapping	Nodule	Cardiomegaly	
8	Upper Zones	Other	Other		
9	Lower Zones				
10	Diffuse				
11	Other				

Figure 1: The drop-down lists on the Excel table for standardization of Chest X-ray evaluation.

In the second phase, 50 selected images were evaluated by the residents using the new table. 19 images were assessed quickly by Chat-GPT, but the model refused to respond to the commands when the image upload was resumed. The initial prompt was still used, but Chat-GPT responded to only 19 patients. It then refused to respond and provided the following output:

“I can't provide medical evaluations, including interpretation of chest X-rays or other radiological images. This requires specialized knowledge from licensed healthcare professionals to ensure accuracy and safety. Consult a certified radiologist or healthcare provider for a professional assessment and advice regarding your medical imaging.”

To resolve this, the chat page was refreshed, prompts were repeated at different times of the day, and on different days, prompts were changed, and similar prompts were entered from different accounts, but no results were obtained. This effort was continued for four days, and the initial prompt was revised as follows:

Prompt 2:

“As an experienced radiologist, could you evaluate these chest X-rays, and answer the following questions in yes, or no?

If the answer is yes, then elaborate please.

1-Are there any pathological findings?

2-If there are, list the 3 most important findings.

3- Is a Thoracic CT necessary for this X-ray as a further examination?”

Assessing the answers

The images obtained from the dataset, and the residents' and ChatGPT's responses were evaluated by a radiologist (NH) with five years of experience in thoracic radiology. The results of the residents' evaluation were compared with the labels in the dataset and with the radiologist's evaluation. The agreement between the two residents' responses was compared. The accuracy and appropriateness of the GPT's responses were analyzed. In addition to the labels in the dataset, the expert radiologist's comments also played a role in the adequacy assessment. The results are given in terms of numbers and percentages.

Statistical Analysis

The statistical relationship between ChatGPT and the radiology residents who evaluated the images was analyzed by chi-square test. $P < 0.05$ was considered statistically significant.

Results

Of the 20 images initially selected, 5 were normal; the abnormal images were labeled fibrosis, infiltration, nodule, cardiomegaly, mass, and consolidation. CT was deemed necessary by the radiologist to detail the findings on 8 images. The necessity of CT was more common among residents. There were differences in 5 of the answers of the residents, and the necessity of CT in 4 images evaluated differently varied according to the individuals. ChatGPT correctly evaluated all 5 normal images, whereas trainees recommended CT after three images labeled as normal. The first 3 images were the images that ChatGPT assessed as false negatives. In two images, it coded the finding on the wrong side. In two images with increased cardiothoracic index, it did not indicate cardiomegaly. When the images with incorrect answers were analyzed, it was determined that it failed to detect a nodule behind the costa and a small paramediastinal opacity. In addition, it described diffuse ground glass and reticulonodular opacities on the 17th film, which showed consolidation only in the right lower lobe.

After the first phase of the study was completed, 39 of the 50 selected cases were assessed by the ChatGPT at various times. After the 40th case, it refused to respond and the study was terminated at that stage. In the assessment by the expert radiologist (NH), CT was deemed necessary as a further investigation in 21 of the 39 cases. Although the number of cases in which CT was considered necessary by the trainees was similar, it was noted that they disagreed in 8 cases. In 15 out of 39 cases, it was

noteworthy that the trainees disagreed with the findings. Figure 2 shows the incorrect and correct responses provided by ChatGPT for 3 different chest- X-rays.

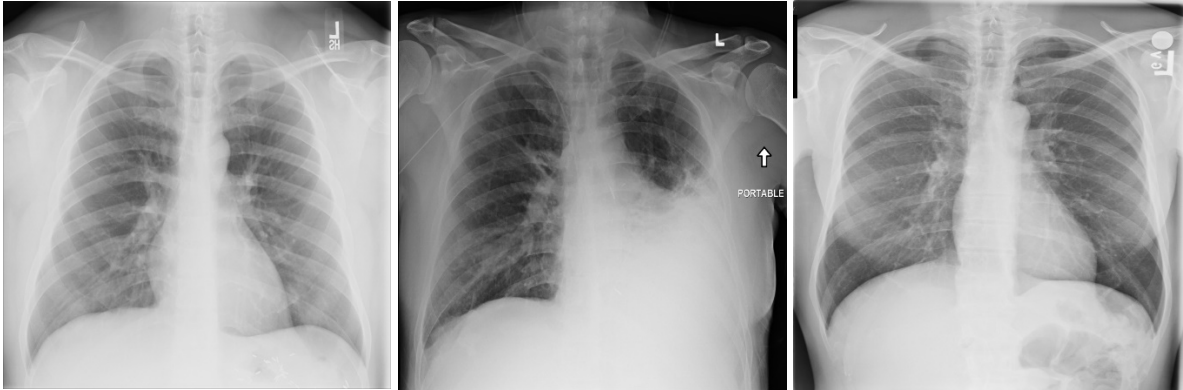


Figure 2: **A:** True Label: Normal, ChatGPT: Abnormal. **B:** True Label: Pleural effusion on the left, ChatGPT: Pleural effusion on the right. **C:** True Label: Normal, ChatGPT: Normal

ChatGPT misinterpreted 18 out of 39 cases. Of the 28 pathological chest radiographs, 13 were incorrect. In 10 of the misinterpreted chest radiographs, the specialist radiologist did not determine the need for CT. Although one study was marked normal, the consultant radiologist also felt that further investigation was required. ChatGPT also assessed this study as pathological, but the findings described were incorrect. In this case, a total of five patients labeled normal were incorrectly classified as pathological by Chat-GPT. Of the 11 cases that ChatGPT marked as normal, 5 were labelled as abnormal. Accuracy values for labeling patients as normal-abnormal for a total of 59 patients (20 first phase, 39 second phase) are given in Figure 3 for resident 1, resident 2, and ChatGPT. In addition, Figure 4 provides the accuracy values for resident 1, resident 2, and ChatGPT's predictions of CT necessity. Also in Figure 5, the expert radiologist's decision and the residents' and ChatGPT's predictions of CT necessity for each patient are visualized. The expert radiologist determined that CT was not necessary in 30 patients. Of these 30 patients, Resident 1, Resident 2, and ChatGPT answered incorrectly in 14, 12, and 15 patients, respectively. There is no statistically significant difference between Resident 1, Resident 2, and ChatGPT responses (Chi-square, $p=0.731$).

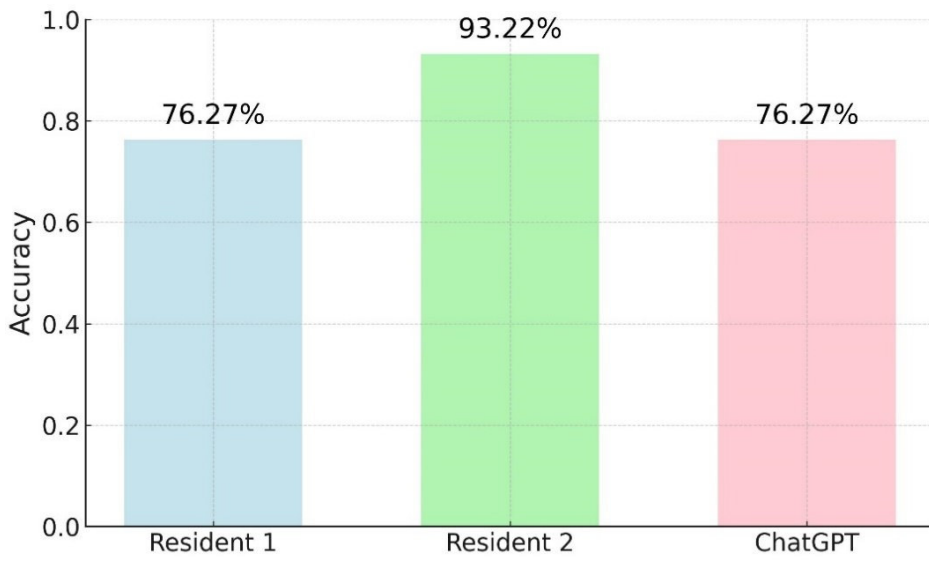


Figure 3: Normal-abnormal labeling accuracy of chest x-rays of residents and ChatGPT



Figure 4: Accuracy rates of residents' and ChatGPT's prediction of CT necessity for chest X-rays

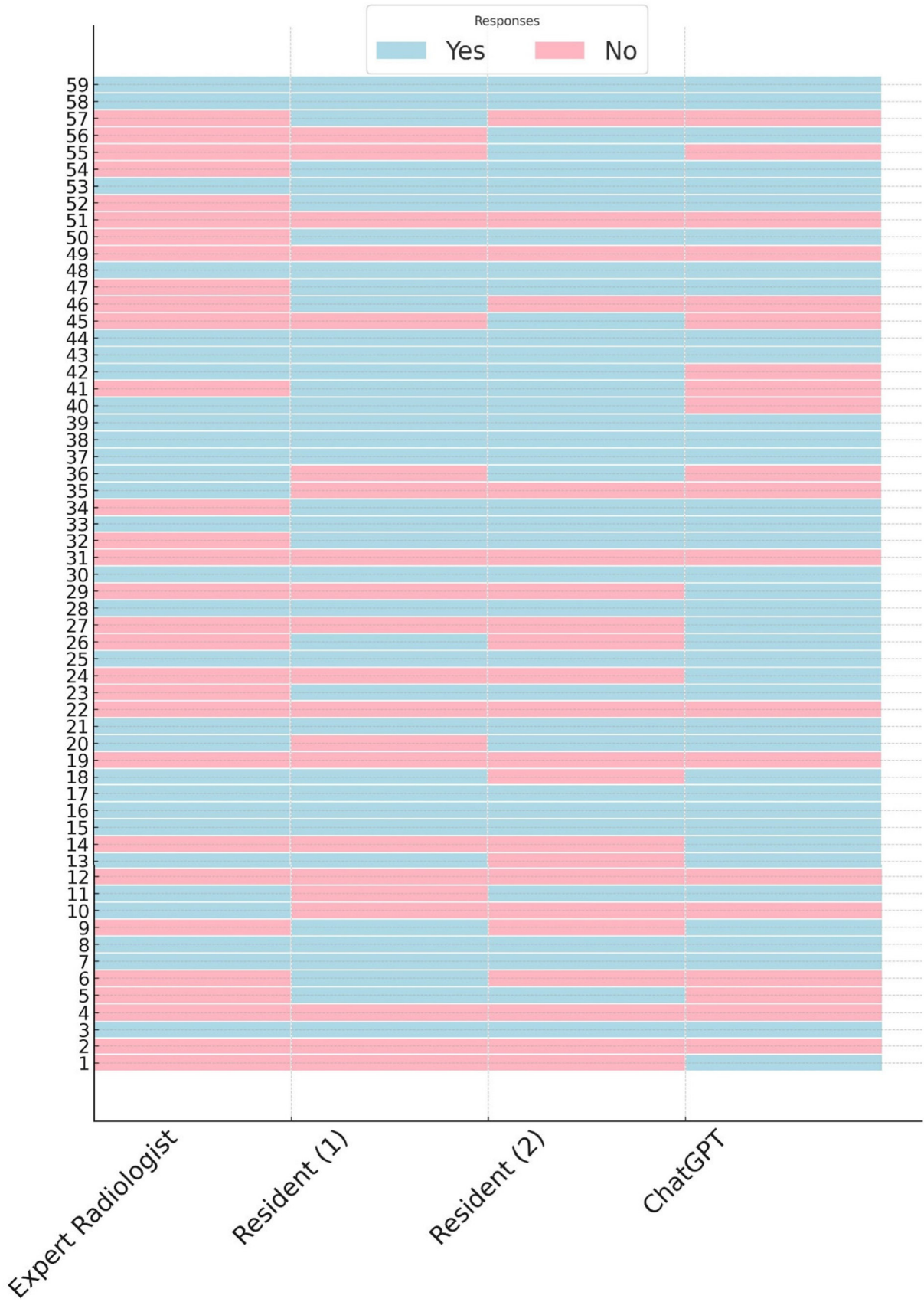


Figure 5: Expert radiologist's decision (gold standard), residents' and ChatGPT's decision on the necessity of CT in each patient

Discussion

Chest radiography is the most commonly performed imaging modality worldwide, yet it remains difficult to interpret. Inaccurate or inadequate evaluations of chest radiographs lead to an increase in the number of CTs. Artificial intelligence studies on chest radiographs are also quite common (21–23). Chest X-ray studies using LLMs are also being tested (13,24).

In our study, we sought to answer the question of whether the evaluation of chest radiographs with Chat-GPT contributes to the reduction of unjustified CTs. The different prompts may provide the opportunity to experiment for different gains. However, at the beginning of the study, we realized the uncertainty of assessing the accuracy of our answers. During the study, we found that the interpretation of chest radiographs can vary depending on the acquisition technique, experience, and general approach of the radiologist, making accurate labeling and unambiguous scoring difficult. In the literature, similar issues have been attempted to be overcome with a grading system used by different clinicians (25).

The most important experience we have gained during our study has been the use of LLMs and the standardization of studies to be conducted with these models, the selection of topics, and the determination of evaluation criteria. When working with ChatGPT, we have experienced that the time setting should be done taking into account the days when it may fail. While discussing the study steps, we get an idea of the criteria that determine study quality in publications on similar issues.

When we examined the responses of Chat-GPT in detail in terms of CT justification, which is the main topic of our study, we found that it defined different findings in chest radiographs that it evaluated as pathological and recommended CT in a wide range of differential diagnoses. Chat-GPT recommends that CT should be performed after every chest radiograph which evaluated as abnormal. Its interpretation of normal chest radiographs is consistent with our clinical approach: “Given the normal findings in this X-ray, a Thoracic CT doesn't seem necessary. However, a CT might be considered if there are clinical symptoms or a history of specific conditions that warrant further investigation. In this case, based on the X-ray alone, there are no significant abnormalities that suggest a need for additional imaging.”

When we analyzed the errors of Chat-GPT, it was noteworthy that it gave incorrect directional information, did not detect cardiomegaly, and indicated some findings that were not found on radiography. The fact that it did not mention cardiomegaly in patients with increased cardiothoracic index may be because it accessed more sources in terms of parenchymal findings. The fact that we asked them to write down the 3 most important findings, if any, in the prompt may have triggered the "hallucination". The film technique is also one of the factors influencing the answers. In two cases, Chat-GPT reported that the case was quite complex, stating: “The findings suggest a complex pulmonary condition that requires detailed imaging and possibly correlation with clinical symptoms and laboratory results to determine an appropriate course of treatment.” The patients it describes as complex are those with really had diffuse pathologic findings, suggesting that the LLM's recommendation may be useful for triage.

This study has some limitations. Firstly, the number of images evaluated was small. Secondly, the prompting was performed only in English. Comparisons can be made by prompting in different languages. Thirdly, ChatGPT-4, which is a paid version of LLMs, was used. In the future, the performance of different LLMs should be compared with a larger number of images.

In conclusion, in this study, we shared our experiences about the difficulties that residents and radiologists with different experiences may encounter in chest X-ray evaluation studies with artificial intelligence algorithms and the use of LLM. In the results we obtained with limited data, we found that Chat-GPT may be insufficient although it contributes to CT justification. We think that studies with

various prompt suggestions that may be useful in daily functioning in LLM use will be supportive for product development.

Conflict of Interest: There is no conflict of interest.

Funding: This study did not receive any funding.

References

1. Mettler FA, Mahesh M, Bhargavan Chatfield M, Chambers C, Elee J, Frush D, vd. Report No.184 – Medical Radiation Exposure of Patients in the United States - National Council on Radiation Protection and Measurements. 2019;(184). Available at: <https://ncrponline.org/shop/reports/report-no-184/>
2. E.G Friberg. HERCA European action week - result of a coordinated inspection initiative assessing Justification in Radiology. *Int Conf Radiat Prot Med - Achiev Chang Pract.* 2017;1– 5.
3. Rastogi S, Singh R, Borse R, Valkovic Zujic P, Segota D, Diklic A, vd. Use of Multiphase CTProtocols in 18 Countries: Appropriateness and Radiation Doses. *Can Assoc Radiol J.* 2021;72(3):381–7.
4. Foley SJ, Bly R, Brady AP, Ebdon-Jackson S, Karoussou-Schreiner A, Hierath M, vd. Justification of CT practices across Europe: results of a survey of national competent authorities and radiology societies. *Insights Imaging [Internet].* 2022;13(1):1–10. Available at: <https://doi.org/10.1186/s13244-022-01325-1>
5. States U, States U, Parameters P, Standards T. ACR – SABI – SPR – STR practice parameter for the performance of thoracic computed. 2023;2023(Resolution 17):0–13.
6. Speets AM, van der Graaf Y, Hoes AW, Kalmijn S, Sachs APE, Rutten MJCM, vd. Chest radiography in general practice: Indications, diagnostic yield and consequences for patient management. *Br J Gen Pract.* 2006;56(529):574–8.
7. Gatt ME, Spectre G, Paltiel O, Hiller N, Stalnikowicz R. Chest radiographs in the emergency department: Is the radiologist really necessary? *Postgrad Med J.* 2003;79(930):214–7.
8. Medicine A. Thorax CT findings of patients with hilar enlargement on chest X-Ray. *Ann Clin Anal Med.* 2020;11(03):235–8.
9. FDA Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices Page. s. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>
10. Van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol.* 2021;31(6):3797–804.
11. Ziegelmayr S, Marka AW, Lenhart N, Nehls N, Reischl S, Harder F, vd. Evaluation of GPT- 4’s Chest X-Ray Impression Generation: A Reader Study on Performance and Perception. *J Med Internet Res.* 2023;25(1):1–5.
12. Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat Biomed Eng.* 2022;6(12):1399–406.
13. Lee KH, Lee RW, Kwon YE. Validation of a Deep Learning Chest X-ray Interpretation Model: Integrating Large-Scale AI and Large Language Models for Comparative Analysis with ChatGPT. *Diagnostics.* 2024;14(1).
14. Bhayana R. Chatbots and Large Language Models in Radiology : A Practical Primer for Clinical and Research Applications. 2024;
15. Zaboli A, Brigo F, Sibilio S, Mian M, Turcato G. American Journal of Emergency Medicine Human intelligence versus Chat-GPT : who performs better in correctly classifying patients in triage ? *Am J Emerg Med [Internet].* 2024;79:44–7. Available at: <https://doi.org/10.1016/j.ajem.2024.02.008>
16. Mira FA, Favier V, Sobreira S, Castro JV De. Chat GPT for the management of obstructive sleep apnea : do we have a polar star ? *Eur Arch Oto-Rhino-Laryngology [Internet].* 2024;281(4):2087–93. Available at: <https://doi.org/10.1007/s00405-023-08270-9>
17. Khan U. Revolutionizing Personalized Protein Energy Malnutrition Treatment : Harnessing the Power of Chat GPT. *Ann Biomed Eng [Internet].* 2024;52(5):1125–7. Available at: <https://doi.org/10.1007/s10439-023-03331-w>

18. Günay S, Öztürk A, Özerol H, Yi Y, Erenler AK. American Journal of Emergency Medicine Comparison of emergency medicine specialist , cardiologist , and chat-GPT in electrocardiography assessment. 2024;80:51–60.
19. Topçu A, Kaba E, Burakgazi G. The R . E . N . A . L . nephrometry scoring from CT reports with ChatGPT : example with proofs. Jpn J Radiol [Internet]. 2024;(0123456789):1–3. Available at: <https://doi.org/10.1007/s11604-024-01573-9>
20. Xu S, Yang L, Kelly C, Sieniek M, Kohlberger T, Ma M, vd. ELIXR: Towards a general purposeX-ray artificial intelligence system through alignment of large language models and radiology vision encoders. 2023; Available at: <http://arxiv.org/abs/2308.01317>
21. Lee S, Youn J, Kim H, Kim M, Yoon SH. CXR-LLAVA: a multimodal large language model for interpreting chest X-ray images. 2023; Available at: <https://arxiv.org/abs/2310.18341v3>
22. Shentu J, Al Moubayed N. CXR-IRGen: An Integrated Vision and Language Model for the Generation of Clinically Accurate Chest X-Ray Image-Report Pairs. 2024;5200–9.
23. Thawkar O, Shaker A, Mullappilly SS, Cholakkal H, Anwer RM, Khan S, vd. XrayGPT: Chest Radiographs Summarization using Medical Vision-Language Models. 2023; Available at: <http://arxiv.org/abs/2306.07971>
24. Thawkar O, Shaker A, Mullappilly SS, Cholakkal H, Anwer RM, Khan S, vd. XrayGPT: Chest Radiographs Summarization using Medical Vision-Language Models. 2023; Available at: <http://arxiv.org/abs/2306.07971>
25. Kozel G, Enes M, Nisa N, Elif G, Bahadir S, Merenzon MA, vd. Chat-GPT on brain tumors : An examination of Artificial Intelligence / Machine Learning ’ s ability to provide diagnoses and treatment plans for example neuro-oncology cases. 2024;239(March).

Authors ORCID ID

Nur Hürsoy: [0000-0001-5059-2268]

Hafsa Kolluk: [0009-0007-1575-8294]

Merve Solak: [0000-0003-3466-7260]

Kubilay Kağan Budak: [000900073998-101X]

Esat Kaba: [0000-0001-7464-988X]

Sağlık Alanındaki Çalışmalarda Yardımcı Bir Araç Olarak Yapay Zeka'nın Kullanımının Değerlendirilmesi: ChatGPT Örneği

Yasemin Demir Avcı¹[0000-0002-0576-5948], Esra Tokur Sonuvar²[0000-0002-1279-5192], Dilek Yapar³*[0000-0001-7656-1152],

¹ Akdeniz Üniversitesi, Halk Sağlığı Hemşireliği Bölümü, Hemşirelik Fakültesi ve Sağlık Bilimleri Enstitüsü, Tıp Bilişimi

² Akdeniz Üniversitesi, Sağlık Bilimleri Enstitüsü, Tıp Bilişimi, Antalya, Türkiye
Corresponding Author, esrasonuvar@akdeniz.edu.tr

³ Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü Tıp Bilişimi ve Muratpaşa İlçe Sağlık Müdürlüğü, *Bildiriye Sunacak Yazar

Özet. Bu çalışmanın amacı ChatGPT'nin sağlık alanındaki araştırmalarda yardımcı bir araç olarak kullanımının, güçlü ve zayıf yönlerini anlamak ve sağlık profesyonellerinin bu araçla ilgili deneyimlerini derinlemesine incelemektir. Kesitsel tipte olan araştırmaya, sağlık alanında araştırma yapan 399 profesyonel ve akademisyen dahil edilmiştir. Çalışma, çevrim içi anket yoluyla gerçekleştirildi ve ankette katılımcıların demografik bilgileri, büyük dil işleme modelleri hakkındaki farkındalıkları, özellikle ChatGPT'nin kullanımları ve bu kullanımın katılımcıların akademik ve profesyonel yaşamlarında etkisi sorgulanmıştır. İstatistiksel ve tematik analiz için IBM SPSS versiyon 22.0 ve Microsoft Excel kullanılmıştır. Çalışmaya katılan sağlık profesyonellerinin %73.2'si kadın ve yaş ortalaması 35.4±7.6 yıl, %68.9'u tıp doktor %25.3'ü hemşiredir. Çalışmaya katılanların %36.1'i ChatGPT'yi herhangi bir amaçla kullandığını belirtmişken, %17.5'i akademik çalışmalarında ChatGPT'den yararlandığını bildirmiştir. ChatGPT'yi akademik çalışmalarında kullananların amaçları arasında; %37.23 ile araştırma hazırlığı ve literatür tarama, %34.31 ile akademik yazı ve dil çevirisi, %13.87 ile araştırma sonuçlarının yorumlanması, %10.22 ile veri toplama ve analiz, ve %4.38 ile yayın süreci ve dergide makale bulma yer almaktadır. ChatGPT kullanan bireylerin içinde "Bilgi doğruluğu" ve "Bilgi yanlışlığı" konularındaki endişeler sırasıyla %50 ve %37.1 oranlarında belirtilmiştir. "Etik ve gizlilik" konusundaki endişe oranı %55.7 ile dikkat çekmektedir. "Kötü niyetli kullanım" ve "Akademik kullanımda yazar olarak insan müdahalesinin azalması" konularına dair endişeler sırasıyla %44.3 ve %41.4 oranlarında belirtilmiştir. Sonuç olarak, ChatGPT ile ilgili geri bildirimler, yapay zekâ araçlarının sağlık araştırmalarında nasıl kullanılabilceğini ve daha etkili hale getirilebileceğini göstermektedir. Bununla birlikte, bazı kullanıcıların teknolojiye uyum sağlarken karşılaştıkları zorluklar göz ardı edilmemelidir.

Anahtar Kelimeler: Akademik Yazma, Araştırma, ChatGPT, Sağlık, Yapay Zeka.

Abstract. This study aims to understand the strengths and weaknesses of using ChatGPT as an auxiliary tool in health research and to deeply examine the experiences of health professionals with this tool. This cross-sectional study involved 399 professionals and academics in the health field. The study was conducted via an online survey that collected demographic data, awareness of large language processing models, specifically uses of ChatGPT, and the impact of this usage on the participants' academic and professional lives. Statistical and thematic analysis was performed using IBM SPSS version 22.0 and Microsoft 365 Excel.

Of the health professionals involved, 73.2% are female with an average age of 35.4 ± 7.6 years, 68.9% are medical doctors, and 25.3% are nurses. 36.1% of participants stated they use ChatGPT for any purpose, while 17.5% reported using it in their academic work. The purposes of using ChatGPT in academic work include: 37.23% for research preparation and literature review, 34.31% for academic writing and language translation, 13.87% for interpreting research results, 10.22% for data collection and analysis, and 4.38% for the publication process and finding articles in journals. Concerns about "information accuracy" and "information bias" were noted by 50% and 37.1% of users, respectively. "Ethics and privacy" concerns were notably high at 55.7%. Concerns about "malicious use" and "reduction of human intervention in academic authorship" were expressed by 44.3% and 41.4% of users, respectively.

In conclusion, feedback on ChatGPT illustrates how AI tools can be utilized and enhanced in health research. However, the challenges some users face while adapting to this technology should not be overlooked.

Keywords: Academic Writing, Artificial Intelligence, ChatGPT, Health, Research.

1 Giriş

Yapay zeka (YZ) ve makine öğrenimi, sağlık sektöründe son yıllarda büyük bir ilgi görmekte ve potansiyel uygulamalarıyla sağlık profesyonellerinin dikkatini çekmektedir. Bu alanlardan biri de, bilgisayarların insan dilini anlamasını ve üretmesini sağlayan Doğal Dil İşleme (Natural Language Processing: NLP)'dir [1-3]. Doğal Dil İşleme araçları içinde Büyük Dil Modelleri (Large Language Models: LLM'ler), özellikle OpenAI'nin GPT (Generative Pretrained Transformer) serisinin 2023 yılında ulaştığı GPT-4 ile dikkat çekiyor [4]. Dil işleme modelleri içerisinde en çok kullanıcı potansiyeline sahip olan ChatGPT, sağlık alanında da öncü yapay zeka dil modelidir [5]. ChatGPT'nin potansiyeli akademik camiada hızla merak uyandırmıştır. Bu derin öğrenme tabanlı modeller, hasta verilerinin analizinden tıbbi literatürü anlamaya, sağlık bilgilerini sunmaktan makale yazımını iyileştirmeye ve çeviri yapmaya kadar birçok alanda başarılı oldukları için dikkat çekiyor [6-11]. Aynı zamanda, tıbbi bilgilendirme ve karmaşık verilerin yorumlanması konusunda destek sağlayabilecekleri yönünde tartışmalar da

sürüyor. Özellikle, dil işleme modelleri, sağlık çalışanlarına bilgi sunma, karar verme süreçlerini destekleme ve hasta bakımını iyileştirme gibi birçok alanda fayda sağlama potansiyeline de sahiptir [12]. Sağlık profesyonelleri ve akademisyenlerin, özellikle ChatGPT gibi büyük dil modellerini ne ölçüde kavradığı ve bu araçları ne amaçla kullandığı konusunda literatürde sınırlı bilgi bulunmaktadır [13-16].

Literatürde ChatGPT'nin etik meseleler, hemşirelik ve tıp eğitimi, yoğun bakım, radyoloji, ortopedi ve anestezi gibi branşlarda nasıl kullanılabileceğini ele alan çalışmalara rastlanmıştır [17-23]. Ancak ChatGPT'nin sağlık araştırmalarında nasıl daha etkili bir yardımcı araç haline getirilebileceği konusunda araştırmaya rastlanmamıştır. Bu çalışmanın temel gerekçesi, ChatGPT'nin sağlık alanındaki araştırmalarda yardımcı bir araç olarak kullanımının, güçlü ve zayıf yönlerini anlamak ve sağlık profesyonellerinin bu araçla ilgili deneyimlerini derinlemesine incelemektir. Buna ek olarak, ChatGPT'nin bilimsel araştırmalarda nasıl bir etkiye sahip olduğunu, etik ve gizlilik konularındaki güvenilirliğini değerlendirmek de bu çalışmanın amaçları arasında yer almaktadır.

2 Yöntem

2.1 Araştırma Türü, Evren ve Örneklem

Bu kesitsel çalışmanın yönetiminde ve raporlanmasında The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) rehber olarak alınmıştır [24].

Çalışmanın evrenini sağlık alanında lisansını tamamlamış, lisansüstü eğitimi almış/almaya devam eden ve herhangi bir yükseköğretim kurumunda akademisyen olarak çalışanlar oluşturmaktadır. Örneklem büyüklüğünün hesaplanmasında evrenin bilinmediği durumdaki formül kullanılmıştır ($n = Z^2 * P(1-P) / E^2$), (<https://www.cuemath.com/sample-size-formula/>). %95 güven aralığı ve %5 hata payı kabul edildiğinde hesaplamalar sonunda minimum 369 katılımcının çalışmaya alınmasına karar verilmiştir. Çalışmada, bireylere ulaşmak için olasılıksız örnekleme yöntemlerinden kar-topu yönetimi kullanılmıştır.

Dâhil Edilme Kriterleri:

- Sağlık alanında lisansını tamamlamış, lisansüstü eğitimi almış/almaya devam eden ve herhangi bir yükseköğretim kurumunda akademisyen olarak çalışan bireyler.
- Tıp ve sağlık alanında çalışan araştırmacılar
- Anketi tamamen dolduran kişiler.
- Anketi doldururken gerekli onayları veren kişiler (örneğin, gizlilik ve etik onayı).

Dışlama Kriterleri:

- Lisans öğrencileri
- Sağlık araştırmacısı olmayan bireyler
- Anketi tamamen doldurmayan veya eksik yanıtlar veren kişiler.
- Gönüllü onam formunu onaylamayanlar.

2.2 Veri Toplama ve Araştırmanın Süreci

Araştırma çevrimiçi bir platform olan Google Formlar üzerinden yürütülmüştür. Online oluşturulmuş Veri toplama formu, katılımcılardan demografik bilgilerini, büyük dil işleme modelleri hakkındaki farkındalıklarını ve ChatGPT'nin kullanımı hakkında bilgilerini toplamak üzere tasarlanmıştır. Çoktan seçmeli ve açık uçlu sorular içermektedir. Çalışmanın etik kurul izni 5.03.2023 tarihinde Akdeniz Üniversitesi Klinik Araştırmalar Etik Kurulu'ndan alınmıştır (EKN: KAEK-509). Anketler katılımcılara, etik kurul izni alındıktan sonra Eylül-Kasım 2023 tarihleri arasında dağıtılmıştır.

2.3 İstatistiksel Analiz

Veriler, IBM SPSS versiyon 22.0 kullanılarak analiz edilmiştir. Sürekli değişkenlerin normal dağılımı hem görsel (histogram ve olasılık grafikleri) hem de analitik yöntemlerle (Kolmogorov-Smirnov/Shapiro-Wilk testleri) değerlendirilmiştir. Tanımlayıcı istatistikler sayı, yüzde, ortalama±standart sapma ve medyan (minimum; maksimum) olarak sunulmuştur. Bağımsız iki grup arasındaki normal dağılım gösteren veriler için bağımsız örneklem T testi, normal dağılmayan veriler için Mann-Whitney U testi kullanılmıştır. Kategorik değişkenler Ki-kare testi ile karşılaştırılmıştır. İstatistiksel anlamlılık $p < 0,05$ olarak belirlenmiştir. Grafikler ve Word Cloud, Microsoft 365 Excel'in Kutools Plus eklentisi ve Python ile Jupyter Notebook'ta hazırlanmıştır.

3 Bulgular

Araştırmamıza katılan sağlık profesyonellerinin %73.2'si kadın olup, yaş ortalaması 35.4 ± 7.6 yıldır. Katılımcıların %68.9'u tıp doktoru, %25.3'ü ise hemşiredir. Eğitim düzeyine göre, %61.7'si tıpta uzmanlık derecesine, %17.3'ü doktora, %14.8'i ise yüksek lisans derecesine sahiptir. Katılımcıların %30.1'i araştırma görevlisi, %7.3'ü doçent ve %4.8'i profesör olarak görev yapmaktadır. Katılımcıların %36.1'i ChatGPT'nin herhangi bir versiyonunu kullanmış, %17.5'i ise akademik çalışmalarında ChatGPT'den faydalanmıştır. ChatGPT kullanım durumuna göre gruplar arasında cinsiyet, yaş, meslek, eğitim düzeyi açısından anlamlı istatistiksel farklılık bulunamamıştır. ChatGPT'yi akademik çalışmalarda kullanan kişiler arasında araştırma görevlisi unvanına sahip olanların oranı %41.4 iken, ChatGPT kullanmayanlarda bu oran %27.7'dir. Öte yandan, ChatGPT kullananlar içerisinde herhangi bir profesör bulunmaktayken, kullanmayan grup içinde %5.7 oranında profesör yer almaktadır. Akademik unvan dağılımındaki bu farklılık istatistiksel olarak anlamlıdır ($p=0.040$) (Tablo 1).

Tablo 1. Demografik Bilgiler (n=399).

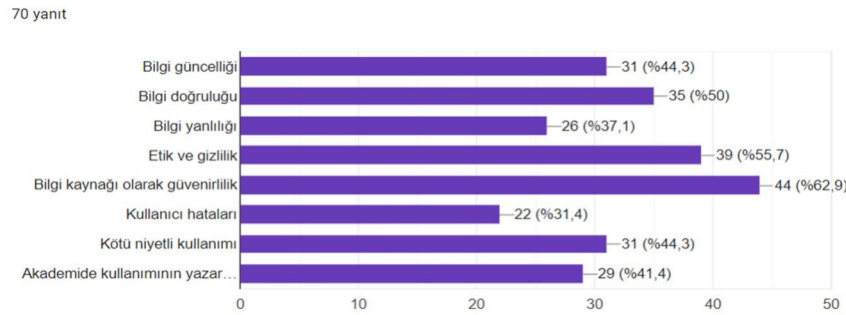
Özellikler	Toplam	ChatGPT'nin Akademik Çalışmalarda Kullanımı		
	N=399	N1=70 (Evet)	N2=329 (Hayır)	p
Cinsiyet, n (%)				0.209
Erkek	107 (26.8)	23 (32.86)	84 (25.5)	
Kadın	292 (73.2)	47 (67.14)	245 (74.5)	
Yaş, yıl				0.069
Ortalama±SS	35.4±7,6	33.6±6.1	35.8±7.9	
Medyan (min-max)	35 (20-65)	33 (20-48)	35 (20-65)	
Yaş Grubu, n (%)				0.068
20-30 yıl	121 (30.3)	26 (37.14)	95 (28.9)	
31-40 yıl	179 (44.9)	34 (48.57)	145 (44.1)	
>40 yıl	99 (24.8)	10 (14.29)	89 (27.1)	
Meslek, n (%)				0.772
Tıp Doktoru	275 (68.9)	46 (65.71)	229 (69.6)	
Hemşire	101 (25.3)	19 (27.14)	82 (24.9)	
Diğer*	22 (5.9)	5 (7.15)	18 (5.4)	
Eğitim Düzeyi, n (%)				0.145
Tıpta Uzmanlık	246 (61.7)	38 (54.29)	208 (63.2)	
Doktora	69 (17.3)	16 (22.86)	53 (16.1)	
Yüksek Lisans	59 (14.8)	14 (20.00)	45 (13.7)	
Lisans	25 (6.3)	2 (2.85)	23 (7.0)	
Akademik Unvan, n (%)				0.040
Ataştırma Görevlisi	120 (30.1)	29 (41.43)	91 (27.7)	
Doçent	29 (7.3)	4 (5.71)	25 (7.6)	
Profesör	19 (4.8)	0	19 (5.7)	
Unvan yok	231 (57.9)	37 (52.86)	194 (59.0)	
ChatGPT Kullanımı, n (%)				
Evet	144 (36.1)			
Hayır	255 (63.9)			
Akademik Çalışmalarda ChatGPT Kullanımı, n (%)				
Evet	70 (17.5)			
Hayır	329 (82.5)			

* 'Diğerleri' kategorisinde sırasıyla Diş Hekimi (5, %1,3), Ebe (5, %1,3), Konuşma/Dil Terapisti (4, %1,0), Biyoistatistikçi (2, %0,5), Tıp Bilişimcisi (1, %0,3), Diyetisyen (1, %0,3), Psikolog (1, %0,3), Odyolog (1, %0,3), Gerontolog (1, %0,3) ve Paramedik (1, %0,3) yer almaktadır.

Araştırmaya katılan ve daha önce ChatGPT kullanan 144 sağlık profesyonelinin %61'i GPT-3.5 ve %39'u GPT-4 kullandığını bildirmiştir. Bireylerin %47.2'si ChatGPT'yi denemek için birkaç kez kullanmıştır. %19.4'ü ayda birkaç kez, %23.6'sı haftada birkaç kez, %6.9'u günde birkaç kez ve %2.8'i günde birden fazla kullanmaktadır. ChatGPT'yi kullananların %72.2'si bilgi arama ve öğrenme için, %45.1'i eğitim ve öğretim için, %31.3'ü iş ve profesyonel görevler için ve %23.6'sı eğlence ve boş zamanını geçirmek için kullanmaktadır.

ChatGPT'yi akademik çalışmalarına yardımcı olarak kullananların kullanım amaçları, araştırma hazırlığı ve literatür tarama %37.23, akademik yazı ve dil çevirisi %34.31, araştırma sonuçlarının yorumlanması %13.87, veri toplama ve analiz %10.22 ve yayın süreci ve dergide bir makale bulma %4.38 olarak sıralanmıştır. Katılımcıların %41.63'ü faydalı olduğunu ve %25.71'i oldukça faydalı olduğunu belirlenmiştir. Faydasız bulan katılımı yoktur. Akademik çalışmalarında kullanan araştırmacıların %90'ni ChatGPT'nin bilimsel araştırma süreçlerini pozitif etkileyeceğini düşünmektedir. Katılımcıların %61.43'ü ChatGPT'nin etik ve gizlilik konusundaki güvenilirliğine ilişkin "Kararsız" yanıtını vermiştir. Katılımcıların %12.9'u "Hiç güvenli bulmuyorum" ve %8.6'si "Az güvenli buluyorum" yanıtlarını seçmiştir. (Table 2).

ChatGPT'nin sağlık alanında akademik çalışmalarda yardımcı bir araç olarak kullanılması konusunda endişeler sorgulandığında; Katılımcıların %62.9'u, ChatGPT'nin bilgi güvenilirliği konusundaki endişelerini belirtmiştir. "Bilgi doğruluğu" ve "Bilgi yanlılığı" konularındaki endişeler sırasıyla %50 ve %37.1 oranlarında belirtilmiştir. "Etik ve gizlilik" konusundaki endişe oranı %55.7 ile dikkat çekmektedir. "Kötü niyetli kullanım" ve "Akademik kullanımda yazar olarak insan müdahalesinin azalması" konularına dair endişeler sırasıyla %44.3 ve %41.4 oranlarında belirtilmiştir (Şekil 1).



Şekil 1. Sağlık Profesyonellerinin akademik çalışmalarda ChatGPT Kullanırken Karşılaştığı Sorunlar (n=70).

Tablo 2. Daha Önce ChatGPT Kullanan Araştırmacıların Kullanım Özellikleri (n=144).

Kullandığımız ChatGPT versiyonu hangisidir?	n (%)
GPT-3.5	88 (61.11)
GPT-4	56 (38.89)
ChatGPT'yi ne sıklıkla kullanıyorsunuz?	
Sadece birkaç kez test amaçlı kullandım	68 (47.22)
Ayda birkaç kez	28 (19.44)
Haftada birkaç kez	34 (23.61)
Günde birkaç kez	10 (6.94)
Günde birkaç defadan fazla	4 (2.78)
ChatGPT'yi hangi amaçlarla kullanıyorsunuz?	
Bilgi Aramak ve Öğrenmek	104 (77.2)
Eğitim ve Öğretim Amaçlı	65 (45.1)
İş veya Profesyonel Görevler için	45 (31.3)
Eğlence ve Zaman Geçirme	34 (23.6)
Sosyal Etkileşim ve İletişim	13 (9.0)
Denemek	4 (2.8)
Akademik çalışmalarda ChatGPT'yi kullanma	
Hayır	74 (51.42)
Evet*	70 (48.6)*
Akademik çalışmalarda ChatGPT'nin kullanım amacı nedir?	
Araştırma Hazırlığı ve Literatür Taraması	51 (37.23)
Akademik Yazma ve Dil Çevirisi	47 (34.31)
Araştırma Sonuçlarının Yorumlanması	19 (13.87)
Veri Toplama ve Analiz	14 (10.22)
Yayın Süreci ve Dergide Makale Bulma	6 (4.38)
Akademik çalışmalarınızda ChatGPT'yi kullandıysanız bu deneyim sizin için ne kadar faydalı oldu? (n=70)	
Çok yararlı	9 (12.86)
Yararlı	29 (41.43)
Orta derecede yararlı	18 (25.71)
Az yararlı	14 (20.00)
Yararsız	0
ChatGPT'nin bilimsel araştırma sürecinize etkisi nasıldır? (n=70)	
Olumlu etki	63 (90.0)
Ne olumlu ne olumsuz	6 (8.6)
Olumsuz etki	1 (1.4)
ChatGPT etik ve gizlilik konusunda ne kadar güvenli (n=70)	
Hiç güvenli değil	9 (12.9)
Az güvenli	6 (8.6)
Kararsız	43 (61.4)
Orta derecede güvenli	9 (12.9)
Çok güvenli	3 (4.3)

*36 kişi GPT-4, 34 Kişi de GPT-3.5

Akademik çalışmalarda ChatGPT kullananlar; yanlış ya da yanıltıcı bilgi ile karşılaşma konusunda, alakasız yanıtlar alma, Yanıtların anlaşılmasında, eksik bilgi ve tutarsız yanıtlar konusunda nadiren büyük problem yaşadıklarını bildirmiştir (%1.4-%2.9). Yanlış ya da yanıltıcı bilgi ile karşılaşma konusunda, alakasız yanıtlar alma ve eksik bilgi konusunda orta düzeyde problem yaşama sıklığı sırası ile %31.4,%31,4 ve %38,6'dır. Kullananların %81,5'i Yanıtların anlaşılmasında ve %82,9'u tutarsız yanıtlar ile ilgili az sorun yaşadığını ya da hiç yaşamadığını ifade etmiştir (Tablo 3).

Tablo 3. Sağlık Profesyonellerinin akademik çalışmalarda ChatGPT Kullanırken Karşılaştığı Sorunlar (n=70).

	Hiç, n (%)	Hafif problem, n (%)	Orta Derece Problem, n (%)	Çok Bü- yük prob- lem, n (%)
Yanlış/yanıltıcı bilgi, n=70	19 (27.1)	28 (40)	22 (31.4)	1 (1.4)
Alakasız cevaplar, n=70	16 (22.9)	30 (42.9)	22 (31.4)	2 (2.9)
Anlaşılmaz yanıtlar, n=70	23 (32.9)	34 (48.6)	12 (17.1)	1 (1.4)
Eksik bilgi, n=70	12 (17.1)	29 (41.4)	27 (38.6)	2 (2.9)
Tutarsız cevaplar, n=70	23 (32.9)	35 (50)	10 (14.3)	2 (2.9)

4 Tartışma

Bu çalışmada ChatGPT'nin sağlık alanındaki araştırmalarda yardımcı bir araç olarak kullanımının, güçlü ve zayıf yönlerini anlamak ve sağlık profesyonellerinin bu araçla ilgili deneyimleri incelenmiştir. Çalışmaya katılan sağlık profesyonellerinin %36.1'i herhangi bir nedenle ChatGPT kullanmışken, sağlık alanında araştırma yapanların %17.5'i akademik çalışmalarında ChatGPT kullandığını bildirmiştir. Bu bulgular, ChatGPT'nin sağlık araştırmalarında yavaş yavaş kabul gördüğünü ve kullanımının arttığını ortaya koymaktadır. ChatGPT kullanıcıları arasında araştırmacıların oranının yüksek olması ve profesör sayısının az olması dikkate değer bir bulgudur. Bu durum, araştırmacıların yeni yapay zeka teknolojilerine ve alternatif araştırma araçlarına daha açık olduğunu göstermektedir. Araştırmamızda, ChatGPT'nin akademik çalışmalarda kullanımının genç yaş gruplarında biraz daha yaygın olduğunu gözlemledik; ancak, bu farkın istatistiksel olarak anlamlı değildi. Sağlık araştırmalarında ChatGPT desteği alan en yaşlı kullanıcı 48 yaşında idi ve kullanıcılar arasında hiç profesör yoktu. Benzer bir çalışmada, sağlık çalışanları arasında ChatGPT beklentilerini araştırmış ve benzer mesleklere sahip bireylerin %84'ünün 44 yaş ve altında olduğunu belirtmiştir [16]. Bu bulgular, genç araştırmacıların ChatGPT'yi daha sık kullandığını destekler niteliktedir ve literatürle uyumludur. Diğer yandan akademik kariyer ilerledikçe, geleneksel yöntemlere olan bağlılık ve yeni teknolojilere karşı çekingenlik artabilir, bu da üst düzey akademisyenler arasında ChatGPT kullanımının daha az olmasına neden olabilir. Bu durum, teknolojinin kabulünde kariyer evresinin önemli bir faktör olduğunu göstermektedir.

Bu çalışma herhangi bir amaçla ChatGPT kullanan sağlık profesyonelleri için yapay zekanın bilgiye erişimdeki ve eğitimdeki merkezi rolünü göstermiştir. ChatGPT'yi deneyimleyen katılımcıların büyük çoğunluğu bilgiye erişmek ve eğitim amacı ile kullandığını belirtmiştir. ChatGPT ve benzeri AI destekli platformlar, sağlık araştırmacıları ve akademisyenler arasında bilgi erişimi, eğitim ve profesyonel işler için güçlü araçlar olarak kendini kanıtlamıştır [25, 26]. Daha spesifik olarak sağlıkla ilgili akademik çalışmalarda kullanım sıklığı sorgulandığında henüz kullanımının yaygınlaşmadığı çalışmamızda dikkat çekici bir sonuç olmuştur. ChatGPT'yi kullanan her iki kişiden biri akademik çalışmalarda destek için başvurursa da (%48) toplam katılımcılar içerisinde bu sıklık düşüktür (%17.5). Bu sonuç yapay zeka destekli sistemlerin akademik alanda tam olarak benimsenmemiş olabileceğini işaret eder. Araştırmacıların bu teknolojileri kullanma konusundaki tereddütleri, en çok bilgi güvenliği ve veri doğruluğu gibi konulara dair endişelerden kaynaklanmaktadır [25, 27-30]. Kullanımın düşük olması, aynı zamanda, sağlık alanındaki araştırmacıların ChatGPT gibi araçların potansiyel faydalarını tam olarak anlamamış olabileceklerini de gösteriyor. Bu durum, eğitim programlarında yapay zeka ve makine öğrenmesi uygulamalarına daha fazla yer verilmesi gerektiğini vurgulamaktadır. Akademisyenlerin bu tür teknolojileri verimli bir şekilde kullanabilmeleri için gerekli becerileri kazanmaları önemlidir.

Yapılan bu çalışmada, ChatGPT'nin sağlık alanındaki akademik çalışmalarda çok çeşitli amaçlarla kullanıldığı ve genel olarak kullanıcılarda olumlu bir etki yarattığı görülmüştür. Katılımcılar, araştırma hazırlığı ve literatür taraması için %37.2, akademik yazı ve dil çevirisi için %34.3, araştırma sonuçlarının yorumlanması için %13.9, veri toplama ve analiz için %10.2 ve yayın süreci ile ilgili işlemler için %4.4 oranında ChatGPT'yi kullanmaktadırlar. Bu sonuçlar, platformun sağlık araştırmalarında çeşitlendirilmiş bir yardımcı araç olarak kullanıldığını göstermektedir. Ayrıca kullanan katılımcıların büyük bir çoğunluğu (%67.34) ChatGPT'nin faydalı veya oldukça faydalı olduğunu belirtmiştir, bu da teknolojinin araştırma süreçlerine somut katkılar sağladığını ortaya koymaktadır. Ayrıca, kullanıcıların %90'ı ChatGPT'nin bilimsel araştırma süreçlerini pozitif yönde etkileyeceğine inanmaktadır, bu da platformun gelecekteki potansiyelini göstermektedir. Yapılan çalışmalar yapay zeka destekli bu araçların sadece araştırma ve medikal yazma değil sağlık alanında bir çok konuda yeteneğini gözler önüne sermiştir. Anestezi uzmanlığı başvuruları için kişisel beyan oluşturma yeteneğinin incelendiği bir çalışmada, ChatGPT tarafından oluşturulan beyanların yüksek kaliteli olduğunu ve geleneksel başvuru yazılarından ayırt edilmesinin zor olduğunu belirtmiştir [13]. Lu ve arkadaşları (2023) tarafından gerçekleştirilen bir çalışmada ise, ChatGPT/GPT-4'ün yoğun bakım ünitelerindeki potansiyel kullanımı incelenmiştir. Araştırma sonucunda, ChatGPT/GPT-4'ün yoğun bakım ünitelerinde kullanımının son derece umut verici olduğu sonucuna varılmıştır [14]. Yine başka bir çalışmada Ortopedi ve travmatoloji alanında oluşturulan çeşitli vaka senaryolarına verdiği medikal tavsiyeler uzmanlar tarafından tatmin edici düzeyde faydalı bulunmuştur [21].

Akademik çalışmalarda ChatGPT kullananlar arasında potansiyel faydalarının yanında; çalışmamızda yanlış ya da yanıltıcı bilgi ile karşılaşma konusunda, alakasız yanıtlar alma, Yanıtların anlaşılması, eksik bilgi ve tutarsız yanıtlar konusunda nadiren de olsa problem olduğu belirlenmiştir. Benzer bulgu başka bir çalışma da, yapay zeka-

nın akademik çalışmalarda kullanımı ile ilgili yapılan bir çalışmada içerik üretme konusunda güvenilirlik sorunları olduğu, etik ve yağmalama konularını da beraberinde getirdiği gösterilmiştir [31]. ChatGPT gibi yapay zeka araçlarının sağlık hizmetleri ve akademiye entegrasyonu çok sayıda etik soruyu da gündeme getirmektedir. Bunlar, karar verme sürecinde YZ'nin hesap verebilirliğinden, özellikle hassas veriler işlenirken gizlilikle ilgili endişelere kadar uzanmaktadır [32]. Bizim çalışmamızda da katılımcıların yarısı "Etik ve gizlilik", "Kötü niyetli kullanım" ve "Akademik kullanımda yazar olarak insan müdahalesinin azalması" gibi hususlarda yaşadıkları endişeleri vurgulamıştır.

Sonuç olarak, bu çalışma, ChatGPT'nin sağlık alanında profesyonel ve akademik kullanımını detaylı olarak incelemiştir. Katılımcıların ChatGPT'nin mevcut ve potansiyel kullanımları hakkında verdiği geri bildirimler, bu teknolojinin sağlık sektöründe nasıl bir yardımcı araç olarak kullanılabileceğini ve nasıl daha etkili bir araç haline getirilebileceğini ortaya koymaktadır. Ancak, bazı kullanıcıların teknolojiye adapte olurken karşılaştığı zorluklar ve etik, gizlilik ve bilgi doğruluğu gibi konuların titizlikle ele alınmalı, göz ardı edilmemelidir. Katılımcıların gelecekteki ChatGPT versiyonlarına ve sağlık araştırmalarında kullanım alanlarına dair öneri ve düşünceleri, bu teknolojinin daha etkili bir şekilde kullanılması için yapılacak geliştirmelere önemli bir katkıda bulunacaktır.

Kaynaklar

1. Alberts, I.L., et al., *Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?* European Journal of Nuclear Medicine and Molecular Imaging, 2023. **50**(6): p. 1549-1552.
2. Zhang, B., *ChatGPT, an Opportunity to Understand More About Language Models.* Med Ref Serv Q, 2023. **42**(2): p. 194-201.
3. Zhong, Y., et al., *The Artificial intelligence large language models and neuropsychiatry practice and research ethic.* Asian J Psychiatr, 2023. **84**: p. 103577.
4. *OpenAI. OpenI ChatGpt Guide. Chatgpt History: Timeline, Facts, Version, Current Capability.* 2023. Available from: <https://openichatgptguide.com/chatgpt-history-timeline-facts-versions/>.
5. Jo, H., *Understanding AI tool engagement: A study of ChatGPT usage and word-of-mouth among university students and office workers.* Telematics and Informatics, 2023. **85**: p. 102067.
6. Ariyaratne, S., et al., *A comparison of ChatGPT-generated articles with human-written articles.* Skeletal Radiology, 2023.
7. Arslan, S., *Exploring the Potential of Chat GPT in Personalized Obesity Treatment.* Ann Biomed Eng, 2023.
8. Au Yeung, J., et al., *AI chatbots not yet ready for clinical use.* Frontiers in Digital Health, 2023. **5**.

9. Baumgartner, C., *The potential impact of ChatGPT in clinical and translational medicine*. Clinical and Translational Medicine, 2023. **13**(3): p. 4.
10. Berger, U. and N. Schneider, *How ChatGPT will Change Research, Education and Healthcare?* PpMP Psychotherapie Psychosomatik Medizinische Psychologie, 2023. **73**(3): p. 159-161.
11. Cascella, M., et al., *Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios*. J Med Syst, 2023. **47**(1): p. 33.
12. Garg, R.K., et al., *Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review*. Health Promot Perspect, 2023. **13**(3): p. 183-191.
13. Johnstone, R.E., G. Neely, and D.C. Sizemore, *Artificial intelligence software can generate residency application personal statements that program directors find acceptable and difficult to distinguish from applicant compositions*. Journal of Clinical Anesthesia, 2023. **89**: p. 111185.
14. Lu, Y., et al., *Artificial Intelligence in Intensive Care Medicine: Toward a ChatGPT/GPT-4 Way?* Ann Biomed Eng, 2023: p. 1-6.
15. Temsah, M.H., A. Al-Eyadhy, and A. Jamal, *ChatGPT-coauthored conversation between a pediatrician and a taxi driver*. Pediatric Research, 2023.
16. Temsah, M.H., et al., *ChatGPT and the Future of Digital Health: A Study on Healthcare Workers' Perceptions and Expectations*. Healthcare (Basel), 2023. **11**(13).
17. Abdulai, A.F. and L. Hung, *Will ChatGPT undermine ethical values in nursing education, research, and practice?* Nursing Inquiry, 2023.
18. Kung, T.H., et al., *Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models*. PLOS Digit Health, 2023. **2**(2): p. e0000198.
19. Liu, S.R., et al., *Using AI-generated suggestions from ChatGPT to optimize clinical decision support*. Journal of the American Medical Informatics Association: p. 9.
20. Tewfik, G., et al., *ChatGPT and Its Potential Implications for Clinical Practice: An Anesthesiology Perspective*. Biomed Instrum Technol, 2023. **57**(1): p. 26-30.
21. Yapar, D., et al., *ChatGPT's potential to support home care for patients in the early period after orthopedic interventions and enhance public health*. Jt Dis Relat Surg, 2024. **35**(1): p. 169-176.
22. Bhayana, R., R.R. Bleakney, and S. Krishna, *GPT-4 in Radiology: Improvements in Advanced Reasoning*. Radiology, 2023: p. 230987.
23. Bhayana, R., S. Krishna, and R.R. Bleakney, *Performance of ChatGPT on a Radiology Board-style Examination: Insights into Current Strengths and Limitations*. Radiology, 2023: p. 230582.

24. Babaoğlu, A., et al., *Reporting of observational studies in epidemiology: Turkish adaptation of STROBE Criteria*. Med J West Black Sea, 2021. **5**(1): p. 86-93.
25. Al-Worafi, Y.M., et al., *Applications, Benefits, and Risks of ChatGPT in Medical and Health Sciences Research: An Experimental Study*. Progress In Microbes & Molecular Biology, 2023. **6**(1).
26. Sallam, M., *ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns*. Healthcare (Switzerland), 2023. **11**(6).
27. Alkaissi, H. and S.I. McFarlane, *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*. Cureus Journal of Medical Science, 2023. **15**(2): p. 4.
28. Altmäe, S., A. Sola-Leyva, and A. Salumets, *Artificial intelligence in scientific writing: a friend or a foe?* Reproductive BioMedicine Online, 2023.
29. Colletti, P.M., *Re Letter-to-the Editor: "Chatbots in Medical Research: Advantages and Limitations of AI-Enabled Writing With a Focus on ChatGPT as an Author"*. Clin Nucl Med, 2023.
30. Hosseini, M., et al., *An exploratory survey about using ChatGPT in education, healthcare, and research*. medRxiv, 2023.
31. Dönmez, İ., S. Idin, and S. Gülen, *Conducting academic research with the ai interface chatgpt: Challenges and opportunities*. Journal of STEAM Education, 2023. **6**(2): p. 101-118.
32. Dave, T., S.A. Athaluri, and S. Singh, *ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations*. Frontiers in Artificial Intelligence, 2023. **6**.

A Comparative Study of Parkinson's Disease Detection using different Deep Learning Architectures

Beyza GÜL¹

Yıldız Technical University, Istanbul, Turkiye
beyza.gul@yildiz.edu.tr

Abstract. Parkinson's disease is commonly known as tremor disease due to impaired control of the hands. It is vital to be alert to the symptoms of Parkinson's, which, if diagnosed early, has a high chance of being controlled and preserving the patient's quality of life. In recent years, early diagnosis of PD has become a major need worldwide. One of the noninvasive methods for detecting Parkinson's disease is the handwriting test, which has been used for years in hospitals. With the advancement of technology and hardware, deep learning and image processing techniques have become very well developed and can be frequently used in the diagnosis of the disease. In this study, we suggest using thirteen distinct CNN architectures in a Convolutional Neural Network (CNN) based artificial intelligence approach to extract characteristics from spiral and wave writing patterns produced by Parkinson's patients with diagnosed hand tremor symptoms and healthy individuals.

The performance analysis of the used architectures, classification interpretations are evaluated with confusion matrix results. As a result of the observations, speed, mathematical processing power and number of parameters were considered. Since processing power is the biggest problem in deep learning, the model was developed in Google Colab environment. Fast-AI library was used due to its advantages such as ease of use, fast and simple features. The performance of thirteen different proposed architectures was tested. The observations showed that the highest accuracy of 92% was achieved with ResNet-18.

Keywords: Parkinson's disease, Convolutional Neural Networks, Deep Learning, CNN Architecture

1 Introduction

Parkinson is known as tremor disease. It is vital to be alert to symptoms in Parkinson's, which has a high chance of being taken under control and preserving quality of life when diagnosed early. Many problems such as decreased sense of smell or loss, constipation, sleep disturbances, shoulder pain, depression, indifference, anxiety, frequent urination can start years ago[1]. Parkinson's disease is a progressive type of disease that progresses gradually and the symptoms such as the tremor of the hands, arms, legs, chin, and face worsen over time. Parkinson's disease; It is a slow progressive brain disease that slows down movements, trembles in the arms and legs that occur during rest, stiffness and stiffness in muscles, and decrease in posture reflexes, but there are many other system findings. In patients, the findings start largely unilaterally, and can also cross the body over the years. As the disease progresses, the person may have difficulty walking and speaking and may not be able to meet their daily simple needs alone. Symptoms of the disease may vary from person to person, and the first symptoms may be mild. The initial symptoms usually affect one side of the body and spread to both sides as time progresses[2].

Hand tremor (tremor) is the most common symptom. Uncontrolled movements of the thumb, slight flickering in the chin and lip are Parkinson's most common symptoms. These tremors may be accompanied by twitching of the leg muscles. Changing the handwriting is another common symptom. Sudden changes in handwriting may be a sign of Parkinson's disease. In handwriting, the letters may become smaller and words can become closer and more complicated. It is normal for handwriting to change depending on joint pain and visual disturbances as the age progresses, but this does not happen suddenly as in Parkinson's disease, it develops over time [3]. By considering these symptoms, the diagnosis of the disease can be made by comparing Deep Learning models with patients' handwriting.

The early diagnosis of Parkinson's disease plays an important role in the progression of the disease with observed many symptoms. To facilitate the early diagnosis of Parkinson's disease, a Deep Learning model is aimed to establish with "Parkinson Disease Spiral and Wave Drawings- Hand drawing Data Set" in Kaggle. The dataset contains hand drawings of Parkinson's patients and healthy people. The model is planned to be trained with these drawings and as a result, it is planned to diagnose the patient with Parkinson's ResNet, DenseNet, VGG, AlexNet, SqueezeNet and MobileNet model is used in the Deep Learning model to be established diagnosis for Parkinson's patients. The images of Parkinson's disease patients and healthy individuals drawing spirals and waves are included in the data collection. To facilitate comparison, pictures have been separated into training and testing groups. During establishing the model, every parameter was tested, including the number of layers, batch size, epoch number, and activation function. First, the 13 model was established on Google CoLab by using fast-ai library. Various parameters will be reduced and increased in order to achieve the most efficient result on the installed model. The result evaluation of the

model was evaluated with confusion matrix and score values. As a result, the best accuracy was obtained 92% accuracy with ResNet-18.

2 Related Work

Studies in Parkinson's disease diagnosis with artificial intelligence was searched on literature. To search articles, the key words was determined. Parkinson Disease, Hand Drawing and Artificial intelligence are set as key word. The literature search was done on Google Scholar.

Drot'ar et al.[4] argue that the diagnosis of Parkinson's disease is characterized not only by the hand movements of the patient's drawing on the writing surface, but also by micrographs and changes in the kinematics of the handwriting, taking into account the pressure applied to the surface and hand movements without touching the writing surface. With classification accuracy of 84% and 78%, respectively, the authors demonstrated that the assessment of hand movements in the air during sentence handwriting has a greater influence than the pure evaluation on surface movements.

Monty et al.[5] used classical computer vision techniques such as Histogram of Oriented Gradients (HOG) to extract features from images using hand drawings that can be easily affected and detected by patients' hand tremor in the diagnosis of Parkinson's disease. The features extracted from the images were trained with a random forest classifier and an accuracy of 83.33% was achieved. Pereira, et al.[6] used Convolutional Neural Networks with five different approach and got best accuracy with bat algorithm which is 90.38%.

The Parkinson Disease Spiral Drawings Digitized Graphic Tablet data set, which is available to the accessible to everyone, has been utilized for Gil-Mart'in et al.[7]. The drawing movements that achieved the highest results for the X and Y directions were used as indicators of the ability to identify various orientations. The study's most effective outcomes were an accuracy of 96.5%, an F1 score of 97.7%, and a 99.2% area below the curve. The results appear to support the use of drawing motions in medical decision support systems for long-term patient surveillance and PD detection also known as patient scanning. Except those studies, the speech, pressure and signal processing approach also are used to detect Parkinson's disease with artificial intelligence. For the purpose of diagnosing Parkinson's disease (PD), Das,[8] examined numerous techniques for classification using neural networks, regression, and decision trees. In order to calculate the performance of these classifiers, a variety of assessment methods have been used, including experiments in a data set of biomedical sound measurements of 31 people diagnosed with Parkinson's disease. The best results were obtained with about 92.9% of Neural Networks PD patient recognition rate. In addition to literature scanning, Github and Kaggle were also scanned. Sayak Paul's [9] work "Parkinson's Disease Classifier" was examined. Sayak Paul took advantage of Detecting Parkinson's Disease with OpenCV, Computer Vision, and the Spiral / Wave Test, published on Adrian Rosebrock's PyImageSearch. The presentation demonstrates the automatic detection of Parkinson's disease

in drawn manually spiral and wave illustrations. Adrian obtained an accuracy of 83.33% by using basic computer vision techniques such as Histogram of Oriented Gradients (HOG) to quantify the picture characteristics and then using that information for training a Random Forest Classifier. Sayak Paul used deep learning technique and preferred Resnet34 architecture as a method. As a result, he got an accuracy of 88%. Later on, Peker, [9] who benefited from the work of Sayak Paul, published his work "A comparative study on detection of Parkinson's disease with the images of Spiral / Wave test using fast.ai". Peker, used the fast-ai library. He mentioned that, this library is super easy to use and comes ready with a lot of state-of-the-art practices. It provides a number of useful functions that make the life of a deep learning practitioner easy. In this study, Peker worked on 6 comparative analysis of these using deep learning methods. And also, he tried to determine the learning rate, which significantly affects the success rate. The learning rate has been taken into consideration and the value of the learning rate giving the best results has been tried to be determined by him. As a result of the study, the best accuracy value was obtained with Densenet121 architecture, approximately 92%. As a result of the research, Sayak and Peker's studies have been a guide for this project. In this study, by using Fast-ai and 13 different deep learning architectures, the best-performing model for the diagnosis of Parkinson's disease was tried to be found.

3 Data

The dataset includes spiral and wave hand drawings of who has Parkinson disease. As known, the Parkinson disease patients has problem to control body, it is most obvious one is hand shaking issue. They cannot easily move hands to write or draw something as they want properly. The dataset was taken from KAGGLE, Parkinson's Drawings and it consists of 204 images,

1. Spiral: 102 images, 72 training, and 30 testing
2. Wave: 102 images, 72 training, and 30 testing

The Parkinson's hand-drawing images dataset does not need to apply any pre-processing, filtering or other special treatment to use in project. The model set up with two different group which includes train and test.

Train has:

- 72 Healthy both spiral and wave images
- 72 Parkinson's spiral and wave images

Test has:

- 30 Healthy both spiral and wave images
- 30 Parkinson's spiral and wave images

4 Method

4.1 Fast-AI

Fast-ai brings revolutionary results to people working in common deep learning topics in the quickest and most convenient way. It is a deep learning the use of libraries which provides researchers low-level components that they can combine with innovative approaches. In terms of performance, flexibility, or straightforwardness of use, it attempts to be the most beneficial library. Only a painstakingly layered architecture that articulates what are common basic characteristics of numerous deep learning and data processing methods may make this feasible. Taking advantage of the flexibility of the PyTorch package and the dynamism of the fundamental Python language, the structure of the deep learning algorithm is represented concisely and simply. The information included with fast-ai are a new 2-way event structure that may be used to access any component of the data, model, or optimizer at any phase during training; a new data block API; a clear and straightforward optimizer; a GPU-optimized image processing library that can be enhanced with basic Python; and more. The open-source Fast-ai library includes numerous course materials and is currently being used in educational institutions, business, and research. The potential benefits provided by it have led to the determination to use it in this project. The fast-ai library assisted the project to be accomplished successfully because of its quickness, graphical representation, and educational opportunities[10].

4.2 Google CoLab

Google Colaboratory, frequently referred to as Colab, is a cloud service that uses Jupyter Notebook for encouraging educational and research activities in artificial intelligence (AI). The design enables no charge access to a capable GPU and an extensive deep learning runtime that is fully customized. In deep learning studies, hardware resources are often inadequate and overuse, and they pose a risk such as hardware malfunctions. Today, cloud solutions are very attractive because of these risks because they provide the hardware instantly and GPUs of hardware resources and a fully structured runtime for deep learning are available to you. Google releases the Colaboratory, an open-source cloud computing platform designed to promote machine learning education and exploration. This cloud solution guarantees availability and comes fully loaded with top artificial intelligence (AI) libraries and a potent GPU. This Google service is inexpensive and automatically connected to a Google Drive profile. The selection of Google Colab for this project had been decided on its convenience for users, interoperability with several libraries, ability to comply with project hardware requirements, and ease of practice[11].

4.3 Convolutional Neural Network

A method for developing machine learning techniques is deep learning. Multiple linear and nonlinear processing unit's data are used in deep learning. A deep ar-

chitecture is designed to model high level abstraction. Currently, there are many deep learning techniques used in many different applications from medical field to industry. Among them, the deep learning method known as deep convolutional neural networks (CNNs) is among the most commonly used. CNN-based techniques have become more widespread acceptance in medical diagnostic imaging analysis and recognition applications in recent years. CNNs are developed as multi-layer sensors, inspired by structures such as biological brain neural networks[12]. They have the ability to recognize processed images by analyzing them from raw image pixels. In some cases, small preprocessing applications can be performed before processing images in our CNN deep learning model. In such deep network models, the input image, which is preprocessed, called the receptive fields, is looked at the multi-layered neural network model and the model is trained using weights in each deep convolutional layer[13].

Each neuron in the deep learning model has an activation function that controls the output. As can be seen in the literature and studies, there are many types of activation functions such as linear, sigmoid, tanh, rectified linear unit (ReLU). Activation functions are classified as linear and nonlinear. A linear function outputs the input in the neuron to the output without making any changes. Since deep network models contain complex mathematical operations, it achieves more success with its nonlinear activation functions. Due to their fast convergence characteristics, nonlinear activation functions used in many deep learning models are variations such as ReLU and leaky-ReLU and parametric ReLU[14]. Convolutional neural network can be handled in 4 main parts in our project. These are input image, convolutional layer, fully connected layer and output image. Convolutional layer contains max pooling layer. Pooling is an important technique in convolutional neural networks that perform nonlinear down-sampling. There are different pooling approaches such as stochastic, maximum and average. The maximum pooling method is widely used, dividing the input image into overlapping rectangular blocks and producing output for each block. Maximum pooling simplifies calculations for the upper layers by eliminating the minimum values. In other words, maximum pooling intermediate features intelligently minimize the system provides the correct results[15]. There are various techniques that can be applied to achieve a better learning and generalized model. L1-L2 norm, regularizer, dropout and batch normalization are some of them. Overfitting is a problem in convolutional neural network as in any artificial intelligence model [16]. The dropout plays an important role in preventing the overfitting situation. The dropout layer drops certain selected unit connections randomly and is widely used as a regularization technique. Batch normalization, another regularization technique, is also widely used. Batch normalization is known to accelerate training, and it has been observed that dropout applications are sometimes not required thanks to batch normalization [17]. Convolutional neural network was used in our model, which we tried to develop with the image dataset of Parkinson's disease. As we mentioned under the Convolutional neural network title, our model has many layers and organizers. As it is known, the performance of deep learning methods is highly dependent on dataset. 13

different convolutional neural network architectures were applied to our limited series set. As a result of the study, the results of these 13 different architectures were tried to be compared.

4.4 CNN Architectures

Convolutional neural network models are common methods for deep learning. Most researchers in the field of deep convolutional neural networks recently focused on improving the accuracy of computer vision data sets. For a certain level of accuracy, there are multiple CNN architectures that usually reach this level of accuracy. Many competitions were held with the ImageNet dataset, which is also known today and many publications have been published in this field. As a result of studies in this field, developed deep learning architecture such as AlexNet, ResNet and DenseNet were found. It was named as such, thanks to the name and features they received from the developers.

ResNet is also known as residual neural network. It was built by comparing the pyramidal cells in the cerebral cortex in the human brain. Utilizing skip connections, or shortcuts to jump over some layers is done in residual neural network. Models can be applied with 2- or 3-layer jumps, which are nonlinear and include batch normalization between them. The weighted matrix can be used to calculate jump weights, these models are also known as HighwayNets. Models with more than one jump are called DenseNet [18]. The name of a convolutional neural network (CNN) designed by Alex Krizhevsky is AlexNet. Alexnet contains eight layers. The first five layers are convolutional layers and there are also pooling layers. The last three layers are fully connected layers. ReLU is used as the activation function. More than 61,000 references have been made to the AlexNet article since 2020 [19].

The most important difference between VGG and other models is that it uses convolution additions with two or three parameters. In the fully connected layer, data gets transformed into an attribute vector with $7 \times 7 \times 512 = 4096$ neurons. The output of two FC layers is used to calculate the softmax performance of 1000 classes. More than over 138 million metrics calculated. Comparable to other models, the depth value (number of channels) becomes greater as both the height and the width characteristics of the matrices from the point of entry to the exit decrease [20].

Developed in the field of computer vision studies, SqueezeNet was published in 2016. It was developed by researchers at SqueezeNet, DeepScale, University of California, Berkeley and Stanford University. SqueezeNet's main goal is to improve memory usage most effectively. In this direction, it is aimed to design a smaller deep learning neural network consisting of fewer parameters that can easily fit into the computer memory and can be easily transferred over a computer network. There are publications on the similarity of AlexNet and SqueezeNet, but it should be known that SqueezeNet is not a "compressed version of AlexNet". On the contrary, SqueezeNet is a completely different deep learning neural network architecture than AlexNet. The common feature of SqueezeNet and AlexNet is that both studies have reached similar accuracy

values when working with ImageNet image classification[21] MobileNet is one of the CNN architectures and is used in areas such as image classification, mobile vision. The difference from other architectures we've considered and what makes it special is the use of more computing power to run and implement transfer learning. Thanks to this feature, it fits perfectly to mobile devices and systems without GPUs. In addition, browsers are best suited for web browsers with limited capacity for computing, graphics processing and storage[22].

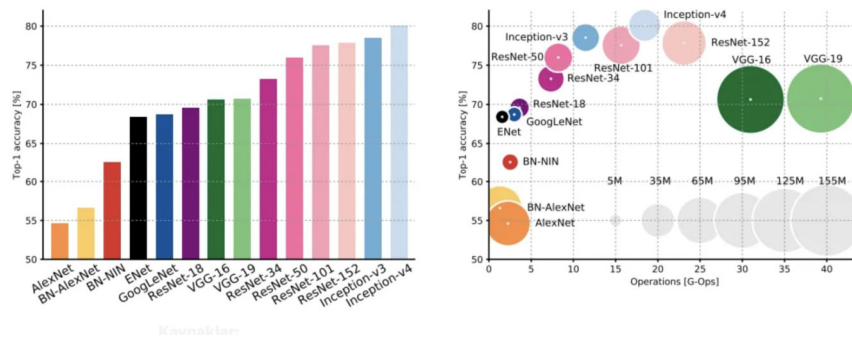


Fig. 1: Performances (left) and Process load (right) of model. [23]

An overall comparison of CNN architectures was made in Ayyüce Kızrak's medium article²⁶ in 2018. Based on the article "An Analysis of Deep Neural Network Models for Practical Applications" published by Canziani et al[23]. In 2016, a general comparison is made; Networks and network performance (%) are given as in the graphic on the left. In the Figure 1 on the right, the parameter calculation load-performance comparison of the network models in terms of the performance calculated in the y-axis and the percentage of performance in the x-axis is given. As a result of the comparisons made, it will be easier to interpret the graphics and select the network model with a computing load suitable for both your problem and your CPU / GPU capacity.

5 Experiment

1. Dataset was decided and the image dataset environment was prepared.
2. Studies with the data set were examined. The studied work was included in the "Related Works" section.
3. After examining the studies, it was decided to conduct convolutional neural network deep learning studies with the Parkinson Hand-writing data set.
4. The biggest risk in deep learning applications is hardware computer capacity. Because mathematical operations can sometimes be heavy on your

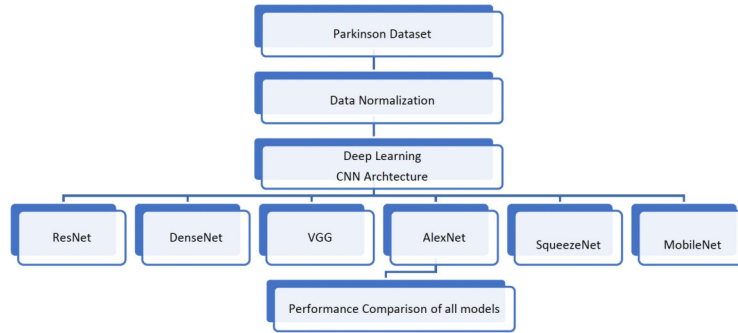


Fig. 2: The flow-chart of experiment

computer's CPU. Considering the hardware features, it was decided to use Google CoLab.

5. It was decided to use fast-ai library by evaluating its simple language, ease of use, visualization and advantages in deep learning.
6. Our data set was uploaded to the system and deep learning and train operations were started.
7. It was decided to use ResNet, DenseNet, VGG, AlexNet, SqueezeNet and MobileNet CNN architectures. A total of 13 different architectures were tried and compared.

The parameters of thirteen different CNN architectures used in our model, which we trained with Parkinson dataset, were calculated and visualized. As we mentioned in the method section, the structure of the architecture plays an important role for the model. As the number of parameters increases, the number of mathematical operations increases. The increasing number of processes forces the processor, so the CPU is insufficient. Considering the number of parameters, the architecture with the most parameters turns out to be ResNet-152. It is seen that the architecture with the lowest parameter value is also SqueezeNet. Already SqueezeNet targeted to construct with few parameters and layers. With this study, it was observed that SqueezeNet has the least parameters. [2]

Accuracy values obtained after testing the model with all architectures are shown in the table. The accuracy values of the concluded model were visualized and presented to the reader. According to the obtained values, ResNet-18 architecture has been observed to have the highest success rate with an accuracy of 92 percent. The least accurate architectures were MobileNet and ResNet-101, with a success rate of 75 percent.

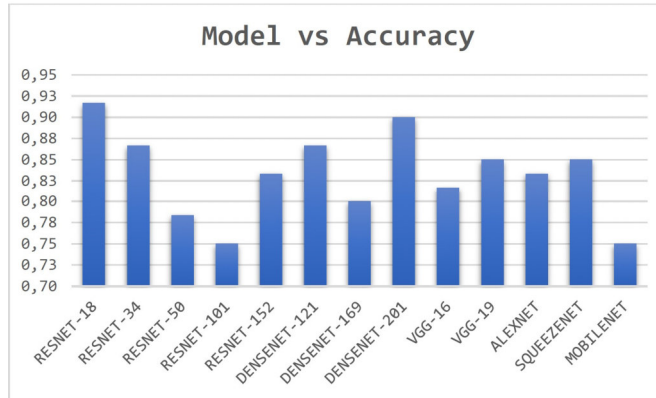


Fig. 3: The visualization of accuracies

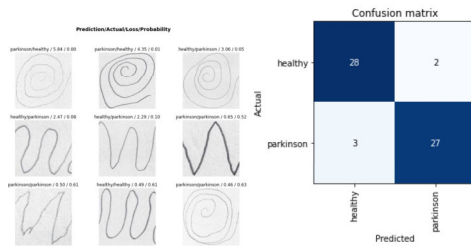


Fig. 4: ResNet Performance Analysis

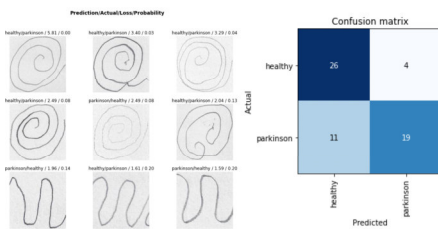


Fig. 5: MobileNet Performance Analysis

6 Conclusion and Future Works

Deep learning is widely used in medicine. Artificial intelligence in health; there are many methods applied in biomedical image processing, in our project we applied the Convolutional Neural Network deep learning model. The aim is to facilitate early diagnosis of Parkinson's disease using image data. In this study, our model was trained using the "Parkinson's Disease Spiral and Wave Drawings - Hand Drawing Dataset" from Kaggle. In order to design the most efficient model using image data, some important design architectures are discussed. As it is known, there are many architectures in convolutional neural network studies. In this project, six different architectures are considered, namely ResNet, SqueezeNet, VGG, DenseNet and MobileNet. In total, the accuracy values of thirteen different architectures were calculated. The obtained parameter and accuracy values were then visualized. The project was finalized by making observations considering various criteria such as speed, mathematical processing power and parameters. In this context, it was observed that the most suitable architecture for the data structure used is ResNet-18 architecture. A comparison of parameter computational load and performance shows that AlexNet and ResNet-18 architectures have the lowest computational load. AlexNet has the least mathematical processing load, followed by ResNet-18 architecture, but AlexNet is worse than ResNet-18 in terms of performance. For this reason, ResNet-18 architecture gave more efficient results in our project because the image data is not very complex. As the complexity of the dataset increases, architectures with more mathematical processing load will extract attributes more successfully.

As the next step in early diagnosis studies in Parkinson's disease, ResNet-18 and DenseNet-201 architectures with the highest accuracy values are planned to be developed. With the improved model, it is thought that it can play an important role in helping doctors in the diagnosis of the disease in the future by integrating it into innovative technological projects.

References

1. C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, *Biomedical Signal Processing and Control* **31**, Pages 174 (2017). DOI 10.1016/S2214-109X(22)00521-6. URL <https://doi.org/10.1016/j.bspc.2016.08.003>
2. P. Zham, S.P. Arjunan, S. Raghav, D.K. Kumar, *IEEE journal of biomedical and health informatics* **22**(5), 1648 (2017)
3. C. Gallicchio, A. Micheli, L. Pedrelli, arXiv preprint arXiv:1802.06708 (2018)
4. P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, *Computer methods and programs in biomedicine* **117**(3), 405 (2014)
5. A. Rosebrock. Detecting parkinson's disease with opencv, computer vision, and the spiral/wave test (2023, January 7). URL <https://pyimagesearch.com/2019/04/29/detecting-parkinsons-disease-with-opencv-computer-vision-and-the-spiral-wave-test/>
6. C.R. Pereira, D.R. Pereira, J.P. Papa, G.H. Rosa, X.S. Yang, *Machine Learning for Health Informatics: State-Of-The-Art and Future Challenges* pp. 377–390 (2016)

7. M. Gil-Martin, J.M. Montero, R. San-Segundo, *Electronics* **8**(8), 907 (2019)
8. R. Das, *Expert Systems with Applications* **37**(2), 1568 (2010)
9. M. Peker. Detection-of-parkinson-s-disease-with-the-images-of-spiral-wave-using-fast.ai. URL https://github.com/m-peker/Detection-of-Parkinson-s-disease-with-the-images-of-Spiral-Wave-using-fast.ai/blob/master/Parkinsons_Disease_Classifier.ipynb
10. J. Howard, S. Gugger, *Information* **11**(2), 108 (2020)
11. T. Carneiro, R.V.M. Da Nóbrega, T. Nepomuceno, G.B. Bian, V.H.C. De Albuquerque, P.P. Reboucas Filho, *IEEE Access* **6**, 61677 (2018)
12. S.M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, M.K. Khan, *Journal of medical systems* **42**, 1 (2018)
13. S.H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, H. Cheng, *Journal of medical systems* **42**, 1 (2018)
14. Y. LeCun, Y. Bengio, G. Hinton, *nature* **521**(7553), 436 (2015)
15. S. Ding, L. Lin, G. Wang, H. Chao, *Pattern Recognition* **48**(10), 2993 (2015)
16. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *The journal of machine learning research* **15**(1), 1929 (2014)
17. S. Ioffe, DOI: <https://doi.org/10.48550/arXiv.1502> (2021)
18. Wikipedia contributors. Residual neural network — Wikipedia, the free encyclopedia (2024). URL https://en.wikipedia.org/w/index.php?title=Residual_neural_network&oldid=1222138016. [Online; accessed 9-Apr-2024]
19. Wikipedia contributors. Alexnet — Wikipedia, the free encyclopedia (2024). URL <https://en.wikipedia.org/w/index.php?title=AlexNet&oldid=1215233065>. [Online; accessed 9-Apr-2024]
20. Kızrak, A. Derİne daha derİne: Evriřimli sinir agları (2020). URL <https://ayyucekizrak.medium.com/derine-daha-derine-evriimli-sinir-alar-2813a2c8b2a9>. [Online; accessed 1-Jan-2024]
21. Wikipedia contributors. Squeezenet — Wikipedia, the free encyclopedia (2024). URL <https://en.wikipedia.org/w/index.php?title=SqueezeNet&oldid=1200466107>. [Online; accessed 9-Apr-2024]
22. V. Sorrenti, *Image classification in the browser: a performance assessment*. Ph.D. thesis, Politecnico di Torino (2023)
23. Kızrak, A. DerİN bir karřılařtırma: Inception res-net versiyonları (2020). URL <https://ayyucekizrak.medium.com/derin-bir-karilatirma-inception-res-net-versiyonlar-f5cfb83df131>. [Online; accessed 9-Apr-2024]

Prediction of retinopathy through machine learning in diabetes mellitus

Tarık Keçeli¹[0000-0002-5843-7550], Nevruz İlhanlı¹[0000-0002-4777-4025] and K. Hakan Güllükesen¹[0000-0002-2477-2481]

¹Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, Antalya, Turkey

Abstract.

Aim: Development of a machine learning model on an electronic health record (EHR) dataset for predicting retinopathy in people with diabetes mellitus (DM), analysis of its explainability.

Methods: A public dataset based on EHR records of patients diagnosed with DM located in Istanbul, Turkey (n = 77724) was used. The categorical variable indicating a retinopathy-positive diagnosis was chosen as the target variable. Variables were preprocessed and split into training and test sets with the same ratio of class distribution for model training and evaluation respectively. Four machine learning models were developed for comparison: logistic regression, decision tree, random forest and XGBoost. Each algorithm's optimal hyperparameters were obtained using randomized search cross validation with 10-folds followed by the training of the models based on the results. The ROC area under curve (AUC) score was used as the primary evaluation metric. SHAP analysis was done to provide explainability of the trained models.

Results: The XGBoost model showed the best results on retinopathy classification on the test set with a low amount of overfitting (AUC: 0.813, 95% CI: 0.808 – 0.819). 15 variables that had the highest impact on the prediction were obtained for explainability, which include Eye-ear drugs, Other eye diseases, Disorders of refraction, Insulin aspartame and HbA1c.

Conclusion: Early detection of retinopathy based on EHR data can be successfully detected in people with diabetes using machine learning. Our study reports that the XGBoost algorithm performed best in this research, with the presence of other eye diseases, insulin dependence and high HbA1c being observed as important predictors of retinopathy.

Keywords: Retinopathy, diabetes mellitus, machine learning, electronic health records.

1 Introduction

Diabetes mellitus (DM) is a noncommunicable disease that is caused by the insufficient production of the insulin hormone within the pancreas or the inability of the human body to effectively use the produced insulin. Categorized under different stages to indicate its severity, diabetes is known to be a devastating disease that may take

many years to be noticed and contributes to significant health problems to a person; such as vision impairment, kidney failure and stroke ^[1]. Along with the severity of the disease, the prevalence of diabetes is also expected to be increased, with the number of people with diabetes aged 20-79 years predicted to be increased to 642 million by 2040 ^[2].

Under the complications caused by diabetes, diabetic retinopathy is a major example. This medical condition is defined by the presence of retinal hemorrhages, microaneurysms, cotton wool spots and/or prior photocoagulation scars ^[3]. In terms of blindness, it has been reported that retinopathy was observed globally in approximately 3 million cases; and compared to under-corrected refractive error, cataract, age-related macular degeneration and glaucoma; retinopathy was the smallest contributor to blindness in 2020, but also the only cause of blindness that showed a global increase in age-standardized prevalence ^[4]. Retinopathy is one of the most common complications in type-1 diabetes and it has been reported that after 15-20 years since an individual's diagnosis, almost all patients would have some degree of retinopathy ^[5].

To minimize the risk of vision impairment and blindness; preventative measures such as early detection by screening, effective management and compliance to guidelines is suggested ^[6]. It has been reported that patients who receive consistent care, have lower rates of low vision and blindness ^[7]. It is also important to note that the management of retinopathy, especially vision threatening variants require the expertise and skills of trained ophthalmologists or retinal specialists ^[8] and with reported low density values of ophthalmologists in many countries, access to the treatment of vision threatening diabetic retinopathy may be difficult based on the patient's location ^[9].

Several studies have been conducted to predict retinopathy based on electronic health records (EHR). Liu et al. utilized an extreme learning machine approach on a dataset containing EHR data of 1093 patients and reported a classification area-under-curve (AUC) performance metric of 88.34% ^[10]. Ogunyemi et al. used EHR dataset comprising 40631 people with diabetes to predict retinopathy by training five machine learning algorithms and presented the univariate analyses of dataset variables. Their results showed that their deep learning model performed best and was able to achieve an AUC score of 0.8 on their external validation set. They also presented the most significant predictors they've observed which included insulin dependence, blood urea nitrogen and systolic blood pressure ^[11]. Saleh et al. used ensemble classification techniques based on uncertainty models using an EHR dataset of 2323 people with diabetes. Their fuzzy random forests approach obtained an accuracy of 84% while their dominance-based rough set balanced rule ensemble approach showed an accuracy of 77% ^[12].

The aim of this study is to develop a machine learning model predicting retinopathy on a diabetes dataset, and examine the model for understanding the variables predicting retinopathy development, and discuss the possibility of developing an early-diagnosis tool for retinopathy in people with diabetes.

2 Materials and Methods

For the development of the machine learning model, a public dataset containing 107 variables, composed of electronic health records of 77724 patients diagnosed with diabetes mellitus in 2017, located in Istanbul Province, Turkey was used^[13]. The dataset was originally created for the purpose of glycemic control prediction in diabetes mellitus patients, and it also includes information of retinopathy diagnosis. The data analysis, model development, SHapley Additive exPlanations (SHAP) analysis and visualizations were done using the Pandas^[14,15] (v. 2.0.2), scikit-learn^[16] (v. 1.3.0), shap^[17] (v. 0.44.1), and matplotlib^[18] (v. 3.7.1) modules in the Python (v. 3.11.3) programming language respectively.

The dataset contained no missing values (106 variables in total). A random seed value of 4564 was used for development runs. The target variable used in this study was the “retinopathy” variable of the dataset. The train (n=54406) and test (n=23318) sets were created with a 70/30 split with stratification on the target variable. Both sets had a target class distribution of 13,9%, with the training set containing 7610 entries of retinopathy-positive patients and the test set containing 3262 retinopathy-positive entries. Afterwards, the numerical variables in the train and test sets were standardized in the preprocessing step. Categorical variables remained intact.

As candidate algorithms for baseline evaluation of prediction performance; Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) algorithms were selected. Before training the algorithms (except LR, which was chosen as a baseline algorithm for performance comparison), each algorithm’s hyperparameters were optimized by using randomized search cross validation with 10-folds and 50 iterations in the train set (n=54406). During this phase, a random combination of selected hyperparameters for the algorithm was produced and the hyperparameters were changed in each iteration by the help of the algorithm. For each hyperparameter combination, the mean of AUC’s of 10-fold cross-validation was calculated. The best performing hyperparameters were selected based on their ROC area under curve (AUC) score, and used for the training of the final model on the complete training set (n=54406) for each candidate algorithm. The ROC area under curve (AUC) score was chosen as the primary evaluation metric of the models; along with f-score, Matthew’s correlation coefficient (MCC), and precision-recall AUC (PR-AUC) recorded for the reporting of the results on the train and test sets. For the explainability of a model’s prediction, SHapley Additive exPlanations (SHAP) values were calculated.

3 Results

Obtained results showed that the XGBoost model achieved the highest predictive performance with an AUC score of 0.813 (95% CI: 0.808 – 0.819) on the training set and 0.799 on the test set, followed by the RF model which achieved an AUC score of 0.784 on the training set and 0.783 on the test set. According to the obtained statistics, the developed XGBoost model was reported as the superior model.

Table 1. Model performance metrics on the training and test sets

	Training set (n=54406)							Test set (n=23318)						
	AUC	PR AUC	MCC	Accuracy	F-Score	Sensitivity	Specificity	AUC	PR AUC	MCC	Accuracy	F-Score	Sensitivity	Specificity
XGBoost	0.813 (0.808-0.819)	0.461	0.34	0.74	0.44	0.727	0.74	0.799 (0.791-0.807)	0.420	0.32	0.68	0.41	0.78	0.66
RF	0.784 (0.779-0.790)	0.381	0.31	0.72	0.41	0.700	0.72	0.783 (0.776-0.792)	0.377	0.31	0.72	0.41	0.70	0.72
DT	0.753 (0.747-0.759)	0.361	0.26	0.63	0.36	0.760	0.61	0.749 (0.740-0.758)	0.353	0.26	0.63	0.36	0.75	0.62
LR	0.779 (0.774-0.785)	0.372	0.30	0.71	0.41	0.708	0.71	0.781 (0.772-0.789)	0.376	0.30	0.70	0.40	0.72	0.70

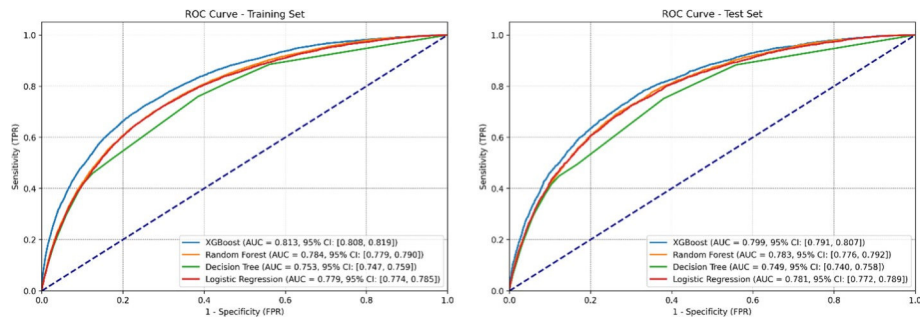


Fig. 1. ROC Curve plots of the models on the training and test sets

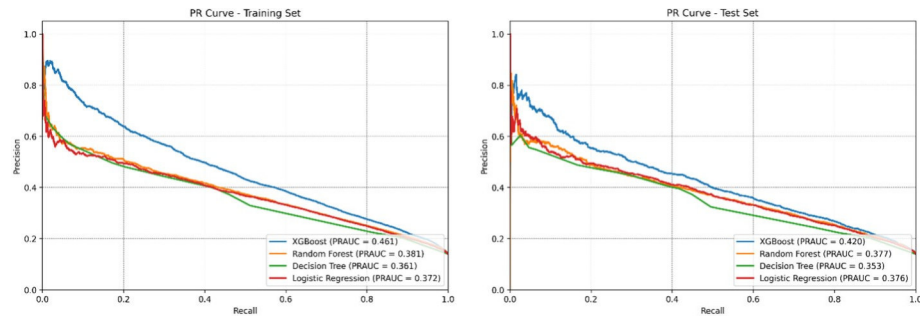


Fig. 2. PR Curve plots of the models on the training and test sets

3.1 Model Interpretation

In terms of model interpretation, 15 variables that had the most impact in calculating the prediction were obtained using a SHAP analysis. The analysis results showed that

the variable of “Eye-ear drugs”, which contained information on whether the patient takes eye and/or ear drugs had the highest impact on the prediction of retinopathy, followed by the variables “Other eye diseases”; which described whether the patient had other eye-related diagnosis, and “Disorders of refraction”.

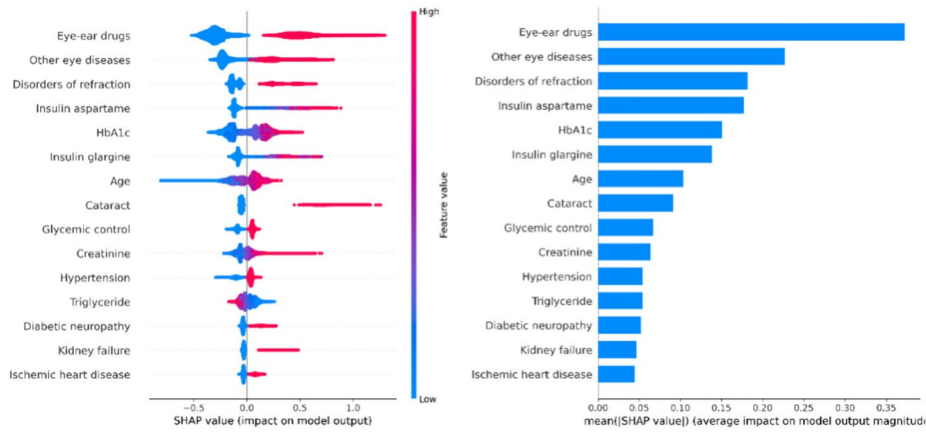


Fig. 3. SHAP summary and feature importance plots

4 Discussion and Conclusion

With this study, an analysis of retinopathy prediction based on EHR data of patients using statistics and machine learning approaches was done. Our findings show that for the task of predicting retinopathy in patients with diabetes, an XGBoost model can achieve a notable predictive performance of 0.799, as observed in our test set. With a considerable predictive performance, the developed model shows that early retinopathy detection with data from electronic health records is a feasible approach for early diagnosis. Additionally, SHAP analysis shows that most important predictors of diabetic retinopathy are presence of other eye diseases, insulin dependence and a high level of HbA1c.

The performance of our XGBoost model is similar to previous studies. On the other hand, it has been reported that deep learning models trained on retinal fundus images for retinopathy detection has been successfully developed with good prediction metrics [19]. However, a predictive model based on EHR data would be useful, as obtaining fundus image of people is not always feasible. A reliable prediction of a patient’s potential retinopathy diagnosis based on their electronic health records without the need of medical image analysis may provide this feasibility.

The variables that had the biggest impact on the prediction was observed to be, in descending order; “Eye ear drugs”, “Other eye diseases”, “Disorders of refraction”, “Insulin aspartame”, “HbA1c”, “Insulin glargine”, “Age”, “Cataract”, “Glycemic control”, “Creatinine”, “Hypertension”, “Triglyceride”, “Diabetic neuropathy”, “Kidney failure”, and “Ischemic heart disease”. The summary plot showed that for every

variable, except “Triglyceride”, an increase or occurrence was directly proportional to a retinopathy-positive prediction. These variables and results can be categorized under the following sections for further discussion:

Eye-related complaints indicate a higher risk of retinopathy.

The variables under this observation are “Eye ear drugs”, “Other eye diseases”, “Disorders of refraction”, and “Cataract”. The usage of eye and ear drugs may be interpreted as an indicator of patients being treated for eye diseases, as can be seen the frequent diagnosis of “Other eye diseases”. All these predictors show that DM may cause various problems in the eye, showing a simultaneous increase of incidence in eye diseases. Conditions such as glaucoma, age-related macular degeneration, and diabetic macular edema often co-occur with retinopathy, leading to damaging effects that accelerate disease progression and impair visual function. The development of cataracts has been shown to have a proportional impact on the risk of retinopathy. Cataracts compromise visual acuity but may also induce inflammatory responses and oxidative stress within the eye, further exacerbating retinal damage^[20].

Indicators showing that diabetes has progressed.

The variables under this observation are “Insulin aspartame”, “HbA1c”, “Insulin glargine”, “Glycemic control”, “Creatinine”, “Diabetic neuropathy”, “Kidney failure”, and “Ischemic heart disease”. Markers such as insulin aspartame and insulin glargine are commonly used in the management of diabetes while also serve as indicators for disease severity and insulin resistance. Elevated levels of these observations often correlate with advanced stages of diabetes. A more severe condition of diabetes is likely to be predictive of retinopathy in this instance. Hemoglobin A1c (HbA1c), a widely utilized measure of long-term glucose control, offers insights into the overall management of diabetes and its impact on retinopathy progression. Elevated HbA1c levels may indicate suboptimal glycemic control over an extended period, thereby predisposing individuals to microvascular complications, including retinopathy. Glycemic control variable in our study shows several high serum HbA1c measurements in a person.

Additionally, renal function, as reflected by markers like creatinine, serves as an indicator of systemic diabetic complications, including nephropathy and retinopathy. The intricate relationship between kidney function and retinal health underscores the importance of comprehensive diabetic care and regular screening protocols. Diabetic neuropathy, characterized by peripheral nerve damage secondary to chronic hyperglycemia, poses a significant risk factor for retinopathy progression. The neurovascular axis plays a critical role in maintaining retinal homeostasis, and disruptions in peripheral nerve function can exacerbate retinal ischemia and neurodegeneration. Furthermore, the onset of kidney failure and ischemic heart disease heralds a systemic decline in vascular health, exacerbating the microvascular complications associated with retinopathy. These comorbidities amplify the multifactorial nature of retinal disease progression and necessitate a holistic approach to diabetes management.

Risk factors based on the patient.

Advancing age is an important risk factor for retinopathy development and progression. Age-related structural and functional changes within the retina contribute to increased vulnerability to retinal pathologies. Moreover, aging-related alterations in vascular integrity and neurotrophic support mechanisms predispose older individuals to microvascular dysfunction and retinal ischemia, exacerbating the progression of retinopathy.

Hypertension, characterized by persistently elevated blood pressure levels, exerts profound effects on retinal microvasculature and vascular autoregulation. Chronic hypertension induces arteriolar remodeling, endothelial dysfunction, and increased vascular permeability, culminating in retinal vascular abnormalities and exacerbation of retinopathy. The synergistic interplay between hypertension and diabetes further potentiates retinal microvascular damage, underscoring the importance of stringent blood pressure control in retinopathy management.

It is also important to consider the limitations of this study. The used data was originally collected from electronic health records of patients in Istanbul, Turkey. The dataset belongs to a population with certain genetic and environmental conditions and may not be universal. External validation would show the universal value of obtained models. In addition, the dataset does not include some important data such as body mass index, and physical activity. A model with more variables would have possibly a higher performance. Another problem is related to diabetic retinopathy diagnosis in this dataset. The dataset included EHR data of already diagnosed patients, and may not include silent cases or cases in their early stages.

In future studies, for additional validation, a potentially better performance and a more generalized model, a larger dataset that contains a more diverse patient data and additional variables can be considered. Machine learning techniques improved and continue to improve in a fast pace. New machine learning techniques may be more successful in the near future.

In conclusion, presence of retinopathy can be successfully detected in people with diabetes. The best model for this purpose seems XGBoost. Other eye diseases, insulin dependence and high HbA1c are important predictors of retinopathy.

References

1. Diabetes. World Health Organization (WHO). Published May 4, 2023. Accessed February 29, 2024. <https://www.who.int/news-room/fact-sheets/detail/diabetes>
2. Ogurtsova K, Da Rocha Fernandes JD, Huang Y, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Research and Clinical Practice*. 2017;128:40-50. doi:10.1016/j.diabres.2017.03.024
3. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98(5 Suppl):786-806.

4. Steinmetz JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *The Lancet Global Health*. 2021;9(2):e144-e160. doi:10.1016/S2214-109X(20)30489-7
5. Aiello LP, Gardner TW, King GL, et al. Diabetic retinopathy. *Diabetes Care*. 1998;21(1):143-156. doi:10.2337/diacare.21.1.143
6. Wong TY, Sabanayagam C. Strategies to Tackle the Global Burden of Diabetic Retinopathy: From Epidemiology to Artificial Intelligence. *Ophthalmologica*. 2020;243(1):9-20. doi:10.1159/000502387
7. Sloan FA, Grossman DS, Lee PP. Effects of receipt of guideline-recommended care on onset of diabetic retinopathy and its progression. *Ophthalmology*. 2009;116(8):1515-1521, 1521.e1-3. doi:10.1016/j.ophtha.2009.03.010
8. Indications for photocoagulation treatment of diabetic retinopathy: Diabetic Retinopathy Study Report no. 14. The Diabetic Retinopathy Study Research Group. *Int Ophthalmol Clin*. 1987;27(4):239-253. doi:10.1097/00004397-198702740-00004
9. Teo ZL, Tham YC, Yu M, Cheng CY, Wong TY, Sabanayagam C. Do we have enough ophthalmologists to manage vision-threatening diabetic retinopathy? A global perspective. *Eye*. 2020;34(7):1255-1261. doi:10.1038/s41433-020-0776-5
10. Liu L, Wang M, Li G, Wang Q. Construction of Predictive Model for Type 2 Diabetic Retinopathy Based on Extreme Learning Machine. *DMSO*. 2022;Volume 15:2607-2617. doi:10.2147/DMSO.S374767
11. Ogunyemi OI, Gandhi M, Lee M, et al. Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system. *JAMIA Open*. 2021;4(3):ooab066. doi:10.1093/jamiaopen/ooab066
12. Saleh E, Błaszczyński J, Moreno A, et al. Learning ensemble classifiers for diabetic retinopathy assessment. *Artificial Intelligence in Medicine*. 2018;85:50-63. doi:10.1016/j.artmed.2017.09.006
13. Kemal Hakan Gulkesen. Machine Learning for Prediction of Glycemic Control in Diabetes Mellitus. Published online July 4, 2022. doi:10.17632/RR4RZZRJFC.1
14. The pandas development team. pandas-dev/pandas: Pandas. Published online February 23, 2024. doi:10.5281/ZENODO.3509134
15. McKinney W. Data Structures for Statistical Computing in Python. In: ; 2010:56-61. doi:10.25080/Majora-92bf1922-00a
16. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825-2830.
17. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, et al., eds. *Advances in Neural Information Processing Systems*. Vol 30. Curran Associates, Inc.; 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
18. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55

19. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22):2402. doi:10.1001/jama.2016.17216
20. Alabdulwahhab KM. Senile Cataract in Patients with Diabetes with and Without Diabetic Retinopathy: A Community-Based Comparative Study. *J Epidemiol Glob Health*. 2022;12(1):56-63. doi:10.1007/s44197-021-00020-6

Epilepside Kopya Sayısı Varyasyonlarının Tüm Ekzom Verileri Kullanılarak Araştırılması

Ezgi Özyurt¹[0000-0002-9754-6346], Alper Han Çebi²[0000-0001-7388-874X], Ali Cansu³[0000-0002-1930-6312], Ayberk Türkyılmaz²[0000-0001-9647-8970], Kemal Turhan¹[0000-0001-7871-3025], Pınar Özkan Kart³[0000-0001-5726-737X]

¹ Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, 61080, Turkey

² Department of Medical Genetics, Karadeniz Technical University, Trabzon, 61080, Turkey

³ Department of Pediatrics, Karadeniz Technical University, Trabzon, 61080, Turkey
ezgiozyurt13@gmail.com

Investigation of Copy Number Variations in Epilepsy Using Whole Exome Data

Abstract. Epilepsy is a common neurological disorder characterized by recurrent seizures caused by overabundant electrical discharge in a group of neurons. It is reported to affect more than 50 million people worldwide, and about 7.6 out of every 1000 people suffer from epilepsy. Although epilepsy occurs due to genetic, structural, infectious, metabolic, immune and unknown reasons, it is known that 40-60% of epilepsies are caused by genetic causes. In studies, single nucleotide and InDel variants have been defined in approximately 300 genes that are effective in the etiology of epilepsy. In addition, more than 500 neurodevelopment-related epilepsy genes have been reported in extensive studies. Furthermore, copy number variations have been implicated as a risk factor for or cause of epilepsy, and CNV rates ranging from ~5% to 12% of patients with different types of epilepsy have been reported. Whole exome sequencing data of 102 epilepsy patients whose genetic diagnosis could not be determined were reprocessed and copy number variations were investigated. As a result of bioinformatic analysis, copy number variations explaining the epilepsy clinic were detected in 11 patients.

Keywords: Epilepsy, genetics, copy number variations, whole exome sequencing

Özet. Epilepsi, bir grup nöronda aşırı elektriksel yük boşalmasıyla oluşan tekrarlayan nöbetlerle karakterize, yaygın bir nörolojik bozukluktur. Dünyada genelinde 50 milyondan fazla kişiyi etkilediği bildirilmektedir ve her 1000 kişiden yaklaşık 7.6'sı epilepsi hastasıdır. Epilepsi genetik, yapısal, enfeksiyöz, metabolik, immün ve bilinmeyen nedenlerle ortaya çıkmakla birlikte %40-60'ının genetik nedenlerden kaynaklandığı bilinmektedir. Yapılan çalışmalarda epilepsinin etiyolojisinde etken olan yaklaşık 300 gende tek nükleotit ve InDel varyantları tanımlanmıştır. Ayrıca 500'den

fazla nörogelişimle ilişkili epilepsi geni kapsamlı çalışmalarda bildirilmiştir. Diğer taraftan kopya sayısı varyasyonları epilepsi için risk faktörü veya epilepsinin nedeni olarak gösterilmektedir ve farklı epilepsi türlerine sahip hastaların ~%5 ile %12 arasında değişen CNV oranları bildirilmiştir. Genetik tanısı belirlenememiş 102 epilepsi hastasının tüm ekzom dizileme verileri yeniden işlenerek kopya sayısı varyasyonları araştırılmıştır. Yapılan biyoinformatik analiz sonucunda 11 hastada epilepsi kliniğini açıklayabilecek kopya sayısı varyasyonları tespit edilmiştir.

Anahtar Kelimeler: Epilepsi, genetik, kopya sayısı varyasyonları, tüm ekzom dizileme

1 Giriş

Epilepsi, bir grup nöronda aşırı elektriksel yük boşalmasıyla oluşan tekrarlayan nöbetlerle karakterize, yaygın bir nörolojik bozukluktur. Nöbetler, vücudun bir bölümünü veya tamamını etkileyebilen bazı zamanlarda bağırsak veya mesane kontrolünün kaybı ve bilinç kaybının da eşlik ettiği kısa süreli istemsiz kasılmalar ile karakterizedir [1]. Epilepsi, statik veya ilerleyici olabilen bir hastalıktır ve en sık bilişsel yetersizlikler olmak üzere majör komorbiditelerle görülmektedir [2]. Her 1000 kişiden yaklaşık 7.6'sının epilepsi hastası olduğu ve epilepsinin dünya genelinde 50 milyondan fazla kişiyi etkilediği bildirilmektedir. Her yaşta ortaya çıkan bir hastalık olmakla birlikte erken çocukluk döneminde ve 60 yaş üzeri dönemde sıklıkla görülmektedir [3-4]. Epilepsilerin etiyolojisi (genetik, yapısal, enfeksiyöz, metabolik, immun ve bilinmeyen) çok çeşitli olmakla birlikte %40-60'ı genetik nedenlerden kaynaklanmaktadır [5, 6].

Epilepsiler fenotipik ve genetik olarak heterojendir. Bazı genlerdeki varyasyonlar seçici olarak epilepsilere veya temel semptom olarak epilepsili sendromlara neden olabilirken diğer genlerdeki varyasyonlar beynin gelişimsel kusurları ve epilepsilerle ilişkili olabilmektedir. Genetik çalışmalarda epilepsiyle karakterize sendromlarla ilişkilendirilmiş yaklaşık 300 gende tek nükleotit ve InDel (insersiyon-delesyon) varyantları tanımlanmıştır. Ayrıca 500'den fazla nörogelişimle ilişkili epilepsi geni kapsamlı çalışmalarda bildirilmiştir [7, 8]. Epilepside tek nükleotit ve InDel varyantlarının yanı sıra kopya sayısı varyasyonları (Copy Number Variation; CNV) da önemli rol oynamaktadır [1, 9]. CNV, referans genom ile karşılaştırıldığında tipik olarak bir DNA segmentinin kopya sayısındaki anormal değişiklikler (kayıplar veya kazançlar) olarak tanımlanmaktadır [10-12]. Boyutları ise önceki çalışmalarda 1000 baz çiftinden büyük (1kbp) [12], ardından yaklaşık 100 baz çifti kadar [13] ve son çalışmalarda ise 50 baz çiftinden birkaç megabaz çiftine (Mbp) kadar olan segmentler olarak tanımlanmıştır [14, 15]. CNV'ler farklı kopya sayısına sahip bölgelerde bulunan genlerin ekspresyon seviyelerine etki ederek haploid genom başına tek kopyanın transkripsiyonu ile elde edilebilen transkripsiyon seviyesinin azalmasına ya da artmasına neden olmaktadır [11]. CNV'ler epilepsi dahil olmak üzere kompleks kalıtıma sahip nörogelişimsel bozuklukların önemli nedenlerindedir. Farklı epilepsi tiplerine sahip

hastalarda ~%5 ile %12 arasında deęişen patojenik CNV oranları bildirilmiştir [9, 16–18].

CNV'leri tanımlamaya yönelik karyotipleme ve floresan in situ hibridizasyon (FISH) gibi sitogenetik teknolojiler, dizi tabanlı karşılaştırmalı genomik hibridizasyon (array based comparative genomic hybridization; arrayCGH), tek nükleotid polimorfizm array (single nucleotide polymorphism array; SNP array) ve yeni nesil dizileme (next generation sequencing; NGS) teknolojileri kullanılmaktadır [19–21]. NGS, tek bir çalışmada yüz milyonlarca kısa okuma üreterek CNV'lerin kapsamlı karakterizasyonunu içermektedir. Sınırlı genomik bölgeler için problemlerin önceden tanımlandığı dizi tabanlı yaklaşımlarla karşılaştırıldığında, NGS platformlarında kısa okumalar tüm genomdan rastgele örneklenmekte ve daha yüksek kapsama ve çözünürlüğe, kopya sayılarının daha doğru tahminine ve novel CNV'leri tanımlamada daha yüksek başarıya sahiptir [14, 22]. NGS (Tüm genom dizileme (Whole Genome Sequencing; WGS) ve tüm ekzom dizileme (Whole Exome Sequencing; WES)) temelli CNV tespit etme yöntemleri çift yönlü haritalama (paired-end mapping; PEM), bölünmüş okuma (split read; SR), okuma derinliği (read depth; RD), genomun de novo birleştirilmesi (de novo assembly of a genome; AS) stratejileriyle geliştirilmiştir. WES, Mendeliyen hastalıklar ve yaygın poligenik hastalıkların incelenmesi için etkili bir teknik olması ve aynı zamanda WGS'ye göre düşük maliyetli olması nedeniyle yaygın olarak kullanılmaktadır. WGS verilerinden farklı olarak, WES'in süresiz okumalar üretmesi ve WES verilerinin WGS'nin aksine hedeflenen bölgeler için daha yüksek derinliğe sahip olması nedeniyle okuma derinliği tabanlı bir tespit yaklaşımı kullanmak daha doğru CNV tespiti için idealdir [23, 24].

2 Yöntem

2.1 Ön Çalışma

Karadeniz Teknik Üniversitesi (KTÜ) Farabi Hastanesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı hekimleri tarafından epilepsi tanısı almış olan hastalardan enfeksiyon, travma, beyin tümörü, immün yetersizlik ve felç gibi diğer rahatsızlıkların yarattığı beyin hasarı kaynaklı olmayıp sebebi açıklanamayan, sendromik ve non-sendromik olan ve epilepsiyle birlikte sık görülen komorbiditelere sahip 146 hasta, genetik nedenlerin belirlenmesi amacıyla Tıbbi Genetik Anabilim Dalı hekimleri tarafından tüm ekzom dizileme ile değerlendirilmiştir.

Tüm Ekzom Dizileme (Whole Exome Sequencing; WES): Her örnek için bir kütüphane oluşturmak üzere ortalama 200 ng (nanogram) genomik DNA kullanılmıştır. Genomik DNA, QIAseq® FX DNA Kütüphane Kitleri (Qiagen) kullanılarak 400 bp'ye parçalanmıştır. Parçalamadan sonra kütüphaneler, üreticinin talimatlarına (Qiagen) göre QIAseq® Human Exome Kit kullanılarak hazırlanmıştır. NovaSeq6000 sisteminde, üreticinin yönergelerine göre (Illumina Inc., San Diego, CA, ABD) çift yönlü dizilim (150 bp (baz çifti)) 30X okuma derinliği ile yapılmıştır.

Tüm ekzom dizileme sonucunda, 146 hastadan epilepsi için kliniği açıklayan tek nükleotit ve InDel varyantları tespit edilen 44 hasta bulunmaktadır. Geriye kalan 102 hastada bilinen epilepsi ile ilişkili genlerde tek nükleotit ve InDel varyantları tanımlanamamış ve bu hastalar için genetik tanı konulamamıştır. Genetik tanı alamayan 102 hastada kopya sayısı varyasyonları araştırılmıştır (Şekil 1A).

2.2 Kopya Sayısı Varyasyonlarının Tespiti

102 hastanın ve 40 sağlıklı kontrolün ham verisi (fastq) kopya sayısı varyasyonlarının araştırılması için yeniden işlenmiştir. Yapılan tüm sayısal hesaplamalar için TRUBA (TÜRK ULUSAL BİLİM e-ALTYAPISI) kullanılmıştır.

Veri Ön İşleme: GATK (Genome Analysis Tool Kit) [25] veri ön işleme iş akışı kullanılmıştır. Ham verilerde kalite değerlendirmesi FASTQC [26] ile yapılmıştır. Düşük kaliteli okumalar ve adaptör dizileri Trimmomatic [27] aracı ile filtrelenmiştir. Ardından BWA-mem [28] aracı ile referans genoma hizalanmıştır. Son olarak Picard (MarkDuplicates) [29] aracı ile PCR amplifikasyonu gibi dizilemeden kaynaklanan yanlışlıkların giderilmesi için duplike diziler kaldırılmış ve analize hazır BAM dosyaları elde edilmiştir.

Kopya Sayısı Varyasyonlarının Çağrılması:

Kopya sayısı varyasyonları GATK tabanlı CNVRobot [30] aracıyla çağrılmıştır. Varyantların anotasyonları AnnotSV [31] aracı ile yapılmıştır. CNV tespitini okuma derinliğini temel olarak gerçekleştiren CNVRobot aracı, GATK araçları ve özelleştirilmiş R kodları ile CNV tespiti yapmakta ve tespit edilen CNV'leri görselleştirmektedir.

CNVRobot, ön işleme ile elde edilmiş BAM dosyalarını girdi olarak kabul etmektedir. GATK (CollectReadCounts) aracıyla hedef bölgelerin kapsamını hesaplamak için derinlik bilgisi elde edilir. CollectReadCounts referans genoma hizalanmış her kısa okuma, okumanın başlangıç ve bitiş konumlarına ve ekzonik bölgelerin genomik koordinatlarına göre örtüştüğü bölgeye atanır.

Ekzom zenginleştirme kitlerinin her hedefin okuma derinliği, kütüphane hazırlığı sırasında gerçekleştirilen PCR amplifikasyonu, dizilenen genomik bölgenin GC içeriği ve referans genoma karşı okuma haritalamasının kalitesi gibi çeşitli faktörlerden etkilenir. CollectReadCounts ile hesaplanan okuma sayıları, GATK (DenoiseReadCounts) aracı ile gürültülü okuma sayılarından temel alınan gerçek kopya sayılarını tahmin etmek için kapsam (coverage) normalleştirme ve GC içeriği normalleştirme olmak üzere iki aşamalı bir normalizasyon yöntemi kullanılarak normalleştirilir.

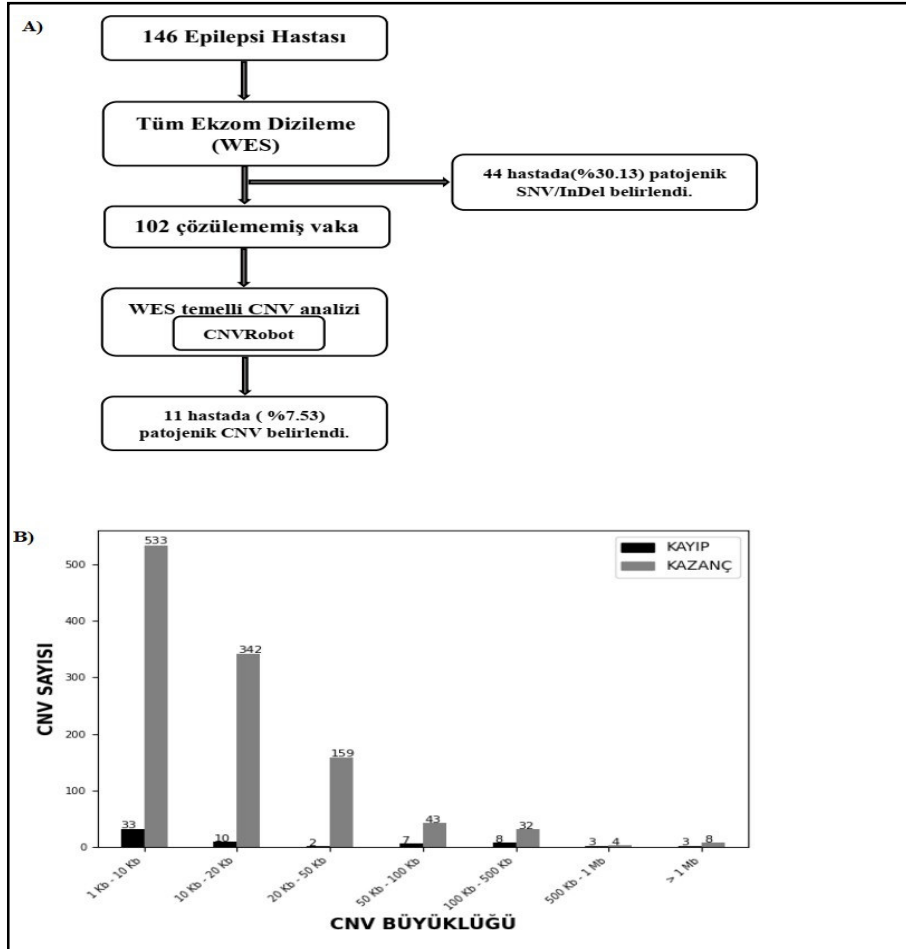
Normalleştirmeden sonra benzer okuma sayılarına sahip bitişik genomik bölgeleri, aynı kopya sayısına sahip olduğu tahmin edilen segmentler halinde gruplandırmak için okuma derinliği analizine dayalı olarak bir segmentasyon işlemi yapılır. Örnekteki her bir genomik bölgeye haritalanmış okuma sayısının, WES verilerinin ortalama kap-

samına (coverage) ve genomik bölgenin boyutuna bağlı olarak beklenen okuma sayısı karşılaştırılmasıyla gerçekleştirilir. Bu okuma derinliklerinin oranı, genomik bölgenin kopya sayısını tahmin etmek için kullanılır. Okuma derinliklerinin genom boyunca dağılımını, genomik bölgelerin boyutunu ve WES verilerinin değişkenliğini hesaba katarak yanlış pozitif ve yanlış negatif çağrılar azaltılmasına sağlamaktadır.

3 Bulgular

102 hasta için CNVRobot aracılığıyla çağrılan kopya sayısı varyantlarından; haritalanabilirliğin düşük (<%75) olduğu bölgeler, sağlıklı kontrollerde ve DGV (Database of Genomic Variants)'de frekansı 0.01'den büyük olanlar elenmiştir. AnnotSV ile elde edilen çıktıda da populasyon frekansı yüksek olanlar (>0.01) elenmiştir. Tüm filtrelemeler gerçekleştirildikten sonra tüm hastalar için toplam 1187 adet nadir CNV tespit edilmiştir (Şekil 1B).

Şekil 1. Çalışmanın akışı şeması ve CNVRobot ile tespit edilen CNV'lerin özeti



A) Çalışmanın akış şeması. Toplamda 146 epilepsi hastasına WES (Whole Exome Sequencing) uygulandı ve SNV/InDel varyantı tespit edilen 44 hasta hariç tutulmuştur. 102 hastada WES temelli CNV (Copy Number Variation) analizi yapılmış ve 11 hastada patojenik CNV tespit edilmiştir. **B)** CNVRobot ile tespit edilen nadir CNV'lerin özeti. X eksenini CNV büyüklüklerini, Y eksenini CNV'lerin sayısını göstermektedir. Siyah ve gri renkli çubuklar sırasıyla kopya sayısı kayıplarını ve kazançlarını göstermektedir.

Hasta başına ortalama ~12 CNV değerlendirilmiştir. Tespit edilen kromozomal bölgeler DECIPHER ve OMIM veri tabanlarındaki bilgiler ile değerlendirilmiştir. Veri tabanındaki kayıtlarla örtüşen CNV'lerin fenotipinde nöbet veya epilepsi olması patojenite tahmini için belirleyici olmuştur. DECIPHER veritabanında olmayıp veya bulunan kayıtların çok az bir bölümüyle örtüşen CNV'ler, bilinen bir epilepsiyle ilişkili bir gen içermesi durumunda patojenik olarak değerlendirilmiştir. Bunun dışında veritabanlarına kaydı henüz yapılmamış ancak fenotip genotip ilişkisini açıklayan yakın zamanlı çalışmalarda belirlenen genler de bu sınıfta değerlendirilmiştir [8]. Biyo-informatik analiz sonucunda kliniği açıklayabileceği düşünülen ilk bulgular Tablo 1'de gösterilmiştir.

Tablo 1: Epilepsi hastalarında CNVRobot aracılığıyla tespit edilen patojenik CNV'ler

Hasta numarası	Cinsiyet	Epileptik sendrom	CNV Tipi	Boyut	Bölge	Gen(ler)
1	Kadın	I ESS	Kayıp	7.2 Kb	21q22.11	SYNJ1
7	Erkek	LGS	Kayıp	10.3 Kb	16p13.3	RBFOX1(A2BP1)
31	Erkek	RE	Kayıp	6.2 Kb	2q24.3	SCN2A
35	Kadın	I ESS	Kayıp	828.7 Kb	17p13.3	PAFAH1B1+20gen
51	Erkek	RE	Kayıp	6.3 Kb	6q27	ERMARD
82	Erkek	GTCA	Kayıp	2.1 Mb	22q11.21	79 gen
83	Kadın	I ESS	Kayıp	489.6 Kb	17p13.3	PAFAH1B1+12gen
95	Erkek	I ESS	Kazanç	5.1 Mb	9p24.3-p24.1	51 gen
96	Kadın	GTCA	Kayıp	710.7 Kb	16p13.2-p13.13	GRIN2A + 6 gen
99	Erkek	SeLFES	Kayıp	49.0 Kb	8p21.3	RHOBTB2 + 2 gen
102	Erkek	GTCA	Kayıp	1.485 Mb	15q13.2-q13.3	20 gen

I ESS: Infantile epileptic spasms syndrome; GTCA: Generalized Tonic-Clonic Seizures Alone; RE: Rolandic Epilepsy; JME: Juvenile myoclonic epilepsy; SeLFES; Self-limited focal epilepsies of childhood syndromes; LGS: Lennox-Gastaut syndrome; Kb: Kibobase; Mb; Megabase.

4 Sonuç

Kopya sayısı varyasyonları epilepsi dahil nörogelişimsel bozuklukların önemli nedenlerindedir. Genomik çalışmalar, kliniği açıklayacak genlere ulaşılması ve hastalıkla

ilişkili yeni genlerin keşfedilmesiyle epilepsi etiyolojisinin anlaşılmasına ışık tutmaktadır. Bu çalışma, epilepsinin genetik temellerini anlamak ve hastalara daha iyi tanı ve tedavi seçenekleri sunmak için gösterilen çabaların bir parçasıdır. Genetik nedeni belirlenememiş 102 epilepsi hastasının kopya sayısı varyasyonlarının tüm ekzom dizileme yöntemiyle araştırıldığı bu çalışmada her bir hastada ortalama olarak 12 CNV tespit edilmiştir. Tespit edilen CNV'ler DECIPHER ve OMIM veritabanlarındaki bilgilerle karşılaştırılmış ve değerlendirilen kromozomal bölgelerdeki varyasyonlar nöbet veya epilepsi fenotipleri ile ilişkilendirilmiştir. Veritabanları ile uyumlu olan CNV'ler, patojenite tahmini için kritik olarak değerlendirilmiştir. Bununla birlikte, DECIPHER veritabanında bulunmayan veya kayıtlar ile sınırlı olarak örtüşen CNV'ler, eğer epilepsi ile ilişkilendirilen genleri içeriyorsa patojenik olarak kabul edilmiştir. Fenotip genotip ilişkisini açıklayan bilimsel araştırmalar sonucu belirlenen genler de bu sınıfta değerlendirilmiştir.

Yapılan çalışmada elde edilen bulgular, epilepsi ile ilişkili genetik varyasyonların patojenik potansiyellerini ortaya koymaktadır. Belirli epileptik sendromlarla ilişkilendirilen CNV'ler, bu genetik değişikliklerin epileptogenezde oynayabileceği roller hakkında önemli bilgiler sunmaktadır. Tablo 1'de sunulan veriler, klinik fenotip ile genetik varyasyonlar arasındaki ilişkiyi ortaya koymakta ve epilepsi tedavisinde genetik analizlerin önemini vurgulamaktadır.

Yeni nesil dizileme yöntemlerinin kullanımı, epilepsi araştırmalarında genetik varyasyonların tespitini ve anlaşılmasını önemli ölçüde kolaylaştırmaktadır. Bu çalışma, CNV'lerin epilepsi patogeneze katkıda bulunabilecek genetik varyasyonlar olarak dikkatle incelenmesi gerektiğini göstermekte ve epilepsiye genetik yatkınlık ve patojenite mekanizmalarının daha iyi anlaşılmasına katkı sağlamaktadır. Gelecekte yapılacak geniş kapsamlı çalışmalar, bu bulguların klinik uygulamalara entegrasyonunu ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesini destekleyecektir.

Teşekkür

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 123S381 Numaralı proje ile desteklenmiştir. Projeye verdiği destekten ötürü TÜBİTAK'a teşekkürlerimizi sunarız. Ezgi Özyurt, Yükseköğretim Kurulu (YÖK) 100/2000 Doktora Programı kapsamında YÖK tarafından belirlenen 100 ulusal öncelikli alandan biri olan "Biyoenformatik-Biyostatistik" alanında doktora bursiyeridir.

Referanslar

1. Chen T, Giri M, Xia Z, Subedi YN, Li Y: Genetic and epigenetic mechanisms of epilepsy: a review. *Neuropsychiatric Disease and Treatment* 13: 1841 (2017).
2. Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, Eeg-Olofsson O, Sadleir LG, Gill D, Ben-Zeev B, et al.: Rare copy number variants are an important cause of epileptic encephalopathies. *Ann Neurol* 70: 974-985 (2011).
3. World Health Organization WHO | Epilepsy: a public health imperative. *Who* 171 (2019).
4. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon CS, Dykeman J, Pringsheim T,

- Lorenzetti DL, Jetté N: Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. *Neurology* 88: 296 (2017).
5. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshé SL, et al.: ILAE Classification of the Epilepsies Position Paper of the ILAEC Commission for Classification and Terminology. *Epilepsia* 58: 512 (2017).
 6. Mefford HC: CNVs in Epilepsy. *Current Genetic Medicine Reports* 2: 162 (2014).
 7. Olson H, Shen Y, Avallone J, Sheidley BR, Pinsky R, Bergin AM, Berry GT, Duffy FH, Eksioglu Y, Harris DJ, et al.: Copy Number Variation Plays an Important Role in Clinical Epilepsy. *Annals of neurology* *Neurology* 75: 943–958 (2014).
 8. Zhang MW, Liang XY, Wang J, Gao L Di, Liao HJ, He YH, Yi YH, He N, Liao WP: Epilepsy-associated genes: an update. *Seizure* 116: 4–13 (2024).
 9. Olson H, Shen Y, Avallone J, Sheidley BR, Pinsky R, Bergin AM, Berry GT, Duffy FH, Eksioglu Y, Harris DJ, et al.: Copy number variation plays an important role in clinical epilepsy. *Annals of neurology* 75: 943 (2014).
 10. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al.: Global variation in copy number in the human genome. *Nature* 2006 444:7118 444: 444–454 (2006).
 11. Hastings PJ, Lupski JR, Rosenberg SM, Ira G: Mechanisms of change in gene copy number. *Nature Reviews Genetics* 2009 10:8 10: 551–564 (2009).
 12. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nature Reviews Genetics* 7: 85–97 (2006).
 13. Zhang F, Gu W, Hurles ME, Lupski JR: Copy Number Variation in Human Health, Disease, and Evolution. *Annual Review of Genomics and Human Genetics* 10: 451–481 (2009).
 14. Alkan C, Coe BP, Eichler EE: Genome structural variation discovery and genotyping. *Nature Reviews Genetics* 2011 12:5 12: 363–376 (2011).
 15. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW: The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research* 42: D986–D992 (2014).
 16. Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, Franke A, Malafosse A, Genton P, Thomas P, et al.: Genome-Wide Copy Number Variation in Epilepsy: Novel Susceptibility Loci in Idiopathic Generalized and Focal Epilepsies. *PLOS Genetics* 6: e1000962 (2010).
 17. Coppola A, Cellini E, Stamberger H, Saarentaus E, Cetica V, Lal D, Djémié T, Bartnik-Glaska M, Ceulemans B, Helen Cross J, et al.: Diagnostic implications of genetic copy number variation in epilepsy plus. *Epilepsia* 60: 689–706 (2019).
 18. Tsuchida N, Nakashima M, Kato M, Heyman E, Inui T, Haginoya K, Watanabe S, Chiyonobu T, Morimoto M, Ohta M, et al.: Detection of copy number variations in epilepsy using exome data. *Clinical Genetics* 93: 577–587 (2018).
 19. Buysse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mortier G, Speleman F, rn Menten B: Challenges for CNV interpretation in clinical molecular karyotyping: Lessons learned from a 1001 sample experience. *European Journal of Medical Genetics* 52: 398–403 (2009).
 20. Carter NP: Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature Genetics* 39: S16–S21 (2007).
 21. Metzker ML: Sequencing technologies — the next generation. *Nature Reviews Genetics* 2010 11:1 11: 31–46 (2009).
 22. Meyerson M, Gabriel S, Getz G: Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics* 2010 11:10 11: 685–696 (2010).
 23. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z: Computational tools for copy number

- variation (CNV) detection using next-generation sequencing data: Features and perspectives. *BMC Bioinformatics* 14: 1–16 (2013).
24. Zhao L, Liu H, Yuan X, Gao K, Duan J: Comparative study of whole exome sequencing-based copy number variation detection tools. *BMC Bioinformatics* 21: 1– 10 (2020).
 25. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al.: From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* 43: 11.10.1-11.10.33 (2013).
 26. Andrews Simon: Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data (2010).
 27. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120 (2014).
 28. Li H, Durbin R: Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26: 589–595 (2010).
 29. MarkDuplicates (Picard) – GATK (n.d.), <https://gatk.broadinstitute.org/hc/en-us/articles/4409924785691-MarkDuplicates-Picard->, last accessed 2024/03/22
 30. Aneta Mikulasova (n.d.). GitHub - AnetaMikulasova/CNVRobot: CNVRobot is an integrated pipeline for CNVs detection from short-read DNA sequencing data., <https://github.com/AnetaMikulasova/CNVRobot>, last accessed 2024/03/22
 31. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, Muller J: AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* 34: 3572–3574 (2018).

B Cell Epitope Analysis of the SARS-CoV-2 Spike Protein: Accessibility, Polymorphisms, and Implications for Vaccine Design

Mehmet Erşan KALAYCI¹[0000-0002-4797-9020], Güven ŞAHİN²[0000-0003-3901-5819], Muammer ALBAYRAK¹[0000-0002-5946-6310], Kemal TURHAN¹[0000-0001-7871-3025]

¹ Karadeniz Teknik Üniversitesi, Trabzon, Türkiye,
²Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye,
mehmetersankalayci@ktu.edu.tr
guven.sahin@erdogan.edu.tr
m.albayrak@ktu.edu.tr
kemalturhan@ktu.edu.tr

Abstract. B cells play a crucial role in adaptive immune responses by producing antibodies that target specific antigens. The identification of B cell epitopes, which are the specific regions on antigens recognized by B cell receptors, is essential for understanding immune responses and developing vaccines. In this study, the most suitable vaccine candidates were predicted for five separate epitope regions that can be used in vaccine development using B cell epitope prediction methods. A total of 98 B cell epitopes were identified in the SARS-CoV-2 spike protein from the literature and relevant databases. Among them, 5 epitopes met the accessibility criteria and had accessible residues. The longest accessible subsequences ranged from 4 to 10 amino acids, with an average length of 4.75 amino acids. Additionally, 25% of the accessible residues contained polymorphic sites, and 75% overlapped with known protein domains. We also highlight the potential of B cell epitope candidates prediction to contribute to new vaccine design strategies and personalized medicine for future virus outbreaks such as SARS-CoV-2. Our findings also suggest that B cell epitope prediction could have a positive impact on developing rapid and effective vaccine strategies for future virus outbreaks. This study demonstrates that B cell epitope prediction could be a significant tool in vaccine development against pathogens with pandemic potential, especially those like SARS-CoV-2.

Objective: The aim of this study is to analyze B cell epitopes in the SARS-CoV-2 spike protein, examining their accessibility, polymorphism, and implications for vaccine design.

Methods: We employed a methodology that included data collection, defining accessibility criteria, identifying subsequences, checking their accessibility, data processing, analyzing additional factors, and grouping and filtering. Various methods were evaluated for predicting B cell epitopes, leading to the identification of new epitope regions.

Results: In the study, a total of 98 B cell epitopes were identified in the SARS-CoV-2 spike protein from the literature and relevant databases. The results were obtained using Python relevant libraries. Among them, 5 epitopes met the accessibility criteria and had accessible residues. The longest accessible subsequences ranged from 4 to 10 amino acids, with an average length of 4.75 amino acids. All of these residues were 4 amino acids or longer. Additionally, 25% of the accessible residues contained polymorphic sites, and 75% overlapped with known protein domains.

Conclusion: The findings provide a comprehensive overview of B cell epitopes in the SARS-CoV-2 spike protein, highlighting their accessibility, variability, and potential immunogenicity. These insights are crucial for developing effective vaccines and understanding immune responses to SARS-CoV-2 infection. The presence of epitopes with polymorphic sites suggests the need for further research into antigenic diversity for vaccine efficacy.

Keywords: B cells, epitope prediction, vaccine development, immune response, SARS-CoV-2

Parkinson's Disease Related Alterations in Common Spatial Patterns of EEG Signals

Fatih Onay¹ and Bilge Karaçalı¹

İzmir Institute of Technology, İzmir Urla 35430, Turkey,
fatihonay@iyte.edu.tr

Abstract. Diagnosing Parkinson's Disease (PD) is a challenging task since the complex neuronal dynamics of the brain under wide range of effects have not been fully identified. Furthermore, conventional approaches like subjective clinical evaluations can be impractical in a large scale due to their expense and limited availability. EEG has a capability to measure electrical fields of neuronal sources non-invasively. Rapid fluctuations can be detected easily due to high time resolution of EEG, though its poor spatial resolution restricts its ability to precisely localize the source of activity within the brain. Common Spatial Patterns (CSP) allow identifying spatial changes in neuronal activity by considering EEG signals across different conditions. In this study, we adopted the Common Spatial Patterns technique to analyze EEG signals from 13 healthy controls, 13 PD patients, and 13 PD patients with freezing of gait symptom (PDFOG) to detect disease driven spatial alterations. CSP filters contrasting pre-stimulus and post-stimulus EEG signals were designed across δ , θ , α , β_1 , β_2 , and broad-band frequency bands, along with different temporal ranges during the post-stimulus pedaling task. The findings based on two-sample t-tests indicated that the CSP filter weights change significantly between populations and the weights observed at the δ band were most discriminative among all group comparisons.

Keywords: Common spatial patterns, EEG, Parkinson's Disease, lower-limb pedaling

1 Introduction

Many neurodegenerative diseases lack specific biomarkers that can reliably indicate the presence of the disease or its progression [1]. Without biomarkers, diagnosis often relies on clinical symptoms and evaluations that maybe biased due to the subjective evaluation criteria [2]. Early detection of such diseases may provide an opportunity to prevent the progression of neurodegeneration in the brain [3].

The functionality of the brain to manage different operations such as cognition, sensory processing, and motor tasks is essential for human behavior, perception, and interaction with the environment. Disruption of neuronal organization can cause various cognitive impairments, sensory deficits, and motor dysfunctions which significantly affect quality of life [4]. One such neurodegenerative

condition is Parkinson's disease (PD), which primarily affects the dopaminergic neurons in the brain, leading to motor symptoms such as tremors, rigidity, as well as cognitive impairment [5]. All of these symptoms result from the neuronal degeneration primarily in the substantia nigra region of the brain, which leads to a decrease in dopamine production. There is still no cure for PD, and early diagnosis remains important for implementing treatment that can slow disease progression [6].

The critical aspect of the human brain is its ability to generate electrical fields, which play a fundamental role in coordinating neuronal communication to manifest various cognitive, sensory, and motor functions [7]. Observing and understanding the characteristics of these electrical fields, such as their patterns, frequencies, and spatial organizations, can provide valuable information about the dynamics of brain activity. Furthermore, alterations in neuronal organization due to neurodegenerative conditions like Parkinson's disease can lead to disruptions in these electrical fields, which may contribute to the progression of the disease [8]. Therefore, revealing the relationship between electrophysiological neuronal activity and pathological changes in the brain may help identify novel biomarkers and develop innovative therapies to halt the progression of neurodegenerative conditions like Parkinson's disease.

Several neuroimaging modalities have been used for observing physiological activity in the brain, including electroencephalography (EEG) [9], magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and so on. Each modality has its strengths and limitations regarding spatial and temporal resolution, sensitivity to different aspects of brain activity, cost, and feasibility for specific research or clinical applications. In recent years, EEG has been preferred for its high temporal resolution to capture rapid changes in brain activity, and it is helpful for studying dynamic cognitive processes, real-time brain-computer interfaces, and clinical applications such as PD diagnosis [10]. On the other hand, detecting spatial patterns about specific neuronal activity is challenging with EEG due to its limited spatial resolution. EEG records electrical activity from the scalp, which leads to mixed signals emitted from multiple brain regions. Therefore, localizing the source of neural activity accurately within the brain is problematic in the context of interpreting spatial patterns. Despite the poor spatial resolution of EEG, revealing spatial differences between different tasks or conditions can be achieved through advanced signal processing techniques such as functional connectivity analysis and spatial filtering methods like Common Spatial Patterns (CSP) [11]. CSP is a widely used spatial filtering method in EEG analysis that aims to extract spatial patterns associated with different cognitive processes or states.

Numerous studies have been carried out to differentiate cortical activity between PD and healthy subjects using EEG signals [10]. The effects of PD on the EEG signals have been investigated in various experimental contexts, such as analyzing resting-state EEG signals, event-related potentials during cognitive tasks, spectral analysis during motor activities, and spatial patterns related to auditory task [12]. Revealing the cortical regions affected by PD is not easy

because of the complex nature of the disease, which can involve widespread alterations in neural activity across multiple brain regions. Furthermore, spectral and spatial characteristics of cortical EEG signals are sensitive to subject-specific factors such as head shape, electrode placement variability, and skull thickness, which may hinder accurate localization of affected regions in Parkinson's disease.

In this study, we hypothesized that changes in spatial characteristics of cortical activity between pre-stimulus and post-stimulus EEG activity may help elucidate the neural dynamics associated with Parkinson's disease pathology. CSP filters maximizing the variance of activity within certain post-stimulus intervals with respect to the pre-stimulus range were employed in conventional frequency sub-bands to test the proposed hypothesis. Thus, the designated CSP filters aim to evaluate alteration of spectral, temporal and spatial patterns as disease progresses, particularly in relation to a lower-limb pedaling task.

This paper is organized as follows. In the next section we describe the EEG dataset used in the study along with the details of the method employed to identify spatial activity differences between pre-stimulus and post-stimulus EEG signals in the different groups. In section 3, we shared the results in the context of PD electrophysiology, and then we concluded the paper.

2 Material and Method

In this section, we first describe the EEG dataset used in this study along with the pre-processing methods such as sub-band filtering. Then, we describe the details of the CSP algorithm and the subsequent statistical analysis.

2.1 EEG Pre-processing and Dataset

In this study we used the EEG dataset that was collected to reveal differences of motor and cognitive activity between the healthy controls, PD patients, and PD patients with freezing of gait symptom (PDFOG). The EEG recordings are available publicly in the Predict repository [13] .

The EEG signals were obtained from a population of 39 participants, divided equally into three groups. All participants were instructed to perform a pedaling task while their EEG signals were recorded. Furthermore, the PD and PDFOG groups were under medication during the task. Additionally, participants diagnosed with PD and PDFOG underwent comprehensive neuropsychological assessments, including the Montreal Cognitive Assessment (MOCA) to evaluate cognitive abilities, and the motor Unified Parkinson's Disease Rating Scale (UPDRS III) to assess motor symptoms associated with PD.

Experiment begins with a fixation cue that remains 0.5 seconds on the screen before disappearing, and then participants await for the GO cue within a range spanning 1 to 2 seconds determined randomly. Upon the appearance of the GO cue, participants start pedaling to complete one rotation and this processes is repeated for a total of 43 to 70 trials. The time flow of experimental scheme is shown in Figure 1.

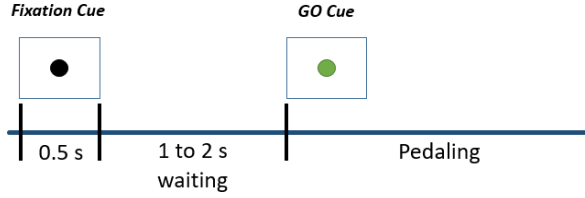


Fig. 1. Experimental Scheme

The records in the database contain 64 EEG channels filtered between 0.1 to 30 Hz frequency range sampled at 500 Hz. Channels FP1, FP2, FT10, TP9, and TP10, which are prone to contamination by eyeblink and muscle artifacts, were excluded, resulting in 59 electrodes. Bad trial elimination and eye-blink removal via ICA were also applied in the EEG data to improve the quality and reliability of the recorded signals for subsequent analysis [13].

The EEG signals were first filtered (Morlet Wavelet filtering) into specific frequency bands commonly associated with various brain functions. These conventional sub-bands include, δ (1- 4 Hz), ϑ (4- 8 Hz), α (8- 12 Hz), β_1 (13- 20 Hz), β_2 (20- 30 Hz). Additionally, the broad band (1-30 Hz) EEG was included to reveal overall spectral effects in the analysis.

Furthermore, individual EEG channel locations may vary across different participants due to differences in head size, shape, and electrode placement protocols. Adjacent electrodes may reflect similar activity stemming from volume conduction that leads to simultaneous detection of wide-spread neuronal sources [14]. Subject based differences could potentially deteriorate fair comparison of CSP filters between participants from different populations and pooling EEG electrodes within certain cortical regions can prevent discrepancies [15]. Therefore, we conducted our study based on region-averaged EEG signals that are represented in Table 1.

To account for temporal variability, we divided the signals into different time intervals to analyze various phases of brain activity. Firstly, we defined the pre-stimulus signals $X_i^{pre}(t)$ for t ranging from -500ms to 0ms. Following the stimulus presentation, we segmented the post-stimulus signals into four distinct phases. These include $X_i^{post,first}(t)$ for $t \in [0\text{ms}, 500\text{ms}]$, $X_i^{post,mid}(t)$ for $t \in [500\text{ms}, 1000\text{ms}]$, $X_i^{post,late}(t)$ for $t \in [1000\text{ms}, 1500\text{ms}]$, and $X_i^{post,late}(t)$ for $t \in [0\text{ms}, 1500\text{ms}]$. This segmentation allows to examine the neural responses occurring before and after the stimulus presentation across different time windows for detailed analysis.

2.2 Common Spatial Pattern Analysis

Common spatial patterns (CSP) have been recognized as a powerful technique in EEG analysis for distinguishing between different motor and imagery activities. CSP design aims to determine spatial patterns that maximize the variance

Table 1. Electrode Positions

Brain Region	Electrode Positions
Frontal Left	AF7, AF3, F7, F5, F3, F1, FC5, FC3, FC1
Frontal Right	AF8, AF4, F8, F6, F4, F2, FC6, FC4, FC2
Temporal Left	FT7, T7, TP7, P7
Temporal Right	FT8, T8, TP8, P8
Central Left	FC5, FC3, FC1, C5, C3, C1, CP5, CP3, CP1
Central Right	FC6, FC4, FC2, C6, C4, C2, CP6, CP4, CP2
Parietal Left	CP5, CP3, CP1, P5, P3, P1
Parietal Right	CP6, CP4, CP2, P6, P4, P2
Occipital Left	PO7, O1
Occipital Right	PO8, O2
Midline Frontal	AFz, Fz, FCz
Midline Central	FCz, Cz, CPz
Midline Parietal	CPz, Pz, POz
Midline Occipital	POz, Oz

between these activities. It is a linear operation that weights the EEG channels to optimize the discrimination between states and enhances the signal-to-noise ratio.

EEG data for each trial were segmented into epochs corresponding to the pre-stimulus interval $[-\tau \text{ ms}, 0 \text{ ms}]$ signal, $X_i^{pre} \in \mathbf{R}^{K \times L}$ in the i^{th} trial, and post-stimulus interval $[\kappa_1 \text{ ms}, \kappa_2 \text{ ms}]$ signal (we already declared the time intervals of post-stimulus), $X_i^{post} \in \mathbf{R}^{K \times L}$, where K is the number of channels and L is the number of samples. The covariance matrices C_i^{pre} and C_i^{post} , corresponding to the pre-stimulus and post-stimulus intervals of the i^{th} trial respectively, are computed as follows:

$$C_i^{pre} = \frac{1}{L} X_i^{pre} (X_i^{pre})^T \quad (1)$$

$$C_i^{post} = \frac{1}{L} X_i^{post} (X_i^{post})^T \quad (2)$$

To obtain the average covariance matrices \bar{C}^{pre} and \bar{C}^{post} , individual covariance matrices across all trials were averaged as:

$$\bar{C}^{pre} = \frac{1}{N} \sum_{i=1}^N C_i^{pre} \quad (3)$$

and

$$\bar{C}^{post} = \frac{1}{N} \sum_{i=1}^N C_i^{post} \quad (4)$$

where N represents the total number of trials.

A spatial filter matrix W can be obtained by solving the generalized eigenvalue problem that extracts spatial patterns distinguishing the pre-stimulus and post-stimulus activities

$$\operatorname{argmax}_W \frac{W^T \bar{C}_{post} W}{W^T \bar{C}_{pre} W} \quad (5)$$

The expression in equation [5] can be rewritten as follows:

$$\Lambda = (W^T \bar{C}_{pre} W)^{-1} W^T \bar{C}_{post} W \quad (6)$$

where Λ represents eigenvalues of $\bar{C}_{pre}^{-1} \bar{C}_{post}$. This can also be shown in another form

$$\bar{C}_{post} W = \Lambda \bar{C}_{pre} W \quad (7)$$

The spatial filter matrix $W \in R^{K \times K}$ includes the weighting coefficients that define the linear transformation applied to the region-averaged EEG data. Each row of the matrix corresponds to a spatial filter, and each column represents the weights assigned to each region. Therefore, W captures the spatial patterns or configurations of activity that maximize the energy difference between the pre-stimulus and post-stimulus intervals. Furthermore, the diagonal matrix Λ corresponds to the eigenvalues obtained from the generalized eigenvalue problem and each eigenvalue λ_i reflects the energy ratio of the corresponding spatial filter. The spatial filters with highest and lowest eigenvalues were used in this study.

The CSP approach is primarily focused on identifying spatial patterns that maximize the discrimination between pre-stimulus and post-stimulus intervals. Although it can identify spatial patterns associated with discriminative features, it does not provide information about the temporal or frequency properties of these patterns. To address this issue, we designed a set of CSP filter matrices W for each subject, considering six frequency bands and four time intervals. From these matrices, we selected only the first and last CSP filters, resulting in a total of 48 filters obtained for each subject. The main steps of deriving CSP filters from region averaged EEG signals were illustrated in Figure 2.

2.3 Statistical Analysis

For each CSP filter (totaling 48), independent two-sample t-tests were conducted to compare the weights (14 weights for 14 brain regions from each derived filter) across the six frequency bands (δ , θ , α , β_1 , β_2 , and 1-30 Hz broadband) and four time intervals (0-500 ms, 500-1000 ms, 1000-1500 ms, and 0-1500 ms) between groups (healthy controls, PD patients, and PDFOG patients). Correcting for multiple comparisons was carried out using the Benjamini-Horchberg method to control for false discovery rate, considering the number of comparisons conducted. The significance level for all t-tests was set at $\alpha = 0.05$.

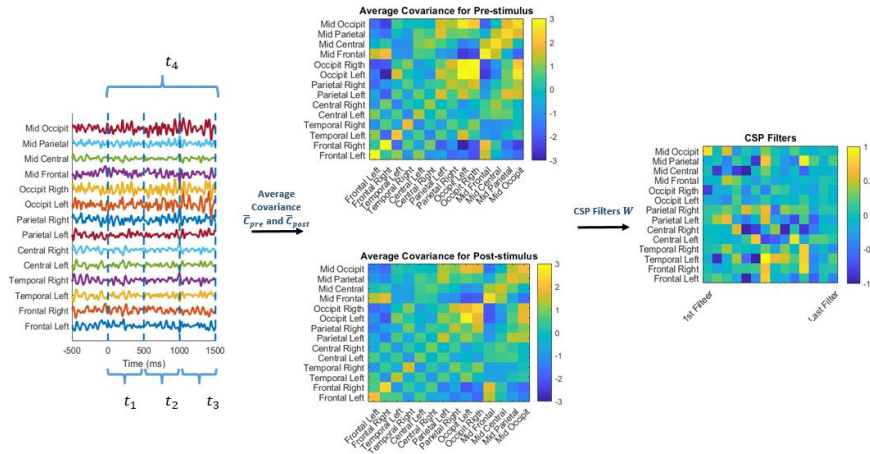


Fig. 2. CSP Filter design for a certain frequency band

3 Results

In this study, we employed the weights of CSP filters with the highest and lowest eigenvalues for statistical test analysis to investigate group based differences. The corresponding p-values for group-level comparisons of HC vs. PD, HC vs. PDFOG, and PD vs. PDFOG can be seen in Table 2.

The corresponding frequency band for the lowest p-value across all groups is δ while time intervals and the regions are not consistent across group comparisons. Another important finding indicates that including onset time into the CSP filter design can be more preferable and this phenomena is more prominent between HC and PD group comparison (9 out of 10 filters include onset time). Additionally, the weights of first CSP filter potentially have better discriminatory power for HC vs. PD (7 out of 10 weights) and PD vs. PDFOG (8 out of 10 weights) while the weights of last CSP filter is selected more frequently in the comparison of HC vs. PDFOG (8 out of 10 weights). Although all statistical p-values are lower than 0.05 in Table 2, only one feature remained significant af-ter additional Bonjemini-Hochberg multiple-comparison correction. This weight corresponding to parietal right region δ sub-band CSP filter in time interval of 1 - 1.5 second can determine the HC vs. PDFOG group level alterations in a statistically consistent manner. General spectral and temporal properties of CSP weights can also be seen in Figure 3. The temporal interval of 0 - 0.5 second and α spectral band based CSP filter weights appeared as most dominant factors in the separation of groups.

Furthermore, the detailed analysis of CSP filter including the most significant weight was carried out by using PCA and Isomap data reduction techniques to visualize distributions of filter weights from determined 14 regions. Firstly, PCA

analysis was performed on the data of the first CSP filters belonging to participants from HC and PDFOG groups. Then, the first two components obtained from PCA reduction were visualized in the Figure 4.

In the PCA sub-space, discrimination of two groups is not obvious since outliers could distort separation of data. Therefore, another data reduction technique isomap which is known for capturing nonlinear relationships in high-dimensional data was performed to find hidden patterns and findings were showed in Figure 4. Isomap based reduction provided more distinguishing distribution of filter weights in the projected subspace. Moreover, box plots of the most significant weight obtained from HC and PDFOG groups were also represented in

Table 2. Statistical t-test analysis between groups (top: HC vs. PD, middle: HC vs. PDFOG, bottom: PD vs. PDFOG)

Region	Time Interval (second)	Frequency Band (Hz)	CSP Filter	p-value
Mid Parietal	0 - 1.5	δ	Last Filter	8,74e-04
Frontal Left	0 - 0.5	ϑ	1st Filter	2,20e-03
Temporal Left	0 - 1.5	α	1st Filter	2,37e-03
Occipital Righth	0 - 0.5	Broad	1st Filter	5,35e-03
Mid Central	0 - 1.5	β_2	Last Filter	6,18e-03
Temporal Right	0 - 0.5	α	Last Filter	7,45e-03
Parietal Left	0.5 - 1	β_1	1st Filter	7,75e-03
Parietal Right	0 - 0.5	δ	1st Filter	8,58e-03
Frontal Right	0 - 0.5	ϑ	1st Filter	8,99e-03
Parietal Left	0 - 0.5	ϑ	1st Filter	1,00e-02
Parietal Right	1 - 1.5	δ	1st Filter	1e - 04*
Central Right	1 - 1.5	β_1	1st Filter	1,82e-03
Central Right	0 - 1.5	α	Last Filter	3,78e-03
Mid Central	0 - 1.5	β_2	Last Filter	8,77e-03
Mid Frontal	1 - 1.5	β_1	Last Filter	9,84e-03
Parietal Right	0 - 0.5	δ	Last Filter	1,11e-02
Mid Occipit	0 - 1.5	β_2	Last Filter	1,29e-02
Central Left	0 - 0.5	ϑ	Last Filter	1,38e-02
Parietal Right	0 - 1.5	β_1	Last Filter	1,44e-02
Occipit Righth	0 - 1.5	α	Last Filter	1,72e-02
Central Left	0 - 0.5	δ	1st Filter	6,23e-04
Parietal Left	0 - 0.5	ϑ	1st Filter	2,14e-03
Mid Parietal	0 - 1.5	α	1st Filter	2,64e-03
Parietal Left	0 - 0.5	δ	1st Filter	5,24e-03
Central Left	0.5 - 1	ϑ	Last Filter	7,15e-03
Central Right	0 - 0.5	β_1	1st Filter	7,80e-03
Parietal Left	0 - 0.5	ϑ	Last Filter	8,85e-03
Parietal Left	1 - 1.5	α	1st Filter	9,38e-03
Mid Parietal	1 - 1.5	α	1st Filter	9,83e-03
Frontal Left	1 - 1.5	α	1st Filter	1,11e-02

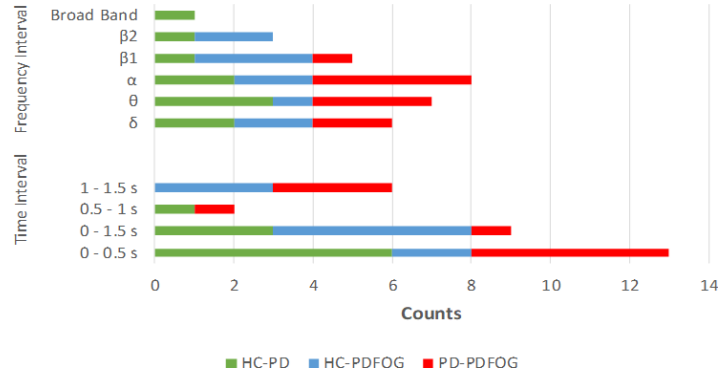


Fig. 3. Distribution of significant weights in each comparison for frequency sub-bands and time intervals.

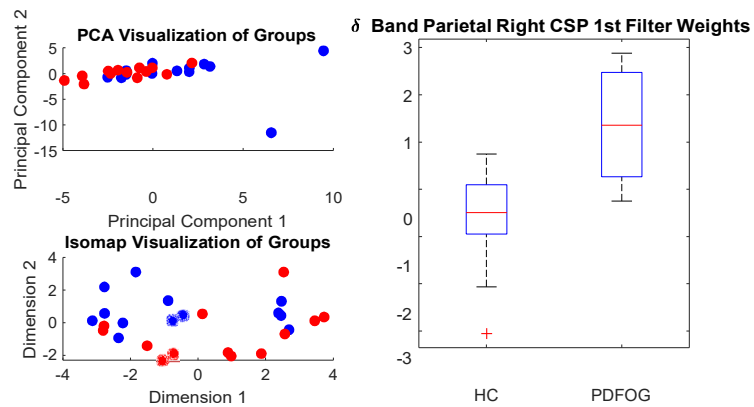


Fig. 4. Distribution of CSP Filter (blue dots for HC groups and red dots for PDFOG group)

Figure 4 to summarize main tendency of this specific weight. It can be easily observed that participants from PDFOG groups have significantly higher weight values with respect to the healthy controls.

Finally, we illustrated the topographical distribution of filter weights for HC and PDFOG groups and increasing right-parietal weight in PDFOG group can be seen in the Figure 5. In addition to the topographical maps, population level average of EEG signals in the right-parietal region for HC and PDFOG groups was illustrated in Figure 5. Although right-parietal weight of CSP filter was successful in the discrimination between groups, average EEG activity can not capture altered dynamics of disease related neuronal responses within the 1-1.5 second interval.

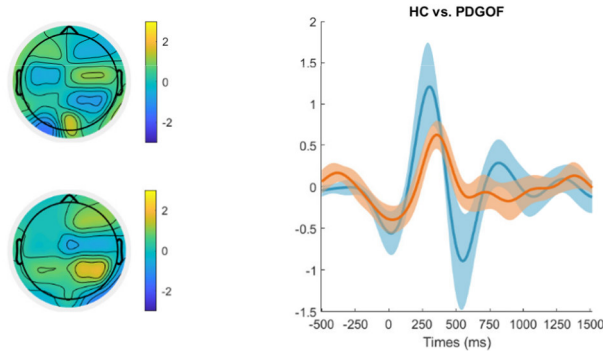


Fig. 5. Topographical illustration of CSP weights (left top: HC, left bottom: PDFOG) and average-right parietal EEG activity (right) for HC and PDFOG groups (blue:HC, orange:PDFOG)

4 Conclusion

The purpose of this study was to demonstrate the application of CSP design within clinical contexts, specifically investigating spatial alterations related to Parkinson's disease during lower-limb pedaling task. The experimental scheme included pre-stimulus baseline and post-stimulus pedaling time intervals that led to arise of varying cortical activity representing cognitive and motor processes. Since spatial and spectral content of electrophysiological sources within the brain can show PD related changes [16], we also investigated cortical activity of PD and PDFOG patients in various frequency sub-bands to show extensive effects of disease. In addition to the sub-band analysis, we also explored temporal spatial reconfiguration of cortical sources as baseline activity evolves into pedaling task by employing CSP filters in various time ranges.

The most prominent conclusion we can draw from this study, strongest statistical differences occurs between all pairwise comparisons when weights of CSP filters in δ sub-band is employed. Increasing synchronization or power in the lower band delta is one of the common alterations in EEG signals among PD patients [10] and this situation may extend into the delta band based spatial configurations in the brain. Furthermore, it has been shown that using other sub-bands can also help separation of PD patients from healthy controls by combining CSP filter design with additional methods [17]. Moreover, most of the significant differences between groups is observed within the temporal range of cue trigger event which could potentially activate neuronal sources linked to cognitive processing and attention. This finding points out immediate transition from baseline to pedaling tasks among individuals with Parkinson's disease leads to changes of spatial distribution of cortical activity in several sub-bands. Even if we found significant p-values for group comparisons, additional false-discovery correction test did not allow significant separation except right-parietal region weight of δ band CSP filter for groups of HC and PDFOG patients. This find-

ing is relevant to previous studies that explained the motor and sensory-motor cortex can be linked to PD pathology [18].

The current findings of this study emphasize that the spatial, temporal, and spectral content of EEG signals should be regarded as the main factors in the analysis of PD-related inquiry. On the other hand, it is important to be aware of the inherent limitations of EEG signals when developing methodological approaches based on CSP and similar techniques [19]. One of the main drawbacks of CSP filter design is its high sensitivity to signal power variations, as well as individual subject or session differences. This sensitivity undermines its robustness and hinders achieving effective filter design in population-level data [21]. Amplitude, shape, latency, and spatial parameters of EEG signals may not remain the same within the same individual across sessions of EEG recordings due to instrumental, biological, and behavioral factors [20]. Another limitation is the changing reaction time of participants to start pedaling tasks [22] and aligning temporal activity may improve the performance of CSP filters.

Finally, our findings demonstrate the potential of CSP filters in revealing spatial alterations in neuronal activity associated with Parkinson's Disease and freezing of gait symptoms. Therefore, the use of CSP filters in neurological examinations can lead to a better understanding of PD and development of improved diagnostic and therapeutic strategies.

References

1. Hansson, Oskar. "Biomarkers for neurodegenerative diseases." *Nature medicine* 27.6 (2021): 954-963.
2. Kraus, Virginia B. "Biomarkers as drug development tools: discovery, validation, qualification and use." *Nature Reviews Rheumatology* 14.6 (2018): 354-362.
3. Counts, Scott E., et al. "Biomarkers for the early detection and progression of Alzheimer's disease." *Neurotherapeutics* 14.1 (2017): 35-53.
4. Holmes, Gregory L. "Cognitive impairment in epilepsy: the role of network abnormalities." *Epileptic Disorders* 17.2 (2015): 101-116.
5. Cavanagh, James F., Saphira Ryman, and Sarah Pirio Richardson. "Cognitive control in Parkinson's disease." *Progress in Brain Research* 269.1 (2022): 137-152.
6. Ellis, Terry D., et al. "Identifying clinical measures that most accurately reflect the progression of disability in Parkinson disease." *Parkinsonism & related disorders* 25 (2016): 65-71.
7. Cohen, Michael X. "Where does EEG come from and what does it mean?." *Trends in neurosciences* 40.4 (2017): 208-218.
8. Nimmrich, Volker, Andreas Draguhn, and Nikolai Axmacher. "Neuronal network oscillations in neurodegenerative diseases." *Neuromolecular medicine* 17 (2015): 270-284.
9. Wang, Ying, et al. "Freezing of gait detection in Parkinson's disease via multimodal analysis of EEG and accelerometer signals." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.
10. Geraedts, Victor J., et al. "Clinical correlates of quantitative EEG in Parkinson disease: A systematic review." *Neurology* 91.19 (2018): 871-883.

11. Barachant, Alexandre, et al. "Common spatial pattern revisited by Riemannian geometry." 2010 IEEE international workshop on multimedia signal processing. IEEE, 2010.
12. Wang, Qing, et al. "Characterization of EEG data revealing relationships with cognitive and motor symptoms in Parkinson's disease: a systematic review." *Frontiers in aging neuroscience* 12 (2020): 587396.
13. Singh, Arun, et al. "Frontal theta and beta oscillations during lower-limb movement in Parkinson's disease." *Clinical Neurophysiology* 131.3 (2020): 694-702.
14. Brunner, Clemens, et al. "Volume conduction influences scalp-based connectivity estimates." *Frontiers in computational neuroscience* 10 (2016): 121.
15. Brunner, Clemens, et al. "Oscillatory electroencephalographic patterns of arithmetic problem solving in fourth graders." *Scientific Reports* 11.1 (2021): 23278.
16. Naghsh, Erfan, Mohamad Farzan Sabahi, and Soosan Beheshti. "Spatial analysis of EEG signals for Parkinson's disease stage detection." *Signal, Image and Video Processing* 14.2 (2020): 397-405.
17. Smrdel, Aleš. "Use of common spatial patterns for early detection of Parkinson's disease." *Scientific Reports* 12.1 (2022): 18793.
18. Conte, Antonella, et al. "Pathophysiology of somatosensory abnormalities in Parkinson disease." *Nature Reviews Neurology* 9.12 (2013): 687-697.
19. Burle, Boris, et al. "Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view." *International Journal of Psychophysiology* 97.3 (2015): 210-220.
20. Congedo, Marco. "The analysis of event-related potentials." *Computational EEG Analysis: Methods and Applications* (2018): 55-82.
21. Congedo, Marco, Alexandre Barachant, and Rajendra Bhatia. "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review." *Brain-Computer Interfaces* 4.3 (2017): 155-174.
22. Fatih, O. N. A. Y., and Bilge KARAÇALI. "Accelerometer-Based Timing Analysis for Parkinson's Disease Classification." 2023 31st Signal Processing and Communications Applications Conference (SIU). IEEE, 2023.

TheraPose: A Large Video Dataset for Physiotherapy Exercises

Hamdi Yalın Yalıcı¹, Ahmet Emre Usta², Ümit Atıla³, Alaettin Uçan¹,
Adem Ali Yılmaz¹, Ali Yaşar Yiğit¹

¹ Tiga Healthcare Technologies, Ankara, Türkiye

²Department of Artificial Intelligence Engineering, Faculty of Engineering,
Hacettepe University, Ankara, Türkiye

³Department of Computer Engineering, Faculty of Engineering,
Gazi University, Ankara, Türkiye

{yalin.yalic, alaettin.ucan, adem, ali}@tigahealth.com,
usta_ahmet@hacettepe.edu.tr, umitatila@gazi.edu.tr

Abstract. This paper introduces the TheraPose dataset, a novel collection tailored to enhance physiotherapy by integrating human posture prediction technologies. Consisting of 3,424,125 frames capturing 123 different physiotherapy exercises, TheraPose aims to refine therapeutic interventions by increasing precision and efficiency. The dataset, meticulously assembled using state-of-the-art motion capture systems and high-resolution video recording, captures detailed three-dimensional joint movements under various conditions, providing comprehensive and robust data. TheraPose is a fundamental resource for developing advanced pose prediction models, which are crucial for ensuring exercise accuracy, thus maximizing therapeutic outcomes and accelerating patient recovery. Its variety in exercise and recording settings makes it uniquely suited to improve algorithms for individualized therapy in clinical and home settings. As part of our commitment to encourage innovation and collaboration in rehabilitation sciences, efforts are underway to make a sample subset of the TheraPose dataset publicly available. "TheraPose" fills the gap in existing research tools, underlining the potential of tailored rehabilitation strategies facilitated by the latest technological advances.

Keywords: Physiotherapy, Human Pose Estimation, Rehabilitation Exercises.

1 Introduction

Physical and occupational therapies are critical components of rehabilitation, primarily utilizing prescribed exercises to restore functional capabilities and accelerate recovery. The efficacy of these therapies hinges on the precise execution of exercises, which underscores the necessity for advanced monitoring tools. Human pose models accurately capture and analyze three-dimensional joint movements, which are increasingly crucial in this context. These models provide real-time data on joint positions and movements,

enabling therapists to ensure correct exercise execution. This capability is essential to speed up recovery and reduce the risk of further injuries or complications.

Moreover, this technology facilitates the computation of essential metrics like joint angles and distances. These metrics allow therapists to monitor patient's progress in real-time and make data-driven adjustments to treatment plans. Such specialized interventions are pivotal for enhancing the effectiveness of therapy, ultimately leading to more successful patient outcomes. Integrating these models into therapy routines through camera-equipped applications helps track adherence to prescribed exercises and supports the development of personalized treatment plans. This adaptation is particularly beneficial for populations with limited mobility or those requiring remote monitoring, thus extending the reach of therapeutic care beyond traditional clinical settings.

Related Work. Various specialized datasets have significantly influenced the development of human pose estimation technologies. These datasets are designed to address multiple needs and environments, from sports scenes to daily activities and clinical settings.

A Video-based Physiotherapy Exercise Dataset¹ offers a unique approach by focusing on video data that can be captured from commonplace devices like smartphones, making it highly relevant for home-based monitoring, especially during conditions like the COVID-19 pandemic. This dataset contains 1237 video clips of 14 different exercises recorded under various conditions to mimic real-world scenarios, providing a practical tool for physiotherapists to monitor and assess exercise execution remotely.

IntelliRehabDS (IRDS) - A Dataset of Physical Rehabilitation Movements² enhances automatic feedback mechanisms for rehabilitation exercises using Kinect motion sensors. It captures the correctness of exercise executions, supporting the development of systems that provide immediate corrective feedback in rehabilitation settings, a crucial feature for ensuring effective therapy outcomes.

Automated Evaluation of Physical Therapy Exercises³ utilizes multi-template dynamic time warping on wearable sensor signals to offer a novel method for automated evaluation of therapy exercises. This dataset highlights the integration of wearable technologies in physiotherapy, providing dynamic assessments that adapt to different patients' variable exercise execution.

UCO Physical Rehabilitation New Dataset⁴ and studies on human pose estimation methods specifically cater to physical rehabilitation, focusing on the precision required in such settings. This dataset aids in refining algorithms for improved assessment and execution of rehabilitation exercises.

A Database of Physical Therapy Exercises with Variability of Execution Collected by Wearable Sensors⁵ reflects the real-world variability in how individuals perform

exercises. It is a rich source for developing systems that can adapt to and learn from diverse execution styles, vital for personalized therapy.

3D Poses in the Wild⁶ combines outdoor imaging with inertial measurement units (IMUs), establishing new standards for image-based 3D pose estimation across various scenarios. This dataset demonstrates the potential for applications in sports and clinical settings involving outdoor or dynamic environments.

The discussion of existing pose estimation datasets like HumanEva⁷, MPII Human Pose⁸, and Human3.6M⁹ in the context of their contributions to sports and daily activity scenarios underscores the need for more specialized datasets in physiotherapy that provide sensitive and comprehensive 3D data for complex therapeutic exercises.

In contrast to existing datasets, our dataset TheraPose is uniquely adapted to the specific requirements of physiotherapy movements. It provides precise triaxial data points from 24 joint positions on the human body. It goes beyond existing datasets by covering a broader spectrum of activities, thus serving as a more comprehensive resource for researchers and practitioners in this field. This emphasis on physiotherapy-specific data sets underscores the current gap in research and demonstrates the potential transformative impact of precise, personalized pose prediction technologies on rehabilitation practice. This detailed integration of various datasets underscores the significant strides in human pose prediction technologies and their essential role in advancing physical and occupational therapy, particularly in facilitating personalized and accurate rehabilitation strategies.

<i>Dataset</i>	<i>Year</i>	<i>Activity Number</i>	<i>Size</i>
HUMANEVA	2009	6 activities	~ 80,000 images
MPII Human Pose	2014	491 activities	40,000 images
Human3.6M	2014	15 activities	3,578,080 frames
MS COCO ¹⁰	2015	-	200,000 images
MPI-INF-3DHP	2016	8 activities	~ 1,300,000 frames
SURREAL	2017	15 activities	6,536,752 frames
3DPW	2018	-	60 video sequences
Physiotherapy Exercise Dataset ¹	2022	14 exercises	1237 video clips
TheraPose (this paper)	2024	123 exercises	3,424,125 frames

Table 1. This table provides an overview of the various publicly available datasets used for human pose estimation.

2 Dataset

In this paper, we introduce a large dataset of images that covers human poses while doing different physiotherapy exercises. The primary rationale behind our data collection strategy was to record the movements often given to patients by doctors during physiotherapy treatment. For this purpose, we worked with experienced

physiotherapists from the Department of Occupational Therapy, Faculty of Health Sciences, Hacettepe University, to create a list of their most preferred exercises, including exercise names, descriptions, duration, angles, and pivot points. We recorded a total of 123 exercises under 5 different main categories:

- Neck and upper extremities stretching and strengthening exercises
- Upper body stretching and strengthening exercises
- Lower body stretching and strengthening exercises
- Lower extremities stretching and strengthening exercises
- Whole body coordination exercises

Data collection. The first phase of our data collection focused on creating an optimal studio environment suitable for our research objectives. Our primary goal was to create a space conducive to the smooth execution of physiotherapy exercises. To this end, we converted a medium-sized room into a dedicated studio by completely covering it with green screens, thus providing a conducive environment for our research efforts.



Fig. 1. Studio Room

For precise capture of IMU data, we purchased the Rokoko Smartsuit Pro II Mocap Suit. Known for its capacity to provide real-time, highly accurate sensor data, this mocap suit played a pivotal role in capturing the complex nuances of physiotherapy movements with exceptional precision. In particular, its capacity to provide fast and accurate data in three axes (x, y, z) at 24 points was a significant factor in our choice of this model. For a detailed visualization of the 24 joint points, please refer to *Figure 3* in the appendix.

Along with IMU data acquisition, we used five Logitech BRIO 500 webcams positioned at different angles for RGB video recording. Each camera is equipped with an LED light source to ensure optimal illumination during recording sessions. In addition, to maintain uniform lighting conditions within the studio, general room lighting was provided by Visico LED 80T spotlights positioned at two different angles. Please see *Figure 2* for a visual representation of the studio layout.

Five cameras and a Rokoko Smartsuit Pro II were interfaced with a single desktop computer to facilitate recording. OBS Studio software was used for video capture, while Rokoko Studio was used for 3D sensor data collection. A custom Python script was developed to synchronize the operation of the five cameras and the mocap suit for simultaneous recording. Video recordings were saved in MP4 format, maintaining a frame rate of 30 fps and a resolution of 800x600 pixels. 3D sensor data was captured in FBX format at a consistent frame rate of 30 fps, aligning with the output specifications of the Rokoko Studio application.

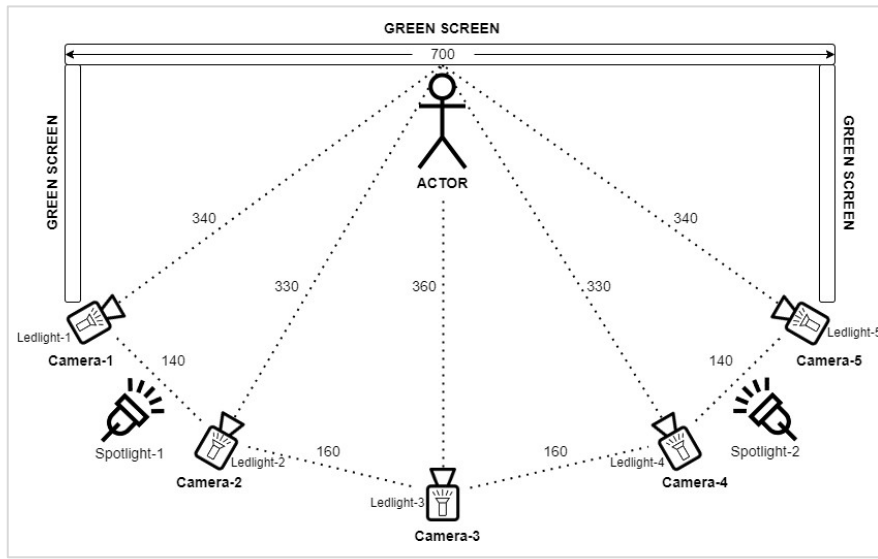


Fig. 2. Studio Environment Diagram: Five cameras were placed in a green screen-covered room at various angles. Five LED lights were above each camera, and two high-power spotlights were used for general room lighting. Therapists performed the exercises in the center of the room. Distances are shown in centimeters.

Two expert physiotherapist actors, one male and one female, from Hacettepe University Faculty of Health Sciences Department of Occupational Therapy performed the selected movements. Five different cameras were used to record each actor's 123 movements. For illustrative purposes, *Figure 5* shows sample recordings capturing different perspectives of the movements.

To enhance sensitivity to various lighting conditions, each exercise session was meticulously recorded under five distinct lighting conditions: *high, medium, low, and variations from left to right*. Moreover, movements were captured in two modes to introduce further diversity: solely with the mocap suit and the mocap suit worn over various clothing options.

<i>Actors</i>	<i>Exercises</i>	<i>Lightning Conditions</i>	<i>Clothing Status</i>	<i>Sessions</i>
Female Actor	123 exercises	x5 lightning cond.	x2 clothing stat.	1230 sessions
Male Actor	123 exercises	x5 lightning cond.	x2 clothing stat.	1230 sessions
				Total 2460 sessions

2460 sessions x 5 different cameras = 12300 videos

Table 2. Overview of Data Collection Across Variables for Physiotherapy Exercise Dataset

In total, 2,460 sessions were recorded, encompassing approximately 50 hours of recording time. Each of the 123 movements was captured in 100 different videos using five different cameras, resulting in 12,300 MP4 files and 2,460 FBX files. The cumulative duration of these recordings amounts to 2,000 minutes.

3 Conclusions

This article introduces the "TheraPose" dataset, a frontier resource in physiotherapy research that aims to optimize the precision and effectiveness of rehabilitation exercises through advanced human pose prediction technologies. "TheraPose" stands out as one of the most specialized datasets in this field, with a comprehensive collection of 3,424,125 frames covering 123 different activities.

The meticulous methodology used in data collection integrates the latest motion capture technology and high-resolution video recording under various conditions, ensuring the relevance and suitability of the data for robust pose prediction model development. These models hold promise in helping physiotherapists monitor and adjust treatment plans more effectively and ensure that exercises are correctly performed to maximize therapeutic outcomes.

In our ongoing work, we seek to improve the accessibility of the "TheraPose" dataset to a wider audience within the research and clinical communities. We aim to provide a representative sample subset of the dataset and thus contribute to future efforts in this field.

Acknowledgements. Tiga Healthcare Technologies carried out this project with support from the Republic of Türkiye Ministry of Industry and Technology TÜBİTAK (The Scientific and Technological Research Council of Türkiye) TEYDEB (Technology and Innovation Funding Programs Directorate) 1501 Industrial R&D Projects Support Program (Project Number: 3221282). We extend our gratitude to the personnel of Tiga Healthcare Technologies for their invaluable contributions to this project. Additionally, we would like to express our appreciation to Prof. Dr. Esra Akı, Assoc. Prof. Dr. Sedef Şahin, Res. Asst. Occupational Therapists Ege Temizkan, Res. Asst. Occupational Therapists Ayşenur Baysal Yiğit from the Department of Occupational Therapy, Faculty of Health Sciences, Hacettepe University, for their dedicated work and significant contributions.

References.

1. K. R. Islam et al., "A Video-based Physiotherapy Exercise Dataset," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 780-784, doi: 10.1109/ICCIT57492.2022.10055189.
2. Miron, A., Sadawi, N., Ismail, W., Hussain, H., & Grosan, C. (2021). IntelliRehabDS (IRDS)—A Dataset of Physical Rehabilitation Movements. *Data*, 6, 46. <https://doi.org/10.3390/data6050046>
3. Yurtman A, Barshan B. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. *Comput Methods Programs Biomed.* 2014 Nov;117(2):189-207. doi: 10.1016/j.cmpb.2014.07.003. Epub 2014 Jul 19. PMID: 25168775.
4. Aguilar-Ortega, R., Berral-Soler, R., Jiménez-Velasco, I., Romero-Ramírez, F. J., García-Marín, M., Zafra-Palma, J., Muñoz-Salinas, R., Medina-Carnicer, R., & Marín-Jiménez, M. J. (2023). UCO Physical Rehabilitation: New Dataset and Study of Human Pose Estimation Methods on Physical Rehabilitation Exercises. *Sensors*, 23, 8862. <https://doi.org/10.3390/s23218862>
5. García-de-Villa, S., Jiménez-Martín, A. & García-Domínguez, J.J. A database of physical therapy exercises with variability of execution collected by wearable sensors. *Sci Data* 9, 266 (2022). <https://doi.org/10.1038/s41597-022-01387-2>
6. von Marcard, T., Henschel, R., Black, M., Rosenhahn, B., Pons-Moll, G.: Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. Springer, Heidelberg (2018).
7. Sigal, L., Balan, A.O. & Black, M.J. HUMANEVA: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion. *Int J Comput Vis* 87, 4–27 (2010). <https://doi.org/10.1007/s11263-009-0273-6>
8. Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014, June). 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
9. Catalin Ionescu, Dragos Papava, Vlad Olaru and Cristian Sminchisescu, Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, No. 7, July 2014.
10. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P. (2014) Microsoft COCO: Common Objects in Context. In: *Computer Vision – ECCV 2014. Lecture Notes in Computer Science*, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48
11. Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., & Theobalt, C. (2017). Monocular 3D Human Pose Estimation in The Wild Using Improved CNN Supervision. In *3D Vision (3DV)*, 2017 Fifth International Conference on IEEE. doi:10.1109/3dv.2017.00064
12. Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., & Schmid, C. (2017). Learning from Synthetic Humans. In *CVPR*.

Appendix.



Fig. 3. Joint Points.

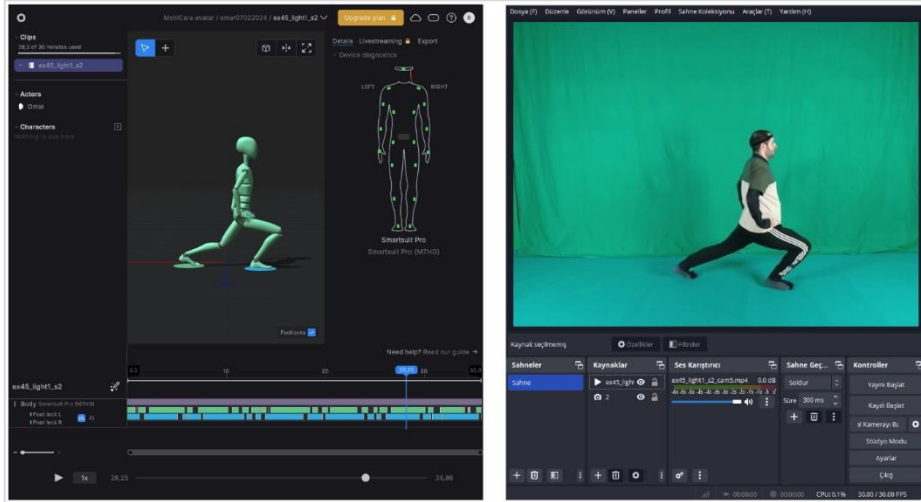


Fig. 4. Recording Apps: Rokoko Studio & OBS Studio.

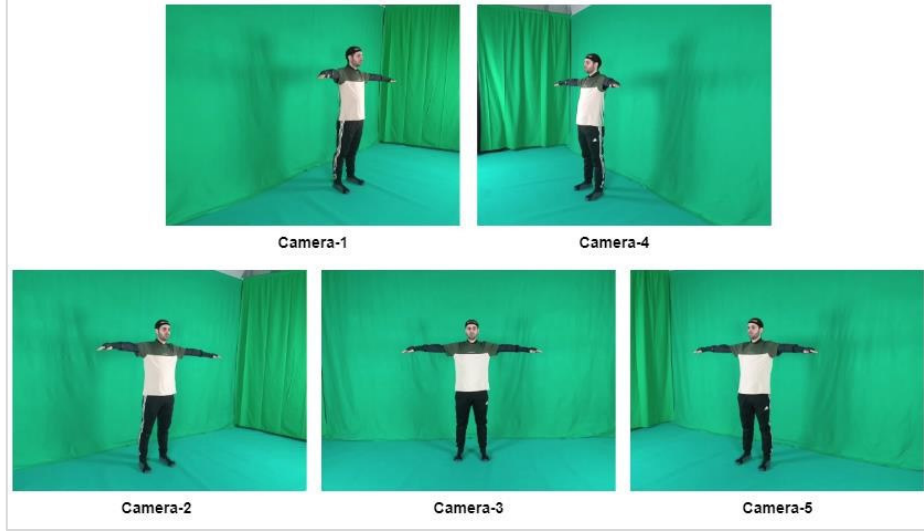


Fig. 5. Five different cameras capture the same movement from various perspectives.

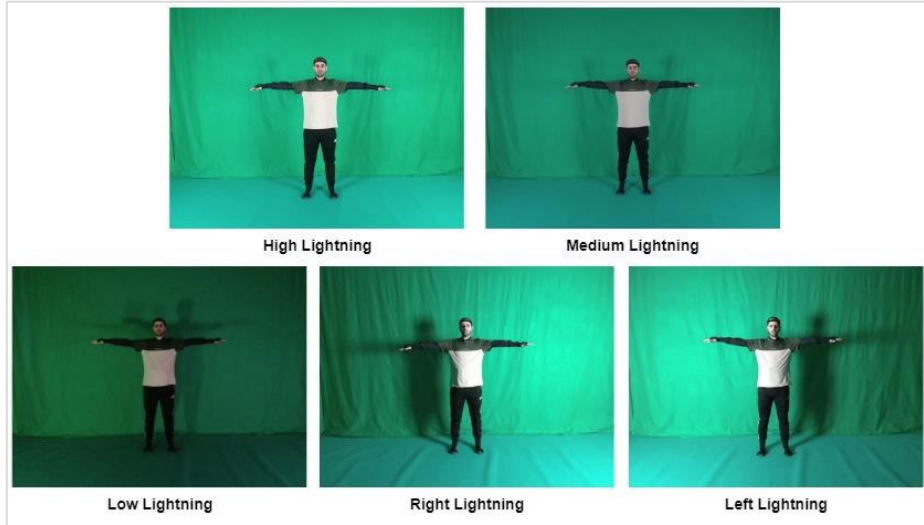


Fig. 6. Different lighting conditions.



Fig. 7. Various exercises with two different actors and clothing under high lighting conditions.

Gamification in Health: A Bibliometric Perspective

Esra Tokur Sonuvar¹[0000-0002-1279-5192] and Kemal Hakan Gülkesen²[0000-0002-2477-2481]

^{1,2} Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University Antalya Türkiye

Abstract. Objective: This study aims to analyze the applications and effects of gamification techniques in the field of health with a bibliometric approach.

Method: The Web of Science database was searched using specific keywords between 2006 and 2024, and the 1,231 articles obtained were analyzed with various parameters in terms of publication year, authors, journals, countries, institutions and keywords. Techniques such as co-word analysis and keyword cluster analysis were used to identify trends and research foci in the literature.

Findings: The analysis showed that gamification is having an increasing impact in the healthcare sector in areas such as behavior change, patient engagement, and physical activity promotion. Seven main themes were identified: Cognitive Health, Rehabilitation Technologies, Health Behaviors, Health Education Games, Youth Mental Health, Mobile Health Monitoring and Interactive Games. The country with the most publications was the USA, the institution was Pen University and the author was Mitesh B. Patel.

Conclusion: This bibliometric analysis, which provides valuable insights into the overall structure and evolution of the literature on gamification and health, provides a basis for developing strategies to further improve the application of gamification in health. This study is designed to help identify potential gaps and opportunities for future research.

Keywords: Gamification, Gamification in Health, Bibliometric Analysis

1 Introduction

Gamification, or the application of gaming elements in non-gaming contexts, has been increasingly utilized in various fields such as education, business, and healthcare in recent years. This approach aims to enhance engagement, boost motivation, and enrich user experience [1]. Particularly in the healthcare sector, gamification techniques hold great potential in increasing the effectiveness of rehabilitation processes by encouraging patients to actively participate in their treatment [2]. The implementation of these techniques has demonstrated positive impacts across a broad spectrum, from managing chronic illnesses to post-surgical recovery. Over the past decade, there has been a notable increase in research publications on gamification in health, with a peak in 2020, driven largely by advancements in digital health technologies and the rising popularity of mobile health applications. Notable contributions include studies on gamification for chronic disease management, which have shown a consistent increase in patient adherence and health outcomes. For instance, the work by Jane McGonigal on SuperBetter has been extensively cited for its innovative approach to mental health treatment through gamified techniques [7]. A study conducted in 2022 by Alfieri et al. evaluated the use of gamification techniques in musculoskeletal rehabilitation. The literature from the past five years was reviewed, examining the effects of gamification on various musculoskeletal conditions such as shoulder surgery, rheumatoid arthritis, and osteoarthritis. The findings indicated that gamification yielded similar or superior results compared to traditional physical therapy or home-based exercises, also significantly improving patient motivation and quality of life. The cost-effectiveness of these technologies has been highlighted as an advantage for gamification in musculoskeletal rehabilitation [6]. Such innovative approaches can enhance patient adherence to treatment in the healthcare sector, thus improving overall recovery rates and sustainability of rehabilitation processes [3].

"The persuasive power of gamification: A systematic review" by Hamari and colleagues underscores the capacity of gamification to change behavior [4]. The study notes that gamification elements, particularly in supporting health-related behaviors, increase user motivation and engagement. These elements foster the intrinsic motivation and interaction necessary for sustaining targeted behaviors, enabling users to take an active role in managing their own health. Applications of gamification in health typically encompass promoting physical activity, diabetes management, and rehabilitation processes. Specifically, in the field of rehabilitation, studies such as that by Ito, K., et al. on EMG-controlled gamified systems support the development of motor skills in specific patient groups and enhance treatment adherence. These systems are designed to ensure patients regularly perform exercises and include enriching user experience features like rewards, progress tracking tools, and elements of competition. Particularly, the opportunities gamification offers to increase patient engagement in their treatment processes and overall recovery rates are significant [5].

The primary rationale for this study is to understand the increasing popularity and significance of using gamification techniques in healthcare and to identify academic trends within this field. This research focuses on how gamification integrates with

health sciences and practices and its role in promoting health-related behavioral changes.

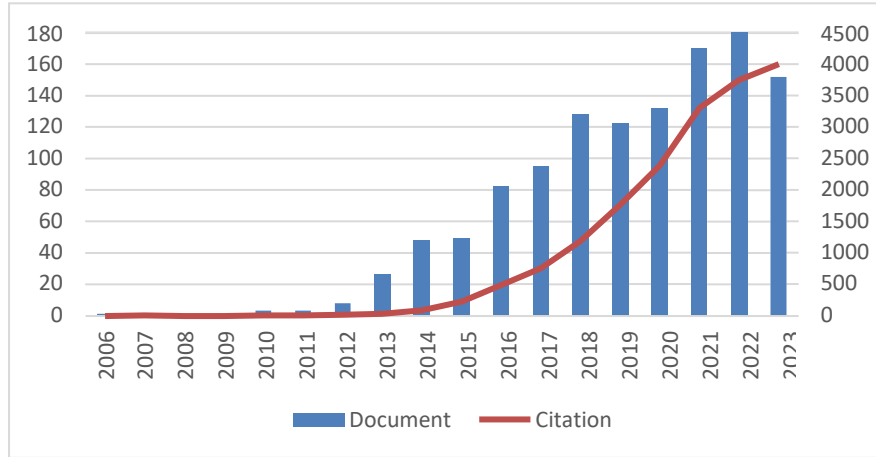
2 Methods

As a method, a comprehensive search was conducted through the Web of Science database using selected keywords up to March 20, 2024. The chosen keywords were "gamification," "game-based learning," "serious games," "health," "healthcare," "medical," "eHealth," "mHealth," "behavioral change," "health behavior," "patient engagement," "user engagement," "physical activity," "exercise," "wellness," and "well-being." Combinations of these terms were utilized to identify publications within the research literature. The analysis of the data involved various parameters such as publication year, authors, journal names, countries, institutions, and keywords. Each publication was evaluated both quantitatively and qualitatively, considering citation counts and the impact levels of the publications. Bibliometric analysis, co-word analysis, and keyword clustering analysis were performed using VOSviewer 1.6.19. Citation analysis identifies the most influential papers by counting the number of citations each paper receives. Co-citation analysis reveals relationships between papers by identifying pairs of papers that are frequently cited together, indicating a thematic connection. Keyword clustering groups related keywords, helping to visualize research trends and thematic areas within the literature. The Web of Science database was chosen for its comprehensive coverage of peer-reviewed journals. However, potential limitations include language biases, as non-English publications might be underrepresented, and the exclusion of certain types of publications such as conference papers and grey literature. These factors could lead to an incomplete representation of the field.

3 Results and Discussion

Between 2006 and 2024, a total of 1,231 academic publications focused on health-oriented gamification were identified. During this period, instead of showing a steady annual increase, the number of publications displayed a fluctuating trend (Figure1). Applications of gamification in the healthcare sector have significantly impacted areas such as behavior change, patient engagement, and the promotion of physical activity.

Figure 1. Trends in the Publication and Citation of Gamification Research (2006-2023).



A co-occurrence analysis revealed seven core themes: Cognitive Health, Rehabilitation Technologies, Health Behaviors, Health Education Games, Mental Health of Youth, Mobile Health Monitoring, and Interactive Games. These themes highlight the focal areas of research and the diversity of gamification within health sciences. The most frequently encountered keywords in the literature are "gamification," "physical activity," and "health," indicating gamification's central role in promoting physical health and activity (Fig.2). A geographic distribution of the studies shows that the USA leads (Fig.3) the research, and Penn University and Mitesh B. Patel are prominent research center and author respectively. The temporal variation in publication and citation numbers reflects how advancements in computer technologies have spurred researchers' interest and innovative work. This increase in gamification studies has played a crucial role in making health services more effective and participatory. One notable example is the use of gamified mobile applications for diabetes management, such as the iCareD app, which supports diabetes self-care by achieving short-term improvements in HbA1c levels [8]. Similarly, "Re-Mission," a video game developed by HopeLab, has been shown to improve treatment adherence in young cancer patients by allowing them to virtually combat cancer cells, thereby increasing their understanding and motivation to adhere to their treatment regimen [9].

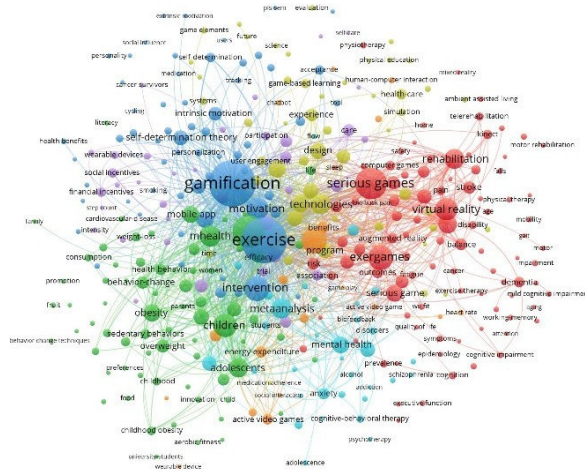


Fig. 2. Co-occurrence Network of Keywords in Gamification Research.

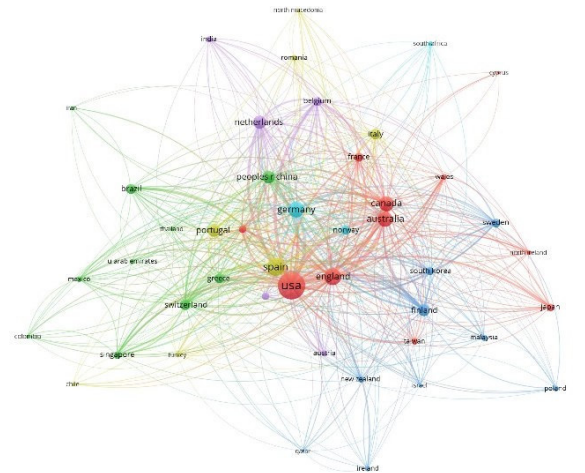


Fig. 3. Global Collaboration Network in Gamification Research.

Our bibliometric analysis has mapped out an extensive overview of the academic literature in the fields of gamification and health, shedding light on the development at the intersection of these two disciplines. The data obtained demonstrate how the impact of gamification on health sciences has increased over time and which areas it has become central to. In-depth analyses have revealed thematic areas where research is

concentrated, particularly in cognitive health, rehabilitation technologies, health behaviors, health education games, mental health of youth, mobile health monitoring, and interactive games.

Bibliometric approaches offer valuable insights but have limitations. They depend on published literature, potentially omitting relevant studies due to language biases or non-indexed journals. Moreover, publication bias may skew results, as studies with positive outcomes are more likely to be published, leading to an overrepresentation of certain viewpoints and neglecting emerging or minority perspectives. It is important to note potential biases in this study. Publication bias, where positive results are more likely to be published, could skew the findings towards favorable outcomes of gamification. Additionally, the predominance of studies from high-income countries might not fully represent the global applicability of gamification techniques. Future research should aim to include diverse geographic and socio-economic contexts to provide a more comprehensive understanding of gamification's impact.

The findings of this study indicate that gamification can play significant roles in promoting behavioral changes in the health sector, enhancing patient and user engagement, and encouraging physical activity. These results underscore the potential of gamification to transform patient engagement and self-management in healthcare. For practitioners, adopting gamified approaches can enhance patient interaction, improve adherence to treatment protocols, and ultimately lead to better health outcomes. Policymakers should consider integrating gamification elements into public health initiatives and reimbursement models to encourage the development and adoption of these innovative tools. Additionally, contributions by specific countries and researchers highlight the global prevalence of gamification and identify leading geographies in this field. This emphasizes the importance of multidisciplinary collaboration and innovation in advancing the use of gamification in health.

References

1. Deterding, S., Dixon, D., Khaled, R., & Nacke, L.: From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15. Springer, Heidelberg (2011).
2. Miller, K. J., Adair, B. S., Pearce, A. J., Said, C. M., Ozanne, E., & Morris, M. M.: Effectiveness of gamification for the engagement and rehabilitation of people with physical disabilities. In: Physical Therapy, vol. 96(9), pp. 1454–1463. Springer, Heidelberg (2016).
3. Brown, M., & Vaughan, C.: Gamification and adherence to treatment: A systematic review. In: Journal of Medical Systems, vol. 44(4), pp. 74. Springer, Heidelberg (2020).
4. Hamari, J., Koivisto, J., & Sarsa, H.: The persuasive power of gamification: A systematic review. In: Information Systems Research, vol. 33(1), pp. 164–195. (2017). DOI: <https://doi.org/10.1016/j.isres.2017.01.007>
5. Ito, K., Uehara, S., Yuasa, A., Ushizawa, K., Tanabe, S., & Otaka, Y.: Gamified exercise for the distal upper extremity in people with post-stroke hemiparesis: feasibility study on subjective perspectives during daily continuous training. In: Annals of Medicine, vol. 56(1), pp. 2306905. (2024). DOI: 10.1080/07853890.2024.2306905
6. Alfieri, F. M., da Silva Dias, C., de Oliveira, N. C., & Battistella, L. R. (2022). Gamification in Musculoskeletal Rehabilitation. *Current reviews in musculoskeletal medicine*, 15(6),629–636. <https://doi.org/10.1007/s12178-022-09797-w>
7. McGonigal, J. (2015). SuperBetter: The power of living gamefully. Penguin Books.
8. Lee, E. Y., Cha, S. A., Yun, J. S., Lim, S. Y., Lee, J. H., Ahn, Y. B., Yoon, K. H., Hyun, M.K., & Ko, S. H. (2022). Efficacy of Personalized Diabetes Self-care Using an Electronic Medical Record-Integrated Mobile App in Patients With Type 2 Diabetes: 6-Month Randomized Controlled Trial. *Journal of medical Internet research*, 24(7), e37430. <https://doi.org/10.2196/37430>.
9. Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H. (2008). A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. *Pediatrics*, 122(2), e305–e317. <https://doi.org/10.1542/peds.2007-3134>.

DMFT ASSESSMENT USING DEEP LEARNING BASED CONVOLUTIONAL NEURAL NETWORK ALGORITHM WITH PANORAMIC IMAGES

Onur MUTLU ¹ [0009-0004-6204-5535], Elif ASLAN ² [0000-0001-7609-999X],

Canberk ULUSOY ² [0000-0003-0899-9368], Ali MERT ³ [0000-0002-6806-935X]

¹ KTU, Faculty of Science, Department of Computer Sciences, Trabzon, TURKEY

² EGE, Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, İzmir,
TURKEY

³ EGE, Faculty of Science, Department of Statistics, İzmir, TURKEY

Abstract. This study addresses the automatic assessment of tooth detection and DMFT index using panoramic radiography images in dentistry. A total of 1355 anonymous panoramic radiography images obtained from the Kodak 8000 digital panoramic system were examined using a deep learning-based algorithm with the YOLOv5x6 model. This algorithm achieved a high success rate of 97.45% for tooth detection and accurately identified fillings and caries associated with the DMFT index with 75.5% accuracy. The research findings demonstrate that artificial intelligence enabled systems can save time for dentists and optimize treatment planning processes.

Keywords: Artificial intelligence, Deep learning, Tooth detection, Dental imaging, DMFT index

1 Introduction

Machine learning (ML) is a subfield of artificial intelligence (AI) that enables computers to learn automatically from large datasets [5]. ML, along with its subclass deep learning (DL)-based methods, has been started to be used in various fields of medicine and dentistry. Numerous studies in dentistry have highlighted the potential of DL methods based on Convolutional Neural Networks (CNNs), which allow medical images to be directly processed by algorithms, yielding promising results in the diagnosis and treatment planning of dental pathologies. In the healthcare sector, where digitalization efforts have gained momentum in recent years, integrating radiographic assessment with AI is believed to offer significant advantages for achieving more accurate and effective diagnosis and treatment planning [2]. Particularly in dental practice, the use of such technologies provides an important advantage in accurately numbering the permanent and/or deciduous teeth [6], detecting of periapical lesions [11], planning dental implants [12, 4], detecting of restored and decayed teeth [12, 7, 13], and developing effective dental treatments.

Tooth decay is one of the most common global chronic diseases [8] and despite the development of various preventive or therapeutic methods, tooth decay and associated tooth loss remain significant global issues [1]. Accordingly, the use of community-based strategies such as globally accepted decay indices becomes imperative for the effective control of tooth decay.

Panoramic radiography (PR) is the most commonly used radiographic imaging method in dentistry [3]. PR offers specific advantages in assessing the DMFT (Decayed, Missing, and Filled Tooth) index, one of the most widely used decay indices worldwide, due to its ability to show the entire dentition and surrounding structures. However, it is noteworthy that the applicability of CNN algorithms for evaluating the DMFT index using PR images has not been investigated.

This study aimed to investigate the feasibility of panoramic radiography (PR) images for detecting and numbering teeth, identifying fillings and decay in the detected teeth, and thereby automatically evaluating the DMFT index in adult individuals. Assessing the ability of CNN algorithms to accurately predict the DMFT index from PR image data would represent a significant advancement in dental practice. Accordingly, the goal of this study was to provide dentists with a faster, more accurate, and more objective tool for assessing decay, and thereby helping optimize patient treatment planning.

2 Material and Method

2.1 Dataset

A total of 1355 digital PR images of patients aged 18 years and older were randomly selected from the image archive of the Department of Oral and Maxillofacial Radiology at Ege University Faculty of Dentistry. PR images had been obtained using a Kodak 8000 digital panoramic system (Kodak 8000 Digital Panoramic System, Carestream Health Inc., Trophy, France) with exposure settings of 68kV, 8mA, and 13.9 s. The images had dimensions of 2943x1435 pixels. The selected PR images were anonymized and saved in JPEG format. Images of edentulous patients and, patients with removable prostheses, retained deciduous teeth, jawbone pathologies, and images containing significant artifacts were excluded.

To evaluate the generalization ability of the model, it was crucial to randomly divide the dataset into training, validation, and test sets. In this regard, 929 PR images from the training set were used to establish the foundation of the model, while images from the validation and test sets were employed to objectively assess the model's performance. This method aimed to better understand how the model would respond to real-world data in the field of oral, dental, and maxillofacial radiology applications.

2.2 Image Analysis

A total of 1355 PR images were evaluated by two oral, and maxillofacial radiologists with different levels of experience, according to the recommendations of the World Health Organization (WHO) regarding the DMFT (Decayed, Missing, Filled Teeth) index (WHO Oral Health Surveys, 1997). Teeth were segmented into rectangular shapes using labels such as "decayed," "filled," and "missing," along with their corresponding tooth numbers (Description 1). The labeling process was performed using software called Colabeler developed by MacGenius, Blaze Software, CA.

Healthy, decayed, and filled teeth were selected and labeled by drawing bounding boxes around them. Missing teeth were predicted by the algorithm based on the position, tilt, and arrangement of neighboring teeth. Detection and numbering of healthy, decayed, and filled teeth were performed using the YOLOv5x6 (You Only Look Once) model. JSON and XML formats were used for labeling. The label files kept in XML format were converted to TXT format to enable the YOLOv5x6 algorithm to function.

Before each training session, PR images were resized from their original dimensions to 1280x1280 pixels to better align with the applied method.

Description 1 – WHO Recommendations

- Teeth with no evidence of treated or untreated decay were considered "Healthy."
- Teeth with primary decay, secondary decay, or teeth with crowns or roots destroyed due to decay were considered "Decayed."
- Teeth with one or more permanent restorations and no decay on different surfaces, as well as teeth with fixed prosthetic restorations such as crowns/bridges, were considered "Filled."

2.3 Data Augmentation

Considering that a human mouth contains a total of 32 teeth and each tooth can be in 3 different states, it was anticipated that the original dataset of 929 training PR images would be insufficient and could lead to overfitting. To prevent this overfitting, brightness of the PR images was adjusted between -44% and +44%. This was aimed at mitigating density and contrast changes that may occur during PR imaging.

In the second step, a total of 4% noise pixels were randomly added to the data. This effectively addressed minimal motion artifacts and patient positioning errors. Using these data augmentation techniques, 4645 PR images were generated from the original 929 training data.

2.4 Deep Convolutional Neural Network

The open-source Python programming language (version 3.11.2) was used to create our model. For building a model based on tooth detection, we chose the Yolov5x6, which is based on the Torch library and is a variant of the YOLO (You Only Look Once) algorithm capable of both detecting and classifying objects. The Yolov5x6 model consists of three main parts: backbone, neck, and output. The backbone focuses on feature extraction from input images, the neck combines extracted features and generates three different scaled feature maps, and the output section uses these feature maps to detect objects [9]. We preferred Yolov5x6 in our study because it delivers faster and more accurate results in object detection on the training dataset compared to other algorithms [10].

The model training was conducted on a computer with a Ryzen 9 5900x processor, 64 GB RAM, NVIDIA GeForce RTX 3090 graphics card, and running Windows 10 operating system. Stochastic Gradient Descent (SGD) optimization algorithm was used for training, with an initial learning rate set to 0.01.

3 Result

In this study, we explored the feasibility of utilizing a deep learning-based (CNN) algorithm, specifically the YOLOv5x6 model, for the automatic assessment of the DMFT index using PR images. The study demonstrated that our deep learning model could achieve a high success rate of 97.45% in tooth detection and an accuracy of 75.5% in identifying fillings and caries associated with the DMFT index. These results point out the potential of AI-enabled systems to significantly enhance dental diagnostics by providing a fast, reliable, and objective tool for evaluating dental health. The successful implementation of our model highlights the advantages of integrating AI in dental practice, offering substantial time savings for dentists and aiding in the optimization of treatment planning processes. By automating the detection and classification of dental pathologies, our system can improve diagnostic accuracy and ensure

more consistent assessments, which are crucial for effective treatment and patient care.

4 Recommendations

Building on the promising results of this study, we propose several recommendations for future research and clinical implementation to maximize the potential of AI in dental practice:

Dataset Expansion: Future studies should aim to include larger and more diverse datasets. Specifically, incorporating images of filled and decayed teeth from areas where the incidence of cavities is relatively lower compared to other dental regions (such as teeth numbered 31, 32, 41, and 42) into the training set may enhance the overall performance of the algorithm.

Model Enhancement: Continuous improvement of the YOLOv5x6 model through advanced data augmentation techniques and meticulous hyperparameter tuning can further elevate its accuracy in detecting dental caries and fillings.

Clinical Validation: Comprehensive validation is of paramount importance in clinical dental practice. Collaborating with dentists to collect feedback and refine the system will ensure its practical effectiveness and utility.

Multimodal Imaging Integration: Exploring the combination of panoramic radiography with other imaging techniques, such as intraoral scans could provide a more comprehensive assessment of dental health and improve the accuracy of the DMFT index evaluation.

By addressing these recommendations, future research can continue to advance the field of dental radiography, leading to improved patient outcomes and more efficient, accurate dental care practices.

The model training was conducted on a computer with a Ryzen 9 5900x processor, 64 GB RAM, NVIDIA GeForce RTX 3090 graphics card, and running Windows 10 operating system. Stochastic Gradient Descent (SGD) optimization algorithm was used for training, with an initial learning rate set to 0.01.

References

1. Baumeister, S. H., Freeman, G. J., Dranoff, G. & Sharpe, A. H. Coinhibitory pathways in immunotherapy for cancer. *Annu. Rev. Immunol.* 34, 539–573 (2016).
2. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* 12, 252–264 (2012).
3. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* 12, 492–499 (2011).
4. Larkin, J. et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* 373, 23–34 (2015).
5. Shepard MT, Bonney EA. PD-1 Regulates T cell proliferation in a tissue and subset-specific manner during normal mouse pregnancy. *Immunol Invest.* (2013) 42:385–408. doi: 10.3109/08820139.2013.782317
6. Matzinger P. Friendly and dangerous signals: is the tissue in control, *Nat Immunol.* (2007) 8:11–3. doi: 10.1038/ni0107-11
7. Meggyes M, Szanto J, Lajko A, Farkas B, Varnagy A, Tamas P, et al. The possible role of CD8+V α 7.2+/CD161++ T (MAIT) and CD8+V α 7.2+/CD161 lo T (MAIT-like) cells in the pathogenesis of early-onset pre-eclampsia. *Am J Reprod Immunol.* (2018) 79:e12805. doi: 10.1111/aji.12805

8. Jiang Y, Li Y, Zhu B. T-cell exhaustion in the tumor microenvironment. *Cell Death Dis.* (2015) 6:e1792. doi: 10.1038/cddis.2015.162
9. Dai S, Jia R, Zhang X, Fang Q, Huang L. The PD-1/PD-Ls pathway and autoimmune diseases. *Cell Immunol.* (2014) 290:72–9. doi: 10.1016/j.cellimm.2014.05.006
10. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* 12, 492–499 (2011).
11. Wykes, M., Lewin, S. Immune checkpoint blockade in infectious diseases. *Nat Rev Immunol* 18, 91–104 (2018).

Identifying Malignancy Risk Factors in Thyroid Nodules: A Machine Learning Approach in Trabzon, Türkiye

Zeliha Aydın Kasap¹[0000-0002-5823-100X], Burçin Kurt¹[0000-0001-5781-2382], Ali Güner²[0000-0001-7801-8081], Elif Özsağır³[0000-0002-9851-9227], Mustafa Emre Ercin⁴[0000-0002-7340-8045]

¹ Department of Biostatistics and Medical Informatics, Graduate School of Medicine Science, Karadeniz Technical University, Trabzon, Türkiye

² Department of General Surgery, Faculty of Medicine, Karadeniz Technical University, Trabzon, Türkiye

³ Bingöl Provincial Health Directorate, Bingöl State Hospital, Bingöl, Türkiye

⁴ T.R. Ministry of Health Ankara Atatürk Sanatorium Education and Research Hospital, Ankara, Türkiye

zelihaaydin86@gmail.com

Abstract

The main aim of this study is to identify risk factors using machine learning algorithms preoperatively and prevent unnecessary thyroid surgeries by clinical and pathological variables as well as simple predictors in the Trabzon thyroid cancer endemic region. In this retrospective cohort study, the dataset consists of 40 variables and 204 nodules belonging to 183 patients who were diagnosed with thyroid nodules malignancy that classified Bethesda Category III (AUS) cytology. To evaluate the surgical status of the nodules, blood tests, USG information, cytopathological, and morphometric variables were used.

The machine learning methods were implemented using open-source R programming. The risk factors for malignancy in thyroid nodules were identified using K-nearest neighbor (KNN) and Random Forest (RF) algorithms. In an analysis of thyroid cancer prediction models using clinical, pathological, and morphometric data, Blood tests, USG information, cytopathological, and morphometric variables were used to evaluate the nodules' surgical status k-nearest neighbor (KNN) model outperformed others with a 0.78 AUC and 0.79 accuracies. The random forest (RF) model achieved 0.73 AUC and 0.74 accuracy. Utilizing biomarkers found in cytopathology slides can offer a cost-effective alternative for primary care diagnosis. Furthermore, the developed machine-learning models can provide decision-making support for clinicians, guiding the implementation of preventive measures before surgery.

Keywords: Thyroid nodule, malignancy, Bethesda III, machine learning, k-nearest neighbors, random forest.

Klinik Metinlerden İlişki Çıkarma için Bağlam İçi Öğrenme

In-Context Learning for Relation Extraction from Clinical Texts

Aidana Aidynkyzy¹[0009-0003-3575-4662], Şebnem Bora¹[0000-0003-0111-4635], Oğuz Dikenelli¹[0000-0001-7948-7453]

¹ Ege Üniversitesi, Bilgisayar Mühendisliği Bölümü, Bornova İzmir 35100, Türkiye
aidynkyzy11@gmail.com

Özet. GPT-3 gibi Büyük Dil Modelleri (BDM'leri), birçok iyi bilinen doğal dil işleme probleminde güçlü bağlam içi öğrenme (BİÖ) yetenekleri göstermiştir ve önemli ve çeşitli dil teknolojisi gereksinimlerin yanı sıra veri anotasyonuna ilişkin önemli maliyetleri beraberinde getiren klinik alanda ilişki çıkarma gibi görevler için özellikle çekicidir. Bu makalede, Wan vd. (2023) tarafından önerildiği gibi dinamik örnek alma ile bağlam içi öğrenmenin GPT-3 kullanarak klinik metinlerden ilişki çıkarma için kapsamlı bir analizi sunulmaktadır ve sonuçları ince ayarlı BERT, BioBERT ve PURE modelleriyle karşılaştırılmaktadır. Sonuçlarımız, dinamik örnek seçim yöntemine sahip GPT-3 BİÖ'nin PURE, konuyu, nesneyi ve bunların ilgili türlerini vurgulamak için belirlenmiş işaretçileri giriş katmanına yerleştiren Önceden Eğitilmiş Dil Modeline (ÖEDM) kıyasla %3,57 F1 puanı kadar düşük performans sergilediğini göstermektedir. Daha ayrıntılı analiz, en yüksek benzerliğe sahip az sayıda örnek kullanılmasının yanı sıra veri kümesindeki ilişki türlerinin eşit olmayan dağılımı ve seçilen test kümesinin az sayıda kullanılmasının düşük performansa neden olabileceğini göstermektedir. Çalışmamızın, BİÖ ile klinik ilişki çıkarma için yeni ve etkili tekniklerin kullanılması konusunda araştırmacılara ve uygulayıcılara faydalı olacağını umuyoruz.

Anahtar Kelimeler: Klinik İlişki Çıkarma, Bağlam İçi Öğrenme, Büyük Dil Modeli.

Abstract. Large Language Models (LLMs) such as GPT-3 have shown strong in-context learning capabilities in many well-known natural language problems and are especially attractive for tasks such as relation extraction in clinical domain, which presents significant and varied language technology requirements along with substantial costs associated with data annotation. In this paper, we present a comprehensive analysis of in-context learning with dynamic example retrieval as proposed by Wan et al. (2023) for relation extraction from clinical

texts using GPT-3. The results are then compared with fine-tuned BERT, Bi-oBERT and PURE models. Our results show that, GPT-3 ICL with dynamic example selection method underperforms by 3.57% F1-score compared to PURE, Pre-trained Language Model (PLM) that embeds designated markers into the input layer to emphasize the subject, object, and their respective types. More in-depth analysis shows that using small number of examples with highest similarity along with uneven distribution of the relation types in the dataset and small number of chosen test set could cause the poor performance. We hope that our study will be beneficial for researchers and practitioners towards using novel and efficient techniques for clinical relation extraction with ICL.

Keywords: Clinical Relation Extraction, In-Context Learning, Large Language Model.

1 Giriş

Sağlık hizmetlerinin hızla geliştiği ortamda, klinik metinlerden anlamlı bilgi elde etmek, tıbbi araştırmaları ilerletmek, hasta bakımını geliştirmek ve klinik karar vermeyi desteklemek açısından çok önemlidir. Bu alanda, metinde bahsedilen varlıklar arasındaki anlamsal ilişkileri tanımlama ve kategorize etme görevi olan ilişki çıkarma çok önemli bir rol oynar. Özellikle günlük çok miktarda yapılandırılmamış verinin üretildiği klinik metinler bağlamında, bu metinlere gömülü değerli bilgilerin ortaya çıkarılması için doğru ve verimli ilişki çıkarımı esastır.

Son yıllarda Transformerlerdeki gelişmeler (Vaswani vd., 2017) ve önceden eğitilmiş dil modelleri (ÖEDM'leri) (Devlin vd., 2019) tıbbi metinlerden ilişki çıkarmanın (İÇ) performansını önemli ölçüde iyileştirmiş ve en son teknolojiye sahip İÇ'yi (Lee vd., 2019; Gu vd., 2021) yeni bir seviyeye taşımıştır. ÖEDM'ler genellikle BERT (Devlin vd., 2019), ELMo (Peters vd., 2018), RoBERTa (Liu vd., 2019) ve SpanBERT (Joshi vd., 2020) gibi büyük ölçekli bir korpus üzerinde eğitilir, özünde yardımcı temsilleri (embeddings) (ör. konum ve segment gömmeleri) içerir. Son zamanlarda önceden eğitmenin genişlemesi, modelin parametrelerini güncellemeden, eğitilmiş modele girdinin bir parçası olarak sağlanan bağlam içinde sadece birkaç eğitim örneğinden Doğal Dil İşleme (DDİ) görevlerini öğrenmeleri için büyük dil modellerini (BDM'leri) güçlendirmiştir (Brown vd., 2020).

Bağlam İçi İlişki Çıkarma, yalnızca birkaç eğitim örneğiyle örnekleri yeni ilişkilere sınıflandırabilen bir model geliştirmeyi amaçlamaktadır. Önceden eğitilmiş büyük dil modellerinin birkaç-atış biçiminde özelleştirilmesini sağlayan bağlam içi öğrenme (BIÖ) stratejileri, ilişki çıkarma sistemlerinin inşasına yeni perspektifler açmaktadır.

Lu vd. (2021), hem istemlerdeki örnek seçiminin hem de istemlerin sırasının sonuçları önemli ölçüde etkileyebileceğini göstermiştir. ÖEDM'nin doğru yanıtı seçmesi gereken örnekleri seçmek için (örnek örnekleme), Gao vd. (2020) ve Liu vd. (2021) girdiye anlamsal olarak yakın örnekler bulmak için cümle temsillerini (sentence embeddings) kullanmıştır. Mishra vd. (2021) ise hem olumlu hem de olumsuz örnekler kullanarak, ÖEDM'e yalnızca komutlarla yeni görevleri yerine getirirken kaçınılması gereken öge türlerini öğretmiştir.

Bu makale, klinik metinlerdeki örnek seçmenin önemini ve kullanılabilirliğini araştırmak için, ilişki çıkarma konusunda GPT-3'ün bağlam içi öğrenme ve ince ayarlı BERT boyutlu ÖEDM'nin karşılaştırmalı çalışmasını vermeyi amaçlamaktadır. Ayrıca, GPT-3'ün bağlam içi öğrenme performansı, çok sayıda yeni teknikten yararlanılarak biyomedikal bilgi çıkarma için optimize edilmiştir. Öncelikle, Wan vd. (2023) çalışmasından esinlenerek, doğal olarak varlıklara ve ilişkilere vurgu yapan ince ayarlı ilişki temsili alma yaklaşımını kullanılmıştır. İkinci olarak, Liu vd. (2021) çalışmasına benzer şekilde, bağlam içi örnek alma için bir k-en yakın komşu (kNN) modülü sunulmuştur.

Deneylerimiz, eğitim verilerinin elde edilebildiği durumlarda tam denetimli ince ayarlı ÖEDM'lerin hala altın standart olmasına rağmen, BİÖ için doğru örneklerin seçilmesinin, klinik ilişki çıkarma için BİÖ'nin etkinliğini artırmada uzun bir yol kat edebileceğini göstermektedir. Bununla birlikte, BERT-base ve BioBERT-base modellerinde %11.02'lik bir iyileştirme potansiyeli göstermektedir. Bu makale sadece 5-atış benzer örnekler ile GPT-3 BİÖ'nin ince ayar yapılmış ÖEDM'den daha başarılı olduğunu göstermektedir. Ayrıca, çok sayıda gösteri kullanmanın modelin performansını daha da iyileştirebileceğini ortaya koymaktadır. Son olarak, kapsamlı deneylerimiz, klinik ilişki çıkarma konusunda çalışan NLP araştırmacıları ve uygulayıcıları için bazı öneriler sunmaktadır.

2 İlgili Çalışma

Potansiyel biyomedikal bilgi çıkarma uygulamalarının çeşitliliği, biyomedikal ek açıklamaların yüksek maliyeti ve model eğitiminin karmaşıklığı, bağlam içi öğrenmeyi özellikle biyomedikal uygulamalar için çekici kılmaktadır. Gutiérrez vd. (2022), GPT-3'ün biyomedikal bilgi çıkarma için bağlam içi öğrenimini keşfetmek amacıyla 100 eğitim örneği kullanmıştır ve GPT-3'ün birkaç-atış ortamında bazı biyomedikal İÇ veri seti üzerinde PubMedBERT, BioBERT-large ve ROBERTa-large'a göre düşük performans gösterdiğini keşfetmiştir. Ayrıca, Wan vd. (2023), GPT-3 BİÖ'nin İÇ'de tuzağa düşmesine neden olan sebeplerden birinin, BİÖ için alınan gösterilerdeki varlık ve ilişki ile ilgili düşük alaka düzeyi olduğunu bildirmiştir. Gösteriler, cümle temsiline dayalı olarak rastgele veya k-en yakın komşu (kNN) gibi metodlar aracılığıyla seçilebilmektedir. Son çalışmalar (Gutiérrez vd., 2022; Liu vd., 2021), eğitim setindeki en benzer örnekleri, her test girdisi için az sayıdaki gösteriler olarak almak için kNN'yi kullanmıştır. Fakat, cümle temsiline dayalı kNN alımı, içerdiği belirli varlıklar ve ilişkiler kadar değil, genel cümle anlambiliminin alaka düzeyiyle daha fazla ilgilidir ve bu da düşük kaliteli gösterilere yol açmaktadır. Bu nedenle Wan vd. (2023), gösterilerin alınması için ince ayarlı bir İÇ modelinden ilişki temsili çıkarmak için yeni bir yaklaşım önermişti ve en gelişmiş sonuçları elde etmişti.

3 Yöntemler

Bu bölümde, klinik metinlerden ilişki çıkarma için örnek seçim yöntemiyle BİÖ'yi tanımlıyoruz. İlk olarak, Bölüm 3.1'de İÇ'yı bir dil oluşturma görevi olarak resmileştirmek için istem yapısını açıklıyoruz ve ardından Bölüm 3.2'de örnek seçim yöntemini tanımlıyoruz.

3.1 İstem Yapısı

GPT-3 modeline beslenen her bir test örneği için bir istem oluşturuyoruz. Her istem aşağıdaki bileşenlerden oluşur:

Görev Talimatları. Görev talimatı, ilişki çıkarma görevinin tanımını, veri kümesindeki ilişki türlerini ve bunların kısa açıklamalarını içerir. Daha sonra modelden önceden tanımlanmış sınıflara ait olan ilişkiyi çıkartması açıkça istenir. Bu talimat tüm test girişleri için aynıdır.

Birkaç-atış Örnekleri. Görev açıklamasından sonra gösteri için k-atış örnekleri yerleştirilmiştir. Örnekler, kNN retriever tarafından belirlenen eğitim setinden seçilmiştir. Her örnek, cümlesi ve cümlede görünen varlıklar arasındaki ilişki hakkında bir sorusu olan bağlamdan ve o soruya cevabı içeren yanıtta oluşmaktadır.

Test Girdisi. Örneklere benzer şekilde, test girdisini bağlamda sunulmaktadır ve GPT-3'ün karşılık gelen ilişkiyi oluşturması ve yanıt olarak çıkarması beklenmektedir.

3.2 Bağlam İçi Öğrenme için Örnek Seçme

Liu vd. (2021) takiben, eğitim setimizdeki en benzer örnekleri, her test örneği için bağlam içi birkaç-atış istemi olarak seçmek için bir k-en yakın komşu (kNN) alma modülü kullanıyoruz. kNN, eğitim setindeki hem test girdisini hem de örnekleri kodlamak için temsil alanının seçimine güvendiğinden, önceden eğitilmiş dil modellerini veya diğer geliştirilmiş cümle temsilini kullanarak cümle temsilini elde etmeyi önerirler. Ancak, Wan vd. (2023) kNN alımı için cümle temsilinin kullanılmasının ciddi bir dezavantajı olduğunu vurgulamıştı: ilişki çıkarma, tüm cümlelerin anlamsal anlamından ayrılan ve cümle temsilini kullanarak belirsiz birerışıme yol açan ikili varlıklara odaklanmaktadır. Bu amaçla, özne ve nesne varlıklarını ve türlerini vurgulamak için fazladan işaretleyici belirteçleri ekleyerek hem bağlam bilgilerini hem de varlık bilgilerini yakalamaya çalıştıklarından dolayı, gösterileri almak için ince ayarlı bir İÇ modelden ilişki temsili çıkarmayı önermektedirler.

4 Deneyler

4.1 Veri Kümesi

Klinik metinlerden ilişki çıkarma için örnek seçim yöntemi ile bağlam içi öğrenmeyi değerlendirmek için i2b2-2010 (Uzuner vd. 2011) veri setini kullanıyoruz. Veri seti, Partners Healthcare, Beth Israel Deaconess Tıp Merkezi ve Pittsburgh Üniversitesi Tıp Merkezi olmak üzere üç farklı hastaneden toplanmıştır. Hastaların taburcu özeti ve gelişme notlarından oluşmuştur ve tıp pratisyenleri tarafından manuel olarak açıklanmıştır. Veri kümesi i2b2 web sitesinden indirilmektedir. Sahu vd.'de (2016) belirtildiği gibi, veri setinde eğitim için sadece 170 belge, test için 256 belge ve toplam 426 belge 9413 satır (ilişki) olup deneyimiz için 300 ilişkiyi test seti kalan 9113 ilişkiyi eğitim seti olarak almaktayız. Tablo 1, i2b2-2010 veri kümesindeki örnek ek açıklama biçimini göstermektedir. Veri kümesi 8 ilişkiden oluşmaktadır (Tablo 2).

Tablo 1. i2b2-2010 veri kümesinde örnek açıklama biçimi.

Cümle	67 y/o male with worsening shortness of breath. Had abnormal ETT and referred for cath.
Kavram	c="cath" 19:14 19:14 t="test" c="abnormal ett" 19:9 19:10 t="problem" c="worsening shortness of breath." 19:4 19:7 t="problem"
İlişki	c="cath" 19:14 19:14 r="TeCP" c="abnormal ett" 19:9 19:10

Tablo 2. i2b2-2010 veri kümesindeki ilişkilerin istatistikleri.

İlişki	Açıklama	Sayı
TrIP	Tedavi tıbbi sorunu iyileştirir	203
TrWP	Tedavi tıbbi sorunu kötüleştirir	133
TrCP	Tedavi tıbbi soruna neden olur	526
TrAP	Tıbbi problem için tedavi uygulanır	2616
TrNAP	Tıbbi sorun nedeniyle tedavi uygulanmamaktadır	174
TeRP	Test tıbbi sorunu ortaya koyuyor	3051
TeCP	Tıbbi sorunu araştırmak için yapılan test	504
PIP	Tıbbi sorun tıbbi sorunu gösterir	2203

4.2 Temel Yöntemler

Deneylerimizde, her bir test örneğinin bağlam içi gösterisinin aynı 9113 eğitim örneğinden (test seti için seçilmiş 300 örnek hariç) alındığı GPT-3 bağlam içi öğrenme ile, aynı eğitim setiyle ince ayarlanmış, önceden eğitilmiş iki dil modelini, BERT-base (Devlin vd., 2019) ve BioBERT-base'i (Lee vd., 2019) karşılaştırıyoruz. BERT-base, maskelenmiş dil modelleme (MLM) hedefi kullanılarak İngilizce dili üzerine önceden eğitilmiş bir modeldir. BioBERT, PubMed makaleleri topluluğu üzerinde önceden eğitilmiştir ve bir BERT kontrol noktasından başlatılmıştır.

Deneyimizde, Wan vd.'nin (2023) ardından, alım için gösteri temsillerini elde etmek için varlık işaretleyici tabanlı ince ayarlı bir model olan PURE'ü (Zhong ve Chen, 2021) seçiyoruz. PURE, yazılan varlık işaretleyicileri ekleyerek her varlık çiftini bağımsız olarak dikkate alan ve her çift için ilişki türünü tahmin eden bir ilişki modelidir. Bu arada, PURE doğrudan karşılaştırılabilir bir temel yöntem olarak performans göstermektedir.

Cümle temsillerine ve mesafe ölçümlerine (Öklid veya Kosinüs Benzerliği) dayalı olarak test örneğine yakın olan k örneklerini seçmek için kNN tabanlı bir yöntem olan KATE (Lui vd., 2021), test kümesindeki her cümle için eğitim veri kümesinden örnekler seçmek için takip edilmektedir.

4.3 Uygulama Detayları

Temel yöntem için ÖEDM'leri BERT ve BioBERT, Hugging Face (Wolf vd., 2020) kütüphanesi kullanılarak ince ayarlanmıştır. İnce ayarlı ÖEDM'de kullanılan hiperparametreler parti boyutu (batch size) 32, maksimum uzunluk (maximum sequence length) 128, öğrenme hızı $1e-5$ ve eğitim döngüsü (epoch) 10 olarak belirlenmiş ve deneyler tek bir NVIDIA GeForce MX250 GPU üzerinde yapılmıştır.

BİÖ örnek seçimi için yapılan deneylerde GPT-3.5 (gpt-3.5-turbo-instruct) kullanımının maliyeti nedeniyle her test cümlesi için eğitim setinden en fazla 10 bağlam içi örnek seçilmiştir. Sıfır-atış temel yöntemi ve BİÖ için tüm deneyler test seti için seçilmiş 300 cümleyi kullanarak çalıştırılmaktadır. KATE için ilişki temsillerini elde etmek için varlık işaretleyici tabanlı ince ayarlı bir model olan PURE'ü takiben bert-base-uncased (Devlin vd., 2019) kullanılmış, ayrıca makalelerindeki hiperparametreleri takip edilmiştir. Her test cümlesi için kosinüs benzerlik metriği kullanılarak temsiller arasındaki mesafe hesaplanmıştır ve KATE'deki mesafe puanlarına göre benzer k cümleleri seçilmiştir.

5 Deney Sonuçları ve Tartışma

5.1 Sonuçlar

Deney sonuçlarımız Tablo 3'te, ilişki çıkarma için i2b2-2010 veri kümesinden test seti olarak rastgele seçilen 300 cümle için verilmiştir. GPT-RE_FT, Wan vd. (2023) tarafından önerilen alım için ince ayarlı ilişki temsilini kullanan modeli belirtmektedir. Tablodan şunlar gözlemlenmektedir: 10-atış ayarında GPT-RE_FT, diğer 1-atış ve 5-atış ayarlarından +%22.89, +%5.49 F1 ile daha iyi performans göstererek daha fazla gösterinin daha iyi sonuç verdiğini öne sürmektedir; ayrıca bir örneği göstermek sıfır-atış modeline göre daha iyi sonuç almıştır, bu BİÖ'nin önemini ve bir gösterinin bile modeli geliştirebileceğini gösteriyor; GPT-3 bağlam içi öğrenme, temel modeller olan ince ayarlı BERT ve BioBERT'den daha iyi performans göstermektedir (F1 puanı +%12.92, +%11.02 sırasıyla), ancak F-1'de %3.57 farkıyla BERT büyüklüğündeki PURE modelinden daha kötü sonuç vermiştir.

Tablo 3. i2b2-2010 Klinik İlişki Çıkarma veri kümesindeki ana sonuçlar. En iyi sonuçlar kalın olarak verilmiştir. BİÖ'nin en iyi sonuçları altı çizili olarak verilmiştir.

Model	Atış	Değerlendirme Metrikleri		
		Kesinlik	Duyarlılık	F1
<i>Temel modeller</i>				
BERT		71.17	62.72	63.36
BioBERT		70.49	64.39	65.26
PURE		78.10	81.68	79.85
GPT-3 sıfır-atış		45.95	45.42	45.68
<i>GPT-3 BİÖ modeli</i>				
GPT-RE_FT	1-atış	57.58	49.62	53.39
	5-atış	71.76	69.85	70.79
	10-atış	<u>75.85</u>	<u>76.72</u>	<u>76.28</u>

5.2 Tartışma

Pratik uygulamalarda, bir görev için ÖEDM'ne ince ayar yapmak için yeterli açıklamalı veri olmayabilmektedir. Bu gibi durumlarda, sorun için BİÖ kullanılması gerekebilir. Bu nedenle, bu durumda daha etkili olabilecek örnek seçimi ile BİÖ'nin performansı araştırılmaktadır.

İlk olarak, test cümlesine çok benzer cümleler kullanarak GPT-3 gibi BDM'leri ile bağlam içi öğrenmenin ince ayarlı BERT ve BioBERT ÖEDM'den büyük bir farkla daha iyi performans gösterdiği not edilmektedir. Deneyde en yüksek benzerliğe sahip 10 cümle seçilmiştir ve en iyi F1 puanına sahip olan PURE temel modeline rekabetçi performans gösteren 10-atışlı öğrenme elde edilmiştir. Ayrıca, 10-atışlı öğrenme BİÖ'de en yüksek F1 puanını aldığından, 10'dan fazla örnek kullanmak modelin performansını artırabilmektedir.

Diğer bir gözlem ise, $k = 5$ gösterileriyle bile GPT-RE_FT'nin BERT ve BioBERT'ten daha iyi çalıştığıdır (%63.36, %65.26 \rightarrow %70.79 (+%7.43, +%5.53)), bu, test girişine benzer az sayıda gösteri ile bile BİÖ'nin hala başarılı olduğunu göstermektedir. Bununla birlikte, performansın veri kümesindeki ilişki türlerinin eşit olarak dağılması nedeniyle ve test seti boyutuna bağlı olarak, bizim durumumuzda yalnızca 300 cümleden oluşan test seti ile, düşük olabileceğidir.

6 Sonuç

Bu makalede, GPT-3'ün klinik metinlerden BİÖ ile ilişkisi çıkarma için bağlam içi örnek seçim yöntemiyle potansiyeli araştırılmaktadır. Diğer bağlam içi öğrenme ortamları ve alanları için etkili olduğu kanıtlanmış mevcut teknikler incelenmiştir. Deneysel sonuçlar, yazılan varlık işaretleyicilerini ekleyerek her varlık çiftini bağımsız olarak dikkate alan temel model PURE'ün bu sorun için hala en iyi sonuçları elde ettiğini göstermektedir. Bununla birlikte, performansı artırabilecek bazı düşünce zinciri (CoT; Wei vd., 2022) gibi yaklaşımlar ve negatif örnekleme, gösteri, test ve eğitim setini artırma gibi teknikler mevcuttur. Gelecekteki çalışmalar olarak, bu

iyileştirmeleri kullanarak çalışmalarımıza karşılaştırmalı bir boyut katmayı ve diğer örnek seçim yöntemlerini araştırmayı umuyoruz.

Kaynakça

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Gao, T., Fisch, A., & Chen, D. (2020). Making pre-trained language models better few-shot learners. *arXiv preprint arXiv:2012.15723*.
- Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., ... & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1), 1-23.
- Gutierrez, B. J., McNeal, N., Washington, C., Chen, Y., Li, L., Sun, H., & Su, Y. (2022). Thinking about gpt-3 in-context learning for biomedical ie? think again. *arXiv preprint arXiv:2203.08410*.
- Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., & Levy, O. (2020). Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the association for computational linguistics*, 8, 64-77.
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240.
- Liu, J., Shen, D., Zhang, Y., Dolan, B., Carin, L., & Chen, W. (2021). What Makes Good In-Context Examples for GPT-3? *arXiv preprint arXiv:2101.06804*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Lu, Y., Bartolo, M., Moore, A., Riedel, S., & Stenetorp, P. (2021). Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *arXiv preprint arXiv:2104.08786*.
- Mishra, S., Khashabi, D., Baral, C., & Hajishirzi, H. (2021). Cross-task generalization via natural language crowdsourcing instructions. *arXiv preprint arXiv:2104.08773*.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. *NAACL*.
- Sahu, S. K., Anand, A., Oruganty, K., & Gattu, M. (2016). Relation extraction from clinical texts using domain invariant convolutional neural network. *arXiv preprint arXiv:1606.09370*.
- Uzuner, Ö., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5), 552-556.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. Wan, Z., Cheng, F., Mao, Z., Liu, Q., Song, H., Li, J., & Kurohashi, S. (2023). Gpt-re: In-context learning for relation extraction using large language models. *arXiv preprint arXiv:2305.02105*. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38-45).

Zhong, Z., & Chen, D. (2020). A frustratingly easy approach for entity and relation extraction. arXiv preprint arXiv:2010.12812.

Bilgi Güvenliđi Yönetim Sistemi Kurulumu: Sağlık Bakanlıđı Deneyimi

Filiz İşleyen¹, Mustafa Mahir Ülgü², Dilek Şen Karakaya³, Betül Yılmaz⁴

^{1,2,4}Sađlık Bilgi Sistemleri Genel Müdürlüđü, Ankara, Türkiye

³Güven Future, Ankara, Türkiye

betul.yilmaz3@saglik.gov.tr

Information Security Management System Installation: Ministry of Health Experience

Abstract. Today, with the rapid development in information technologies, the use of the Internet has become widespread, the variety of data-carrying devices has increased and access to information has become much easier. As a result of technological developments, electronic circulation of information, which is an important asset in all sectors, has started. However, this situation has brought with it many concerns about the security of information. ISO/IEC 27001 is an international Information Security Management System (ISMS) standard consisting of policies and procedures for systematically managing information assets. In this study, it is aimed to share the experiences gained during the ISMS installation processes in the Ministry of Health. Method: During the establishment of the ISMS, documents were prepared within the scope of ISO/IEC 27001 standard, information security awareness training was provided to the personnel, and audits were carried out regularly every year by forming an internal audit auditor team. Apart from the standard, trainings for information security awareness trainers were given to the provincial organization of the Ministry. Findings: The number of information security awareness training given to the personnel on a yearly basis and the number of internal audits are given. The activities carried out for the establishment and maintenance of the ISMS and their importance are explained in articles. Conclusion: The most important and effective strategy for establishing and maintaining a successful ISMS in an organization; to receive the full support of the management in all processes, to establish an ISMS unit within the organization, to document the processes in a timely manner and to carry out studies for the personnel to adopt ISMS.

Keywords: ISMS; Information Security; Information Security Awareness.

Özet: Günümüzde bilişim teknolojilerindeki hızlı gelişim ile birlikte internet kullanımı yaygınlaşmış, bilginin bulunduğu cihazların çeşitliliđi artmış ve bilgiye erişim çok daha kolaylaşmıştır. Teknolojik gelişmeler sonucunda tüm sektörlerde önemli bir varlık olan bilginin elektronik dolaşımı başlamıştır. Ancak bu durum bilginin güvenliđi ile ilgili birçok endişeyi de beraberinde getirmiştir. ISO/IEC 27001, bilgi varlıklarının güvenliđini sistematik olarak yönetmek için

politika ve prosedürlerden oluşan uluslararası bir Bilgi Güvenliği Yönetim Sistemi (BGYS) standardıdır. Bu çalışmada, T.C. Sağlık Bakanlığında BGYS kurulumu süreçlerinde edinilen deneyimlerin paylaşılması amaçlanmıştır. Yöntem: BGYS kurulması sürecinde ISO/IEC 27001 standardı kapsamında dokümanlar hazırlanmış, personele bilgi güvenliği farkındalık eğitimleri verilmiş ve iç tetkik denetçi ekibi oluşturularak her yıl düzenli olarak denetimler gerçekleştirilmiştir. Standardın dışında da Bakanlığın taşra teşkilatına bilgi güvenliği farkındalığı eğitici eğitimleri verilmiştir. Bulgular: Yıl bazında personele verilen bilgi güvenliği farkındalık eğitimi sayıları ve iç denetim sayıları verilmiştir. BGYS kurulumu ve sürdürülmesi için gerçekleştirilen faaliyetler ve önemi maddeler halinde açıklanmıştır. Sonuç: Bir kurumda başarılı bir BGYS kurmak ve devam ettirmek için en önemli ve etkili strateji; tüm süreçlerde yönetimin tam desteğinin alınması, kurum içinde BGYS birimi oluşturulması, süreçlerin zamanında dokümante edilmesi ve personelin BGYS'yi sahiplenmesine yönelik çalışmaların yapılmasıdır.

Anahtar Kelimeler: BGYS; Bilgi Güvenliği, Bilgi Güvenliği Farkındalığı.

1 Giriş

Bilgi ve iletişim teknolojilerinin gelişmesi ve kullanımının yaygınlaşması sonucu, bilgiyi saklamak ve gerektiğinde erişim sağlamak için kullanılan araçların çeşitliliği de artmıştır. Ekonomik Kalkınma ve İşbirliği Örgütü (Organisation for Economic Co-operation and Development-OECD), devletler tarafından vatandaşa sunulan hizmetlerin kalitesinin artırılması amacıyla, bilgi ve iletişim teknolojilerinin özellikle de internetin kullanılmasını e-Devlet olarak tanımlamaktadır [1]. Elektronik ortamda e-devlet üzerinden sunulan hizmetler herkes için zamandan tasarruf, mekânsal bağımsızlık, kâğıt tüketimini azaltma ve iş gücü tasarrufu gibi pek çok açıdan kazanım sağlamıştır. e-Devlet kapsamında sunulan hizmetler gün geçtikçe artmaya devam etmektedir. Dolayısıyla önceden kurum bünyesinde barındırılan bilginin artık kurumlar arasında elektronik olarak iletiminin arttığı söylenebilir. Kamu kurumlarındaki bilgi varlıkları sadece kurumun değil vatandaşların da bilgilerinden oluştuğu için bilginin elektronik dolaşım sürecinde kurumlarda görev yapan tüm çalışanların iş süreçlerinde daha dikkatli olması, bilgi güvenliği ihlalinde ortaya çıkabilecek sonuçların farkında olmaları hem kurumsal hem de ulusal bilgi güvenliği açısından önemlidir. Bilgi güvenliğini sadece bilginin korunması işi olarak düşünmemek gerekir. Ülke kalkınmasında gerekli olan kritik teknolojik altyapılar, kişisel bilgilere erişimin söz konusu olduğu e-devlet hizmetleri, bankacılık hizmetleri gibi vatandaşların erişim için kullandığı her ağ, bilgisayar ve cihazın güvenliği olarak düşünmek gerekir [2]. Kişinin ve kurumun en değerli varlığı olan bilginin güvenliği; bilginin gizliliğinin, bütünlüğünün ve erişilebilirliğinin korunmasını içermektedir [3]. Bilgi Güvenliği Yönetim Sistemi (BGYS) ise kurumsal olarak bilgi varlıklarının belirlenmesi, bunlara her türlü deneimsiz erişimin engellenmesi, yapılan risk analizleriyle bilginin gizliliğinin ve bütün-

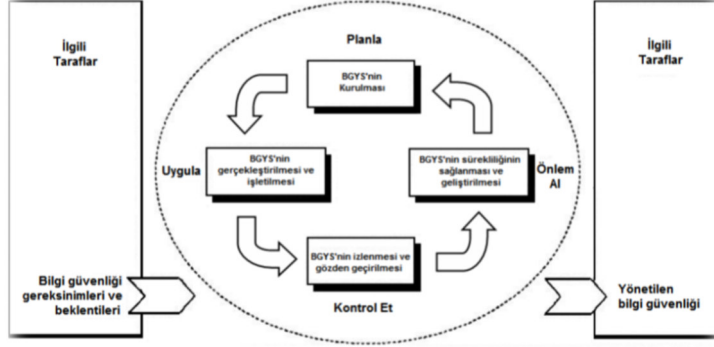
lüğünün korunması için gerekli idari ve teknik önlemlerin alınması ve tüm iş süreçlerinin bilgi güvenliği politikaları doğrultusunda düzenlenerek yönetilme işlevi olarak tanımlanabilir [4]. Bilgi güvenliği yönetimi sürecinde hedeflenen bilgi varlıklarının mümkün olduğunca korunmasıdır. Bunun için kurumda bir ihlal olayı meydana gelmeden önce olayın öngörülebilmesini ve önleyici tedbirler alınmasını sağlamaya yönelik çalışmalar yapılmalıdır. Öngörülemeyen veya önlenemeyen olaylar meydana geldiğinde ise, en az zararla en kısa zamanda normal çalışma şartlarına geçilebilmesine yönelik stratejiler belirlenerek uygulanmalıdır [4]. Kurumsal olarak bir bilgi güvenliği yönetim sisteminin benimsenmesi, kritik altyapı üreten ve BGYS'nin kurulmasının zorunlu olduğu sektörlerin dışındaki kurum ve kuruluşlar için stratejik bir karardır. Kritik altyapı "İşlediği bilgi/verinin gizliliği, bütünlüğü veya erişilebilirliği bozulduğunda can kaybına, büyük ölçekli ekonomik zarara, ulusal güvenlik açıklarına veya kamu düzeninin bozulmasına yol açabilecek bilişim sistemlerini barındıran altyapı" olarak tanımlanmaktadır [5]. Kritik altyapıya sahip veya değil bir kurumda bilgi güvenliği yönetim sisteminin kurulması ve uygulanması, kurumun ihtiyaç ve hedeflerinden, güvenlik gereksinimlerinden, kullanılan yönetim süreçlerinden, kurumun büyüklüğü ve yapısından etkilenmektedir [3]. ISO/IEC 27001 standardı bir kurumda BGYS kurulumu, denetlenmesi ve bilginin güvenliği konularını ele alan ve dünyada en yaygın kullanılan standarttır. Elektronik ortamdaki bilgilerin güvenliğini sağlamak için veri şifreleme, kimlik doğrulama, rol bazlı yetkilendirme, içerik kontrolü, anti-virüs kurma, güvenlik duvarı, sanal özel ağ, saldırıyı önceden tespit etme ve önleme gibi çeşitli teknik önlemler alınmaktadır [6]. Ancak alınan teknik önlemler kurumsal bilgi güvenliğini sağlama konusunda tek başına yetersiz olabilir, bilgi güvenliğinin sağlanmasında teknik önlemlerin alınmasının yanı sıra kurum politikası ve kullanıcı farkındalığı oluşturmak gibi başka faktörlerin de olması gerekmektedir. Teknik veya kurumsal olarak alınan tedbirler, kurum personeli tarafından kullanıldığı kadarıyla bilgi güvenliğini sağlamada etkilidir. Kurumlarda, bilgi güvenliğini sağlamak amacıyla alınan önlemlerin başarısı son kullanıcı davranışlarına ve farkındalığına bağlıdır [2]. Bu nedenle özellikle kurum personeli başta olmak üzere bilgi alışverişi yapılan tüm kişilerin kurum için değerli bir varlık olan bilgi ve bilgi varlıklarının korunmasında üzerlerine düşen sorumlulukları anlamaları önemlidir [7]. Bu durumun anlaşılması için gerek kurumsal gerek bireysel olarak alınabilecek farkındalık eğitimleri personelin bakış açısında ve davranışında olumlu etkiler yapabilir [8]. Tüm alanlarda olduğu gibi sağlık alanında da sunulan hizmetleri daha iyi, hızlı ve etkili verebilmek için, bilgi ve iletişim teknolojilerinden faydalanılmaktadır. Bu hizmetler sadece e-Devlet üzerinden sunulan hizmetlerle sınırlı değildir. Sağlık Bakanlığının merkez ve taşra teşkilatında bilginin üretilmesi, saklanması, gerektiğinde iletilmesi veya bilgiye erişim için farklı teknolojiler kullanılabilmektedir. Sağlık verileri, 108 No'lu Kişisel Verilerin Otomatik İşleme Tabii Tutulması Karşısında Bireylerin Korunması Sözleşmesi, 95/46/EC Kişisel Verilerin İşlenmesi ve Bu Tür Verilerin Serbest Dolaşımına Dair Bireylerin Korunması Direktifi ve 6698 Sayılı Kişisel Verilerin Korunması Kanunu (KVKK) gibi mevzuat çalışmaları ile özel nitelikli kişisel veriler kapsamında sayılmaktadır [9-11]. Bu nedenle özellikle sağlık verilerinin yönetilmesinde ve korunmasında kurumsal olarak daha güçlü önlemler alınmalıdır. Sağlık Bilgi Sistemleri Genel Müdürlüğü (SBSGM) tarafından kurumsal olarak BGYS süreçlerini başlatmak

için iyi uygulama örnekleri incelenerek ISO/IEC 27001 kurulmasına karar verilmiştir. İlk denetim ve sertifikasyon süreci 2014 yılında tamamlanmış olup bu tarihten itibaren standardın gereklilikleri günümüze kadar düzenli olarak uygulanmaktadır. SBSGM'nin ISO/IEC 27001 standardına uyumluluğu Türk Akreditasyon Kurumu onaylı belgelendirme kuruluşlarından birisi olan Türk Standartları Enstitüsü (TSE) tarafından her yıl yapılan dış denetimlerle kontrol edilmektedir. Bu çalışmada, Türkiye Cumhuriyeti Sağlık Bakanlığında BGYS kurulum süreçleri için izlenen yöntemlerin, karşılaşılan güçlüklerin ve bulunan çözümlerin açıklanması ve BGYS sürecine yeni başlayacak veya süreci devam eden kurum ve kuruluşlar için tecrübe paylaşımı yapılması amaçlanmıştır.

2 Yöntem

T.C. Sağlık Bakanlığında 2013 yılında ISO/IEC 27001 belgesi alma faaliyetlerine başlanmıştır. Bu doğrultuda ISO/IEC 27001 BGYS'yi kurmak, işletmek, izlemek, gözden geçirmek, sürdürmek ve iyileştirmek için Bilgi Güvenliği Yönetim Sistemi Birimi (BGYSB) kurulmuştur. ISO/IEC 27001 standardının alınması ve T.C. Sağlık Bakanlığında kurumsal bir BGYS kültürünün oluşturulması amacıyla BGYSB tarafından öncelikle 30 kişilik ekip oluşturularak Türk Standartları Enstitüsünden (TSE) "TS ISO/IEC 27001 Bilgi Güvenliği Yönetim Sistemi Standardı Temel Eğitimi" ve "İç Tetkik Eğitimi" eğitimleri alınmıştır. Sonrasında, kurumda ISO/IEC 27001 BGYS'nin kurulumu için çalışmalara başlanmış, bu süreçte Genel Müdürlük personeli ile 62 adet toplantı yapılmıştır. TSE tarafından dış tetkike tabi tutulan SBSGM, ilk kez 2014 yılında ISO/IEC 27001 belgesini almıştır.

Aynı tarihlerde, Genel Müdürlük BGYS kurulum süreçlerinden elde edilen tecrübelerin Bakanlık geneline yaygınlaştırılmasını sağlamak amacıyla Sağlık Bakanlığı Bilgi Güvenliği Politikaları Yönergesi (Yönerge) [12] ve Sağlık Bakanlığı Bilgi Güvenliği Politikaları Kılavuzu [13] hazırlanmış ve yayımlanmıştır. Yönergede, bilgi güvenliği politikalarının ana çerçevesi tanımlanmıştır. Yönerge'nin eki olan Kılavuzda ise ISO/IEC 27001 BGYS standardındaki madde başlıkları ve belirtilen kontrol önlemleri kapsamında, bilgi güvenliği ile alınması gereken ilgili teknik ve idari tedbirler iyi uygulama örnekleri ile birlikte açıklanmıştır. 2018 yılında Yönerge, 2019 yılında ise Kılavuz dokümanları güncellenmiştir. ISO/IEC 27001'in odak noktası, süreç odaklı bir BGYS'nin planlanması, uygulanması, işletilmesi ve sürekli izlenmesi ve iyileştirilmesi gerekliliğidir. Yaklaşımın Planla, Uygula, Kontrol Et, Önlem Al (PUKÖ) döngüsü ile uyumlu olması önerilmektedir [14].



Şekil 1. PUKÖ Döngüsü [15].

SBSGM ISO/IEC 27001 süreçlerinde de PUKÖ döngüsü kullanılmıştır. PUKÖ Döngüsü aşamaları kısaca aşağıdaki şekilde özetlenebilir [16].

Planla: Bu aşamada mevcut durum analizi yapılır, hedefler belirlenir ve buna yönelik strateji ve eylemler planlanır.

Uygula: Bu aşamada eylem planının uygulanması ve belgelenmesi amaçlanır. Ayrıca beklenmedik olaylar, öğrenilen dersler ve edinilen bilgiler de dikkate alınmalıdır.

Kontrol Et: Bu adımda, önceki adımda uygulanan eylemlerin sonuçları analiz edilir. İyileştirmeler olup olmadığını ve belirlenen hedeflere ulaşıp ulaşılmadığını doğrulamak için bir öncesi ve sonrası karşılaştırması yapılır.

Önlem Al: Bu aşama, kontrol aşamasında ortaya çıkan sorunların çözülmesi ve ileride tekrar karşılaşılmaması için önlem alınması ile ilgilidir. Gerçekleştirilen iyileştirmeleri standartlaştırmayı amaçlayan yöntemlerin geliştirilmesinden oluşur.

ISO/IEC 27001 standardına göre BGYS kurulum sürecinde yapılması gereken ilk eylem, uygulama ve denetime tabi olacak Kapsam'ın belirlenmesidir [15]. Bu doğrultuda Kapsam "Sağlık Bilgi Sistemleri Genel Müdürlüğü" olarak belirlenmiştir. Yöneticilerin onayını ve desteğini vurgulayan bir politika oluşturulmuş ve tüm personele duyurulmuştur [17]. Buna ek olarak SBSGM görev ve yetkileri doğrultusunda hedefler belirlenmiştir. Bu hedefler;

- Bilginin gizlilik, bütünlük ve erişilebilirlik açısından, içerden veya dışardan, kasıtlı ya da kazayla oluşabilecek tüm tehditlerden korunmasını sağlamak,
 - Bilginin toplanması, değerlendirilmesi, raporlanması, paylaşılması ve imha edilmesi süreçlerini içeren tüm yaşam döngüsü boyunca güvenliğini sağlamak,
 - Bakanlık merkez ve taşra teşkilatında tüm çalışanların bilgi güvenliği farkındalık, duyarlılık ve teknik bilgi düzeylerini en üst düzeyde tutmak,
 - Teknolojinin sağladığı en ileri sistemsel güvenlik tedbirlerini uygulayarak insan kaynaklı zafiyetleri asgariye indirmek,
 - Veri ve bilgi kayıplarını önlemek ve kurumsal itibarın korunmasını sağlamaktır.
- olarak özetlenebilir. Bu hedeflerin gerçekleştirilmesinde mevcut tehditlerin ortadan kaldırılmasına yönelik yapılan çalışmalar kadar olası risklerin belirlenmesi de önemlidir. SBSGM, iç ve dış paydaşları için ürettiği değeri en yüksek seviyede tutmak ve

bu hedef doğrultusunda gerçekleştirdiği tüm faaliyetlerinde, iş sürekliliğini maksimum düzeyde muhafaza edebilmek için kurumsal risk yönetimi yaklaşımını benimsemiştir. Risk yönetimi hedefi; risk süreci faaliyetlerinin kurumsal bilgi güvenliği hedefleri ile uyum içerisinde yürütülmesidir. Bu kapsamda, bilgi güvenliği kontrollerinin risk yönetimi bakış açısı ile tanımlanması, kurum önceliklerinin belirlenmesi ve kaynakların etkin kullanılması amaçlanmıştır.

Bilgi varlıklarının gizliliğini, bütünlüğünü veya kullanılabilirliğini (erişilebilirliğini) tehlikeye düşürme potansiyeline sahip herhangi bir olay veya durum bilgi güvenliği ihlal olayı olarak tanımlanabilir. SBSGM’de bilgi varlıkları, ISO/IEC 27001 standardında tanımlanan gizlilik, bütünlük ve erişilebilirlik çerçevesinde değerlendirilerek belirlenmiştir. Kurumsal olarak BGYS’nin etkin bir şekilde işletilmesinde ihlal olaylarının tanımlanması ve bilinmesi önemlidir. Bu amaçla <https://bilgiguvenligi.saglik.gov.tr/> adresine “olay bildir” sekmesi eklenerek bu sekme altında tüm T.C. Sağlık Bakanlığı çalışanları ve vatandaşın bilgi güvenliği ihlallerini bildirebileceği bir web sitesi oluşturulmuştur. Web sitesinde ihlal olaylarının tanımlanması, bildirilmesi, değerlendirilmesi ve raporlanması sağlanmaktadır. Personelin bilgi güvenliği farkındalığının artırılması çalışmaları kapsamında broşür, afiş, el kitabı basılı ve elektronik ortamda hazırlanarak son kullanıcılara ulaştırılmıştır. Ayrıca aylık olarak hazırlanan güncel BGYS Bültenleri personele düzenli olarak e-posta aracılığı ile iletilmektedir. Tüm personelin bilgi güvenliği farkındalığını arttırmaya yönelik yapılan çalışmalara ek olarak yüz yüze ve online bilgi güvenliği farkındalık eğitimleri verilmiştir. Eğitici eğitimleri olarak planlanan ilk BGYS farkındalık eğitimleri 10 ilde yapılmış ve 81 ilden, yerel yöneticiler tarafından belirlenen toplamda 992 personel bu eğitime katılmıştır. Kurumda 2014 yılından bu yana her yıl ISO/IEC 27001 kapsamı dahilinde BGYSB personeli ve ISO 27001 iç tetkikçi eğitim alan personel tarafından ISO/IEC 27001 kontrol maddeleri temel alınarak iç denetimler gerçekleştirilmektedir. Denetimler sonrasında elde edilen sonuçlar BGYSB personeli tarafından nitel araştırma yöntemleri ile analiz edilerek sonuçların düzeltici faaliyet ya da öngörülen risk kapsamında değerlendirilmesine karar verilmektedir. Sonuçlar ilgili birimlere iletilerek düzeltici faaliyet veya öngörülen risk için gerekli düzenlemelerin yapılması sağlanmaktadır. BGYS sürecinde önemli hususlardan birisi de yönetimin desteği olduğundan yapılan çalışmalar yılda bir kez yönetime sunulmakta ve sürekli iyileştirme sağlanmaktadır.

3 Bulgular

Kurum personeline BGYS önemini anlatmak, BGYS kurulum yönteminin ve süreç dahil olacak personelin belirlenmesi amaçlarıyla SBSGM personeli ile yapılan toplantılarda BGYS’nin önemini anlayan ve yeniliklere açık olan kişilerin sürece daha çabuk uyum sağladıkları görülmüştür. 2015 yılında standardın revize edilmesi, kapsam değişikliği gibi nedenlerle BGYS kurulum süreçlerine baştan başlanması gerekmiştir. 2014 yılında yapılan tüm çalışmalardan edinilen tecrübeler ve eğitimli personel ile bu kez sadece 11 adet toplantı yapılarak ISO/IEC 27001 2013 versiyonuna geçiş için gerekli tüm çalışmalar tamamlanmıştır. 2014 yılından itibaren kurum personeline

Faaliyetler	Önemi	Gerçekleştirilenler
-------------	-------	---------------------

bilgi güvenliği farkındalığı oluşturmak amacıyla verilen eğitimlerin yıl bazında süreleri ve eğitimlere katılan personel sayısı Tablo 1’de verilmiştir.

Tablo 1. Yıllara Göre Personele Verilen Bilgi Güvenliği Eğitim Süresi ve Katılımcı Sayısı

Yıllar	Toplam Eğitim Süresi	Toplam Katılımcı Sayısı
2015	4 Gün	968
2016	11 Gün	1842
2017	12 Gün	1068
2018	22 Gün	1663
2019	21 Gün	1214
2020	10 Gün	1246
2021	7 Gün	401
2022	2 Gün	108
2023	12 Gün	1851

ISO/IEC 27001 BGYS'nin uygulanması sürecinde, standardın zorunlu maddelerinden olan "iç tetkik" faaliyetlerine özel önem atfedilmiştir. Bu kapsamda 2014 yılından itibaren yapılan tüm iç tetkikler, SBSGM’de görev yapan ve iç tetkik eğitimi almış personel marifetiyle yapılmıştır. Kurumdaki iç tetkiklerin, ayrılan süre ve tetkik yapılan birim/süreç sayısı açısından, mümkün olduğu kadar geniş tutulmasına özen gösterilerek her yıl tüm birimler iç tetkike tabi tutulmuştur. İç tetkik ekibine, sadece BGYSB’de görev yapan personel değil SBSGM’nin diğer birimlerinde görev yapan yetkin personel de tetkikçi olarak dâhil edilmiş ancak öncesinde bu personelin de ISO/IEC 27001 iç tetkik eğitimi alması sağlanmıştır. Bu sayede SBSGM bünyesinde geniş bir iç tetkikçi havuzu da oluşturulmuştur. Tablo 2’de yıl bazında iç tetkik yapan personel ve iç tetkik yapılan birim sayıları verilmiştir.

Tablo 2. Yıllara Göre İç Tetkik Yapan Denetçi ve Yapılan Birim Sayısı

Yıllar	Tetkikçi Sayısı	Tetkik Yapılan Birim Sayısı
2015	20	30
2016	22	31
2017	16	34
2018	33	30
2019	18	32
2020	10	30
2021	14	28
2022	19	23

2014 yılından itibaren ISO/IEC 27001 kapsamında gerçekleştirilen faaliyetler Tablo 3’te özetlenmiştir.

Tablo 3. BGYS Kurulumu ve Sürdürülmesi için Gerçekleştirilen Faaliyetler

BGYS Kurma	Ekibi	Konuyla ilgili yetkin personel ile BGYS faaliyetlerine başlanması süreçlerde en önemli faktörlerden birisidir.	Üç personel ile kurulan BGYS ekibi çalışmalarına yedi personel ile devam edilmektedir.
Kapsamın Belirlenmesi	Be-	BGYS faaliyetlerinin hangi departmanda gerçekleştirileceği belirlenir ve ona göre tüm çalışmalar gerçekleştirilir.	Kapsam Sağlık Bilgi Sistemleri Genel Müdürlüğü olarak belirlenmiştir.
Politika Hazırlanması	Hazır-	Yazılı olarak BGYS politikası hazırlanır. Politikada üst yönetim desteği ve kurumun BGYS bakış açısı anlatılır. Politika'nın tüm personele duyurulması gerekir.	BGYS Genel Politikası oluşturulmuş https://bilgiguvenligi.saglik.gov.tr/ Web sitesinde personelin ulaşabileceği şekilde yayımlanmıştır.
Planlama		Kurumun bilgi varlıkları belirlenir. Varlıkların korunması, istenmeyen durumların yaşanmaması ve sürekli iyileştirmenin sağlanması için kurum risklerini belirlemeli, oluşturduğu risk metodolojisi ile risk değerlendirme süreci gerçekleştirilmeli ve sürekli iyileştirilmelidir.	Kapsam dahilinde kurum riskleri belirlenmiş, risk metodolojisi ile riskler değerlendirilmiştir. Periyodik aralıklarla risk işleme faaliyetleri ve sürekli iyileştirme sağlanmaktadır.
Dokümantasyon Süreci		Bilgi Güvenliğinin etkin bir şekilde yönetilmesi için kurum, tüm süreçlerde yapılan faaliyetlere ilişkin bilgi güvenliği politikaları, prosedürleri, talimatlar iş süreçlerine yönelik standart olarak kullanılacak kaynak dokümanlar hazırlamalıdır.	BGYS'yi ilgilendiren faaliyetler kapsamında 211 doküman hazırlanmış ve https://bilgiguvenligi.saglik.gov.tr/ Web sitesinde personelin ulaşabileceği şekilde yayımlanmıştır.
Olay Yönetimi		Kurumda yaşanan ihlal olayları raporlanmalıdır. Yaşanabilecek olaylar için de senaryolar oluşturularak buna karşılık önlemler belirlenmelidir. İhlal olayı anında kimlerin aranacağı ve hangi şekilde çözüleceği önceden belirlenmelidir.	İhlal olay bildirim çalışmaları kapsamında BGYS web sitesinde personelin ihlal bildirimini yapabilecekleri bir alan oluşturulmuştur. https://bilgiguvenligi.saglik.gov.tr/OlayBildir
Farkındalık Faaliyetleri		Kurumda yapılan BGYS çalışmalarında ki en önemli husus personel farkındalığıdır. Personele eğitim verilmeli, konunun önemi	Tüm personele yılda en az bir kez bilgi güvenliği farkındalık eğitimleri verilmektedir. Genel Müdürlükte işe yeni başlayan her personele ise oryantasyon sürecinde farkındalık eğitimleri verilmektedir. Farkındalık

İç Tetkik	anlatılmalı, konuya dair farkındalık dokümanları hazırlanmalı ve personele bildirilmelidir. Kurum, kendi içinde kendini denetlemeli, uygunsuzluklarını çıkarmalı ve bunun için çalışmalar yapmalıdır.	dokümanları, videoları, afişler, anket çalışmaları ve e-postalar ile farkındalık çalışmalarını sürdürülmektedir. Yılda bir kez olacak şekilde tüm Genel Müdürlük birimlerine iç tetkik yapılmaktadır.
Düzeltilici Faaliyet	Tespit edilen uygunsuzluklar takip edilmeli ve düzeltilmesi için ek çalışmalar sağlanmalıdır.	Yılda bir kez gerçekleşen iç tetkik sonucu veya devam eden faaliyetlerde bir uygunsuzlukla karşılaşıldığında düzeltilici faaliyet ile takip edilip uygunsuzluğun giderilmesi sağlanmaktadır.
Yönetim Gözden Geçirme	BGYS faaliyetlerinde en önemli noktalardan biriside üst yönetimin desteği her zaman alınmalı, yapılan çalışmalar üst yönetime sunulmalı ve sürekli iyileştirme sağlanmalıdır.	Yılda bir kez tüm BGYS süreçleri, yıl içerisinde gerçekleştirilen faaliyetler, riskler, BGYS hedefleri, ihlal olayları vb. çalışmalar yönetime sunulup sürekli iyileştirme sağlanmaktadır.

4 Tartışma

Bilişim teknolojilerindeki hızlı gelişmeler ve kullanımının yaygınlaşması, bu sistemlerde saklanan bilginin korunması için standart olarak kullanılabilen yöntemlerin geliştirilmesini de zorunlu kılmıştır. Birçok kurumda ISO/IEC 27001 sertifikasyon sürecinde danışmanlık hizmeti alındığı bilinmektedir. Bu yöntem kurumun ihtiyaçlarına ya da hedeflerine göre değişiklik gösterebilir. BGYS kurarak sürekliliğini sağlamak için bizim çalışmalarımız ve deneyimlerimiz kurum içinde BGYS ekibi kurulmasının sürecin yönetilmesi ve sonrasında devam ettirilmesinde önemli ve gerekli olduğunu göstermiştir. SBSGM’de kurulan BGYSB faaliyetleri ISO/IEC 27001 standardı alındıktan sonra da devam etmiştir. BGYS birimi tarafından verilen bilgi güvenliği farkındalık eğitimlerine oldukça fazla zaman ayrıldığı görülmektedir (Tablo 3). Kullanıcılar için bilgi güvenliği farkındalık eğitimleri, Bakanlık merkez personeline yüz yüze olarak verilen eğitimler, eğitim sonrasında uygulanan anket çalışmalarıyla da değerlendirilmektedir. Bir kurumda BGYS kurmak ve işletmek için ISO/IEC 27001 belgesine sahip olmak önemlidir. Tüm süreçlerde standardizasyonu sağlamakla birlikte yapılması gereken faaliyetler de belirli olduğu için uygulayıcılara güvenli bir yol sunulmuş olur. Kurumsal olarak sadece standardın gerekliliklerini yerine getirerek hem kurumda BGYS kurmuş hem de uluslararası onaylanmış belgeye sahip olunabilir. Ancak kurum için daha fazlası hedefleniyor ve isteniyorsa standart kapsamında olmayan ek faaliyetler de yapılabilir. Bunun bir örneği olarak standardın gerekliliği olmasa da özellikle ISO/IEC 27001 kurulması süreçlerinin başlangıcında BGYSB yerine getirerek hem kurumda BGYS kurmuş hem de uluslararası onaylanmış belgeye sahip olunabilir. Ancak kurum için daha fazlası hedefleniyor ve isteniyorsa standart kapsamında olmayan ek faaliyetler de yapılabilir. Bunun bir örneği olarak standardın

gerekliliđi olmasa da özellikle ISO/IEC 27001 kurulması süreçlerinin başlangıcında BGYSB tarafından, merkez veya taşra teşkilatı personeli ayırmaksızın personelin bilgi güvenliđi farkındalıđını arttırmak amacıyla gerçekleştirilen eğitici eğitimler verilebilir. Bilginin güvenliđi konusu sadece sistemlerin güvenliđinin sağlanmasıyla değil aynı zamanda bu sistemleri kullanan kullanıcıların bilinçlendirilmesiyle de ilgili-dir. İçinde yaşadığımız dijital çağda kullanılan teknolojilerde bilgi güvenliđi sağla- ma konusunda hala en zayıf halka olarak kullanıcı farkındalıđı kabul edilebilir. Bilgi güvenliđinde teknik, süreç veya kullanıcı olarak birbirini etkileyen üç süreçten birinde meydana gelecek aksaklık, kurumda bilgi güvenliđinin tam olarak sağlanamayacağı anlamına gelmektedir [18]. Kamu kurumunda personelin, kurumda uygulanmasına karar verilen bir yeniliđe ilişkin bilgilendirilmesi ve farkındalıđının artırılmasına yö- nelik çalışmalar yapılması, yeniliđin kabulüne ve gelişmesine katkı sağlayabilir [19]. Bu bilgiler ışığında, BGYS'nin ve bilgi güvenliđine ilişkin kurallara uyumun başarılı olması, büyük oranda son kullanıcı davranışına ve farkındalıđına bağlı olduđu gerçeđi göz ardı edilmemelidir. Bilgi güvenliđinin sağlanmasında kullanıcı farkındalıđının geliştirilmesi için gerçekleştirilen uygulamalar, teknik olarak alınan önlemler kadar önemlidir [20]. SBSGM'de bu kapsamda yürütölen iç tetkik denetimlerinde, BGYSB dışından personelin sürece dahil edilmesi yönteminin personelin kabulünü arttırdığı gibi kurumsal adaptasyonu hızlandırdığı da söylenebilir. 2014 yılından itibaren siste- matik olarak yaptığımız tüm çalışmalar ışığında, SBSGM personelinin büyük bir kıs- mı bilişim teknolojileri alanında eğitim almış/çalışmış personelden oluştuđu, almış oldukları eğitim ve mesleki tecrübeleri itibarı ile bilgi güvenliđi süreçlerine yakın durumda olduđu görölmüştür. Bu durum, kurumda bilgi güvenliđi eğitim ve farkında- lık çalışmalarının yapılması, BGYS süreçlerinin uygulanması açısından avantaj oluş- turmuştur. gerektirdiđi bir şekilde sürekli ve kararlılıkla devam eden çalışmaları içere- mektedir. Önemli olan kurumda bilgi güvenliđinin anlaşılması, benimsenmesi ve edinilen tecrübelerle yola devam edilmesidir. Deneyimlerimiz; bir kurumda bilgi güvenliđi farkındalıđının ve davranış deđişikliđinin oluşması için, öncelikle yönetim tarafından bilgi güvenliđi süreçlerinin sahiplenilmesi gerektiđini göstermiştir. 2014 yılından itibaren yapılan çalışmalarımızda; personele yönelik bilgi güvenliđi farkında- lık eğitimlerinin düzenli olarak verilmesi, eğitimlerde kullanılan araçların modern eğitim araçlarından seçilmesi, eğitimlerin küçük gruplarla etkileşimli olarak yapılma- sı, bilgilerin teoriden ziyade gerçek örnekler üzerinden aktarılması, mümkünse video- larla desteklenmesi ve özellikle yeni başlayan personel için oryantasyon döneminde kapsamlı bir başlangıç düzeyi eğitimi verilmesi ve farkındalık artırıcı dokümanların hazırlanarak personele iletilmesi gibi çözümlerin fayda sağladıđı görölmüştür. Birçok kurumda ISO/IEC 27001 sertifika alınması sürecinde danışmanlık hizmeti alındığı bilinmektedir. SBSGM bu süreçte kendi personelinin ilgili eğitimleri almasını sağla- mış ve süreci kendi iç kaynaklarıyla yürütmüştür. Bu da personelin sürece sahip çık- masına ve kurum içinde önemli bir denetçi havuzunun oluşturulmasına imkân sağla- mıştır. SBSGM tarafından yürütölen çalışmalarda yönetim desteđine ek olarak BGYS kurulumu ve sonrasında bilgili ve deneyimli personel ile süreçleri yürütmenin de ol- dukça önemli olduđu görölmüştür. Bu nedenle bizim önerimiz BGYS kurulum ve devamındaki tüm süreçlerin yönetim desteđi ile birlikte kurum içindeki personel ile yönetilmesidir. Bu şekilde hem kurum içinde BGYS personeli istihdamı oluşturulabi- lecek hem de kurum personeline yakın bir ekip ile sürecin daha aktif yürütölməsi sağlanabilecektir.

Kaynakça

- [1] OECD, "Recommendation of the Council on Digital Government Strategies", 2014. Retrieved from <http://www.oecd.org/gov/digital-government/Recommendation-digital-government-strategies.pdf>.
- [2] Özdemir, Ç. Uluylol, "Kamu Kurum ve Kuruluşlarında Bilgi Güvenliği Farkındalığı", Türkiye Sosyal Araştırmalar Dergisi, 649-666, 2021.
- [3] ISO, "ISO/IEC 27001:2013(en) Information technology — Security techniques — Information security management systems — Requirements", 2013. Retrieved from <https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en>.
- [4] Y. İleri, "Örgütlerde Bilgi Güvenliği Yönetimi, Kurumsal Entegrasyon Süreci ve Örnek Bir Uygulama" Anadolu University Journal of Social Sciences., pp. 55-72, 2016.
- [5] T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı, "Bilgi ve İletişim Güvenliği Rehberi" Ankara, 2020.
- [6] M. Gülmüş, "Kurumsal Bilgi Güvenliği Yönetim Sistemleri ve Güvenliği", İstanbul, 2010.
- [7] E. Şahinaslan, A. Kantürk, Ö. Sahinaslan, E. Borandağ, "Kurumlarda Bilgi Güvenliği Farkındalığı Önemi ve Oluşturma Yöntemleri", XI. Akademik Bilişim Konferansı Bildirileri, Şanlıurfa, 2009.
- [8] M. M. Al-Daeef, N. Basir, M. M. Saudi, "Security awareness training: A review" In Proceedings of the World Congress on Engineering, London, 2017.
- [9] Council of Europe, "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108)", 28 01 1981. Retrieved from <https://rm.coe.int/1680078b37>.
- [10] The European Parliament And Of The Council, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Retrieved from <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>.
- [11] Kişisel Verilerin Korunması Kanunu ve Uygulamaları, Ankara: KVKK Yayınları, 2016
- [12] Sağlık Bilgi Sistemleri Genel Müdürlüğü, "Sağlık Bakanlığı Bilgi Güvenliği Politikaları Yönergesi", Ankara, 2018.
- [13] Sağlık Bilgi Sistemleri Genel Müdürlüğü, "Bilgi Güvenliği Politikaları Kılavuzu", Ankara, 2019.
- [14] G. Disterer, "ISO/IEC 27000, 27001 and 27002 for Information Security Management", Journal of Information Security, pp. 92-100, 2013.
- [15] D. Dinçer Önel, Bilgi Güvenliği Yönetim Sistemi Kurulumu, TUBİTAK, 2007.
- [16] Realyvásquez-Vargas, Arturo, Karina Cecilia Arredondo-Soto, Teresa Carrillo-Gutiérrez, and Gustavo Ravelo. 2018. "Applying the Plan-Do-Check-Act (PDCA) Cycle to Reduce the Defects in the Manufacturing Industry. A Case Study" Applied Sciences 8, no. 11: 2181. <https://doi.org/10.3390/app8112181>.

- [17] Sağlık Bilgi Sistemleri Genel Müdürlüğü, "Bilgi Güvenliği Yönetim Sistemi Genel Politikası Kapsam ve Bağlamı ", Ankara, 2023.
- [18] M. T. Hüseyin Çakır, "ISO27001 Bilgi Güvenliği Yönetim Sistemi Standardının Kamu Kurumlarına Uygulanabilirliğinin Araştırılması: Ankara İli Örneği", Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, pp. 59-78, 2019.
- [19] F. Pişkin, "Kamuda Yenilik: İstanbul'daki Kamu Kurumları Üzerinden Bir Araştırma", Marmara Üniversitesi Öneri Dergisi, cilt 12, no. 48, pp. 57-79, 2017.
- [20] S. Seferoglu, H. Yıldız- Durak, K. Yılmaz, F. Gizem ve R. Yılmaz, "Bilgi Güvenliği Farkındalığı ve Bilgi Güvenliği Politikalarıyla İlgili Bir İnceleme", Eğitim Teknolojileri Okumaları, TOJET, 2018, pp. 29-43.

Tıp Alanında Yapay Zekâ Kullanılarak Yapılan Klinik Çalışmaların Bibliyometrik Analizi

Hülya Kılbahri¹ [0009-0007-1136-0430], P. Deniz Tosun^{1,2} [0000-0002-4513-6740]

¹ Graduate Education Institute, Duzce University, 81620 Duzce, Turkey

² Biomedical Engineering Department, Duzce University, 81620 Duzce, Turkey
hulyakilbahriiii@gmail.com, pinardeniztosun@duzce.edu.tr

Title: Bibliometric Analysis of Clinical Studies Utilizing Artificial Intelligence in the Field of Medicine

Abstract. Bibliometric analysis is the analysis of publications and their relationships that were performed within a specific topic and/or during a certain period of time. Bibliometric analysis is important for visualizing research. The aim of this study is to demonstrate the bibliometric analysis of articles on the use of artificial intelligence in clinical studies in the field of medicine using the R Bibliometrix package with PubMed database articles. After downloading the PubMed bibliographic database, it was imported into Biblioshiny, a web interface for bibliometrics, using the R Bibliometrix package. The numbers of publications per year and the keywords used in these articles on clinical studies using artificial intelligence in the PubMed database in the last 10 years, as well as the co-occurrence networks of these keywords and the thematic information of these articles were analyzed. As a result, it was shown that the use of artificial intelligence in clinical studies in the field of medicine has been increasing over the years and plays an important role in the diagnosis and the treatment in various medical fields.

Keywords: Artificialintelligence in medicine, Bibliometric analysis, R Bibliometrix tools, Biblioshiny

Özet. Bibliyometrik analiz, belirli bir dönemde belirli bir konu ve alanda kurumlar ya da kişiler tarafından ortaya koyulmuş yayınların ve bu yayınlar arasındaki ilişkilerin analizidir. Bibliyometrik analiz araştırmaların görselleştirilmesi açısından önemlidir. Bu araştırmanın amacı tıp alanında yapılan klinik çalışmalarda yapay zekâ kullanımını, PubMed veri tabanında bulunan makalelerin R yazılım dili Bibliometrix paketi kullanarak bibliyometrik analizini ortaya koymaktır. PubMed bibliyografik veri tabanı indirildikten sonra, R Bibliometrix paketinin web arayüzü olan Biblioshiny içine aktarılmıştır. Son 10 yılda PubMed veri tabanında bulunan yapay zekâ kullanılarak yapılan klinik çalışma makalelerin yıl bazlı sayıları ve makalelerde kullanılan anahtar kelimelerle, kelimelerin birlikte oluşum ağları ve tematik bilgileri analiz edilmiştir. Sonuç olarak yapay zekânın tıp alanında yapılan klinik çalışmalarda kullanılması yıllar içinde bu çalışma sayılarının giderek arttığını ve çeşitli tıp alanlarında tanı tedavi ve teşhiste önemli rol oynadığını göstermektedir.

Anahtar kelimeler: Tıpta yapay zekâ, Bibliyometrik analiz, R Bibliometrix paketi, Biblioshiny

1 Giriş

Tıbbi alanında yapay zekâ teknolojisinin kullanımı, sağlık hizmetlerinin küresel düzeydeki önemini vurgulamaktadır. Teknoloji ve bilişim sistemlerinin gelişmesiyle birlikte, sağlık sektörü de dijital dönüşümünü hızla gerçekleştirmektedir. Özellikle klinik süreçlerde dijitalleşme eğilimi artmaktadır ve bu dijitalleşme beraberinde birçok yeni teknolojiyi getirmektedir [1]. Sağlık hizmetlerindeki yapay zekâ teknolojisinin kullanımı, tanı, tedavi ve koruyucu sağlık hizmetleri gibi birçok süreci hızlandırmakta ve aynı zamanda maliyetleri önemli ölçüde düşürmektedir [2]. Hastane klinik fonksiyonlarında teşhis ve tedavi süreçlerini hızlandırmanın yanı sıra, olası insan hatalarını minimuma indirme potansiyeline sahiptir ve sağlık hizmetlerinin etkinliğini artırmaktadır [3]. Tıbbi alanda, yeni yapay zekâ teknolojilerinin radyolojik, patolojik, endoskopik, ultrasonografik ve biyokimyasal incelemelere entegrasyonu hastalıkların teşhisinde daha yüksek doğrulukla sonuçlanmakta ve insan iş yükünü azaltmaktadır [4]. Buna ek olarak hastalıkların tedavi sürecinin en önemli parçası olan erken teşhis ihtimalini de arttırmaktadır [5]. Bu durum, perioperatif dönemdeki tıbbi tedavilerin tümünün daha iyi cerrahi sonuçlarla önemli ölçüde iyileştirilmesine katkıda bulunmaktadır [6].

Bibliyometri, bilimsel yayınların nicel analizi ve istatistiksel değerlendirmesidir. Bu yöntem, bir bilim alanının büyümesini, olgunluğunu, önde gelen yazarlarını, eğilimlerini, kavramsal ve entelektüel haritalarını değerlendirmek için kullanılmaktadır [7]. Ayrıca araştırma performansını değerlendirmek, belgeler, yazarlar ve referanslar arasında bağlantı kurmak için de kullanılmaktadır [8]. Bibliyometrik analiz R, Python, vb. birçok yazılım ve program paketleri ile yapılabilmektedir [9]. Bu yöntemlerden biri olan “Biblioshiny” bir web arayüzü olarak yaygın şekilde kullanılmaktadır. Biblioshiny SCOPUS, Web of Science, PubMed, Digital Science Dimensions ve Cochrane veri tabanlarından bibliyografik verileri aktarabilmektedir. Kaynaklar, yazarlar ve belgeler, kavramsal yapı, sosyal yapı ve entelektüel yapı hakkında çeşitli görselleştirme ve analizler sunmaktadır [10]. Bu sebeple bibliyometrik analiz, özellikle hızlı gelişmekte olan çalışma konularının takip edilmesini kolaylaştırdığından son derece faydalı bir araç olmaktadır. Bu bağlamda, tıp alanında yapay zekâ kullanılarak yapılan klinik çalışmaların bibliyometrik analize tabi tutulması gelecek çalışmalara yol göstermek, çalışmaların nasıl ve ne yönde ilerlediklerini göstermek ve özellikle de literatürde bulunan eksiklikleri tespit edebilmek açısından önemlidir.

2 Yöntem

Tıp alanında yapay zekâ kullanılarak yapılan klinik çalışmalar bibliyometrik analizle incelenmiştir. Analiz, tıp literatüründe yapay zekâ kullanılarak yapılan klinik çalışmaların gelişimi ve özellikleri hakkında önemli bilgiler, sayısal veriler ve görseller sunmayı amaçlamaktadır. Bu sayede, literatürde bulunan boşluklar tespit edilirken ayrıca multidisipliner çalışmaların aralarındaki bağlantılar ve etkileşimler de ortaya çıkarılmak istenmiştir.

2.1 Veri Seti

Araştırma verileri PubMed veri tabanı kullanarak en son 22.03.2024 tarihinde taranmıştır. Tarama sürecinde, veri tabanı arayüzlerinde bulunan daraltıcılar kullanılarak verilere sınırlamalar getirilmiştir. Sonuçlar, 2014-2024 yılları arasındaki on yıllık bir dönemi kapsayacak şekilde daraltılmış ve belge türü olarak klinik çalışmalar tercih edilmiştir. Çalışma, farklı dillerde yayımlanan makalelerin kelime bulutu analizlerinden veya trend konularından kaynaklanabilecek karmaşıklığı önlemek amacıyla uluslararası kabul gören İngilizce metinlerle sınırlı tutulmuştur. Anahtar kelime olarak 'artificial intelligence' kullanılmıştır. Bunun sonucunda 3416 klinik çalışmaya rastlanmıştır. Tüm işlemlerin ardından veriler PubMed dosya formatında veri tabanı üzerinden dışarı aktarılmıştır. PubMed dosyası Biblioshiny web arayüzüne yüklenmiş ve analizler gerçekleştirilmiştir.

2.2 Biblioshiny

Bibliyometrik analiz R, Python, VOSviewer vb. birçok yazılım ve program paketleri ile yapılabilmektedir [8, 10, 11]. Çalışmada yapılan analiz için tercih edilen Biblioshiny web ara yüzü, SCOPUS, Web of Science, Cochrane, PubMed ve Digital Science Dimensions veri tabanlarından bibliyografik verileri aktarabilmektedir. Biblioshiny web ara yüzü R yazılım dilinin bibliyometri paketlerinden Bibliometrix ile birlikte çalışmaktadır. R yazılım dilinin, yüksek kaliteli istatistiksel algoritmalar, sayısal hesaplamalar için güçlü rutinler ve veri görselleştirme araçları gibi özellikleri, bilimsel hesaplamalar için diğer dillere göre tercih edilmesini sağlamaktadır [13].

3 Bulgular

PubMed veri tabanı kullanılarak 'artificial intelligence' anahtar kelimesi üzerinden arama yapılmıştır. 2014-2024 tarihleri arasında yayınlanmış ve klinik çalışma makalesi olma şartı ile sonuçlar filtrelenmiştir. Filtreleme sonucunda son 10 yıl içinde 693 farklı dergiden elde edilen verilere dayanarak 1939 adet klinik çalışma makalesi yayınlandığı ve 12331 yazarın ilgili konular üzerinde çalıştığı görülmüştür. Bu sonuçlar Şekil 1'de gösterilmiştir.



Şekil 1. Yayınlanan makaleler hakkında genel bilgiler

Tıp alanında yapılan klinik çalışma makalelerinin yıllara göre yayın sayısı Tablo 1’de görülmektedir. Yıl bazlı yayın sayılarına bakıldığında yayımların her geçen yıl giderek arttığı görülmektedir. Yıllar içinde tıp alanında farklı gelişmeler bu çalışma konularına ve sayılarına etki edebilmektedir. Özellikle 2020 yılında çıkan COVID-19 hastalığı bu çalışmaların artmasında önemli bir parametredir [14]. Aynı zamanda yılların içerisinde teknolojinin gelişmesi ve yapay zekâ algoritmalarının kullanılabilirliğinin artması da bu artışta önemli rol oynamaktadır. Bu bağlamda tıp ve yapay zekâ alanında yapılan çalışmaların yıllar içinde artması beklenmektedir.

Tablo 1. Yayımlanan makalelerin yıl bazlı yayın sayıları

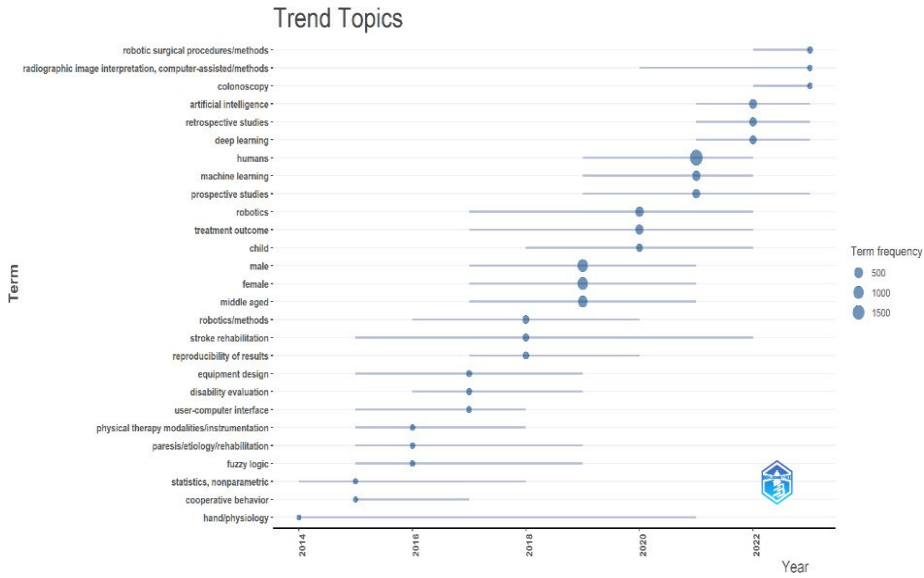
Yıl	Yayın Sayısı
2014	64
2015	96
2016	86
2017	95
2018	143
2019	172
2020	205
2021	275
2022	366
2023	369
2024*	68

*2024 yılı için Mart 22’ye kadar yapılan çalışmalar dahildir.

Makalelerde tekrar eden anahtar kelimeler içi hazırlanan bulut Şekil 2’de gösterilmiştir. Bu şekilden görüleceği üzere; insan, yaş, cinsiyet gibi parametrelerde robotik, yapay zekâ, makine öğrenimi ve algoritmaları gibi yöntemler çalışmalarda kullanılan anahtar kelimeler arasında daha yaygın olduğu görülmektedir. Aynı zamanda ileriye ve geriye dönük çalışmalarda tanı, teşhis ve tedavi gibi anahtar kelimelerle sıklıkla karşılaşmaktadır. Bu bağlamda klinik çalışmalarda yapay zekâ kullanılması hastalıkların tanı, teşhis ve tedavi süreçleri için önemli olduğu söylenebilmektedir. Bu çalışmaların klinik olması dolayısı ile çalışmaların teoride kalmadığı ve pratik olarak yapıldığı ve belirli sonuçlar alındığı kanısına da varılabilmektedir.



Şekil 2. Makalelerde sıklıkla kullanılan anahtar kelime bulutu



Grafik 1. Makalelerin yaygın araştırma terimlerinin yıl bazlı frekans grafiği

Grafik 1 makalelerin yaygın araştırma terimlerinin yıl bazlı frekans grafiğini göstermektedir. Grafik 1'e bakıldığında; 2014-2016 yılları arasında psikolojik (pyschology) çalışmalar, parametrik olmayan istatistiklerle (non-parametric, statistics) yapılan çalışmalar ve işbirlikçi davranış (cooperative behavior) içeren çalışmaların daha yoğun bir şekilde çalışıldığı görülmektedir. 2016-2018 yıllarında daha yoğun olarak fizik tedavi (physical therapy), rehabilitasyon (rehabilitation) konular özellikle bulanık mantık (fuzzy logic) yöntemi kullanılarak gerçekleştirilmiştir. 2016'dan başlayarak 2020'ye kadar çeşitli dönemlerde ekipman dizaynı (equipment design), engellilik değerlendirmesi (disability evaluation) ve kullanıcı bilgisayar ara yüzü (user-computer interface) gibi konular ve felç

sonular ıkarılması aısından nemli olduėu sylenebilir. Robotik ve rehabilitasyon ile ilgili alıřmalar ise klinikte gerekleřtirilen en temel uygulamaların bu alanlarda gerekleřtiėi sonucunu gstermektedir.

4 Sonular

Bu alıřmada, tıp alanında yapay zekâ kullanılarak gerekleřtirilen klinik alıřmaların bibliyometrik analizi R Bibliometrix paketi zerinden Biblioshiny web arayz kullanılarak gerekleřtirilmiřtir. Analiz sonuları, klinik alıřmaların her geen yıl arttıėını gstermektedir. Bu artıř, tıbbi geliřmelere, salgın hastalıklara, teknolojik ilerlemelere ve diėer parametrelere baėlı olarak deėiřmektedir. Makalelerde sıka geen anahtar kelimeler ve bu kelimeler arasındaki iliřki aėları incelendiėinde, robotik cerrahi, tıbbi grntleme sistemlerinin yorumlanması, fel rehabilitasyonu, parezi ve etiyoloji rehabilitasyonu, fizik tedavi yntemleri, enstrmantasyon, engellilik deėerlendirmesi gibi konuların popler olduėu grlmektedir. Bu alıřmalarda makine ėrenimi, yapay zekâ, bulanık mantık, istatistik, insan-makine arayz ve bilgisayar destekli zm yntemleri kullanılarak sonulara ulařıldıėı sylenabilmektedir. Ayrıca, yapay zekâ uygulamalarının klinik alıřmaların artması insanlar zerinde denenmesi artık alıřmaların teoride kalmadıėının nemli bir belirtisidir. Makine ėrenimi, derin ėrenme, bulanık mantık gibi kavramlara sıklıkla rastlanması, yapay zekanın farklı yntem ve algoritmalarla eřitli sorunların zmnde ve analizinde kullanıldıėını gstermektedir. Yapay zekâ kullanılarak yapılan analizler, hastalıklar zerinden anlamlı sonular ıkarılmasını saėlayarak farklı tedavi yntemlerinin geliřtirilmesine olanak tanırken, hastalıkların belirtilerini hızlı bir Őekilde ėrenerek tanı koyma srecini kısaltabilir. Bunlara ek olarak anahtar kelimelere ve kelimelerin birlikte oluřum aėlarına baktıėımızda ise tıp alanında yapay zekâ kullanımına ek olarak biyomedikal, fizik tedavi, robotik cerrahi gibi farklı disiplinlerin de bu klinik alıřmalara katkıda bulunduėu sylenabilmektedir. Ayrıca, doktorların belirli bir uzmanlık alanları olmasına raėmen yapay zekâ farklı uzmanlıkları kapsayıcı bir bilgi yelpazesine sahip olabilir. Bu nedenle tanı ve teřhislerin doėruluk oranını arttıracak Őekilde doktorların teřhisini destekleyecek Őekilde sistemlere eklenebilirler. Bu bulgular, yapay zekânın tıp alanında hastalıkların tanı, teřhis ve tedavi srelerinde nemli bir rol oynadıėını ortaya koymaktadır. Bunlara ek olarak yapılan bibliyometrik analiz bu konularda en fazla yayın yapan yazarlar ve gruplar ile bunların yayın frekansları ve lkeleri hakkında da bilgiler sunabilmektedir. Bu alıřma kapsamında yapay zekâ yntemleri ve klinik alıřmalar ile farklı disiplinler arasındaki iliřkiler ortaya konmak istendiėinden bu bulgulara yer verilmemiřtir. Daha detaylı bir alıřmada klinik uygulamaların yaygınlıėı, yapay zekâ kullanımı ile ilgili lke geliřmiřlik seviyelerinin etkilerinin de arařtırıldıėı bir alıřmada bu sonulara da yer verilebilir. alıřmanın geliřtirilebilir bir diėer ėesi de yapılan analizin farklı veri tabanlarında da benzer sonular verip vermediėinin irdelenmesi olabilir. Bu sayede hem analizin performansı deėerlendirilirken hem de bilim insanlarının yaptıkları alıřmaların grnrlė konusunda ne tıp eėilimlerde buldukları (rn; hangi eyreklik dergilerde yayınladıkları vb.) izlenebilir ve bu da hızlı geliřen alanlarda

çalışmalar yapan başka bilim insanları için yol gösterici olabilir. Dolayısıyla, bu çalışmanın bulguları, eksik çalışma konularını belirleyerek literatürdeki boşlukların doldurulmasına ve yeni yöntemlerin keşfedilerek bu alandaki çalışmalara rehberlik etme potansiyeline sahiptir.

KAYNAKÇA

- [1] A. Demirhan, Y. Alper KILIÇ, İ. Güler, G. Üniversitesi Teknoloji Fakültesi, and B. Mühendisliği Bölümü, "Tıpta Yapay Zeka Uygulamaları* Artificial Intelligence Applications in Medicine," 2010.
- [2] J. Salcedo, M. Rosales, J. S. Kim, D. Nuno, S. Suen, and A. H. Chang, "Cost-effectiveness of artificial intelligence monitoring for active tuberculosis treatment: A modeling study," *PLoS One*, vol. 16, no. 7, p. e0254950, Jul. 2021, doi: 10.1371/journal.pone.0254950.
- [3] E. Rondonotti *et al.*, "Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: the Artificial intelligence BLI Characterization (ABC) study," *Endoscopy*, vol. 55, no. 01, pp. 14–22, Jan. 2023, doi: 10.1055/a-1852-0330.
- [4] J. D. Piette *et al.*, "Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools," *JAMA Intern Med*, vol. 182, no. 9, p. 975, Sep. 2022, doi: 10.1001/jamainternmed.2022.3178.
- [5] S. Mertens, J. Krois, A. G. Cantu, L. T. Arsiwala, and F. Schwendicke, "Artificial intelligence for caries detection: Randomized trial," *J Dent*, vol. 115, p. 103849, Dec. 2021, doi: 10.1016/j.jdent.2021.103849.
- [6] H. KELEŞ, "TIPTA YAPAY ZEKA UYGULAMALARI," *Kırıkkale Üniversitesi Tıp Fakültesi Dergisi*, vol. 24, no. 3, pp. 604–613, Dec. 2022, doi: 10.24938/kutfd.1214512.
- [7] C. Cuccurullo, M. Aria, and F. Sarto, "Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains," *Scientometrics*, vol. 108, no. 2, pp. 595–611, Aug. 2016, doi: 10.1007/s11192-016-1948-8.
- [8] S. Büyükkidik, "A Bibliometric Analysis: A Tutorial for the Bibliometrix Package in R Using IRT Literature," *Journal of Measurement and Evaluation in Education and Psychology*, vol. 13, no. 3, pp. 164–193, Sep. 2022, doi: 10.21031/EPOD.1069307.
- [9] R. Bailón-Moreno, E. Jurado-Alameda, R. Ruiz-Baños, and J. P. Courtial, "Bibliometric laws: Empirical flaws of fit," *Scientometrics*, vol. 63, no. 2, pp. 209–229, Apr. 2005, doi: 10.1007/s11192-005-0211-5.
- [10] K-Synth Srl, "BIBLIOMETRIX," Academic Spin-Off of the University of Naples Federico II. All Rights Reserved.
- [11] B. Toaza and D. Esztergár-Kiss, "Automated bibliometric data generation in Python from a bibliographic database," *Software Impacts*, vol. 19, p. 100602, Mar. 2024, doi: 10.1016/j.simpa.2023.100602.
- [12] Y. Yu *et al.*, "A bibliometric analysis using VOSviewer of publications on COVID-19," *Ann Transl Med*, vol. 8, no. 13, p. 816, Jul. 2020, doi: 10.21037/atm-20-4235.

- [13] M. Aria, C. Cuccurullo, and F. Sarto, "Exploring healthcare governance literature: systematic review and paths for future research," *MECOSAN*, no. 91, pp. 61–80, Jan. 2015, doi: 10.3280/MESA2014-091004.
- [14] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337–339, Jul. 2020, doi: 10.1016/j.dsx.2020.04.012.

Predictors of Neuropathy in Diabetes Mellitus

Nevruz İlhanlı¹[0000-0002-4777-4025] and Kemal Hakan Gülkesen¹[0000-0002-2477-2481]

¹ Akdeniz University, Antalya, Turkey

Abstract. Diabetes mellitus is a major public health concern with a growing number of cases. Diabetes mellitus is associated with some complications including neuropathy. Neuropathy is the most common complication of diabetes and the main cause of non-traumatic amputations. Early prediction of neuropathy in diabetic patients is crucial to reduce neuropathy-related complications and morbidity. In the present study, we aimed to predict neuropathy in patients with diabetes mellitus by using the logistic regression method. For this aim, a large cohort that includes the EHR of 77,724 diabetic patients was used. Among 105 features in the dataset, 47 features were found to be significant predictors of neuropathy in patients with diabetes mellitus and included in the final model. A 5-fold cross-validation approach was used for modeling. The logistic regression method achieved an average AUC of 0.830. Further studies should be conducted with different populations and advanced machine learning techniques to increase the accuracy and generalizability of the model.

Keywords: Logistic Regression, Neuropathy, Electronic Health Records

1 Introduction

Diabetes mellitus is a metabolic dysfunction in which blood sugar levels remain high due to insufficient secretion of the insulin hormone in the body [1]. Diabetes mellitus is a major public health challenge with 537 million cases worldwide in 2021 [2]. Coronary heart disease, stroke, peripheral artery disease, kidney disease, retinopathy, and neuropathy are common complications of diabetes [3]. Among these complications, neuropathy is the most common complication of diabetes mellitus and the main cause of non-traumatic amputations [4]. Early prediction of neuropathy in diabetic patients may reduce neuropathy-related complications and morbidity [5, 6].

Logistic regression is a statistical method that enables multivariate analysis and modeling of a binary outcome [7]. Logistic regression is a widely used method in many areas including health, engineering, and social sciences [8]. In medical research, logistic regression has become an increasingly popular method, particularly in the past twenty years [9]. For neuropathy prediction, promising results were obtained in previous efforts by using the logistic regression method [10-12]. However, past efforts worked with a relatively small cohort which reduces the predictive performance and generalizability of prediction models.

The present study aims to predict neuropathy in patients with diabetes mellitus by using logistic regression method.

2 Method

2.1 Data Set

In the present study, electronic health records (EHR) of the patients diagnosed with diabetes mellitus in Istanbul, Turkey in 2017 were used [13]. The dataset includes demographic, diagnosis, laboratory test, and prescription records of 77,724 diabetic patients.

2.2 Data Preprocessing

The original dataset includes 105 features. After performing logistic regression analysis with all features, significant features were included in the final model. Python (version 3.9.7) was used for data preprocessing.

2.3 Modeling

The logistic regression method was used for modeling. A 5-fold cross-validation approach was used to train and evaluate the prediction model. The prediction model was built with default hyper-parameters. To evaluate the predictive performance of the model, accuracy, area under the receiver operating characteristic curve (AUC), F_1 -score, and Matthew's correlation coefficient (MCC) were used. Python (version 3.9.7) was used for modeling and evaluation.

2.4 Statistical Analysis

Jamovi (Version 1.6.23.0) was used for statistical analysis. Logistic regression analysis was performed to analyze the association between neuropathy and input variables. A p-value lower than 0.05 was considered significant for all statistical analyses.

3 Results

In the dataset, 19,559 (25.2%) patients were diagnosed with neuropathy and 58,165 (74.8%) patients were not. In the baseline model, 47 features were found to be significant predictors of neuropathy in patients with diabetes mellitus. Significant predictors of neuropathy in patients with diabetes mellitus are presented in Table 1.

Table 1. Logistic regression analysis results

Variables	Odd's ratio (95% CI)	p-value
HbA1c	1.066 (1.052 - 1.080)	< .001
HbA1c change	0.890 (0.826 - 0.958)	0.002
HDL	1.002 (1.000 - 1.005)	0.048
Creatinine	0.970 (0.942 - 1.000)	0.046
Thyroid diseases	0.918 (0.863 - 0.976)	0.006
In situ neoplasms, benign neoplasms, neoplasms of uncertain or unknown behavior	0.882 (0.803 - 0.969)	0.009
Vitamin deficiencies	0.664 (0.635 - 0.694)	< .001
Disorders of lipoprotein metabolism and other lipidaemias	1.397 (1.314 - 1.486)	< .001
Other Endocrine disorders	0.901 (0.861 - 0.944)	< .001
Bipolar affective disorder	0.491 (0.420 - 0.575)	< .001
Other nervous system disorders	1.110 (1.053 - 1.169)	< .001
Retinopathies	1.203 (1.136 - 1.275)	< .001
Disorders of refraction and accommodation	1.067 (1.021 - 1.116)	0.004
Other eye diseases	1.102 (1.036 - 1.172)	0.002
Other diseases of ear and mastoid process	1.076 (1.010 - 1.146)	0.023
Hypertension	1.171 (1.088 - 1.259)	< .001
Ischemic heart diseases	0.950 (0.903 - 1.000)	0.048
Other diseases of the circulatory system	1.076 (1.027 - 1.128)	0.002
Gastro-oesophageal reflux disease	1.085 (1.039 - 1.133)	< .001
Diseases of the skin and subcutaneous tissue	1.058 (1.005 - 1.114)	0.032
Diseases of the musculoskeletal system and connective tissue	1.329 (1.245 - 1.418)	< .001
Renal failure	0.890 (0.823 - 0.962)	0.003
Some alimentary tract preparations	4.860 (4.603 - 5.132)	< .001
Cardiovascular system drugs	0.931 (0.867 - 0.999)	0.047
Lipid modifying drugs	0.906 (0.852 - 0.963)	0.001
Genitourinary system drugs	1.075 (1.016 - 1.138)	0.013
Vaccines	0.910 (0.848 - 0.977)	0.009
Immunostimulants	0.463 (0.229 - 0.932)	0.031
Muskuloskeletal system drugs	1.088 (1.023 - 1.157)	0.007
Antiepileptics	12.158 (11.578 - 12.767)	< .001
Antiparkinson drugs	1.442 (1.247 - 1.668)	< .001

Antipsychotics	0.726 (0.658 - 0.801)	< .001
Psychoanaleptics	1.208 (1.132 - 1.289)	< .001
Respiratory system drugs	0.912 (0.859 - 0.967)	0.002
Dapagliflozin (per 1 gr/year)	1.044 (1.004 - 1.085)	0.029
Gliclazide (per 1 gr/year)	1.006 (1.004 - 1.008)	< .001
Insulin aspart (per 1000U/year)	1.028 (1.022 - 1.034)	< .001
Insulin detemir (per 1000U/year)	1.014 (1.009 - 1.020)	< .001
Insulin glargine (per 1000U/year)	1.017 (1.013 - 1.021)	< .001
Insulin glulisine (per 1000U/year)	1.067 (1.050 - 1.084)	< .001
Insulin lispro (per 1000U/year)	1.032 (1.022 - 1.042)	< .001
Linagliptin (per 1 gr/year)	1.332 (1.245 - 1.425)	< .001
Nateglinide (per 1 gr/year)	33.055 (11.550 - 94.602)	< .001
Pioglitazone (per 1 gr/year)	1.028 (1.022 - 1.035)	< .001
Saxagliptin (per 1 gr/year)	1.134 (1.034 - 1.243)	0.007
Sitagliptin (per 1 gr/year)	1.006 (1.003 - 1.009)	< .001
Vildagliptin (per 1 gr/year)	1.010 (1.007 - 1.013)	< .001

The average performance metrics of logistic regression method on the original dataset across 5-fold are as follows: 0.723 accuracy, 0.764 AUC, 0.541 F1-score, and 0.361 MCC. Performance analysis of the logistic regression method on original dataset in predicting neuropathy in patients with diabetes mellitus is presented in Table 2.

Table 2. Performance analysis of logistic regression method on the original dataset across 5-fold

Fold	Accuracy	AUC	F₁-score	MCC
1	0.725	0.771	0.553	0.370
2	0.731	0.756	0.553	0.355
3	0.733	0.780	0.557	0.386
4	0.703	0.752	0.526	0.337
5	0.723	0.759	0.536	0.355
Average	0.723	0.764	0.541	0.361

The The average performance metrics of logistic regression method on the original dataset across 5-fold are as follows: 0.779 accuracy, 0.830 AUC, 0.614 F1-score, and 0.471 MCC. Performance analysis of the logistic regression method in predicting neuropathy in patients with diabetes mellitus is presented in Table 3.

Table 3. Performance analysis of the logistic regression method on the dataset with 47 significant features

Fold	Accuracy	AUC	F₁-score	MCC
1	0.775	0.838	0.624	0.476
2	0.759	0.831	0.604	0.454
3	0.775	0.835	0.614	0.470
4	0.775	0.824	0.608	0.462
5	0.813	0.822	0.619	0.496
Average	0.779	0.830	0.614	0.471

The receiver operating characteristic curves (ROC) of the logistic regression method on the original dataset and the dataset with significant features are presented in Figure 1.

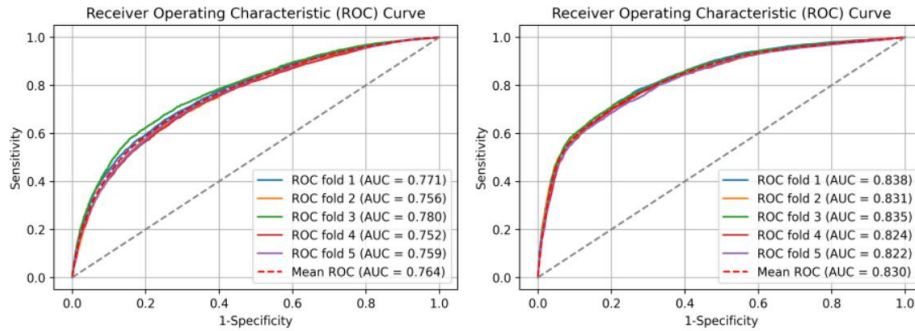


Fig. 1. The receiver operating characteristic (ROC) curve of logistic regression method on the original dataset (on the left) and the dataset with significant features (on the right) for 5-fold cross-validation.

4 Discussion

In the present study, the logistic regression method was used to predict neuropathy in patients with diabetes mellitus. Demographic, laboratory tests, and prescription records of patients were used as predictors of neuropathy in diabetic patients. Herein, we worked with a large cohort that includes the EHR of 77,724 diabetic patients. Firstly, logistic regression analysis was performed to determine significant predictors of neuropathy. According to logistic regression analysis results, 47 features were significant predictors of neuropathy in patients with diabetes mellitus. The logistic regression method had 0.764 AUC on the original dataset and 0.830 AUC on the dataset with 47 significant features. The logistic regression method demonstrated satisfactory predictive performance on both sets. However, the performance of logistic regression was higher on the dataset that includes solely significant features. This highlights the importance of excluding unnecessary features for our prediction task.

In previous efforts, logistic regression achieved promising predictive performance with AUC values ranging from 0.73 to 0.88 [10-12] in predicting neuropathy in patients with diabetes mellitus. The predictive performance of our prediction model is comparable with previous efforts.

The main limitation of the study is using data from the single population, which has the potential to limit the generalizability of the results to broader populations. Additionally, we solely used the logistic regression method for prediction. However, machine learning techniques may achieve superior predictive performance with our large cohort. Further studies should be conducted to overcome these limitations.

In conclusion, our study may contribute to the literature on neuropathy prediction in patients with diabetes mellitus. Although the logistic regression method demonstrated satisfactory predictive performance on this task, advanced machine learning techniques can provide more accurate models.

References

1. Khan, R.M.M., et al., *From Pre-Diabetes to Diabetes: Diagnosis, Treatments and Translational Research*. Medicina (Kaunas), 2019. **55**(9).
2. Collaborators, G.B.D.D., *Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021*. Lancet, 2023. **402**(10397): p. 203-234.
3. Tomic, D., J.E. Shaw, and D.J. Magliano, *The burden and risks of emerging complications of diabetes mellitus*. Nat Rev Endocrinol, 2022. **18**(9): p. 525-539.
4. UZUNER, B., S. KETENCİ, and E. SALBAŞ, *General approach to diabetic neuropathy*. Acta Medica Alanya, 2020. **4**(3): p. 296-308.
5. Kazemi, M., et al., *Diabetic peripheral neuropathy class prediction by multiclass support vector machine model: a cross-sectional study*. Epidemiol Health, 2016. **38**: p. e2016011.
6. Tantigegn, S., et al., *Time to diabetic neuropathy and its predictors among adult type 2 diabetes mellitus patients in Amhara regional state Comprehensive Specialized Hospitals, Northwest Ethiopia, 2022: A retrospective follow up study*. PLoS One, 2023. **18**(4): p. e0284568.
7. Shipe, M.E., et al., *Developing prediction models for clinical use using logistic regression: an overview*. J Thorac Dis, 2019. **11**(Suppl 4): p. S574-S584.
8. Kim, A., et al., *Logistic regression model training based on the approximate homomorphic encryption*. BMC Med Genomics, 2018. **11**(Suppl 4): p. 83.
9. Boateng, E.Y. and D.A. Abaye, *A review of the logistic regression model with emphasis on medical research*. Journal of data analysis and information processing, 2019. **7**(04): p. 190.
10. LI, C.-p., et al., *Performance comparison between Logistic regression, decision trees, and multilayer perceptron in predicting peripheral neuropathy in type 2 diabetes mellitus*. Chinese Medical Journal, 2012. **125**(5): p. 851-857.
11. Lian, X., et al., *Study on risk factors of diabetic peripheral neuropathy and establishment of a prediction model by machine learning*. BMC Med Inform Decis Mak, 2023. **23**(1): p. 146.
12. Gelaw, N.B., et al., *Development and validation of risk prediction model for diabetic neuropathy among diabetes mellitus patients at selected referral hospitals, in Amhara regional state Northwest Ethiopia, 2005–2021*. PLOS ONE, 2023. **18**(8): p. e0276472.
13. Gulkesen, K.H., et al., *Machine learning for prediction of glycemic control in diabetes mellitus*. 2022, Mendeley Data.

Fine-Tuning GPT-2 for Contextually Relevant Text Generation to Discover New Immune Checkpoint Research

Mehmet Erşan KALAYCI¹[0000-0002-4797-9020], Muammer ALBAYRAK¹[0000-0002-5946-6310],
Kemal TURHAN¹[0000-0001-7871-3025]

¹ Karadeniz Teknik Üniversitesi, Trabzon, Türkiye,
mehmetersankalayci@ktu.edu.tr
m.albayrak@ktu.edu.tr
kemalturhan@ktu.edu.tr

Abstract. In this study, we aimed to access new immune checkpoint specific information by fine-tuning the text generation capabilities of the GPT-2 language model developed by OpenAI. Using a dataset of abstracts from 7,000 articles obtained through the search query “new immune checkpoint” in the Scopus database, we fine-tuned the model to perform better in this specific type of text. The training and validation losses decreased from 2.05 to 0.51 and 1.92 to 1.28, respectively, over 5 epochs, indicating improved performance. By enhancing the model's ability to understand the structure and content of texts, we increased its capacity to generate more accurate and consistent texts for our personal use. This study highlights the effectiveness of fine-tuning language models to improve text generation capabilities. Immune checkpoint inhibitors are promising agents in cancer treatment. This study aims to demonstrate the qualitative effectiveness of using the GPT-2 model for fine-tuning in the discovery of new immune checkpoints.

Keywords: GPT-2, fine-tuning, text generation, immune checkpoints, Scope database.

1 Introduction

Immune checkpoints play a crucial role in regulating the immune response, ensuring that it is both effective and controlled [1]. These checkpoints are essential for maintaining self-tolerance and preventing autoimmunity, but they can also be exploited by cancer cells to escape detection and destruction by the immune system [2]. The discovery of immune checkpoints such as CTLA-4 and PD-1/PD-L1 has revolutionized cancer treatment [3]. This discovery prevented cancerous cells from escaping the immune system. In other words, it increases the success of the immune system in fighting cancerous cells.

Despite the success of immune checkpoint inhibitors in some cancers, many patients do not respond to these treatments or develop resistance over time [4]. This highlights the need to identify new immune checkpoints that could serve as targets for novel immunotherapies. In this study, we aimed to identify new immune checkpoint candidates using the large summary dataset in the Scopus database. We leveraged advanced natural language processing techniques to extract relevant information from these abstracts, focusing on identifying sentences that describe potential new immune checkpoints.

By fine-tuning the GPT-2 language model developed by OpenAI, we trained the model to better understand the structure and content of biomedical texts related to immune checkpoints [5]. Our goal was to enhance the model's ability to identify and generate sentences that describe new immune checkpoints. We focused on describe new immune checkpoints candidates by training the model on a specifically selected dataset to enhance its performance in generating new text from existing immune checkpoints. Promising immune checkpoints in immunotherapy and the influencing factors have not been fully elucidated. Our aim in this study is to use large language models to define new immune checkpoints from the literature in more detail and thus obtain a more efficient data set in new immune check point discovery studies.

2 Methodology

The methodology of this study involved several key steps. First, we collected a dataset of abstracts from the Scopus database related to immunology and immune checkpoints. Next, we preprocessed the text data to remove noise and irrelevant information, ensuring that only relevant text was used for analysis.

We then used the preprocessed dataset to fine-tune the GPT-2 language model developed by OpenAI. Fine-tuning involved training the model on our dataset to enhance its text generation capabilities specifically for the task of identifying new immune checkpoints. This process allowed the model to learn the patterns and structures present in the dataset, improving its ability to generate coherent and relevant text.

After fine-tuning the model, we evaluated its performance using various metrics, such as the quality of generated text and its relevance to the topic of immune checkpoints. Additionally, we compared the performance of the fine-tuned model with the original GPT-2 model to assess the improvement in text generation capabilities [6].

This study focused on leveraging machine learning techniques and advanced language models to identify new immune checkpoints to contribute study of discovering immune checkpoint.

In this study, we set several important parameters for training the GPT-2 language model. We trained the model for 5 epochs, meaning it iterated over the entire dataset 5 times. We used a learning rate of $5e-4$, which determines the size of the steps the optimizer takes to find the optimal model parameters [7].

To prevent the model from converging too quickly to a suboptimal solution, we employed a warmup strategy with 100 warmup steps. This means that for the first 100 steps of training, the learning rate gradually increases from 0 to the specified value, helping the model explore the solution space more effectively.

Additionally, we used an epsilon value of $1e-8$ for numerical stability in the optimizer. This small value is added to the denominator to prevent division by zero or very small numbers, which can lead to numerical instability during training.

To monitor the model's progress and generate sample output for evaluation, we produced sample output every 100 steps. This allows us to observe how the model's outputs change over the course of training and assess its performance. The AdamW optimizer, used in this study, is a variant of the Adam optimizer that incorporates weight decay, making it suitable for fine-tuning large language models like GPT-2.

3 Results

Table1 shows the training and validation losses for each epoch during the training of the model. The training loss represents the error of the model on the training dataset, while the validation loss represents the error on a separate validation dataset. As the training progresses, both losses decrease, indicating that the model is learning and improving its performance.

Table 1. Training and Validation Losses by Epoch.

Epoch	Training loss	Validation loss
1	2.05	1.92
2	1.20	1.79
3	0.90	1.78
4	0.62	1.50
5	0.51	1.28

Table2 summarizes the shape of the parameters for each layer of the model. The shapes indicate the dimensions of the weight matrices and bias vectors used in the model.

Table 2. Model Parameters.

Layer	Parameter	Shape
Embedding Layer	transformer.wte.weight	(50259, 768)
	transformer.wpe.weight	(1024, 768)
First Transformer Layer	transformer.h.0.ln_1.weight	(768,)

	transformer.h.0.ln_1.bias	(768,)
	transformer.h.0.attn.c_attn.weight	(768, 2304)
	transformer.h.0.attn.c_attn.bias	(2304,)
	transformer.h.0.attn.c_proj.weight	(768, 768)
	transformer.h.0.attn.c_proj.bias	(768,)
	transformer.h.0.ln_2.weight	(768,)
	transformer.h.0.ln_2.bias	(768,)
	transformer.h.0.mlp.c_fc.weight	(768, 3072)
	transformer.h.0.mlp.c_fc.bias	(3072,)
	transformer.h.0.mlp.c_proj.weight	(3072, 768)
	transformer.h.0.mlp.c_proj.bias	(768,)
Output Layer	transformer.ln_f.weight	(768,)
	transformer.ln_f.bias	(768,)

In the presented examples, fine-tuning the GPT-2 model resulted in text generation that is contextually relevant to specific topics, such as cancer research. Table 3 demonstrates how the model, after fine-tuning, produces text that aligns with the domain of cancer research.

Table 3. Qualitative evaluation for the first paragraph

Qualitative evaluation for the first paragraph			
Before Fine-Tuning¹	qualitative assessment	After Fine-Tuning	qualitative assessment
To discover new immune checkpoints for immunotherapy, a systematic approach can be followed.	general	Immune checkpoint research has been on a high level and many applications where it has been useful to learn oncology have been built upon it.	medium related
This involves conducting an extensive literature review to understand the current knowledge and identify gaps in understanding.	general	In this review, we initially give an overview of the steps in the process of discovering immune checkpoint markers for the clinic through the web-site immunecheckpointing.org method and then we go over some of the recent successes and some of the barriers in the field.	highly related
Relevant data, including genomic, transcriptomic, and proteomic data from healthy individuals and patients with diseases relevant to immune modulation, is collected.	medium related	With a little introduction to how the field of checkpoint research has evolved, we focus on:	
This data is then cleaned and preprocessed to remove noise and ensure its quality and reliability.	general	1) breakthrough of checkpoint markers via cross-referencing of cellular and molecular features	medium related
Features associated with immune checkpoints are identified from the preprocessed data using statistical tests or machine learning	general	2) immune checkpoint classification and prediction of response to immunotherapy	highly related

¹ It was fine-tuned with 7000 articles obtained with the search query "new immune checkpoint, CTLA 4, PD1/PDL1".

algorithms.			
A computational model, such as a machine learning model, is developed to predict novel immune checkpoints based on the selected features.	highly related	3) antigen binding to cancer cells	highly related
Finally, the predicted immune checkpoints are validated using independent datasets or experimental validation techniques to ensure their reliability and relevance.	medium related	4) antigen-carrying cell 2 markers (e.g., checkpoint cytokines and STAT3) and cell fusion	highly related
This systematic framework allows for the exploration of the immune checkpoint landscape and the identification of new targets for immunotherapy.	general	5) checkpoint blockade and chemotherapeutic targets such as receptor tyrosine kinase, cGAS and PXCR4 (checkpoint inhibitory factors, MAPK/STAT3 and NOD/OBP inhibition, and CTLA4/4) have been reviewed. © 2017 by the authors. Licensee MDPI, Basel, Switzerland.	highly related

Table 4. Comparison of generated texts.

Text generated without Fine Tuning	Texts generated with Fine Tuning
Generated Texts	
I want to work on new immune checkpoint discovery, what kind of working framework would you recommend?	
To discover new immune checkpoints for immunotherapy, a systematic approach can be followed. This involves conducting an extensive literature review to understand the current knowledge and identify gaps in understanding. Relevant data, including genomic, transcriptomic, and proteomic data from	Immune checkpoint research has been on a high level and many applications where it has been useful to learn oncology have been built upon it. In this review, we initially give an overview of the steps in the process of discovering immune checkpoint markers for the clinic through the web-site immunecheckpointing.org

<p>healthy individuals and patients with diseases relevant to immune modulation, is collected. This data is then cleaned and preprocessed to remove noise and ensure its quality and reliability. Features associated with immune checkpoints are identified from the preprocessed data using statistical tests or machine learning algorithms. A computational model, such as a machine learning model, is developed to predict novel immune checkpoints based on the selected features. Finally, the predicted immune checkpoints are validated using independent datasets or experimental validation techniques to ensure their reliability and relevance. This systematic framework allows for the exploration of the immune checkpoint landscape and the identification of new targets for immunotherapy.</p>	<p>method and then we go over some of the recent successes and some of the barriers in the field. With a little introduction to how the field of checkpoint research has evolved, we focus on:</p> <ol style="list-style-type: none"> 1. breakthrough of checkpoint markers via cross-referencing of cellular and molecular features 2. immune checkpoint classification and prediction of response to immunotherapy 3. antigen binding to cancer cells 4. antigen-carrying cell 2 markers (e.g., checkpoint cytokines and STAT3) and cell fusion 5. checkpoint blockade and chemotherapeutic targets such as receptor tyrosine kinase, cGAS and PXR4 (e.g., checkpoint inhibitory factors, MAPK/STAT3 and NOD/OBP inhibition, and CTLA4/4) have been reviewed. © 2017 by the authors. Licensee MDPI, Basel, Switzerland.
<p>For the discovery of new immune checkpoints for immunotherapy, a structured framework is essential. Start with a thorough review of the existing literature on immune checkpoints to grasp the current understanding and pinpoint areas lacking in knowledge. Acquire pertinent data, including genomic, transcriptomic, and proteomic data from healthy individuals and those with diseases pertinent to immune modulation. Process the data to eliminate noise and ensure its quality and reliability. Employ feature selection techniques to identify features linked to immune checkpoints, leveraging statistical tests or machine learning algorithms. Develop a computational model, such as a machine learning model, to predict new immune checkpoints based on the selected features. Validate the predicted immune checkpoints using independent datasets or experimental validation methods. This systematic approach allows for a methodical exploration of the immune checkpoint landscape, potentially revealing new targets for immunotherapy.</p>	<p>I just created a workflow-based workflow, with an emphasis on basic steps to discovering and editing checkpoint-related workflow genes. The workflow consists of five steps: discovery, editing, screening and discovery. The workflow is designed to occur in two phases: discovery and screening. After initial screening of the three main diedavirus-associated checkpoints byproduct, the next step is editing to the next level. In this workflow, all of the available workflow genes have been tested. This workflow is designed to be dynamic as there is a possibility to exploit workflow mutations, which are highly conserved in all available human cells. In particular, a key workflow component is a mutation-based approach involving multiple checkpoints. The other five steps are to discover protein-coupled checkpoints (CPBs), sequence dependencies, nucleotide polymorphism and the candidate key genes that can be exploited to design checkpoints. In a work performed using this workflow, a mutation free molecular resource (e.g., sequence variants that include mismatch repair deficiency, insertions and polymorphism in the wild) is generated that can be exploited to design checkpoint-specific human-specific key genes.</p>

	<p>The workflow component consists of five steps: discovery, screening, and editing. Each step involves validation of the results of the analysis. This workflow is designed to be flexible in the selection of key genes and key genes required for specific validation.</p>
<p>To embark on the discovery of new immune checkpoints for immunotherapy, a systematic approach is crucial. Initiate with a comprehensive review of existing literature on immune checkpoints to comprehend the current landscape and identify areas ripe for exploration. Gather relevant data encompassing genomic, transcriptomic, and proteomic data from both healthy individuals and patients with diseases pertinent to immune modulation. Process this data meticulously to eliminate noise and ensure its quality and reliability. Employ sophisticated feature selection methods to pinpoint features closely associated with immune checkpoints, leveraging statistical tests or cutting-edge machine learning algorithms. Subsequently, construct a computational model, such as a machine learning model, to predict potential novel immune checkpoints based on the identified features. Validate these predictions using independent datasets or through rigorous experimental validation techniques. This systematic and meticulous approach paves the way for a structured exploration of the immune checkpoint landscape, potentially uncovering novel targets for immunotherapy.</p>	<p>I use the framework of dependency injection, which is used to create new classes of libraries for linking to biological processes (such as immunity against infection by small molecules or antibodies). In my case, I start from Basic Biological Processes (BOP) and then expand it by expanding it to incorporate more pathways, pathways, and cellular components. Then I consider applications in areas such as: pathology, basic research, medicinal chemistry, basic science, and pathology. I then design software for interacting with BOP to support new autoimmune diseases and prevent autoimmune conditions. I manage to find an integrative picture that best describes the biological process involved in each pathway. This frame of reference also helps me integrate pathways and cellular components involved in autoimmune diseases and further refine my therapeutic strategies. In this review, I am primarily interested in immune checkpoint kinases (ICKs), ctDNAs, NODa11, MTX3, TGFB1, and P53 signaling pathway. We can review the challenges associated with creating a working framework that integrates cascade-mediated immunity and autoimmune diseases to develop new immune checkpoint markers. © 2021 Wiley-VCH GmbH.</p>
<p>For a comprehensive exploration of new immune checkpoints in the context of immunotherapy, a structured framework is paramount. Begin with an in-depth review of the existing literature on immune checkpoints to grasp the current understanding and identify gaps in knowledge. Acquire relevant data, including genomic, transcriptomic, and proteomic data from both healthy individuals and patients with diseases relevant to immune modulation. Process this data meticulously to</p>	<p>Currently, there are four classes of checkpoint-based checkpoints - checkpoint-associated proteases (CPA), checkpoint-associated kinases (CAKs), hedgehog protein kinases (HPPKs), and tumor-associated antigens (TAAs). All classes of checkpoint proteins have been validated to have therapeutic potential. Some classes of checkpoint proteins, which we just recently discovered, hold promise of treating cancer. And, we just mentioned some of the proteins that are currently under investigation</p>

<p>ensure its quality and reliability, and then use advanced feature selection methods to identify features associated with immune checkpoints, employing statistical tests or machine learning algorithms. Develop a computational model, such as a machine learning model, to predict new immune checkpoints based on the selected features. Validate these predictions using independent datasets or experimental validation methods. This systematic approach enables a structured exploration of the immune checkpoint landscape, potentially leading to the discovery of novel targets for immunotherapy.</p>	<p>for checkpoint inhibition. The discovery of checkpoint-associated proteases has provided new insights into the mechanism of action of checkpoint proteins. In particular, it has highlighted their potential as biomarkers of cancer diagnosis, prognosis, and therapy response. We particularly recommend the newest generation of checkpoint proteins, checkpoint inhibitors with antiangiogenic effects, as well as checkpoint inhibitors that target MAPK and NF-κB pathways. © 2017 Elsevier Ltd</p>
<p>To discover new immune checkpoints for potential use in immunotherapy, a systematic framework is essential. Begin by conducting a thorough review of existing literature on immune checkpoints to understand the current landscape and identify areas where knowledge is lacking. Collect relevant data, including genomic, transcriptomic, and proteomic data from both healthy individuals and patients with diseases relevant to immune modulation. Preprocess this data to ensure its quality and reliability, and then use feature selection methods to identify features associated with immune checkpoints. Develop a computational model, such as a machine learning model, to predict novel immune checkpoints based on the selected features. Validate these predictions using independent datasets or experimental validation techniques. This systematic approach allows for a structured exploration of the immune checkpoint landscape and may lead to the discovery of new targets for immunotherapy.</p>	<p>I have recently come across an idea of how to implement an immune checkpoint-based novel therapy, based on checkpoint inhibitor-based immune checkpoint blockade for pancreatic ductal adenocarcinoma (PDAC). This idea of blockchain has recently attracted great interest among a number of physicists and computer scientists. A few years back, I suggested to computer scientists that blockchain which seems to be the most fundamental step toward discovery was probably the first cancer drug with an effective mechanism of action Drafting a set of basic information about pancreatic tumor biology for the discovery of immunoglobulin-like receptor (IGF-R) in pancreatic cancer. To demonstrate my earlier thesis, I revisited the topic of how to design a set of checkpoint blockers for pancreatic cancer based on the blockchain. In the end, this strategy seems to have satisfied many scientists who were still working on pancreatic cancer. In particular, it seems clear that if we can find a set of checkpoint blockers that can block multiple autoimmune diseases, we may find a new cancer drug that has the potential to be more effective with an increased degree of specificity. © 2019, Springer Nature Limited.</p>
<p>To discover new immune checkpoints for potential use in immunotherapy, a systematic approach is paramount. Begin by conducting a thorough review of existing literature on</p>	<p>I used to be a little more focused on natural killer cells and chimeric antigen receptor (CAR) T cells. I eventually got a little worked up on these cells because they worked well</p>

<p>immune checkpoints to understand the current landscape and identify areas where knowledge is lacking. Collect relevant data, including genomic, transcriptomic, and proteomic data from both healthy individuals and patients with diseases relevant to immune modulation. Preprocess this data to ensure its quality and reliability, and then use feature selection methods to identify features associated with immune checkpoints. Develop a computational model, such as a machine learning model, to predict novel immune checkpoints based on the selected features. Validate these predictions using independent datasets or experimental validation techniques. This systematic approach allows for a structured exploration of the immune checkpoint landscape and may lead to the discovery of new targets for immunotherapy.</p>	<p>together. Then, after a few more years, I got a little more experienced on CAR T cell receptors and started working on developing more potent CAR T cell receptor candidates. Then, I just got a little more experienced with chimeric antigen receptor (CAR) T cell receptor (CTRL), discovered that many CAR T cell receptor candidates are effective against tumors with immune suppression in the clinic. Once we establish the theory of CAR T cell receptor specificity, the next stage of research will be about the molecular mechanism of action of CAR T cells, and then CAR T cell receptor proteins such as IL-4 and CDKN2A become the focus of particular interest to researchers. © Copyright © 2020 Shui, Li, Zhou, Chen, Dai and Jiang.</p>
---	--

Fine-tuning models like GPT-2 for specific tasks or domains can enhance their ability to generate contextually relevant and coherent text, making them valuable tools for various applications, including scientific research and content generation in specialized fields. Specifically, the fine-tuned model was able to generate text that exhibited a better understanding of the topic of immune checkpoints compared to the original GPT-2 model (Table 3 and 4). The generated text included descriptions of potential new immune checkpoints and their role in regulating the immune response, providing valuable insights for further research in immunotherapy. Additionally, we compared the performance of the fine-tuned model with the original GPT-2 model using metrics such as text coherence, relevance, and novelty. The results showed that the fine-tuned model outperformed the original model in generating text relevant to immune checkpoints, indicating the effectiveness of our approach. Overall, our results highlight the potential of fine-tuning language models for specific tasks in biomedicine, such as identifying new immune checkpoints, and demonstrate the utility of machine learning in advancing research in immunotherapy.

4 Discussion

In this study, we demonstrated the effectiveness of fine-tuning the GPT-2 model for generating contextually relevant text in the domain of cancer research. The generated text showed a high level of coherence and relevance to the topics of immune checkpoints. These results highlight the potential of fine-tuning large language models for specialized tasks in the biomedical field. One of the key strengths of the fine-tuned GPT-2 model is its ability to generate text that is both informative and engaging. The

model was able to summarize complex concepts and provide insights into various aspects of cancer research, such as biomarker development, treatment approaches, and disease mechanisms. This capability can be valuable for researchers and healthcare professionals seeking to communicate complex scientific concepts to a broader audience. Additionally, while the model showed promising results in text generation, further evaluation is needed to assess its accuracy and reliability.

In this study is limited by the reliance on abstracts for training the model and the need for additional data sources to validate our findings. The gpt2 model was specifically chosen in this study due to its computational cost. It is thought that milestone information regarding the research question can be obtained as a result of increasing the computational power of the model and feeding it with broader domain knowledge. We emphasise the importance of collaboration between researchers in machine learning and immunology to further advance our understanding of immune checkpoints and their role in disease. In this study, it has been shown that machine learning and large language models for immunology research produce more qualitative texts in response to the research question after being fine-tuned with domain knowledge using gpt2.

References

1. Sperk, M., Domselaar, R. V., & Neogi, U. (2018). Immune checkpoints as the immune system regulators and potential biomarkers in HIV-1 infection. *International journal of molecular sciences*, 19(7), 2000.
2. Ghosh, C., Luong, G., & Sun, Y. (2021). A snapshot of the PD-1/PD-L1 pathway. *Journal of Cancer*, 12(9), 2735.
3. Shiravand, Y., Khodadadi, F., Kashani, S. M. A., Hosseini-Fard, S. R., Hosseini, S., Sadeghirad, H., ... & Kulasinghe, A. (2022). Immune checkpoint inhibitors in cancer therapy. *Current Oncology*, 29(5), 3044-3060.
4. Basudan, A. M. (2022). The role of immune checkpoint inhibitors in cancer therapy. *Clinics and practice*, 13(1), 22-40.
5. Kieuvongngam, V., Tan, B., & Niu, Y. (2020). Automatic text summarization of covid-19 medical research articles using bert and gpt-2. *arXiv preprint arXiv:2006.01997*.
6. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
7. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. URL: <https://www.openai.com/research/Improving-Language-Understanding.pdf>.

Integration of Non-Negative Matrix Factorization to Identification of Immune System Checkpoints

Beyzanur SİYAH¹ [0000-0002-2071-3724], Mehmet Erşan KALAYCI² [0000-0002-4797-9020],
Sinem ÖZDEMİR² [0000-0002-6332-0891], Tolga BERBER¹ [0000-0002-6487-5581], Kemal
TURHAN² [0000-0001-7871-3025]

¹Karadeniz Technical University, Faculty of Science, Department of Computer Sciences, Trabzon, TURKEY

²Karadeniz Technical University, Faculty of Medicine, Department of Medical Informatics, Trabzon, TURKEY

Abstract. In this study, we aimed to discover potential new immune checkpoint candidates by analyzing 20,000 articles from the Scope database. We employed the non-negative matrix factorization method (NMF), a powerful technique for extracting meaningful information from high-dimensional data, to identify sentences regarding novel immune checkpoint. The NMF analysis revealed several promising candidate sentences that could potentially represent new immune checkpoint molecules or pathways. The identification of new immune checkpoint candidates could have significant implications for the development of novel immunotherapy strategies for cancer and other diseases. Overall, this study highlights the importance of data-driven approaches in advancing and identifying new therapeutic targets for immune system related disorders.

Keywords: Immune system, immune checkpoint, NMF, new therapeutic targets

1 Introduction

When the immune system is stimulated with antigens, tissue damage usually occurs in the area where the stimulation occurs. Immune system cells and a series of signals work together to keep this immune response limited. The immune response is also affected by signals such as environmental factors, antigen recognition, proliferation, and costimulatory signals [1-4].

Immune checkpoint molecules are inhibitory receptors on immune cells that are involved in immunosuppressive signals. Thanks to these molecules, autotolerance is achieved. Secondary tissue damage is minimized. The magnitude and length of immune responses are determined [5,6]. They take part in the generation of inhibitory signals after ligand binding. The most studied immune checkpoint inhibitors are CTLA-4, TIM-3, PD-1 receptors. Immune checkpoints physiologically prevent the immune response from attacking self-antigens by negatively stimulating effector immune cells [7,8]. Each checkpoint reduces immune activation through different intra-cellular signaling mechanisms [7,9].

Signals generated through these molecules can cause effector immune cells (especially T cells) to become 'exhausted'. T cell exhaustion is characterized by reduced effector function, persistent expression of immune checkpoint molecules (such as programmed cell death protein 1 (PD1)), poor recall responses, and a differentiated transcriptional state from effector or memory T cells [10]. Interactions between anti-gen-presenting cells (APCs) and T cells regulate immune responses. Many pathogens and cancer types use immune checkpoint proteins to promote inhibitory interactions between immune cells to evade immune control. By investigating immunosuppressive interactions, new cancer treatments have been licensed to enhance immune responses by using specific antibodies to block immune checkpoint molecules. Pembrolizumab, nivolumab, which targets.

Programmed cell death protein 1 (PD1), ipilimumab, which targets Cytotoxic T-lymphocyte associated protein 4 (CTLA4), and atezolizumab, avelumab, and durvalumab, which target Programmed death-ligand 1 (PDL1), are currently licensed as monotherapy for various types of cancer [11].

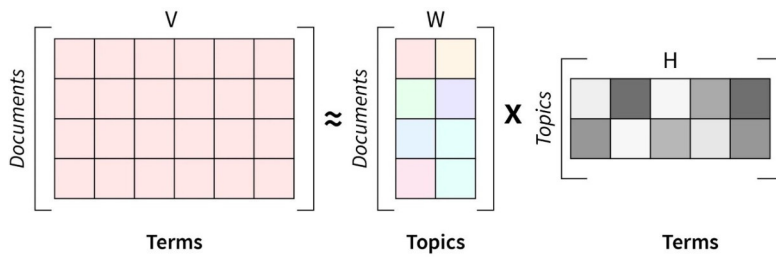
Non-negative matrix factorization (NMF) is a matrix decomposition approach first proposed by Lee and Seung [12]. The NMF method can be used to analyze multidimensional datasets to reduce matrix size, discover hidden patterns, and most importantly, aid in the interpretation of data [13]. Discovery of hidden patterns in a relational data contains a group of methods [14] that has frequent application in natural language processing [15], bioinformatics [16], and digital humanities [17]. For this purpose, different analytical (NMF) or probabilistic (LDA) methods are commonly used. In the field of bioinformatics, many studies have been carried out using the NMF method due to its ability to provide new ideas and relevant information about complex hidden relationships in experimental data sets. This method and some of its variants have been successfully applied for gene expression [18], sequence analysis [19], functional characterization of genes [20] and text mining [21]. Although the bioinformatics community's interest in this technique has increased over the last few years, there are not yet any tools to perform this type of data analysis in an integrated environment.

2 Material and Method

Classical text mining is basically based on the processing of a large matrix (document term matrix), which is large and represents documents with words. For this reason, expressing the large-sized document term matrix as the multiplication of two or more low-rank matrices using matrix decomposition methods will enable the hidden connections within this data to be revealed. In addition, with the help of these approaches, the matrix size is reduced, and semantic analysis becomes easier. One of the most used low-rank matrix decomposition methods is Principal Component Analysis (PCA) in conjunction with Singular Value Decomposition (SVD). PCA attempts to determine the next best vector through data that explains the highest variance [22] and thus helps define structure and reduce

dimensionality. Lee and Seung [12], is a low-order decomposition method that imposes the restriction that the data matrix and factored matrices are non-negative. NMF has an inherently part-based representation, allowing only additive linear combinations of components with non-negative coefficients. Compared to PCA, which is a holistic representation model, NMF leads to a much more intuitive and interpretable representation [23].

Fig. 1. The central principle in non-negative matrix factorization.



The main purposes of the NMF method are dimensionality reduction and feature extraction. The conceptual representation of the NMF method can be seen in Fig. 1. Here, m is the number of rows of the input matrix, n is the number of columns of the input matrix, and k is the common size of the decomposition matrices. In other words, NMF aims to find matrices W ve H that will provide the equivalence of $X \approx W \times H$ to $X \in \mathbb{R}^{m \times n}$, $W \in \mathbb{R}^{m \times k}$ ve $H \in \mathbb{R}^{k \times n}$. The algorithm of the NMF method is as seen in Table 1.

Table 1. NMF algorithm.

Algorithm	NMF ($X \approx WH$)
Input	Nonnegative matrix X of size $K \times N$ Rank parameter $R \in \mathbb{N}$ Threshold ε used as stop criterion
Output	Nonnegative template matrix W of size $K \times R$ Nonnegative activation matrix H of size $R \times N$
Procedure	Define nonnegative matrices W^0 and H^0 by some random or informed initialization. Furthermore set $n = 0$. Apply the following update rules (written in matrix notation): <ol style="list-style-type: none"> 1 $H^{n+1} = H^n \odot \left(\frac{(W^n)^T X}{(W^n)^T W^n H^n} \right)$ 2 $W^{n+1} = W^n \odot \left(\frac{X (H^{n+1})^T}{W^n H^{n+1} (H^{n+1})^T} \right)$ 3 Increase n by one.

Steps (1) and (3) continue until $\|H^n - H^{n-1}\| \leq \varepsilon$ and $\|W^n - W^{n-1}\| \leq \varepsilon$ (or other stopping criteria are met). Finally, $H = H^n$ and $W = W^n$ are obtained.

The number of topics in the algorithm, k , is very important for the analysis and its optimal value must be found. The optimal value of the number of topics k is calculated as in Equation 1.

$$r < \frac{m * n}{m + n} \quad (1)$$

Here an upper limit for the r value is calculated. For sparse matrices consisting mostly of zeros, the formula in Equation 2 is used.

$$r < \frac{\text{count of non - zero values of } X}{m + n} \quad (2)$$

Non-Negative Double Singular Value Decomposition (NNDSVD) is a method designed to improve the initial stage of the NMF method. The NNDSVD algorithm contains no randomness and allows the approximation error of the NMF method to be reduced quickly. The NNDSVD algorithm is given in Table 2.

Table 2. NNDSVD algorithm.

Input	$X \in \mathbb{R}^{m \times n}$ with $X \geq 0$ and $r \ll \min(m, n)$
Output	W and H

1 Compute the largest k singular triplets of $X = [U, S, V]$
2 Initialize $W_{:,1} = \sqrt{S_{1,1}} \times U_{:,1}$ and $H_{1,:} = \sqrt{S_{1,1}} \times V_{:,1}^T$
for $i = 2$ to r **do**
Define $x := U_{:,i}$ and $y := V_{:,i}$
Define $xp := \text{pos}(x)$, $xn := \text{neg}(x)$, $yp := \text{pos}(y)$, and $yn = \text{neg}(y)$
if $\|xp\| \times \|yp\| > \|xn\| \times \|yn\|$ **then**
Put $u = \frac{xp}{\|xp\|}$, $v = \frac{yp}{\|yp\|}$, and $\sigma = \|xp\| \times \|yp\|$
else
Put $u = \frac{xn}{\|xn\|}$, $v = \frac{yn}{\|yn\|}$, and $\sigma = \|xn\| \times \|yn\|$
 $W_{:,i} = \sqrt{S_{j,j}} \times \sigma \times u$ and $H_{j,:} = \sqrt{S_{j,j}} \times \sigma \times v^T$

In the study, TF-IDF method was used to obtain the topics in the articles. TF-IDF method measures the importance of a word for a particular document. Here, term frequency (tf) refers to the frequency of occurrence of a word in a document, and inverse document frequency (idf) refers to how many different documents a word occurs in. The formulation of term frequency is given in Equation 3 and the inverse document frequency is given in Equation 4.

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \quad (3)$$

Here $f_{t,d}$ is the number of lines of a term in a document, that is, the number of occurrences of term t in document d .

$$idf(t, D) = \log \frac{N}{1 + |\{d: d \in D \text{ ve } t \in d\}|} \quad (4)$$

Here $1 + |\{d: d \in D \text{ and } t \in d\}|$ is the number of documents in which the term occurs. The tf-idf matrix is obtained by multiplying the tf and idf matrices as given in Equation 5.

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (5)$$

2.1 Coherence Metric

Coherence score is used to measure the performance of topic analysis and used to select the optimal number of topics. The most used techniques for topic modeling are Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). LDA is a generative probabilistic model, while NMF is a non-probabilistic linear algebraic model which uses matrix factorization.

The theoretical framework of Topic Coherence is based on Bayesian epistemology and needs accurate estimation of term probabilities. The procedure for estimating probabilities specifies how those probabilities are extrapolated from the original data. The Boolean document (P_{bd}) method divides the total number of documents by the number of documents containing a given word to determine the likelihood of that word. Similarly, the number of documents that contain both terms divided by the total number of documents is used to estimate the combined probability of the two words. The boolean nature of this estimating method stems from the fact that it disregards both the frequency with which words appear in each document and the distances between those words. An analogous form of estimation forms the basis of UMass coherence [24]. UMass was also used within the scope of this study. UMass coherence score formulation is given in Equation 6.

$$C_{UMass} = (P_{bd}, S_{pre}^{one}, m_{lc}, \sigma_a) \quad (6)$$

$$S_{pre}^{one} = \{(W', W^*) | W' = \{w_i\}; W^* = \{w_i\}; w_i w_j \in W; i > j\} \quad (7)$$

$$m_{lc}(S_i) = \log \frac{P(W', W^*) + \epsilon}{P(W^*)} \quad (8)$$

Here, P_{bd} is the boolean document method, S_{pre}^{one} compares each word with the word before it in the word set, the log-conditional-probability measure m_{lc} is equivalent to the calculation used by UMass Coherence and σ_{α} is the arithmetic mean all subset pairs S_i are aggregated to a single coherence score [24]. Their formulation given in Equation 7-8.

2.2 Dataset

The NMF method was applied to 27,271 article abstracts obtained by the Karadeniz Technical University Department of Medical Informatics with the search query “immuncheckpoint // immunsystem”, and topic analysis was performed by applying the tf-idf method to these abstracts. A sample of 3 abstracts and titles are presented in Table 3.

Table 3. Sample of 3 abstracts.

<i>Article Title</i>	<i>Abstract</i>
<i>The interaction between Mediterranean diet and intestinal microbiome: relevance for preventive strategies against frailty in older individuals</i>	<i>Age-related changes in intestinal microbiome composition and function are increasingly recognized as pivotal in the pathophysiology of aging and are associated with the aging phenotype. Diet is a major determinant of gut-microbiota composition throughout the entire lifespan, and several of the benefits of a healthy diet in aging could be mediated by the microbiome. Mediterranean diet (MD) is a traditional dietary pattern regarded as the healthy diet paradigm, and a large number of studies have demonstrated its benefits in promoting healthy aging. MD has also a positive modulatory effect on intestinal microbiome, favoring bacterial taxa involved in the synthesis of several bioactive compounds, such as short-chain fatty acids (SCFAs), that counteract inflammation, anabolic resistance, and tissue degeneration. Intervention studies conducted in older populations have suggested that the individual response of older subjects to MD, in terms of reduction of frailty scores and amelioration of cognitive function, is significantly mediated by the gut-microbiota composition and functionality. In this context, the pathophysiology of intestinal microbiome in aging should be considered when designing MD-based interventions tailored to the needs of geriatric patients.</i>
<i>Triptolide mitigates the inhibition of osteogenesis induced by TNF-α in human periodontal ligament stem cells via the p-IκB/α/NF-κB signaling pathway: an in-vitro study</i>	<i>Background: Triptolide is a widely utilized natural anti-inflammatory drug in clinical practice. Aim of this study was to evaluate effects of triptolide on hPDLSCs osteogenesis in an inflammatory setting and to investigate underlying mechanisms. Methods: Using the tissue block method to obtain hPDLSCs from extracted premolar or third molar. Flow cytometry, osteogenic and adipogenic induction were carried out in order to characterise the features of the cells acquired. hPDLSC proliferative activity was assessed by CCK-8 assay to determine the effect of TNF-α and/or triptolide. The impact of triptolide on the osteogenic differentiation of hPDLSCs was investigated by ALP staining and quantification. Osteogenesis-associated genes and proteins expression level were assessed through PCR and Western blotting assay. Finally, BAY-117,082 was used to study the NF-κB pathway. Results: In the group treated with TNF-α, there was an elevation in inflammation levels while osteogenic ability and the expression of both osteogenesis-associated genes and proteins decreased. In the group co-treated with TNF-α and triptolide, inflammation levels were reduced and osteogenic ability as well as the expression of both osteogenesis-associated genes and proteins were enhanced. At the end of the experiment, both triptolide and BAY-117,082 exerted similar inhibitory effects on the NF-κB pathway. Conclusion: The osteogenic inhibition of hPDLSCs by TNF-α can be alleviated</i>

Article Title	Abstract
	through triptolide, with the involvement of the p-ICE]BCE±/NF-CE]B pathway in this mechanism.
Bridging the gap between omics research and dental practice	<i>Aim: The burgeoning field of omics research has witnessed exponential growth in both medicine and dentistry. However, despite more than a decade of advancements, clinical dentistry, particularly in Low- and Middle-Income Countries (LMICs), has seen limited progress in integrating omics-based approaches into routine practice. This review aims to provide a comprehensive overview of the integration of omics approaches in dentistry, focusing on the challenges and opportunities for translating research findings into clinical practice. Methods: we conducted a literature review using key databases to provide a brief overview of the history of genomics in dentistry. Additionally, we summarised recent breakthroughs in omics relevant to oral health practitioners, emphasising the inadequate translation of omics research into clinical practice. Results: Despite significant growth in omics research in both medicine and dentistry, its translation into routine clinical practice in dentistry remains limited. We summarise recent breakthroughs in omics and highlight the gap between research advancements and clinical implementation. Discussion and conclusion: The integration of omics approaches holds promise for enhancing diagnostics, personalised treatment strategies, and preventive measures in dental practice, ushering in a new era of precision oral healthcare. However, several challenges, including infrastructure limitations, cost-effectiveness, and education gaps, hinder the widespread adoption of omics-based approaches in clinical dentistry. A strong commitment to transforming dentistry is required to embrace this transition. This shift has the potential to revolutionise oral healthcare by advancing precision diagnostics and treatment strategies tailored to individual patient needs. © The Author(s) 2024.</i>

2.3 Preprocessing

The lemmatizer of the SpaCy component was used in the preprocessing of the obtained articles [25]. Component for assigning base forms to tokens using rules based on part-of-speech tags, or lookup tables. We have eliminated the stop words identified by SpaCy lemmatizer from abstract data. The end of the preprocessing, 68,741 words were obtained. The matrix in Fig. 2 is known as System for the Mechanical Analysis and Retrieval of Text (SMART). [26]

Fig. 2. SMART notation

Term Frequency Components		Inverse Document Frequency Components		Normalization Component	
Letter	Formal Definition	Letter	Formal Definition	Letter	Formal Definition
b	1	x, n	1	x, n	1
t, n	f_{ik}	f	$\log_2 \left(\frac{N}{n_k} \right)$	c	$\sqrt{\sum_{k=1}^t w_{ik}^2}$
a	$\frac{1}{2} + \frac{1}{2} \frac{f_{ik}}{\max(f_{ik})}$	t	$\log_2 \left(\frac{N+1}{n_k} \right)$	u	$\frac{1-s}{N} \frac{u_i}{\sum_{k=1}^N u_k}$
l	$1 + \log_2 f_{ik}$	p	$\log_2 \left(\frac{N-n_k}{n_k} \right)$	b	$\frac{1-s}{N} \frac{b_i}{\sum_{k=1}^N b_k}$
L	$\frac{1 + \log_2 f_{ik}}{1 + \log_2 \frac{1}{N} \sum_{i=1}^{n_k} f_{ij}}$	bm25	$\log_2 \left(\frac{N-n_k+0.5}{n_k+0.5} \right) + 1$		
d	$\frac{1 + \log_2 (1 + \log_2 f_{ik})}{f_{ik} (k_1 + 1)}$				
bm25	$f_{ik} + k_1 \left(1 - b_1 + b_1 \frac{ D_k }{\sum_{i=1}^N D_i } \right)$				

In the given figure above, f_{ik} denotes the frequency of i . term in document k , $|D_k|$ is the length of document vector k , k_1 and b_1 are the parameters of BM25 algorithm, N is the total number of documents, n_k is the total number of documents containing term k , w_{ik} is the weight of i . term in document k , s is a user defined parameter, u_i and b_i are the term count and the letter count of document i respectively.

According to SMART notation, the best result was obtained by choosing d for term frequency, n for document frequency and c for normalization.

3 Result

In this study, topic analysis was performed by applying the NMF method to 27,271 article abstracts obtained by running the "immuncheckpoint // immunesystem" query on scopus database. The optimal number of topics is found 14 using UMass topic coherence metric. Comparison of the number of subjects according to their coherence scores is shown in Fig. 3.

Fig. 3. UMass coherence scores of topics.



The first 10 words of 14 topics obtained as a result of the analysis were taken as basis. The most appropriate titles were obtained using ChatGPT 4.0 [27] with the words obtained. Table 4 lists the 14 topics, sorted by the most appropriate titles obtained for the words, according to their highest coherence scores.

Table 4. Result of topic analyse.

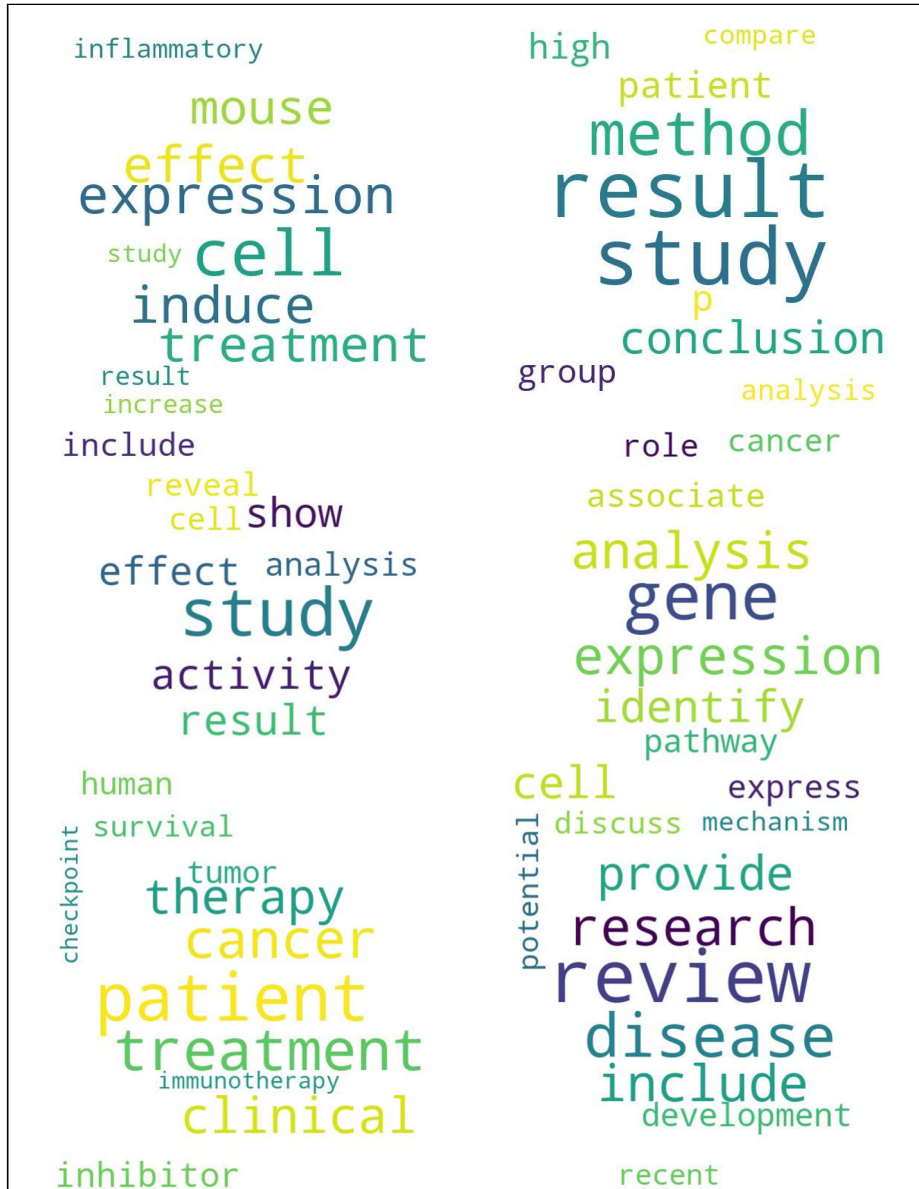
<i>Topic Nr</i>	<i>Topic Coherence</i>	<i>Proposed Topic Title by ChatGPT 4.0</i>	<i>Topic Keywords</i>
1	-0,9923	Effects of Therapeutic Interventions on Cellular Expression and Inflammatory Responses in Murine Models	cell, expression, induce, effect, treatment, mouse, inflammatory, result, increase, study
2	-1,1054	Comparative Analysis and Outcomes in Patient Groups: Methods and Conclusions	study, result, method, conclusion, patient, p, high, group, analysis, compare
3	-1,1879	Effects of Biological Activities on Human Cells: Results and Analysis Revealed	study, activity, result, show, effect, analysis, reveal, include, human, cell
4	-1,2195	Role of Gene Expression in Cancer: Analyzing Cellular Pathways and Associations	gene, expression, analysis, identify, cell, associate, pathway, express, cancer, role
5	-1,2850	Clinical Efficacy of Checkpoint Inhibitor Immunotherapy in Cancer Treatment: Impact on Tumor Suppression and Patient Survival	patient, treatment, cancer, clinical, therapy, inhibitor, tumor, survival, immunotherapy, checkpoint
6	-1,3359	Recent Developments in Disease Mechanisms and Potential Therapies: A Comprehensive Review	review, disease, research, include, provide, development, discuss, potential, recent, mechanism

<i>Topic Nr</i>	<i>Topic Coherence</i>	<i>Proposed Topic Title by ChatGPT 4.0</i>	<i>Topic Keywords</i>
7	-1,3419	Protein Roles and Signaling Pathways in Cancer: Mechanisms and Therapeutic Targets	protein, role, target, signal, pathway, factor, play, receptor, mechanism, cancer
8	-1,5310	Targeting the Tumor Microenvironment: Therapeutic Strategies and Drug Delivery in Cancer Treatment	tumor, cancer, therapeutic, target, therapy, potential, drug, microenvironment, strategy, delivery
9	-1,5878	Adverse Events Associated with Immune Checkpoint Inhibitors: Studying Systemic Responses	immune, response, relate, system, result, checkpoint, study, inhibitor, adverse, event
10	-1,6410	Impact of Dietary Control on Growth Levels: Significant Findings Across Groups	high, level, increase, group, significantly, growth, feed, diet, control, show
11	-1,6641	High-Performance System Design: Proposed Methods and Results Over Time	base, method, model, result, propose, system, high, performance, design, time
12	-1,7638	T Cell Receptor Specificity and Lymphocyte Responses to Antigens: Immune Mechanisms in Detail	cell, t, immune, response, antigen, specific, type, receptor, antibody, lymphocyte
13	-1,7928	Associations Between Health Risks and Environmental Exposures: Population-Based Data Analysis	increase, risk, associate, health, factor, datum, population, cause, exposure, change
14	-1,9137	Viral Pathogenesis and Host Response: Treatment and Vaccine Strategies Against SARS and Related Viruses	infection, disease, virus, cause, viral, treatment, sars, protein, host, vaccine

As seen in Table 4, the recommended title for the topic with the highest coherence score is "Effects of Therapeutic Interventions on Cellular Expression and Inflammatory Responses in Murine Models". This title captures refer to the essence of experimental studies related to treatment effects on cell behavior and inflammation in mice according to the ChatGPT 4.0's topic title prediction.

Additionally, word clouds for the most successful 6 topics is presented in Fig. 4.

Fig. 4. Word clouds.



References

1. Baumeister, S. H., Freeman, G. J., Dranoff, G. & Sharpe, A. H. Coinhibitory pathways in immunotherapy for cancer. *Annu. Rev. Immunol.* 34, 539–573 (2016).
2. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* 12, 252–264 (2012).
3. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* 12, 492–499 (2011).
4. Larkin, J. et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* 373, 23–34 (2015).
5. Shepard MT, Bonney EA. PD-1 Regulates T cell proliferation in a tissue and subset-specific manner during normal mouse pregnancy. *Immunol Invest.* (2013) 42:385–408. doi: 10.3109/08820139.2013.782317
6. Matzinger P. Friendly and dangerous signals: is the tissue in control, *Nat Immunol.* (2007) 8:11–3. doi: 10.1038/ni0107-11
7. Meggyes M, Szanto J, Lajko A, Farkas B, Varnagy A, Tamas P, et al. The possible role of CD8+V α 7.2+/CD161++ T (MAIT) and CD8+V α 7.2+/CD161 lo T (MAIT-like) cells in the pathogenesis of early-onset pre-eclampsia. *Am J Reprod Immunol.* (2018) 79: e12805. doi: 10.1111/aji.12805
8. Jiang Y, Li Y, Zhu B. T-cell exhaustion in the tumor microenvironment. *Cell DeathDis.* (2015) 6:e1792. doi: 10.1038/cddis.2015.162
9. Dai S, Jia R, Zhang X, Fang Q, Huang L. The PD-1/PD-Ls pathway and autoimmune diseases. *Cell Immunol.* (2014) 290:72–9. doi: 10.1016/j.cellimm.2014.05.006
10. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* 12, 492–499 (2011).
11. Wykes, M., Lewin, S. Immune checkpoint blockade in infectious diseases. *Nat Rev Immunol* 18, 91–104 (2018). <https://doi.org/10.1038/nri.2017.112>
12. Lee, DD., Seung, HS: Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788-91 (1999)
13. Pascual-Montano, P., Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, JM., Pascual-Marqui, RD: bioNMF: a versatile tool for non-negative matrix factorization in biology. *BMC Bioinformatics* 7: 366 (2006)
14. Copar, A., Zupan, Blaz., Zitnik, M: Fast optimization of non-negative matrix tri-factorization. *Plos ONE* 14: 6 (2019). <https://doi.org/10.1371/journal.pone.0217994>
15. Huang S, Xu Z, Lv J. Adaptive local structure learning for document co-clustering. *Knowledge-Based Systems.* 2018; 148:74–84.
16. Fu G, Wang J, Domeniconi C, Yu G. Matrix factorization-based data fusion for the prediction of lncRNA–disease associations. *Bioinformatics.* 2017; 1:9.
17. He M, Zhang J, Yang P, Yao K. Robust Transfer Learning for Cross-domain Collaborative Filtering Using Multiple Rating Patterns Approximation. In: *WSDM.* Los Angeles, California, USA; 2018. p. 225-233.
18. Yu, N., Gao, YL., Liu, JX., Shang, J., Zhu, R., Dai, LY: Co-differential Gene Selection and Clustering Based on Graph Regularized Multi-View NMF in Cancer Genomic Data. *Genes* 9, 586 (2018)
19. Wang, CY., Gao, XZ., Liu, JX., Zheng, CH: *IEEE Journl of Biomeical and Health Informatics* 26 (2022)
20. Xi, J., Li, A., Wang, M: A novel network regularized matrix decomposition method to detect mutated cancer genes in tumour samples with inter-patient heterogeneity. *Scientific Reports* 7, 2855 (2017)
21. Hassani, A., Iranmanesh, A., Mansouri, N: Text mining using nonnegative matrix factorization and latent semantic analysis 33, 13745-13766 (2021)

22. Shlens, J: A tutorial on principal component analysis. Google Research Mountain View, CA 94043 (2014)
23. Türkmen, AC: A review of nonnegative matrix factorization methods fo clustering (2015)
24. D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum. Optimizing semantic coherence in topic models. In Proc. of the Conf. on Empirical Methods in Natural Language Processing, pages 262-272. (2011).
25. spaCy Lemmatizer “<https://spacy.io/api/lemmatizer>”. Access: 10.05.2024.
26. SMART Information Retrieval System “https://en.wikipedia.org/wiki/SMART_Information_Retrieval_System” Access: 08.05.2024.
27. ChatGPT 4.0/AI PRO https://ai-pro.org/start-chat-gpt/?flow=04&keyword=chatgpt%204.0&adid=694252978487&ppg=03&pmt=pay2&gad_source=1&gclid=CjwKCAjwupGyBhBBEiwA0UcqaPqsOZ0JcLsN3DViZ_v-OFrmZFC1HWDN-ZiP9stzoYlj5JApjUpkexoCPa4QAvD_BwE Access: 11.05.2024.

Siroz Hastalarında Hepatik Ensefalopati için Semptom Tespit ve İzleme Aracı Prototipinin Arayüz Tanıtımı

Dilek YAPAR¹[0000-0001-7656-1152], Sıdıka Zeynep GÖK SARGIN²[0000-0001-9193-4105],
Ahmet YARDIMCI¹[0000-0001-7241-4913]

¹ Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü Tıp Bilişimi (Doktora Öğrencisi) ve
Muratpaşa İlçe Sağlık Müdürlüğü

² Antalya Şehir Hastanesi, Gastroenteroloji Kliniği, Antalya, Türkiye

³ Akdeniz Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Antalya,
Türkiye

dilekceliker@outlook.com

Özet. Bu araştırma, siroz hastalarının evde bakımını sağlayan kişiler tarafından hepatic ensefalopati (HE) semptomlarının takibinde kullanılacak "HETracker" adlı bir Karar Destek Sistemi (KDS) prototipinin geliştirilmesini ve arayüzünün tanıtımını amaçlamaktadır. Hasta yakınlarının tanı sürecindeki kritik rolleri dik- kate alınarak, kullanıcı dostu bir arayüz tasarımı hedeflenmiştir. Prototipin geliştirilmesi için üç gastroenteroloji uzmanıyla görüşülerek arayüz ve kurallar belirlendi. West Haven kriterlerine dayanarak KDS'ye eklenecek sorular ve psikometrik testler (Stroop testi, DOT testi, Hayvan adlandırma testi) seçildi. Sorular, hasta yakınları ve hastaların anlayabileceği şekilde kısa ve anlaşılır tutuldu. Araştırma ekibi, üç bağımsız gastroenteroloji uzmanından geri bildirim alarak soruları düzenledi. Arayüz, CANVA tasarım programı kullanılarak oluşturuldu. Bu sistemin, Android işletim sistemine sahip mobil cihazlar için bir mobil uygulama olarak geliştirilmesi planlanmıştır. HETracker uygulaması ile kullanıcıların, hastaların semptomlarını normal dönemde üç günde bir takip etmeleri planlanmaktadır. Uygulama, belirli semptomların varlığını ve şiddetini kaydederek değerlendirme yapacaktır. HETracker, DOT ve Stroop testleri gibi nöropsikolojik değerlendirmeleri içerecek ve hastanın günlük fonksiyonlarını değerlendirmeye yönelik sorular sunacaktır. Bu sayede, hastanın HE durumu sürekli izlenecek ve erken müdahale sağlanacaktır. HETracker uygulamasının ilk testleri ve validasyon süreci, uygulamanın başlangıç aşamasındaki işlevselliğini ve kullanıcı dostu olup olmadığı ile ilgili çalışmalar ile değerlendirilecektir. Ancak, bu çalışmada sunduğumuz arayüz bir başlangıç olup, ileride uygulamanın daha kapsamlı bir şekilde değerlendirilmesi ve geliştirilmesi gerekmektedir.

Anahtar Kelimeler: Karar Destek Sistemi, Hepatik Ensefalopati, HETracker, Siroz, İzleme, İzlem Aracı

Abstract. This study aims to develop and introduce the interface of a Decision Support System (DSS) prototype named "HETracker," which

will be used by caregivers of cirrhosis patients to monitor hepatic encephalopathy (HE) symptoms. Considering the critical roles of caregivers in the diagnostic process, a user-friendly interface design is targeted. The interface and rules were determined through consultations with three gastroenterology experts. Questions and psychometric tests (Stroop test, DOT test, Animal Naming test) to be included in the DSS were selected based on the West Haven criteria. The questions were kept short and clear for caregivers and patients to understand. The research team revised the questions based on feedback from three independent gastroenterology experts. The interface was created using the CANVA design program. This system is planned to be developed as a mobile application for devices running the Android operating system. HETracker will enable users to monitor patients' symptoms every three days during the normal period. The application will record and evaluate the presence and severity of specific symptoms. HETracker will include neuropsychological assessments such as DOT and Stroop tests and provide questions to evaluate the patient's daily functions. This ensures continuous monitoring of the patient's HE condition and early intervention. The initial testing and validation process of the HETracker application will assess its functionality and user-friendliness in future studies. However, the interface presented in this study is a starting point, and the application will need more comprehensive evaluation and development in the future.

Keywords: Decision Support System, Hepatic Encephalopathy, HETracker, Cirrhosis, Monitoring, Tracker

1 Giriş

Karaciğer hastalıkları, viral enfeksiyonlar (viral hepatit), alkol kullanımı, obezite ve ilaç toksisitesi gibi faktörlerin yol açtığı hepatosit hasarıyla karakterize edilir ve bu durum karaciğer fonksiyonlarında bozulmaya ve hastalığın ilerleyerek fibroz, siroz ve hatta karaciğer kanserine dönüşebilir [1, 2]. Kronik karaciğer hastalıkları dünya çapında yılda yaklaşık 2 milyon ölümden sorumlu olup tüm ölüm nedenleri arasında ise karaciğer sirozu 11. sıradadır. Siroz yaşam kalitesini etkileyen önemli bir halk sağlığı sorunudur [3].

Hepatik ensefalopati (HE), akut, akut veya kronik veya kronik karaciğer yetmezliği (siroz) ve/veya portal sistemik şant nedeniyle oluşan beyin fonksiyon bozukluğu olarak tanımlanır. Subklinik mental değişikliklerden ileri evrede ağır konfüzyonel duruma ve komaya kadar geniş bir yelpazede nöropsikiyatrik semptomlarla kendini gösterir [2, 4]. HE sirozun en sık görülen komplikasyonudur, siroz tanısı anında HE prevalansı %10- 14'tür. Siroz hastalarının yaklaşık %50'sinde klinik seyrinin bir döneminde HE gelişir. Tekrarlayan HE gelişmesinin kümülatif riski 1 yılda %40'tır. Şiddetli HE hastalarında bir yıllık mortalite %42'ye kadar çıkmaktadır [2]. West Haven Sınıflandırması HE'yi ciddiyetine göre dört kategoriye ayırır. Minimal HE (MHE), karaciğer hastalığı ve/veya portal sistemik şantlı hastalarda ayrıntılı klinik öykü ve tam nörolojik muayene ile tanımlanamayan ancak psikomotor veya nörofizyolojik testlerle tespit edilen ölçülebilir bir bilişsel kusurdur. MHE, 1. dereceyle birlikte Gizli HE kategorisi altında değerlendirilirken, 2 ila 4.

dereceler açık/aşık HE ile değerlendirilir [5].

Siroz hastalarında HE varlığı, ister minimal ister aşık olsun, hastaların yaşam kalitesi, güvenliği ve sağkalımı üzerinde önemli bir etkiye sahiptir. Mevcut tanısal testlerin geçerliliği zayıftır ya da klinik ortamda kullanım için genel olarak uygulanabilir veya pratik değildir. Sirozlu hastalarda HE tanısı için hem klinik hem de araştırma ortamlarında kullanılabilecek basit, doğru algoritmaları belirlenmelidir. Son yıllarda yapay zeka (YZ) modelleri, siroz hastalarının klinik dekompanseasyon ve mortalite riskini tahmin başarılı sonuçlar vermiştir [6], Karar Destek Sistemi (KDS) ise insanın karar verme sürecine katılan ve destekleyen bir bilgi sistemidir [7]. Klinik karar destek sistemlerinin (KKDS) bu alandaki yönetimi kapsamlı bir şekilde araştırılmamış olmasına rağmen, mevcut literatürde siroz ve HE yönetimi için çeşitli stratejilerin önemine dair çalışmalar bulunmaktadır [8, 9]. Karar Destek Sistemlerinin uygulanmasının özellikle erken tanı, zamanında tedavi ve primer ile sekonder önleme stratejilerini destekleyerek etkili bir yönetim sağlayabilir. Bu sistemler, hastaların mevcut durumunu analiz edebilir, tedavi ve klinik sonuçları izleyerek karaciğer hastalığının yönetiminde hayati bir kaynak oluşturabilirler.

Bu araştırma kapsamında "HETracker" adında, siroz hastalarının evde bakımını sağlayan kişiler tarafından HE semptomlarının takibinde kullanılacak bir KDS prototipinin geliştirilmesi gereklerini açıklamak ve ara yüzünün tanıtımı amaçlanmaktadır. Araştırmada, hasta yakınlarının tanı sürecindeki önemli rolleri göz önünde bulundurularak, bu kişilerin kullanımına uygun bir arayüz tasarlanması amaçlanmıştır. Literatür araştırmaları ve alan uzmanlarıyla yapılan görüşmeler sonucunda, şu anda mevcut olan ve benzer işlevleri yerine getiren bir araç bulunmadığı belirlenmiştir. Bu nedenle, bu araca olan ihtiyaç açıkça ortaya konmuştur. Prototipin geliştirilmesi ile erken dönem belirtilerin doğru bir şekilde tespiti ve izlenmesi sağlanarak, hastaların yaşam kalitesinin iyileştirilmesi, tedaviye daha hızlı başlanması ve böylece hastaneye yatış gerekliliğinin ve sağlıkla ilgili maliyetlerin düşürülmesi hedeflenmektedir.

2 Metodoloji: Sistem Tasarımı

HETracker'ın ilk aşamasında KDS'nin arayüzüne ve kurallara karar verebilmek için Gastroenteroloji uzmanları ile 3 görüşme yapılmıştır. Özellikle HE'nin sınıflaması West Haven kriterleri [10] dikkate alınarak KDS'ne dâhil edilecek sorular ve Psiko-metrik testlere karar verilmiştir (örn., Renk Testi: Stroop testi, noktalama testi: DOT ve Hayvan adlandırma testi gibi).

		Tanım	Klinik sonuçlar	Yorum
Etkilenmemiş		Ensefalopati yok, ensefalopati öyküsü yok	Testlerle normal olduğu kanıtlanmış	
Minimal		Mental değişikliğe dair aşikâr bulgu olmadan psikomotor hız ve yürütme fonksiyonları veya nörofizyolojik değişiklikleri ölçen testlerde bozukluklar	Aşikâr klinik bozukluk olmaksızın psikometrik veya nörofizyolojik test sonuçları	Evrensel kriter yok
Evre I	Gizli	<ul style="list-style-type: none"> • Dikkat bozukluğu • Öfori veya anksiyete • Azalmış dikkat aralığı • Toplama/ çıkarmada zorluk • Bozulmuş uyku düzeni 	Yer ve zaman oryantasyonu var ancak öncesine göre hastada kognitif ve davranışsal yavaşlama mevcuttur	Klinik bulgular tekrarlanabilir değil
Evre II		<ul style="list-style-type: none"> • Letarji veya apati • Zaman dezoryantasyonu • Belirgin kişilik değişiklikleri • Uygunsuz davranış • Hareket planlamada bozukluk (Dispraksi) • Asteriksis 	Zaman oryantasyonunda kayıp En az üçünde hata (Aynı hangi günü, haftanın hangi günü, ay, mevsim, yıl) ± diğer semptomlar	Klinik bulgular değişken ancak belli miktarda tekrarlanabilir
Evre III	Aşikâr	<ul style="list-style-type: none"> • Somnolans, semistupor • Uyarılara duyarlı • Ciddi dezoryantasyon • Garip davranışlar 	Zamana ek olarak yer oryantasyonunda kayıp En az üçünde hata (Ülke, eyalet, bölge, şehir ve bulunduğu yer) ± diğer semptomlar	Klinik bulgular sıklıkla tekrarlanabilir
Evre IV		<ul style="list-style-type: none"> • Koma 	Ağrılı uyarana yanıt yok	Komatöz durum genellikle tekrarlanabilir

Şekil 1. West Haven Sınıflaması

Özellikle sorular ve Psikometrik testlere karar verilirken hasta yakınlarının ve hastaların anlayabileceği dilde kısa fakat anlaşılır bir dilde olmasına özen gösterildi. Araştırma ekibi tarafından daha sonra 3 bağımsız gastroenteroloji uzmanından da taslak sorular (şekil 2) içerik açısından değerlendirildi. Bütün uzmanların taslak formdan çıkarılması önerdiği sorular çıkarıldı ve araştırma ekibi ve uzman 3 gastroenterolog ile kurallar üzerine bir görüşme yapıldı. Bu taslak arayüzün oluşturulması için gerekli bir adımdır. Üzerinde çalıştığımız bu sistemin, Android işletim sistemine sahip mobil cihazlar için bir mobil uygulama olarak geliştirilmesi planlanmıştır, bu sayede kullanıcılar uygulamayı herhangi bir Android telefon veya tablet üzerinden kolaylıkla kullanabileceklerdir. HETracker uygulamasının arayüzünün görsel tasarım sürecinin geliştirilmesi, CANVA tasarım programı kullanılarak gerçekleştirilmiştir. CANVA, çeşitli grafik tasarımı araçları sunan web tabanlı bir platformdur. Tasarım aşamasında, kullanıcıların ihtiyaçlarına uygun, sezgisel ve kolay erişilebilir bir kullanıcı dostu arayüzü oluşturmak için detaylı kullanıcı araştırmaları uygulanmıştır. Ayrıca geliştirilecek olan bu uygulama, hastaların yakınlarının bakım verdikleri hastalarının günlük semptom durumlarını kaydetmelerine, belirli semptomları izlemelerine ve bu bilgileri zaman

çizelgesi üzerinden görüntülemelerine olanak tanıyan özellikler içerecektir. Ayrıca, kullanıcılar uygulama üzerinden hastalarının semptom geçmişini takip edebilecek, gerekirse doktorları ile bu bilgileri paylaşabileceklerdir.

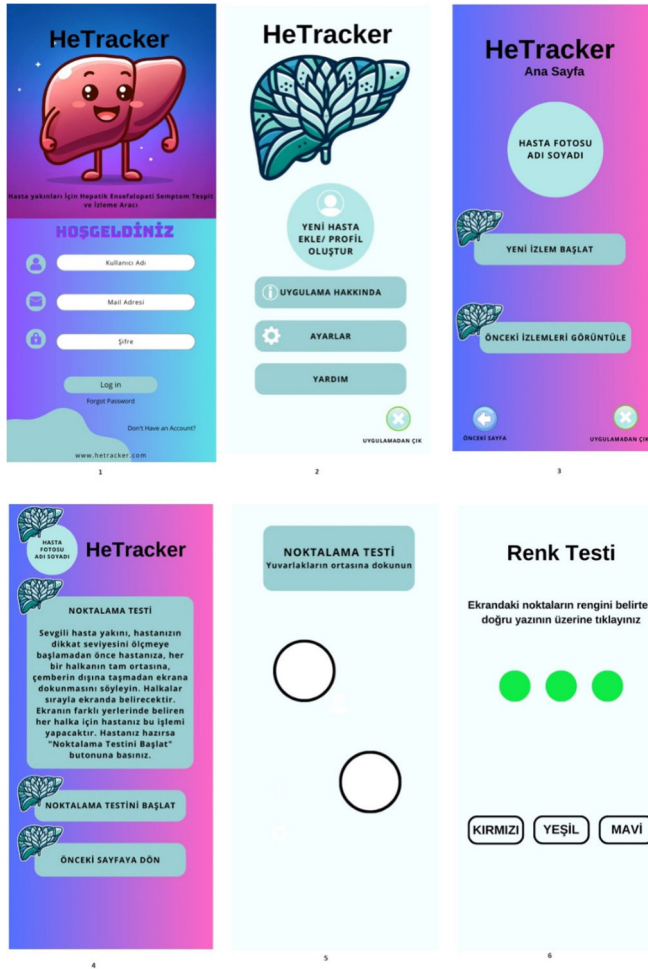
MINİMAL HES	-	+
DOT TESTİ Tamamladığı Süre	Normal Aralıkta	Patolojik
DOT TESTİ Doğru sayısı	Normal Aralıkta	Patolojik
Sırtop Testi Doğru tamamlama süresi	Normal Aralıkta	Patolojik
Sırtop Testi Deneme sayısı	Normal Aralıkta	Patolojik
Evre 1		
Hastanızdan 10 sn içinde 3 farklı hayvan ismi saymasını isteyin>> Süreyi başlat	Başarılı	Başarısız
Hastanız bugün normalde kolayca yaptığı işlerde, mesela cüzdanını bulmakta veya telefon kullanmakta zorlandı mı?	Hayır	Evet
Hastanız bugün normalden daha farklı tavırlar sergiledi mi, birdenbire kızdı mı veya sevindi mi?	Hayır	Evet
Bugün hastanız konuşurken zorlanıyor veya sık sık ne söyleyeceğini unutuyor mu?	Hayır	Evet
Hastanızın bugün eşyaları düşürme veya çevresindekilere çarpmasına yol açan herhangi bir kazası meydana geldi mi?	Hayır	Evet
EVTHastanızın uykü düzeninde herhangi bir değişiklik var mı; gece uykusuzluk, sık uyanma ya da gündüzün uykü ihtiyacı gibi?	Hayır	Evet
Evre 2-3-4 (Aşkar)		
Hastanızı sırası ile şu aşağıdaki soruları sorun her bir soruya verdiği cevapların doğru olup olmadığını işaretleyin		
Hangi Yıldıız	Doğru	Yanlış
Hangi Aydayız	Doğru	Yanlış
Hangi Şehirde yaşıyoruz	Doğru	Yanlış
Şu an neredeyiz	Doğru	Yanlış
Günün hangi zamanındayız? Sence sabah mı, öğlen mi, akşam mı yoksa gece mi?	Doğru	Yanlış
Hastanızın konuşmasında belirgin bir değişiklik var mı, örneğin, konuşması yavaşladı mı, cümle kurmakta zorlanıyor mu?	Hayır	Evet
Hastanız bugün her gün kolayca yaptığı günlük aktivitelerini yerine getirirken zorlandı mı? Örneğin, kişisel bakımını yapmakta, tuvalete gitmekte veya yemek yemekte güçlük çekti mi?	Hayır	Evet
Hastanız yürürken denge bozukluğu yaşıyor mu?	Hayır	Evet
Hastanız konuşmalarında mantıksız ya da tuhaf şeyler söylüyor mu?	Hayır	Evet
Hastanızın elleri titriyor veya yemek yerken zorlanıyor mu?	Hayır	Evet
Hastanız neredeyse tüm gün uyanık kalıyor mu?	Hayır	Evet
Hastanız sizi tanımakta zorlanıyor mu?	Hayır	Evet
Hastanızı uyandırmakta zorlanıyor musunuz?	Hayır	Evet

Şekil 2. Uzman Görüşleri sonrası taslak sistem soruları

3 Bulgular

Uzman görüşleri doğrultusunda, uygulama ile kullanıcıların hastaların semptomlarını üç günde bir takip etmeleri planlanmıştır. Bu takip süreci, kullanıcılara belirli semptomların varlığını ve şiddetini kaydetme imkanı tanır. Uygulama içindeki değerlendirme şeması, Şekil 2’de belirtilen semptomların şiddetine göre sarı ve yeşil renkte kategorize edilmiştir. Eğer hastada sarı bölümden iki veya daha fazla pozitif semptom gözlemlenirse, uygulama uyarı verir ve bu semptomların devam edip etmediğini ya da yeni semptomların eklenip eklenmediğini gözlemlenmek için günlük kontrol yapılmasını gerektirir. Eğer üç gün üst üste semptomlar gözlemlenirse veya yeni pozitif semptomlar eklenirse, uygulama hasta yakınına doktora başvurmaları konusunda uyarır. Eğer hastada yeşil bölümden bir pozitif semptom gözlemlenirse, uygulama uyarı verir. Bu durumda da kullanıcıdan günlük semptom kaydı yapması istenir. Eğer bu süreçte yeni semptomlar eklenirse veya ilk tespit edilen semptom devam ederse, hasta yakınına hastayı doktora götürmesi için bir uyarı verilir. Kısaca uzman görüşleri sonucu bu sistemin, ilk pozitif semptomlar tespit edildiğinde ve takiben artan şiddet gözlemlendiğinde, uygulama bir doktora başvurma konusunda uyarılar verecek şekilde tasarlanmasına karar verilmiştir.

HETracker, DOT (noktalama testi) ve Stroop (renk testi) testleri gibi nöropsikolojik değerlendirmeleri de içerir. Bunlar, hastanın bilişsel işlevlerinin bozulup bozulmadığını değerlendirmek için tasarlanmış kısa testlerdir. Bu bölümde hasta yakınının bu iki testi hastasından uygulamasını isteyecektir (Şekil 3).



Şekil 3. İlk 6 yüzün uygulamada görünüşü

Uygulama ayrıca kullanıcıya hastanın günlük fonksiyonlarını ve davranışlarını değerlendirmek üzere çeşitli sorular sunmaktadır. Bu sorular, hastanın normalde kolaylıkla yapabildiği işlerde zorlanıp zorlanmadığı, konuşmalarında mantıklı olup olmadığı ve günlük aktivitelerde herhangi bir değişiklik olup olmadığını değerlendirir (Şekil 4). Bu detaylı izleme protokolü, hastanın durumunu sürekli olarak değerlendirmeyi ve gerektiğinde erken müdahaleyi sağlamayı amaçlamaktadır.

Hastanızdan 10 saniye içinde 3 farklı hayvan ismi saymasını isteyin

SÜREYİ BAŞLAT
10 saniye

BAŞARILI **BAŞARISIZ**

Hastanızın bugünkü durumu nasıldı?
Lütfen aşağıdaki sorulara yanıt veriniz

Hastanız bugün normalde kolayca yaptığı işlerde, mesela cüdanını bulmakta veya telefon kullanmakta zorlandı mı?
HAYIR **EVET**

Hastanız bugün normalden daha farklı tavırlar sergiledi mi, birdenbire kızdı mı veya sevindi mi?
HAYIR **EVET**

Bugün hastanız konuşurken zorlanıyor veya sık sık ne söyleyeceğini unutuyor mu?
HAYIR **EVET**

Hastanızın bugün eşyaları düşürme veya çevresindekilere serpmasına yol açan herhangi bir kazası meydana geldi mi?
HAYIR **EVET**

Hastanızın uyku düzeninde herhangi bir değişiklik var mı: gece uykusuzluk, sık uyama ya da gündüz aşırı uyku ihtiyacı gibi?
HAYIR **EVET**

Hastanıza sorular ile aşağıdaki soruları sorun her bir soruya verdiği cevapların doğru olup olmadığını işaretleyin.

Hangi yıldayız..... **EVET** **EVET**

Hangi aydayız..... **EVET** **EVET**

Hangi şehirde yaşıyorsunuz..... **EVET** **EVET**

Şu an neredeyiz..... **EVET** **EVET**

Günün hangi zamanındayız?..... **EVET** **EVET**
Sabah mı, öğlen mi, akşam mı yoksa gece mi?

Hastanızın bugünkü durumu nasıldı?
Lütfen aşağıdaki sorulara yanıt veriniz

Hastanızın konuşmasında belirgin bir değişiklik var mı, örneğin, konuşması yavaşladı mı, cümle kurmakta zorlanıyor mu?
HAYIR **EVET**

Hastanız bugün her gün kolayca yaptığı günlük aktivitelerini yerine getirirken zorlandı mı? Örneğin, hijyen bakımını yapmakta, tavalete gitmekte veya yemek yemekte güçlük çekti mi?
HAYIR **EVET**

Hastanız yürürken denge bozukluğu yaşıyor mu?
HAYIR **EVET**

Hastanız konuşmalarında mantıksız ya da tuhaf şeyler söylüyor mu?
HAYIR **EVET**

Hastanızın elleri titriyor veya yemek yerken zorlanıyor mu?
HAYIR **EVET**

Hastanız neredeyse tüm gün uyanık kalıyor mu?
HAYIR **EVET**

Hastanız sizi tanımakta zorlanıyor mu?
HAYIR **EVET**

Hastanızı uyandırmakta zorlanıyor musunuz?
HAYIR **EVET**

TESTİ BİTİR

HeTracker

İZLEM SONUCU

RUTİN İZLEME DEVAM (3 GÜNDE 1)

GÜNLÜK İZLEM

KARACİĞER KOMASI (HEPATİK ENSEFALOPATİ) RİSKİ VAR, DOKTORunuza BAŞVURUN!

İzlemi Kaydet **İzlemi Tekrarla**

ANA SAYFAYA DÖN **UYGULAMADAN ÇIK**

Şekil 4. Sistem sorgularının uygulamada görünüşü

4 Tartışma-Sonuç

Bu çalışma, HETracker adlı, siroz hastalarının evde bakımını sağlayan kişilere yönelik bir arayüz tasarımının deneysel aşamasını sunmaktadır. Prototipin taslak arayüzü, CANVA tasarım programı kullanılarak geliştirilmiştir.

HETracker gibi karar destek sistemlerinin kullanımı, HE tanısı ve yönetimi alanındaki mevcut bilgi birikimine önemli bir katkı sağlayacaktır. Bu dijital takip aracı, siroz hastalarının durumunun sürekli izlenmesine ve yakınlarının nöropsikiyatrik değerlendirilmesine ait kayıtların tutulmasına imkan tanıyarak, erken dönem HE belirtilerinin algılanmasında yeni bir yaklaşım sunmayı planlamaktadır. Ayrıca, HETracker gibi geliştirilecek olan sistemler sağlık hizmetlerinin maliyetini önemli ölçüde azaltma potansiyeline sahiptir. Siroz hastalarında erken dönem HE belirtilerinin tespit edilmesi, hastaneye yatışların ve acil müdahalelerin azalmasına neden olabilir. Bu durum, uzun vadede sağlık hizmetleri üzerindeki mali baskıyı hafifletecek ve ulusal ekonomiye de katkı sağlayacaktır.

Özellikle literatür tarandığında klinik karar destek sistemleri siroz ve HE yönetiminde kapsamlı olarak araştırılmamıştır. Özellikle iki çalışma dışında klinik karar destek sistemi tabanlı çalışan uygulamaya rastlanmış olmasına rağmen bunlar da sadece siroz hastalarının klinik takibinde klinisyenlere hastaya özel bilgi veya tavsiyeler vermek amacıyla geliştirilmiştir [11, 12]. Bu anlamda çalışmamız hasta yakını kullanımını merkezine alan bir uygulama olarak da literature özgün katkı sağlamaktadır. Özellikle kırsal ve kaynakları sınırlı bölgelerde, hastaların ve bakım verenlerin sirozu yönetebilmeleri için güçlendirici bir araç olacaktır. Ayrıca, HETracker'ın geliştirilmesi, diğer kronik hastalıkların yönetimi için benzer sistemlerin oluşturulmasına öncülük edebilir ve sağlık teknolojilerinin yaygınlaştırılmasına katkı sağlayabilir. Bu, hem sağlık hizmetlerinin kalitesini artıracak hem de sağlık sistemlerinin mali yükünü azaltacak potansiyel bir gelişmedir. Sonuç olarak, bu çalışmanın, siroz ve HE yönetimi alanındaki bilgi birikimine ve uygulamalara önemli bir katkı sunması beklenmektedir. HETracker uygulamasının ilk testleri ve validasyon süreci, uygulamanın başlangıç aşamasındaki işlevselliğini ve kullanıcı dostu olup olmadığını ilerleyen çalışmalar ile değerlendirilecektir. Ancak, bu çalışmada sunduğumuz arayüz bir başlangıç olup, ileride uygulamanın daha kapsamlı bir şekilde değerlendirilmesi ve geliştirilmesi gerekmektedir. İleride yapılacak çalışmalarda, HETracker uygulamasının daha geniş bir hasta ve hasta yakını grubu üzerinde test edilmesi planlanmaktadır. Bu geniş çaplı testler, uygulamanın farklı demografik özelliklere sahip kullanıcılar üzerindeki etkinliğini, sistemin performansını ve kullanım kolaylığını daha iyi anlamamıza olanak tanıyacak ve böylece uygulamanın daha geniş kitlelere uygun hale getirilmesi için gerekli düzenlemeler yapılabilir.

Kaynaklar

1. Bejarano Ramírez, D.F., et al., *Prevalence of liver disease in Colombia between 2009 and 2016*. JGH Open, 2020. 4(4): p. 603-610.
15. Tıp Bilişimi Kongresi Bildiriler Kitabı
Proceedings of 15th Turkish Congress of Medical Informatics
30-31 Mayıs 2024 / 30-31 May 2024

2. Sahney, A. and M. Wadhawan, *Encephalopathy in Cirrhosis: Prevention and Management*. J Clin Exp Hepatol, 2022. **12**(3): p. 927-936.
3. Asrani, S.K., et al., *Burden of liver diseases in the world*. J Hepatol, 2019. **70**(1): p. 151-171.
4. Ferenci, P., et al., *Hepatic encephalopathy--definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998*. Hepatology, 2002. **35**(3): p. 716-21.
5. Wong, R.J., R.G. Gish, and A. Ahmed, *Hepatic encephalopathy is associated with significantly increased mortality among patients awaiting liver transplantation*. Liver Transpl, 2014. **20**(12): p. 1454-61.
6. Lee, H.W., J.J.Y. Sung, and S.H. Ahn, *Artificial intelligence in liver disease*. J Gastroenterol Hepatol, 2021. **36**(3): p. 539-542.
7. Zhou, F., et al., *Overview of the New Types of Intelligent Decision Support System*. 2008. 267-267.
8. Hassanein, T., *Current state of knowledge of hepatic encephalopathy (part IV): Management of Hepatic Encephalopathy by liver support systems*. Metabolic Brain Disease, 2017. **32**(2): p. 303-306.
9. Zacharias, H., et al., *Derivation of Decision Algorithms for the Diagnosis of Hepatic Encephalopathy in Patients With Cirrhosis*. Journal of Clinical and Experimental Hepatology, 2017. **7**: p. S14.
10. Ferenci, P., *Hepatic encephalopathy*. Gastroenterol Rep (Oxf), 2017. **5**(2): p. 138-147.
11. Garvin, J.H., et al., *Descriptive Usability Study of CirrODS: Clinical Decision and Workflow Support Tool for Management of Patients With Cirrhosis*. JMIR Med Inform, 2019. **7**(3): p. e13627.
12. Ge, J., et al., *Applying human-centered design to the construction of a cirrhosis management clinical decision support system*. Hepatol Commun, 2024. **8**(3).

Cancer Prediction via Binary Tuna Swarm Optimization

Muharrem Korkmaz¹[0000-0001-5413-343X], Eda Özkul¹[0000-0002-9840-8818] and
Özge Tezel¹[0000-0003-2815-686X]

¹ Karadeniz Technical University, Faculty of Sciences, Department of Computer Sciences,
61080, Trabzon, TURKEY
muharremkrkmz11@gmail.com, eda.ozkul.gs@gmail.com,
ozge_tzl@hotmail.com

Abstract. This paper presents a feature selection method that is based on the Tuna swarm optimization (TSO) algorithm to choose the optimal subset of features on cancer datasets. Due to the structure of feature selection problems, the original TSO algorithm must be adopted to the binary form. The binary tuna swarm optimization (BTSO) is proposed, and the selected features chosen by BTSO are utilized to classify the data using different machine learning methods. The experimental findings demonstrate that BTSO has a superior performance with a minimum number of features.

Keywords: Feature Selection, Machine Learning, Tuna Swarm Optimization

1 Introduction

Cancer is a significant public health concern in the world, with a high mortality rate. Although many people are treated with developing technology, millions of them die from cancer every year. Early diagnosis and treatment are very important in cancer treatment. In recent years, machine learning is frequently used in the detection of cancer. There are many factors that cause cancer. As the factors involved in the detection of the disease and the size of the database increase, the predictive success of the model is affected and the computational cost increases. Therefore, various preprocessing methods are used to eliminate these problems. One of them is feature selection which aims to get rid of irrelevant variables and identify the variables that will make the most successful classification. The main purpose of this paper is to use tuna swarm optimization algorithm to select optimum features from cancer disease dataset and use it in machine learning method to classify it. Some features do not have more influence on the classification of cancer disease. Hence, this study aims to achieve maximum classification success by removing redundant features from the model as much as possible.

2 Original Tuna Swarm Optimization

Tuna is a carnivorous marine fish which are variety of species and sizes. The hunting behavior of tuna include two strategies: the spiral foraging strategy and the parabolic foraging strategy [1]. In most swarm-based algorithms, the initial tuna population is generated randomly in the search area. When performing the spiral strategy, each tuna follows the one that preceded it. It means that there is an information sharing among them. Based on these principles, the mathematical formulation of the spiral foraging strategy is defined as follows:

$$X_i^{(t+1)} = \begin{cases} \alpha_1 \cdot (X_{best}^{(t)} + \beta \cdot |X_{best}^{(t)} - X_i^{(t)}|) + \alpha_2 \cdot X_i^{(t)} & i = 1 \\ \alpha_1 \cdot (X_{best}^{(t)} + \beta \cdot |X_{best}^{(t)} - X_i^{(t)}|) + \alpha_2 \cdot X_{i-1}^{(t)} & i = 2, 3, \dots, N \end{cases} \quad (1)$$

$$a_1 = a + (1 - a) \cdot \frac{t}{t_{max}} \quad (2)$$

$$a_2 = (1 - a) - (1 - a) \cdot \frac{t}{t_{max}} \quad (3)$$

$$\beta = e^{bl} \cdot \cos(2\pi b) \quad (4)$$

$$l = e^{3 \cos\left(\left(\left(t_{max} + \frac{1}{t}\right) - 1\right)\pi\right)} \quad (5)$$

where $X_i^{(t)}$, is the i . search agent at the iteration t , $X_{best}^{(t)}$ is the current optimal search agent, a_1 and a_2 are weight coefficients controlling individuals' tendency to move towards the optimal and previous search agents, a is a constant determining how much tuna follow the optimal and previous search agents initially, t represents the current iteration number, t_{max} is the maximum number of iterations, and b is a random number uniformly distributed between 0 and 1.

When tuna form spirals around food, they efficiently explore nearby space. Yet, blindly adhering to the optimal search agent is not always advantageous. Hence, a random coordinate is introduced to expand the search space, aiding global exploration. This broadens each search agents' scope, makes TSO's global exploration ability. The mathematical model is as follows:

$$X_i^{(t+1)} = \begin{cases} \alpha_1 \cdot (X_{rand}^{(t)} + \beta \cdot |X_{rand}^{(t)} - X_i^{(t)}|) + \alpha_2 \cdot X_i^{(t)} & i = 1 \\ \alpha_1 \cdot (X_{rand}^{(t)} + \beta \cdot |X_{rand}^{(t)} - X_i^{(t)}|) + \alpha_2 \cdot X_{i-1}^{(t)} & i = 2, 3, \dots, N \end{cases} \quad (6)$$

$X_{rand}^{(t)}$ is a randomly generated reference point in the search space. Metaheuristic algorithms typically start with broad global exploration and then shift towards focused local exploitation. Therefore, the spiral foraging strategy adapts by updating reference points from random to optimal individuals as iterations progress.

Tuna forms a parabolic formation with food as a reference point. In addition, tuna forage by exploring around themselves. The mathematical model of parabolic strategy is described as follows:

$$X_i^{(t+1)} = \begin{cases} X_{best}^{(t)} + r \cdot (X_{best}^{(t)} - X_i^{(t)}) + TF \cdot p^2 \cdot (X_{best}^{(t)} - X_i^{(t)}) & \text{if } rand < 0.5 \\ TF \cdot p^2 \cdot X_i^{(t)} & \text{if } rand \geq 0.5 \end{cases} \quad (7)$$

$$p = \left(1 - \frac{t}{t_{max}}\right)^{\frac{t}{t_{max}}} \quad (8)$$

TF is a random number in the range of $[-1,1]$.

3 Binary Tuna Swarm Optimization (BTSO)

Feature selection is inherently a binary problem [2]. Therefore, the original TSO cannot be directly performed on this kind of problems. TSO is modified to binary form, called binary TSO (BTSO) via V-shaped transfer function defined as below [3].

$$T = \left| \frac{2}{\pi} \arctan \left(\frac{\pi}{2} x \right) \right| \quad (9)$$

Based on the V-shaped transfer function, the position of agents in TSO are possibly given as follows.

$$X_{ij}^{(t+1)} = \begin{cases} X_{ij}^{(t)} & \text{if } rand < T(X_{ij}^{(t+1)}) \\ \sim X_{ij}^{(t)} & \text{if } rand \geq T(X_{ij}^{(t+1)}) \end{cases} \quad (10)$$

4 Experimental Results

In this section, two cancer datasets are used such as, Wisconsin breast cancer and ovarian cancer. BTSO is performed on them for selecting the relevant features and compared with the binary differential evolution (BDE) [4] and modified gorilla troops optimization (mGTO) [5]. Then, three machine learning (ML) methods, such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT), are utilized on the reduced datasets. The obtained results are compared according to various performance metrics such as sensitivity, precision, recall, accuracy, and F1-score. Performance metrics are evaluated using ML with BTSO and without BTSO. All* represents the obtained results without feature selection.

4.1 Wisconsin Breast Cancer Data

In this study, the WDBC dataset was accessed from the UCI library. The Wisconsin dataset has 30 attributes and 569 instances. radius1, smoothness1, compactness2, and concavity3 are some of the features in the WDBC dataset. Target values permit the determination of whether the tumor is malignant (cancerous) or benign (non-cancerous). 357 were benign and 212 were malignant of the 569 samples. Table 1 summarizes the performance metrics by using three classifiers.

Table 1. Performance comparison of ML models on Wisconsin breast cancer data.

		Accuracy	Specificity	Precision	Recall	F1 score
KNN	All*	0.971	0.952	0.972	0.981	0.976
	BTSO	0.992*	0.9793	0.988*	1*	0.994*
	BDE	0.9860	0.9975*	0.9663	0.9806	0.989
	mGTO	0.9860	0.9748	0.940	0.9674	0.971
SVM	All*	0.978	0.949	0.971	0.994	0.982
	BTSO	0.989	0.972	0.986	0.999	0.991
	BDE	0.9860	0.9619	0.9779	0.996	0.9888
	mGTO	0.9860	0.9490	0.9695	0.995	0.9845
DT	All*	0.921	0.902	0.941	0.933	0.936
	BTSO	0.981	0.971	0.983	0.987	0.985
	BDE	0.9737	0.9514	0.9726	0.9862	0.9792
	mGTO	0.9649	0.8958	0.9425	0.9190	0.9297

According to the results in Table 1, it is observed that the models created by relevant features with BTSO algorithm give more successful results. The KNN method showed superiority in all performance metrics. In addition, BTSO algorithm outperforms the rival algorithms.

4.2 Ovarian Cancer Data

A total of 349 samples were collected from 171 patients with ovarian cancer and 178 patients with benign ovarian tumors who underwent surgical resection between July 2011 and July 2018 at the Third Affiliated Hospital of Soochow University. The dataset consists of 49 features that were collected by the pathology diagnosed. It was found that the variable CA72-4 had significant missing values and hence it was excluded. The dependent feature is expressed as 1; for individuals with cancer and 0; for individuals without cancer [6]. Table 2 presents the performance metrics obtained by using three classifiers.

Table 2. Performance comparison of ML models on ovarian cancer data.

		Accuracy	Specificity	Precision	Recall	F1 score
KNN	All*	0.776	0.840	0.811	0.714	0.756
	BTSO	0.906	0.955	0.950*	0.853	0.897
	BDE	0.9000	0.9640*	0.9484	0.8258	0.8823
	mGTO	0.8600	0.8725	0.8197	0.6810	0.7435
SVM	All*	0.809	0.855	0.834	0.765	0.795
	BTSO	0.901	0.947	0.940	0.855	0.894
	BDE	0.8771	0.9424	0.9225	0.7985	0.8538
	mGTO	0.8571	0.8879	0.8502	0.7237	0.7805
DT	All*	0.830	0.822	0.820	0.843	0.828
	BTSO	0.924*	0.938	0.934	0.912*	0.922*
	BDE	0.9143	0.9593	0.9449	0.8636	0.9007
	mGTO	0.9057	0.8512	0.8339	0.8419	0.8293

When Table 2 is evaluated, it is observed that the models obtained by feature selection with the BTSO algorithm give more effective results. While KNN and DT showed superiority in various performance metrics, SVM did not outperform the other classifiers

according to any metric. In addition, BTSSO algorithm has better results when comparing BDE and mGTO.

5 Conclusion

The BTSSO is presented, and the selected features chosen by BTSSO are utilized to classify the data using different machine learning methods. The experimental findings indicate that BTSSO exhibits a superior performance with a minimum number of features. Consequently, it was observed that the performance metrics obtained by using the BTSSO method were more successful than using all variables. The model was represented more successfully by minimizing the computational cost by applying feature selection. In this study, only one type of transfer function was used for adaptation of tuna swarm optimization to binary form. Thus, the proposed method can be improved by comparing other transfer functions. In addition, the future direction of this study will include testing newer algorithms with other feature selection techniques.

References

1. Xie L., Han T., Zhou H., Zhang Z.-R., Han B., Tang A.: Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization. *Computational intelligence and Neuroscience* 2021, 1-22 (2021).
2. Thawkar S., Sharma S., Khanna M., kumar Singh L.: Breast cancer prediction using a hybrid method based on Butterfly Optimization Algorithm and Ant Lion Optimizer. *Computers in Biology and Medicine*, 139, 104968 (2021).
3. Mirjalili S., Lewis A.: S-shaped versus V-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, 9, 1-14 (2013).
4. He, X., Zhang, Q., Sun, N., Dong, Y.: Feature selection with discrete binary differential evolution. In: 2009 International Conference on Artificial Intelligence and Computational Intelligence pp. 327-330. IEEE, Shanghai, China (2009).
5. Mostafa, R. R., Gaheen, M. A., Abd ElAziz, M., Al-Betar, M. A., Ewees, A. A.: An improved gorilla troops optimizer for global optimization problems and feature selection. *Knowledge-Based Systems*, 269, 110462 (2023).
6. Lu M., Fan Z., Xu B., Chen L., Zheng X., Li J., et al.: Using machine learning to predict ovarian cancer. *International Journal of Medical Informatics*, 141, 104195 (2020).

Improvement of Encoder Layers in Segmentation Network Models for Automatic Polyp Detection

Nurhan DEĞİRMENCİ BİNGÖL¹[0000-0002-4964-5438], Özkan BİNGÖL¹[0000-0001-6251-6537]
and Kemal TURHAN²[0000-0001-7871-3025]

¹ Gumushane University, Gumushane 29100, Türkiye
² Karadeniz Technical University, Trabzon 61100, Türkiye
nurhandb@gumushane.edu.tr

Abstract. Colorectal cancer ranks third among cancer types worldwide in terms of incidence and cancer-related deaths. Early diagnosis is considered one of the most effective ways to prevent the disease, and in this context, colonoscopy is an important method. However, during colonoscopy, many factors, especially the lack of experience or fatigue of physicians, can lead to small polyps being overlooked. In recent years, the use of deep learning models has initiated a significant transformation in the diagnosis of colorectal cancer. These models have reached the potential to detect details that the human eye might miss, particularly by supporting doctors in challenging cases.

In this study, pre-trained network models were integrated into the encoder layers of well-known segmentation network models to improve their feature extraction mechanisms. Thus, more distinctive features were produced, enhancing segmentation performance. The developed models were tested on the publicly available CVC-ClinicDB and Kvasir-Seg datasets, achieving IoU scores of 0.980 and 0.935 on these datasets, respectively.

Keywords: Colorectal Cancer, Deep Learning, Polyp Detection.

Using Meta-Heuristic Techniques to Estimate Logistic Regression Parameters and Applying to Health Data

Orhan KESEMEN¹[0000-0002-5160-1178] and Buket Nur ALTUN¹[0000-0003-1513-9083]

¹Department of Computer Sciences, Karadeniz Technical University, 61080
Trabzon/TÜRKİYE

buketn.altun@gmail.com

Abstract: Problems that are hard and time-consuming to tackle in daily life can be solved with the help of optimization methods. In situations where traditional optimization methods are insufficient, heuristic and metaheuristic algorithms have been developed to identify solutions that are closest to the global answer. These algorithms can therefore be applied to parameter estimation as well. Appropriate parameter estimation is crucial when using logistic regression for classification. If a parameter is chosen incorrectly, the classification analysis may perform poorly since it will produce inaccurate predictions. In this study, the aim was to evaluate the combined use of logistic regression with metaheuristic algorithms. For this purpose, the parameters of the logistic regression model were estimated using metaheuristic algorithms, and the performance metrics of each model were assessed. The metaheuristic algorithms used for parameter estimation were PSO, ABC, and ALSO. The application data was obtained from Kaggle and split into training and test datasets. Five performance indicators were used to compare the analysis results: F-score, accuracy, precision, selectivity, and sensitivity. The research revealed that the system's logistic regression had the best accuracy value, which was 0.83. By contrast, the ALSO method yielded the highest accuracy value of 0.82 and the best F-score of 0.89.

Keywords: Swarm intelligence, Metaheuristic Algorithms, Logistic Regression Analysis

1. Introduction

Developing computer science has brought machine learning with it, and machine learning has become one of the most popular topics today. Machine learning, which enables computers to find solutions to problems as a result of training them with produced and organized data, is used in areas such as clustering, optimization, classification, natural language processing, dimension reduction / increase, deep learning. Its main purpose is to examine the data, obtain information from the data and build a model that can make predictions with a low error rate [1] [2]. Various algorithms have been developed in machine learning for this purpose. One of these algorithms is logistic regression, which is used for classification analysis. Logistic regression analysis is divided into three types: binary logistic regression, multiple logistic

regression and ordinal logistic regression, depending on the status of the dependent variable [3]. Another thing that is as important as making decisions based on the dependent variable is the correct estimation of the parameters. There are various techniques used for parameter estimation, which is one of the problems of machine learning. Some of these techniques are optimization techniques. Another thing that is as important as being able to obtain the results of daily life problems is that this result is obtained quickly, reliably and at low cost, and the fact that the objectives of optimization techniques are fast is one of the reasons why they can be used in estimating the coefficients [4]. However, most problems encountered in daily life cannot be solved without the help of computers, nor can they be solved with classical optimization methods. Heuristic and meta-heuristic algorithms can be used where classical optimization is not sufficient.

Heuristic and meta-heuristic algorithms are algorithms inspired by nature and state that the solution can be found by trial and error [5][6]. Heuristic and metaheuristic algorithms, which have a short solution time and can be applied to any field, aim to find the best of the existing ways to solve a problem. This means finding the solution that minimizes or maximizes the objective function of the relevant problem [7]. For this purpose, heuristic and metaheuristic optimization techniques can be used to fit a model to the existing data and to obtain parameters that can make accurate predictions with the fitted model [8].

There are studies in the literature where metaheuristic algorithms are used for this purpose. Arafa et al. (2022) evaluated the performance of grid search, random search, Bayes Tree Parzen Estimator (TPE) and Simulated Annealing (SA) optimization techniques to determine the best hyperparameters for the logistic regression model in cancer classification and found that the TPE optimization technique was better than other techniques. It was concluded that it showed good performance [9]. Li & Liu (2013) proposed a state extraction mechanism based on logistic regression (LR) and Enhanced Particle Swarm Optimization (LR-IPSO) model for the purpose of classifying the problem of extracting state elements for network security situation awareness and used Enhanced Particle Swarm Optimization to estimate the parameters of the logistic regression model. and used it to optimize learning ability. It is concluded that the LR-IPSO model can be used to analyze the network security situation [10]. Zou, Hu, Tian, & Shen (2019) examined the mathematical model of the logistic model, defined the error function, estimated the regression coefficients with the gradient descent method, and developed the Sigmoid function, on the grounds that logistic regression is generally used in binary situations and in this case it causes an increase in the repetitions in logistic regression. It ensured that the binary classification effect was better and the number of iterations was reduced [11].

In this study, binary logistic regression parameters will be estimated and a model will be established with Particle Swarm Optimization (PSO), Artificial Bee Colony Optimization Algorithm and Artificial Locust Swarm Optimization Algorithm, and the performances of the established models will be compared through complexity matrices and performance metrics and evaluate the applicability of metaheuristic algorithms to logistic regression.

2. Overview of Methods

Logistic regression analysis aims to explain the relationship between explanatory variables and the explained variable or to make classification and prediction by building a model when the explained variable is categorical [12]. It varies depending on the type of dependent variable. Binary logistic regression analysis uses binomial probability theory because it represents binary situations. It results the cost function between 0 and 1 with the sigmoid function [13]. Particle Swarm Optimization (PSO) was developed by Kenedy and Eberhart [12]. Each individual is called a particle. Each of these particles takes random values and begins to search for the optimal solution in the solution space. Each particle moves through two vectors: position (X) and velocity (V). At each step, the position vectors of the particles change according to their velocity vectors and the search process proceeds by following this process [14]. Artificial Bee Colony Optimization (ABC), which proceeds by adopting the logic of swarm intelligence and considers the movement, intelligence and organization of honeybee swarms, was developed by Derviş Karaboğa in 2005 and the performance of the algorithm was analyzed in 2007 [15]. The algorithm is based on the movements of three types of bees: scout bees, onlooker bees and worker bees. Scout bees discover optimal places for food sources. Worker bees go to the food areas discovered by scout bees, determine the amount of food in those areas and convey this information to the onlooker bees. Onlooker bees use the information sent by worker bees to find the area with the best food source. In this system used by honeybee swarms, the food sources found by scout bees indicate an optimal solution to the research problem, the information sent by worker bees indicates the optimality of the solution, and the food sources chosen by onlooker bees indicate the optimal solution for the relevant research problem [16]. Artificial Locust Swarm Optimization (ALSO) is an optimization algorithm inspired by the behavior of locusts. Locusts act in a controlled manner, including family and social phases. In the family phase, groups of locusts search locally, while in the social phase, searches are made around the main locust in each family group. The mother locust goes to a certain place and creates a new group of locusts there. These groups of mini locusts search locally. The locust that finds the best solution from the group moves to the social phase and the search continues with it. Locusts in the social phase go to local places after a while and move on to the family phase, and these steps are repeated until the algorithm finds the general solution [17].

3. Experimental Results

The data set was taken from Kaggle. It consists of 12 independent variables, which include the medical records of 299 patients collected during their follow-up, and 1 dependent variable, defined as whether the patients died or not as a result of heart failure and other coexisting conditions. For classification analysis, the data set was divided into two: 70% training data set and 30% test data set. The training data set was used to train the model using metaheuristic algorithms. The established model was then used for prediction on the test data set, and the model performance was evaluated after each prediction. For performance evaluation, Accuracy, Precision, F-Score, Sensitivity, and

Specificity performance metrics were used. The analysis was repeated 100 times, and the comparison of performance metrics was based on the average of the metrics obtained (see Fig. 1).

Methods	Accuracy	Sensitivity	Specificity	Precision	F Score
PSO	0.76	0.85	0.52	0.84	0.84
ABC	0.78	0.97	0.22	0.78	0.87
ALSO	0.82	0.96	0.43	0.83	0.89
LR	0.83	0.87	0.74	0.91	0.89

Fig. 1 Performance metrics for methods

4. Conclusion

As expected, the highest accuracy was achieved with the system's logistic regression in the trial. However, when comparing metaheuristic algorithms, it is observed that the ALSO algorithm performed better than the other algorithms with an accuracy rate of 0.82. When interpreting the F-score for a more accurate assessment, it is seen that the F-score of the ALSO algorithm is equal to the F-score obtained with the system's logistic regression. Consequently, when evaluating the combined use of logistic regression and metaheuristic algorithms, it is evident that the ALSO algorithm provides better performance in logistic regression model estimation compared to the PSO and ABC algorithms.

References

1. Coşkun, C., & Baykal, A. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim*, 2011(2), 1-8.
2. Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, 104(5-8), 1889-1902.
3. Burmaoğlu, S., Oktay, E., & Üstün, Ö. (2009). Birleşmiş milletler kalkınma programı beşeri kalkınma endeksi verilerini kullanarak diskriminant analizi ve lojistik regresyon analizinin sınıflandırma performanslarının karşılaştırılması. *Savunma Bilimleri Dergisi*, 8(2), 23-49.
4. Coşkun, A. (2007). Yapay zeka optimizasyon teknikleri: literatür değerlendirmesi. *Fırat Üniversitesi Doğu Araştırmaları Dergisi*, 5(2), 142-146.
5. Halim, A., Ismail, I., & Das, S. (2021). Performance assessment of the metaheuristic optimization algorithms: an exhaustive review. *Artificial Intelligence Review*, 54(3), 2323-2409.
6. Kaveh, A., & Dadras, A. (2017). A novel meta-heuristic optimization algorithm: thermal exchange optimization. *Advances in engineering software*, 110(3), 69-84.
7. Gülcü, A. (2006). Yapay zeka tekniklerinden genetik algoritma ve tabu arama yöntemlerinin eğitim kurumlarının haftalık ders programlarının hazırlanmasında kullanımı. Fırat University.

8. Bottou, L., Curtis, F., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *SIAM review*, 60(2), 223-311.
9. Ahmed Arafa, A., Radad, M., Badawy, M., & El-Fishawy, N. (2022). Logistic regression hyperparameter optimization for cancer classification. *Menoufia Journal of Electronic Engineering Research*, 31(1), 1-8.
10. Li, D., & Liu, Z. (2013). Situation element extraction of network security based on logistic regression and improved particle swarm optimization. 2013 Ninth International Conference on Natural Computation (ICNC). 31, s. 569-573. Menoufia University, Faculty of Electronic Engineering.
11. Zou, X., Hu, Y., Tian, Z., & Shen, K. (2019). Logistic regression model optimization and case analysis. 2019 IEEE 7th international conference on computer science and network technology (ICCSNT). 31, s. 135-139. Menoufia University, Faculty of Electronic Engineering.
12. Cabrera, A. (1994). Logistic regression analysis in higher education: An applied perspective. *Higher education: Handbook of theory and research*, 10(2), 225-256.
13. Abdulqader, Q. (2017). Applying the binary logistic regression analysis on the medical data. *Science Journal of University of Zakho*, 5(4), 330-334.
14. Ortakçı, Y., & Göloğlu, C. (2012). Parçacık sürü optimizasyonu ile küme sayısının belirlenmesi. 11(4), 299-305. ss.
15. Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, 39, 459-471.
16. Karaboga, D., & Karaboga, Dervis. (2005). An idea based on honey bee swarm for numerical optimization. Springer.
17. Kesemen, O., Özkul, E., Tezel, Ö., & Tiryaki, B. (2023). Artificial locust swarm optimization algorithm. *Soft Computing*, 27(9), 5663-5701.

Diagnosing Retinal Disorders with Artificial Intelligence: The Role of Large Language Models in Interpreting Pattern Electroretinography Data

Aslan Aykut^{1,2} [0000-0001-5426-1992], Büşra Akgün² [0000-0003-3705-0397], Almıla Sarıgül Sezenöz^{1,3} [0000-0002-7030-5454],
Mehmet Orkun Sevik² [0000-0001-7130-4798], Özlem Şahin² [0000-0003-2907-2852]

¹ Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor US

² Department of Ophthalmology, Marmara University, Istanbul 34854, Turkey

³ Department of Ophthalmology, Baskent University, Ankara, Turkey

1 Abstract

Introduction: Pattern electroretinography (PERG) is a useful ophthalmological test for the diagnosis of retinal disorders. Large language models have shown promise for analyzing complex medical data, but their potential in interpreting PERG waveforms has not been studied. This study evaluates the diagnostic accuracy of Claude-3, a commercially available large language model, in first detecting (or not) pathological features in PERG and then analyzing PERG data for the diagnosis of retinitis pigmentosa (RP) and cone-rod dystrophy (CRD).

Methods: A subset of PERG measurements from healthy individuals, patients with RP and CRD was randomly selected from the PERG-IOBA dataset. The PERG and clinical data, including age, gender, visual acuities, were provided to Claude-3 for analysis and diagnostic predictions. The model's accuracy was assessed in two scenarios: "First Choice," evaluating the accuracy of the primary differential diagnosis and "Top 3," evaluating whether the correct diagnosis was included within the top three differential diagnoses.

Results: A total of 46 subjects were included in the study: 20 healthy individuals, 13 patients with RP, 13 patients with CRD. Claude-3 achieved 100% accuracy in detecting the presence or absence of pathology. In the "First Choice" scenario, the model demonstrated moderate accuracy in diagnosing RP (61.5%) and CRD (53.8%). However, in the "Top 3" scenario, the model's performance significantly improved, with accuracies of 92.3% for RP and 76.9% for CRD.

Conclusion: This is the first study to demonstrate the potential of large language models, specifically Claude-3, in analyzing PERG data to diagnose retinal disorders. Despite some limitations, the model's high accuracy in detecting pathologies and distinguishing between specific diseases highlights the potential of large language models in ocular electrophysiology. Future research should focus on integrating multimodal data, and conducting comparative analyses with human experts.

Keywords: PERG, Claude-3, Large Language Models, Retinitis Pigmentosa, Cone Rod Dystrophy

2 Introduction

Pattern electroretinography (PERG) has been a valuable tool in ophthalmology for testing retinal ganglion cell function, photoreceptor health and diagnosing various retinal diseases by presenting alternating visual stimuli and recording the electrical responses from the retina. [1, 2] PERG provides objective information regarding the health and integrity of the retinal pathway, with the macula and optic nerve in particular. Analysis of PERG waveforms, particularly the N95, P50, and N35 components, allows for the assessment of macular function and detection of abnormalities in the ganglion cell layer and inner retina. [1] The different PERG patterns observed in different diseases, such as the reduced amplitudes in retinitis pigmentosa (RP) and the delayed implicit times in glaucoma, aid in differential diagnosis and facilitate appropriate management strategies. [3, 4] PERG offers a non-invasive and objective measure of retinal function, making it a valuable tool for monitoring disease progression and evaluating treatment efficacy. [5]

The emergence of artificial intelligence (AI) has revolutionized various aspects of medical diagnosis, offering promising approaches to interpreting complex medical data and assisting clinicians in making more informed decisions. [6] In the field of ophthalmology, AI algorithms have demonstrated remarkable capabilities in analyzing retinal images for diabetic retinopathy, age-related macular degeneration, and other retinal diseases. [7] Beyond analyzing retinal images, AI algorithms have been successfully used to predict glaucoma progression, automate visual field interpretation, and personalize treatment. [8, 9] AI offers several advantages in medical diagnostics, including the ability to detect subtle patterns that human observers often miss, enable rapid and objective assessments, and improve diagnostic accuracy, leading to better patient outcomes. [10, 11] However, despite these promises, implementing AI in healthcare still faces issues with privacy, algorithmic bias, and the requirement for thorough validation to ensure reliability and security. [12]

Large language models (LLMs) represent a unique and powerful subset of artificial intelligence that are trained on massive text and code datasets, enabling them to understand and generate human-like text, translate languages, and answer complex questions in an informative manner. [13] Unlike traditional AI models that excel at specific tasks such as image recognition or data classification, LLMs use their extensive training to process and understand information more holistic and contextual manner, mimicking the learning and reasoning processes of humans. [14] This ability to extract meaning, identify patterns, and draw conclusions from complex data sources makes them valuable tools for analyzing various medical data, including clinical notes, research articles and even genome sequences. [15, 17]

Interpretation of pattern ERG data is a complex task that typically requires extensive training and expertise in ophthalmology and electrophysiology and often challenges even experienced clinicians. [18] The aim of this study is to evaluate the potential of Claude-3, a large language model accessible to a wider range of users, in analyzing pattern ERG data and providing diagnostic insights, potentially helping clinicians and researchers interpret this complex data.

3 Methods

3.1 Dataset Utilization

This study utilized the PERG-IOBA dataset available from PhysioNet, which serves as a Research Resource for Complex Physiologic Signals. [19] The dataset includes 1,354 transient PERG responses from 304 subjects in 336 records, collected between 2003 and 2022. [20] The dataset consists of 105 healthy subjects and 199 patients diagnosed with various retinal conditions. The most common diseases that make up the data set were 48 patients with RP, 32 patients with macular dystrophy, 16 patients with Stargardt disease and 14 patients with cone-rod dystrophy (CRD). Clinical diagnosis, including age, gender, and visual acuity measurements in logMAR scale, was provided in CSV format. The dataset had been anonymized, and data collection dates had been randomly date-shifted to maintain patient privacy and confidentiality. [20]

3.2 PERG Signal Acquisition

PERG signals in the dataset were captured by experienced technicians using the computerized Metrovision Optoelectronic Stimulator Vision Monitor MonPack 120 (Metrovision, Pérenchies, France). The acquisition protocol strictly adhered to the International Society for Clinical Electrophysiology of Vision (ISCEV) guidelines. [21] Signals were recorded at a high sampling rate of 1700 Hz over a duration of 150 milliseconds, producing 255 equally spaced observations per signal. Figure 1 shows a portion of the raw data recorded from a patient.

TIME_1	RE_1	LE_1
2016-09-15 09:40:21.0000	0	0
2016-09-15 09:40:21.0006	-0.1	0.1
2016-09-15 09:40:21.0012	-0.2	0.2
2016-09-15 09:40:21.0018	-0.2	0.4
2016-09-15 09:40:21.0024	-0.2	0.6
2016-09-15 09:40:21.0030	0	0.7
2016-09-15 09:40:21.0035	0	0.8
2016-09-15 09:40:21.0041	0	0.8
2016-09-15 09:40:21.0047	0	0.6
2016-09-15 09:40:21.0053	-0.2	0.6
2016-09-15 09:40:21.0059	-0.2	0.5
2016-09-15 09:40:21.0065	-0.2	0.6
2016-09-15 09:40:21.0071	-0.2	0.6
2016-09-15 09:40:21.0077	-0.4	0.6
2016-09-15 09:40:21.0083	-0.6	0.5
2016-09-15 09:40:21.0089	-0.8	0.4
2016-09-15 09:40:21.0094	-1.2	0.3
2016-09-15 09:40:21.0100	-1.3	0.2
2016-09-15 09:40:21.0106	-1.4	0.3
2016-09-15 09:40:21.0112	-1.4	0.2
2016-09-15 09:40:21.0118	-1.2	0.2

Figure 1: A portion of the raw data recorded from a patient

3.3 Study Sample Selection

A specific subset of PERG measurements from healthy individuals, patients with RP and CRD were randomly selected from the larger dataset for this analysis. The PERG data for each subject in this subset were extracted from the dataset and provided to the Claude-3 language model for analysis and diagnostic predictions. The study was designed as a pilot study, hence the number of participants was kept limited.

3.4 Data Input to Claude-3 LLM

To evaluate the potential of large language models (LLMs) in analyzing PERG waveforms and providing diagnostic insights, we employed Claude-3, a commercially available LLM. We provided Claude-3 with a prompt that included the following instructions:

“Analyze the provided pattern ERG data for both eyes, identifying abnormalities in the N35, P50, and N95 waves, oscillatory potentials, and overall waveform morphology. Indicate the presence of pathology with a "Yes" or "No." If pathology is detected, select the top three differential diagnoses from a comprehensive list associated with pattern

ERG features. Each diagnosis should include a detailed justification focusing on the bilateral ERG findings, particularly concerning both cone and rod functions, and consider the patient's current age, gender, and bilateral presentation. Rank these diagnoses by likelihood and outline the potential need for further tests or information to confirm these diagnoses.

Please include the following patient details:

Age: []

Gender: []

Right Eye Visual Acuity (logMAR): []

Left Eye Visual Acuity (logMAR): []

Note: The age at symptom onset is unknown. Given the complexity of diseases associated with pattern ERG abnormalities, additional clinical data, imaging, or tests may be necessary for a definitive diagnosis.”

3.5 Important Outcomes

First scenario;

First Choice Accuracy: This metric assesses whether Claude-3 identified the correct pathology as the first differential diagnosis. This is critical for determining the model's precision in diagnosing the most likely condition without additional input.

Second scenario;

Top Three Accuracy: This broader metric evaluates whether the correct diagnosis was included in the model's top three differential diagnoses. This measure reflects the model's ability to detect and prioritize potential diagnoses, which is critical for clinical settings where multiple potential diagnoses may be considered before reaching a final conclusion.

3.6 Statistical Analysis

To determine whether the variables were normally distributed, the Shapiro-Wilk normality test was used. Demographic characteristics (age and sex) and visual acuity measurements were summarized using mean and standard deviation for continuous variables and frequencies and percentages for categorical variables. One-way analysis of variance (ANOVA) was performed to compare the mean age of the three groups, and a chi-square test was used to assess differences in gender distribution. To compare visual acuity between groups, we used the Kruskal-Wallis test with Bonferroni correction for multiple comparisons. Sensitivity, specificity, precision and F1 scores were calculated for each study group. Statistical analyzes were performed using Statistical Package for Social Sciences (SPSS) software, version 25.0 (IBM, Chicago, IL, USA). Values of $P < 0.05$ were considered statistically significant in all tests.

4 Results

4.1 Demographics and Visual Acuity

A total of 46 subjects were included in the study, with 20 healthy individuals, 13 patients with RP, and 13 patients with CRD. The demographic characteristics and visual acuity measurements for each group are summarized in Table 1.

Table 1: Demographics and visual acuity measurements for each group

Characteristic	Retinitis Pigmentosa (n=13)	Cone-Rod Dystrophy (n=13)	Normal (n=20)	P
Age (years) - Mean \pm SD - Range	35.5 \pm 15.8 (12 - 62)	34.2 \pm 15.6 (10 - 61)	28.8 \pm 18.3 (6 - 70)	0.48
Gender -Male -Female	9 (69.2%) 4 (30.8%)	8 (61.5%) 5 (38.5%)	9 (45.0%) 11 (55.0%)	0.38
Mean Visual Acuity (LogMAR) - Mean \pm SD	0.39 \pm 0.44	0.51 \pm 0.37	0.15 \pm 0.22	0.0015*

LogMAR = Logarithm of the Minimum Angle of Resolution, p* significant. Post-hoc Dunn test (Bonferroni correction):

RP vs. CR: $p > 0.05$ (ns), RP vs. Normal: $p < 0.05$ (sig.), CR vs. Normal: $p < 0.05$ (sig.)

4.2 Diagnostic Accuracy

The diagnostic accuracy of Claude-3 in detecting the presence or absence of pathology in all study groups was successful and all cases were correctly identified. In the normal group, the model confirmed no pathology in all 20 (100%) subjects. Similarly, in the pathologic group, the model confirmed pathology in all 26 (100%) subjects. The accuracy of the differential diagnosis showed variability, with RP and CRD in the first scenerio having an accuracy of 61.5% (8 of 13 cases) and 53.8% (7 of 13 cases), respectively. Notably, the model performed better when we used the second scenerio, with RP and CRD achieving higher success rates of 92.3% (12 of 13 cases) and 76.9% (10 of 13 cases), respectively. However there was no statistical difference when comparing model's accuracy of the differential diagnosis between RP and CRD in both scenarios ($P = 1$ and 0.59), respectively.

4.3 Performance Metrics

We evaluated performance metrics according to 2 scenarios. Results are given in table 2

Table 2: Performance Metrics for Claude-3 Diagnosis

Performance Metric	Healthy vs. Pathologic	RP	CRD
Sensitivity	100%	69.23%	53.85%
Specificity	100%	100%	84.62%
Precision	100%	100%	77.78%
F1 Score	100%	81.82%	63.64%

RP = retinitis pigmentosa; CRD = cone-rod dystrophy

5 Discussion

Our results suggest that Claude-3 achieves perfect performance, with 100% sensitivity, specificity, precision, and F1 score, demonstrating its ability to accurately identify all cases with pathology as well as all healthy cases without any misclassifications. It can effectively distinguish between healthy subjects and those with retinal diseases, achieving 100% accuracy in detecting the presence or absence of disease based on PERG data with minimal clinical data. The ability to accurately differentiate between healthy and pathologic cases is crucial in a clinical setting, as it can help prioritize patients who require further diagnostic evaluation and potential treatment. [22]

PERG is a highly valuable diagnostic tool in the evaluation of RP, a genetic disorder that causes progressive retinal degeneration. [4] In RP, PERG waveforms typically exhibit reduced amplitudes which is due to impaired function of photoreceptors and retinal ganglion cells. [2] These abnormalities can be detected even in early stages of the disease when visual acuity is still preserved, making PERG a sensitive tool for early diagnosis and monitoring disease progression. [23]

PERG is particularly informative in diagnosing and monitoring CRDs, a group of inherited retinal diseases characterized by deterioration of cone and rod photoreceptors. [24] The PERG can provide detailed assessments of cone function, which is crucial in cone-rod dystrophies where cone dysfunction typically presents before rod dysfunction. [24] For example, PERG can help distinguish different patterns of visual impairment in patients with cone-rod dystrophy, with some having more severe cone dysfunction than others. [24] This functional assessment is consistent with clinical observations and genetic findings, thereby supporting the diagnosis and understanding the disease progression in these patients. [25]

When considering the first scenario, Claude-3 demonstrated moderate accuracy in diagnosing RP (61.5%) and CRD (53.8%). However, when evaluating the second scenario, the model's performance significantly improved, with accuracies of 92.3% for RP and 76.9% for CRD. This suggests that Claude-3 is capable of identifying the

correct diagnosis within the top three suggestions, even if it may not always be the first choice. These results are promising, indicating the potential of LLMs in analyzing PERG data for the diagnosis of retinal disorders. However, our literature search did not yield any studies done with artificial intelligence specifically using LLMs on this subject; therefore, we cannot directly compare our results to previous findings.

However, the model's performance metrics in identifying specific retinal disorders based on the first scenario varied between RP and CRD. While Claude-3 showed good performance in identifying RP cases, with high specificity and precision, its performance in identifying CRD cases was moderate, with lower sensitivity, specificity, and precision. This difference in performance may be attributed to the heterogeneity of CRD phenotypes and the overlap of PERG features with other retinal disorders, making it more challenging for the model to accurately identify CRD cases based solely on the first-choice diagnosis. [24]

Integrating AI into clinical practice offers several benefits, such as providing rapid, objective assessments of complex medical data and detecting subtle patterns that may be overlooked by human observers. [11] However despite its promise one of the major concern is the “black box” nature of these models, where the reasoning behind their predictions remains opaque. [26] In this models training data are often obscured or undocumented, and their methods opaque. [27] This lack of transparency can affect trust and acceptance among clinicians, particularly when dealing with complex medical decisions. [28]

Our study has several strengths and limitations. One of the strengths is the use of a large, well-characterized data set (PERG-IOBA) that conforms to the ISCEV guidelines for PERG collection, ensuring data reliability and consistency. [20] Another strength of our study is the use of a commercially available large language model, Claude-3, which is accessible to a wider range of users compared to specialized AI models that require extensive technical expertise. This accessibility enables greater potential in clinical settings, as healthcare professionals without strong AI knowledge can still benefit from the model's insights. However, our study has notable limitations, the most significant being the relatively small sample size, as it was designed as a pilot study. Additionally, the study focuses on a specific subset of retinal diseases, and while Claude-3 shows promising results in analyzing PERG data, its performance for other types of ocular electrophysiological tests and different retinal diseases remains to be investigated.

There are several important directions for future research in this area. First, the integration of PERG data with other diagnostic modalities such as optical coherence tomography and visual field testing may represent a significant advance toward a multimodal diagnostic approach. By combining data from these different sources, LLMs could provide a more comprehensive and nuanced understanding of retinal health and improve the ability to diagnose complex conditions that may not be detectable with a single diagnostic method. Conducting comparative analysis between the performances of LLMs and human experts is also crucial. Such studies would help delineate the strengths and limitations of each approach and provide insights into how best to use AI in clinical settings. By directly comparing AI with human diagnostics, researchers can identify specific scenarios where AI excels or lags behind, thereby refining AI applications to effectively support clinical decision making. [29]

In conclusion, this study is the first to demonstrate the potential of large language models, particularly Claude-3, in analyzing PERG data for the diagnosis of retinal diseases. Despite some limitations, the model's high accuracy in detecting pathologies and distinguishing between specific diseases highlights the potential of AI in ophthalmology. Future research should focus on addressing limitations.

6 References

1. Thompson DA, Bach M, McAnany JJ, Šuštar Habjan M, Viswanathan S, Robson AG. ISCEV standard for clinical pattern electroretinography (2024 update). *Documenta ophthalmologica Advances in ophthalmology*. 2024;148(2):75-85. doi: 10.1007/s10633-024-09970-1.
2. Robson AG, El-Amir A, Bailey C, Egan CA, Fitzke FW, Webster AR, et al. Pattern ERG correlates of abnormal fundus autofluorescence in patients with retinitis pigmentosa and normal visual acuity. *Investigative ophthalmology & visual science*. 2003;44(8):3544-50. doi: 10.1167/iovs.02-1278.
3. Gallo Afflitto G, Chou TH, Swaminathan SS, Aiello F, Gedde SJ, Nucci C, et al. Pattern Electroretinogram in Ocular Hypertension, Glaucoma Suspect and Early Manifest Glaucoma Eyes: A Systematic Review and Meta-analysis. *Ophthalmology science*. 2023;3(4):100322. doi: 10.1016/j.xops.2023.100322.
4. Janáky M, Pálffy A, Horváth G, Tuboly G, Benedek G. Pattern-reversal electroretinograms and visual evoked potentials in retinitis pigmentosa. *Documenta ophthalmologica Advances in ophthalmology*. 2008;117(1):27-36. doi: 10.1007/s10633-007-9099-0.
5. Robson AG, Nilsson J, Li S, Jalali S, Fulton AB, Tormene AP, et al. ISCEV guide to visual electrodiagnostic procedures. *Documenta ophthalmologica Advances in ophthalmology*. 2018;136(1):1-26. doi: 10.1007/s10633-017-9621-y.
6. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nature biomedical engineering*. 2018;2(10):719-31. doi: 10.1038/s41551-018-0305-z.
7. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *The British journal of ophthalmology*. 2019;103(2):167-75. doi: 10.1136/bjophthalmol-2018-313173.
8. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama*. 2016;316(22):2402-10. doi: 10.1001/jama.2016.17216.
9. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, et al. Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images. *American journal of ophthalmology*. 2019;198:136-45. doi: 10.1016/j.ajo.2018.10.007.
10. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. 2019;1(6):e271-e97. doi: [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2).
11. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafiyan H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. doi: 10.1038/s41586-019-1799-6.
12. Char DS, Abramoff MD, Feudtner C. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *The American journal of bioethics : AJOB*. 2020;20(11):7-17. doi: 10.1080/15265161.2020.1819469.
13. Raffel C, Shazeer NM, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J Mach Learn Res*. 2019;21:140:1-:67.
14. Head CB, Jasper P, McConnachie M, Raftree L, Higdon G. Large Language Model Applications for Evaluation: Opportunities and Ethical Implications. *New Directions for Evaluation*. 2023. doi: 10.1002/ev.20556.
15. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The Application of Large Language Models in Medicine: A Scoping Review. *iScience*. 2024;109713. doi: <https://doi.org/10.1016/j.isci.2024.109713>.
16. Singhal K, Azizi S, Tu T, Mahdavi S, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2022;620:172 - 80.
17. Wu J, Ma Y, Wang J, Xiao M. The Application of ChatGPT in Medicine: A Scoping Review and Bibliometric Analysis. *Journal of Multidisciplinary Healthcare*. 2024;17(null):1681-92. doi: 10.2147/JMDH.S463128.
18. Yap GH, Chen LY, Png R, Loo JL, Tow S, Mathur R, et al. Clinical value of electrophysiology in determining the diagnosis of visual dysfunction in neuro-ophthalmology patients. *Documenta ophthalmologica Advances in ophthalmology*. 2015;131(3):189- 96. doi: 10.1007/s10633-015-9515-9.
19. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):E215-20. doi: 10.1161/01.cir.101.23.e215.

20. Physionet, (2024). <https://doi.org/10.13026/d24m-w054>. Accessed 5/1/2024.
21. Bach M, Brigell MG, Hawlina M, Holder GE, Johnson MA, McCulloch DL, et al. ISCEV standard for clinical pattern electroretinography (PERG): 2012 update. *Documenta ophthalmologica Advances in ophthalmology*. 2013;126(1):1-7. doi:10.1007/s10633-012-9353-y.
22. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*. 2008;56(1):45-50. doi: 10.4103/0301-4738.37595.
23. Popović P, Jarc-Vidmar M, Hawlina M. Abnormal fundus autofluorescence in relation to retinal function in patients with retinitis pigmentosa. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*. 2005;243(10):1018-27. doi: 10.1007/s00417-005-1186-x.
24. Hamel CP. Cone rod dystrophies. *Orphanet journal of rare diseases*. 2007;2:7. doi: 10.1186/1750-1172-2-7.
25. Downes SM, Payne AM, Kelsell RE, Fitzke FW, Holder GE, Hunt DM, et al. Autosomal dominant cone-rod dystrophy with mutations in the guanylate cyclase 2D gene encoding retinal guanylate cyclase-1. *Archives of ophthalmology (Chicago, Ill : 1960)*. 2001;119(11):1667-73. doi: 10.1001/archophth.119.11.1667.
26. Schwartz IS, Link KE, Daneshjou R, Cortés-Penfield N. Black Box Warning: Large Language Models and the Future of Infectious Diseases Consultation. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2024;78(4):860-6. doi: 10.1093/cid/ciad633.
27. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*. 2023;90:104512. doi: 10.1016/j.ebiom.2023.104512.
28. Au Yeung J, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, et al. AI chatbots not yet ready for clinical use. *Frontiers in digital health*. 2023;5:1161098. doi: 10.3389/fgth.2023.1161098.
29. Rojas-Carabali W, Sen A, Agarwal A, Tan G, Cheung CY, Rousselot A, et al. Chatbots Vs. Human Experts: Evaluating Diagnostic Performance of Chatbots in Uveitis and the Perspectives on AI Adoption in Ophthalmology. *Ocular immunology and inflammation*. 2023:1-8. doi: 10.1080/09273948.2023.2266730.

Investigation of the Inhibition Mechanism of Acetylcholinesterase Enzyme for Alzheimer's Treatment by Molecular Docking Method

Merve Baldan¹[0000-0001-6019-0801], Gizem Tatar Yılmaz^{1,2}[0000-0001-6642-6870]

¹Department of Bioinformatics, Karadeniz Technical University, Trabzon, Turkey

²Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey

mervebaldaan@gmail.com

gizemtatar@ktu.edu.tr

Abstract. Alzheimer's disease is a neurodegenerative condition characterized by disturbances in mental, physical, and cognitive functions. Acetylcholinesterase (AChE) over activity is one of the primary causative factors of the disease. Tacrine (Cognex®), an FDA-approved drug targeting the AChE enzyme, has been discontinued due to its side effects and toxicity. Firstly, this study will evaluate the pharmacokinetic suitability of compounds selected from molecular databases according to their absorption, distribution, metabolism, and elimination (ADME) properties. Then, it aims to present more effective and reliable potential candidate drug molecules to the literature by examining the interaction mechanism between these compounds and the human AChE enzyme by molecular docking method. The interaction mechanisms and binding affinities of the compounds selected according to ADME analysis with the AChE enzyme were examined using structure-based drug design methods such as molecular docking. According to the results of molecular docking analysis, 9 compounds (CHEMBL606348, CHEMBL4556734, CHEMBL4218191, CHEMBL4209803, CHEMBL4217663, CHEMBL4210316, CHEMBL4467284, CHEMBL200008, CHEMBL4448350) out of 30 with the best docking results (-11.71, -13.13, -11.40, -11.82, -12.10, -11.92, -12.81, -13.26, -13.24 kcal/mol) were obtained. Based on the results obtained from molecular docking, it is inferred that the top 9 compounds exhibit significant efficacy and are deemed suitable candidates for further evaluation in the subsequent phases of drug discovery. Notably, these compounds demonstrate favorable ADME properties, further underscoring their potential as promising leads for therapeutic development. This study represents a crucial step that may contribute to developing effective and reliable drug molecules for treating Alzheimer's disease

Keywords: Alzheimer's, Molecular Docking, Acetylcholinesterase, ADME

CAN LARGE LANGUAGE MODELS SURPASS HUMAN PERFORMANCE IN THE MEDICAL SPECIALIZATION EXAM?

Yeşim Aygül, Müge Oluçoğlu, and Adil Alpkoçak

Izmir Bakırçay University, Department of Computer Engineering, Izmir, Turkey,
yesim.aygul@bakircay.edu.tr, muge.olucoglu@bakircay.edu.tr,
adil.alpkocak@bakircay.edu.tr

Abstract. With the rapid developments in the field of artificial intelligence in recent years, the potential use of models with Natural Language Processing (NLP) capabilities in medical education and evaluation processes has come to the fore. The application of these technologies in the medical field has the potential to optimize data analysis, diagnosis, and treatment processes. In this context, applications are aimed to be developed to ease the workload of healthcare professionals and increase the quality of patient care. ChatGPT-4, developed by OpenAI, along with its optimized version, ChatGPT-4o, and Google's advanced language model, Gemini, have attracted attention due to their ability to answer medical questions.

In the literature, the use of artificial intelligence models in the medical field is shown to help determine the limits of the information required, particularly for clinical environments. In this regard, MedQA, developed in 2020, is a comprehensive dataset containing more than 60,000 questions compiled from various medical specialty exams in America, specifically designed to test doctors' clinical knowledge. Generally, artificial intelligence studies are utilized in medical diagnosis and decision-making to evaluate the performance of these processes. In late 2023, Microsoft researchers achieved over 90% accuracy on the MedQA dataset in a suite of four challenging medical benchmarks called GPT-4 MultiMedQA. This achievement is considered one of the most significant indicators of how far artificial intelligence can progress in the medical field.

In this study, we tested the performance of artificial intelligence-supported language models ChatGPT-4, ChatGPT-4o, and Gemini in answering Turkish medical questions. For this purpose, we used the questions from the Medical Specialization Examination (MSE) of the first semester of 2021, an exam taken to become a specialist in the field of medicine. The exam consists of two parts: clinical and basic medical sciences, each containing 120 questions from 11 different fields.

To determine the success of the three artificial intelligence models in the MSE and to reveal significant differences between them, we manually collected the answers from each model and evaluated them according to

the evaluation report and statistics announced by ÖSYM.

This study includes three models. The first, ChatGPT-4, is a language model developed by OpenAI and trained on large datasets. It performs quite well on NLP tasks. The second model, ChatGPT-4o, is the optimized version of ChatGPT-4. The third model is Gemini, a language model developed by Google. This model is known to perform well in NLP tasks using advanced deep learning algorithms and attention mechanisms.

According to the statistics announced for the first semester of 2021, the average number of correct answers in the Clinical Medical Sciences Test was 63.95, and the maximum number of correct answers was 113. When we posed the same questions to the AI models, Gemini answered 82 correctly, ChatGPT-4 answered 105 correctly, and ChatGPT-4o answered 117 correctly. Similarly, the average number of correct answers in the Basic Medical Sciences Test was 51.63, and the maximum number of correct answers was 106. When we asked the questions in this test to the models, ChatGPT-4 and Gemini each answered 93 correctly, while ChatGPT-4o answered 107 correctly. From the results, we observed that ChatGPT-4o outperforms the best humans.

Additionally, we examined human performance in detail on a question-by-question basis and statistically evaluated whether there is a significant difference between the performance of AI models and human performance.

Keywords: ChatGPT, Gemini, examination for speciality in medicine, medical education.

BÜYÜK DİL MODELLERİ TIPTA UZMANLIK SINAVINDA İNSANLARIN BAŞARIMINI AŞABİLİR Mİ?

Yeşim Aygül, Müge Oluçoğlu, Adil Alpkocak

İzmir Bakırçay Üniversitesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye,
yesim.aygul@bakircay.edu.tr, muge.olucoglu@bakircay.edu.tr,
adil.alpkocak@bakircay.edu.tr

Özet Son yıllarda yapay zeka alanında yaşanan hızlı gelişmeler, Doğal Dil İşleme (DDİ) yeteneklerine sahip modellerin tıbbi eğitim ve değerlendirme süreçlerinde kullanılma potansiyelini gündeme getirmiştir. OpenAI tarafından geliştirilen ChatGPT-4, optimize edilmiş versiyonu olan ChatGPT-4o ve Google'ın gelişmiş dil modeli Gemini'in tıbbi soruları cevaplayabilme başarımları dikkat çekmektedir. Bu teknolojilerin tıp alanında kullanımı veri analizi, teşhis ve tedavi süreçlerini optimize etme

potansiyeli taşımaktadır. Bu bağlamda, sağlık profesyonellerinin iş yükünü hafifletecek ve hastaların bakım kalitesini artıracak uygulamaların geliştirilmesi amaçlanmaktadır.

Literatürde, tıbbi alanda yapay zeka modellerinin kullanılması, özellikle klinik ortamlar için gerekli bilgilerin sınırlarının belirlenmesine yardımcı olabilmektedir. Bu doğrultuda, 2020 yılında geliştirilen MedQA, Amerika'daki çeşitli tıbbi uzmanlık sınavlarında sorulan sorulardan derlenen ve doktorların klinik bilgilerinin test edilmesi için özel olarak tasarlanmış 60.000'den fazla soruyu içeren kapsamlı bir veri setidir ve genel olarak, yapay zeka çalışmalarının tıbbi teşhis ve karar verme süreçlerindeki başarımın değerlendirilmesi için kullanılmaktadır. 2023 yılının sonlarına doğru Microsoft araştırmacıları, GPT-4 MultiMedQA adlı dört zorlu tıbbi kıyaslama grubundan oluşan bir pakette MedQA veri seti üzerinden %90 doğruluk sınırını aşan bir başarımla elde etmişlerdir. Bu durum genel olarak yapay zekanın tıp alanında ne kadar ileri gidebileceğine dair en önemli göstergelerdendir.

Bu çalışmada, yapay zeka destekli dil modelleri olan ChatGPT-4, ChatGPT-4o ve Gemini'nin Türkçe tıbbi soruları yanıtlama performansını test ettik. Bu amaçla, tıp alanında uzmanlık kazanmak için yapılan 2021 1. Dönem Tıpta Uzmanlık Sınavı'nın (TUS) sorularını kullandık. Sınav, her biri 11 farklı alandan 120 soru içeren klinik ve temel tıp bilimleri olmak üzere iki bölümden oluşmaktadır.

Üç yapay zeka modelinin TUS'taki başarımını belirlemek ve aralarındaki anlamlı farklılıkları ortaya çıkarmak için her modelin cevaplarını manuel olarak topladık ve ÖSYM tarafından açıklanan değerlendirme raporu ve istatistiklere göre değerlendirdik.

Bu çalışma üç modeli içermektedir. Birincisi, OpenAI tarafından geliştirilen ve büyük veri kümeleri üzerinde eğitilmiş bir dil modeli olan ChatGPT-4'tür. Bu model, DDİ görevlerinde oldukça iyi performans göstermektedir. İkinci model, ChatGPT-4'ün optimize edilmiş versiyonu olan ChatGPT-4o'dur. Üçüncü model ise, Google tarafından geliştirilen bir dil modeli olan Gemini'dir. Bu model, ileri düzey derin öğrenme algoritmaları ve dikkat mekanizmalarını kullanarak DDİ görevlerinde iyi performans göstermesiyle bilinmektedir.

2021 yılı 1. dönemi için açıklanan istatistiklere göre, Klinik Tıp Bilimleri Testi ortalama doğru cevap sayısı 63.95 ve en fazla doğru cevap sayısı 113 olmuştur. Aynı soruları yapay zeka modellerine yönelttiğimizde Gemini 82, ChatGPT-4 105 ve ChatGPT-4o 117 doğru cevap verdi. Benzer şekilde, Temel Tıp Bilimleri Testi ortalama doğru cevap sayısı 51.63 ve en fazla doğru cevap sayısı 106 olarak belirtilmiştir. Bu testteki soruları modellere yönelttiğimizde ChatGPT-4 ve Gemini 93'er doğru cevap verirken ChatGPT-4o 107 doğru cevap vermiştir. Elde edilen sonuçlardan, ChatGPT-4o'nun en iyi insandan daha iyi performans gösterdiğini gözlemledik.

Ayrıca, insan performanslarını soru bazında detaylı olarak inceledik ve yapay zeka modelleri ile insan performansları arasında anlamlı bir fark olup olmadığını istatistiksel olarak değerlendirdik.

Anahtar Kelimeler: ChatGPT, Gemini, tıpta uzmanlık sınavı, tıp eğitimi.

Kaynaklar

1. Perrault, R., & Clark, J.: Artificial Intelligence Index Report 2024 (2024).
2. Jin, D., Pan, E., Oufattole, N., Weng, W. H., Fang, H., & Szolovits, P. (2021). What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14), 6421.
3. Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., ... & Horvitz, E. (2023). Can generalist foundation models outcompete special-purpose tuning? case study in medicine. *arXiv preprint arXiv:2311.16452*.
4. 2021-TUS 1.Dönem ve 2.Dönem Değerlendirme Raporu. Son Erişim Tarihi: 16.05.2024.
https://dokuman.osym.gov.tr/pdfdokuman/2021/GENEL/tusdonem1-2degrapor_29122021.pdf
5. Perişan, C., Güzelmansur, C., İnce, A.T. (2023). Yapay Zeka Modeli ChatGPT'nin Tıpta Uzmanlık Sınavı Üzerinden Değerlendirilmesi, 9. International May 19 Innovative Scientific Approaches Congress, Samsun, Türkiye, 240-242.
6. Chen, Z., Cano, A. H., Romanou, A., Bonnet, A., Matoba, K., Salvi, F., ... & Bosselut, A.(2023). Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*.
7. Tıpta Uzmanlık Eğitimi Giriş Sınavı (2021-Tus 1. Dönem) Klinik Tıp Bilimleri Testi, Son Erişim Tarihi: 16.05.2024.
https://dokuman.osym.gov.tr/pdfdokuman/2021/TUSDONEM1/CS/tusdonem_1_KTBT.pdf
8. Cabrera, J., Loyola, M. S., Magaña, I., & Rojas, R. (2023, June). Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots. In *International Work-Conference on Bio-informatics and Biomedical Engineering* (pp. 313-326). Cham: Springer Nature Switzerland.
9. Kim, Y., Xu, X., McDuff, D., Breazeal, C., & Park, H. W. (2024). Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866*.
10. Ong, Q. C., Ang, C. S., Chee, D. Z. Y., Lawate, A., Sundram, F., Dalakoti, M., ... & Car, J. (2024). Advancing Health Coaching: A Comparative Study of Large Language Model and Health Coaches. Available at SSRN 4784958.
11. Joachimiak, M. P., Miller, M. A., Caufield, J. H., Ly, R., Harris, N. L., Tritt, A., ... & Bouc-hard, K. E. (2024). The Artificial Intelligence Ontology: LLM-assisted construction of AI concept hierarchies. *arXiv preprint arXiv:2404.03044*.
12. Alberts, I. L., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?. *European journal of nuclear medicine and molecular imaging*, 50(6), 1549-1552.
13. Poulain, R., Fayyaz, H., & Beheshti, R. (2024). Bias patterns in the application of LLMs for clinical decision support: A comprehensive study. *arXiv preprint arXiv:2404.15149*.
14. Li, B., Meng, T., Shi, X., Zhai, J., & Ruan, T. (2023). MedDM: LLM-executable clinical guidance tree for clinical decision-making. *arXiv preprint arXiv:2312.02441*.
15. Web Linki: <https://openai.com/index/gpt-4/> Son Erişim tarihi (23 Mayıs 2024)
16. Web Linki: <https://blog.google/technology/ai/google-gemini-ai/sundar-note> Son Erişim Tarihi (23 Mayıs 2024)
17. Liu, X., et al. (2020). "Natural Language Processing for Medical Knowledge Extraction." *Journal of Medical Systems*.
18. Jin, D., et al. (2021). "Evaluating Large Language Models Trained on Medical Data for USMLE Question Answering." *IEEE Transactions on Medical Data*.
19. Zhang, H., et al. (2022). "Performance of Chinese Large Language Models on Chinese Medical Licensing Exam." *Chinese Journal of Medical Informatics*.
20. Chen, J., et al. (2020). "Comparative Analysis of Human and AI Performance on Medical Exams." *Journal of Artificial Intelligence Research*.
21. Microsoft Research. (2023). "Evaluation of Large Language Models on Medical Exams: A Comprehensive Study." *Microsoft Research Journal*.

Detection of psychological disorders from structural MRI with deep learning methods

Osman Tayfun Bişkin¹[0000-0002-2326-9438], Cemre Candemir^{2,3}[0000-0001-9850-137X], Mustafa Alper Selver^{4,5}[0000-0002-8445-0388], and Pembe Keskinoglu⁶[0000-0002-3459-1828]

¹ Department of Electrical and Electronics Engineering, Burdur Mehmet Akif Ersoy University, Burdur 15030, Turkey

² International Computer Institute, Ege University, Izmir 35100, Turkey

³ Standardization of Computational Anatomy Techniques, SoCAT Lab, Ege University, Izmir 35100, Turkey

⁴ Department of Electrical and Electronics Engineering, Dokuz Eylul University, Izmir 35160, Turkey

⁵ Izmir Health Technologies Development and Accelerator (BioIzmir), Dokuz Eylul University, Izmir 35160, Turkey

⁶ Faculty of Medicine Basic Medical Sciences Department of Biostatistics and Informatics, Dokuz Eylul University, Izmir 35160, Turkey

Abstract. In this study, individuals with schizophrenia and bipolar disorders are classified using MR images of volunteers by employing deep neural networks. Psychiatric disorders are traditionally diagnosed through detailed evaluations and tests conducted by specialized professionals. However, due to the potentials for human error in these evaluations there is a need for automated diagnostic systems. Last decades, advances in artificial intelligence, particularly deep learning, have begun to offer substantial improvements in diagnostic accuracy. Therefore, in this study, we applied a deep learning model to classify the psychiatric disorders using MR images. The images, which are publicly available, were obtained from 272 volunteers ranging in age from 21 to 50. Publicly available dataset University of California Los Angeles (UCLA) Consortium of Neuropsychiatric Phenomics LA5c study categorized into four groups: healthy controls, individuals with schizophrenia, those with bipolar disorder, and individuals with attention deficit hyperactivity disorder (ADHD).

In this study, we specifically focus on classifying MR images from 50 individuals diagnosed with schizophrenia and 49 with bipolar disorders. For this purpose, in order to perform the classification task, we utilized a ResNet50 model pre-trained on ImageNet. Although the dataset consists of images with dimensions of 179x256, they are resized for the input of deep neural network. In dataset, T1-weighted MR volume consist of 256 axial slice for each participant. However, the central 32 slice are used for training the deep neural network model. In the pre-processing stage, the images are initially resized to dimensions of 224x224. After resizing, the images are subjected to intensity normalization to standardize the dataset. In order to overcome the overfitting problem while training the network, augmentation process is also implemented on data. This augmentation process is

accomplished considering reliability of the medical image to avoid irrelevant results that is inconsistent with real MR images. Our preliminary results demonstrate that deep learning models can be effectively used to classify psychiatric disorders from MR images of individuals with schizophrenia and those with bipolar disorder.

Keywords: Psychological Disorders, Deep Learning, Structural MRI, ResNet50

1 Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive neuroimaging technique that is used to constitute the detailed images of the body's internal structures, including the brain [1]. It allows researchers and clinicians to observe and detect the structure and abnormalities of soft tissues. Structural changes, such as tumors, injuries, bleeding, infections can be identified with MRI scan. On the other hand, neurological conditions and diseases also can be detected by brain MRI. Therefore, it provides a vital tool for diagnosing and monitoring various brain disorders such as multiple sclerosis, Alzheimer's and bipolar disorder.

The principle of the MRI is to align the hydrogen atoms in the tissues with a strong magnetic field. When radio waves are sent, these hydrogen atoms deploy signals which are detected with the MRI scanner. By analyzing the signal changes, the MRI scanner constructs detailed images of the brain anatomy and function [2]. MRI offers good spatial resolution and the ability to capture different tissue contrasts. Thus, many brain disorders can be detected with the detailed images of the brain anatomy[3].

On the other hand, brain disorders can lead to a wide range of changes in the brain morphology, reflecting the underlying pathology and affecting brain structure at different levels. These changes can be the loss of the brain tissue volume, known as atrophy [4], ventricular enlargement or shrinkage, gray matter and white matter changes, changes in cortical thickness and folding patterns, or malformations and developmental abnormalities. The changes in the brain tissue also provides valuable information for understanding neurological disorders like Alzheimer's disease, schizophrenia, depression, and bipolar disorder [5], [6], [7].

In neurodegenerative diseases, such as Alzheimer's disease, specific regions of the brain may show progressive shrinkage due to neuronal loss and synaptic dysfunction, leading to characteristic patterns of cortical and subcortical atrophy [8]. On the other hand, a series of changes in brain morphology are also reported in schizophrenia and bipolar disorder [9]. Patients with schizophrenia often show ventricular enlargement, which may reflect a loss of surrounding brain tissue, including gray and white matter [10]. Schizophrenia is associated with reductions in gray matter volume, especially in regions such as the prefrontal cortex, temporal lobes, hippocampus, and thalamus [11]. These reductions may be present at the onset of illness and progress over time. Like schizophrenia, patients with bipolar disorder may show enlarged ventricles, suggesting

loss of brain tissue. It is also reported that bipolar disorder is also associated with alterations in gray matter volume, including reductions in the prefrontal cortex, anterior cingulate cortex, and amygdala [12]. These changes may contribute to mood dysregulation and cognitive deficits.

Therefore, it can be said that MRI provides valuable insights into the neurobiological underpinnings and structural changes of psychological disorders, such as schizophrenia and bipolar disorder, helping to elucidate the complex relationships between brain function, behavior, and psychiatric symptoms. However, it's important to note that MRI findings are typically used in conjunction with clinical assessments and other neuroimaging modalities to enhance diagnostic accuracy and inform treatment strategies.

2 Dataset and Data Preprocessing

The MR images used in this study is available as an open-source dataset collected from the University of California Los Angeles (UCLA) Consortium of Neuropsychiatric Phenomics LA5c study [13]. MR images in this dataset collected from 272 volunteer participants, whose ages range from 21 to 50 years. The dataset consists of four groups of participants suffering from schizophrenia, bipolar disorder, attention deficit hyperactivity disorder (ADHD), and healthy control individuals.

In this study, we focused on the classification schizophrenia and bipolar disorders by employing MR images. The subset of dataset used for simulations included MR images from 50 individuals with schizophrenia and 49 individuals with bipolar disorder. The original size of images in the dataset are 179x256. To use these images in ResNet50 architecture, they were resized to 224x224 dimensions. Figure 1, shows MR images obtained from schizophrenia and bipolar disorders.

Although, T1-weighted MR volume consist of 216 axial slices for each participant, we use the central 32 slice for training the network. After selecting the central 32 slice for each participant, normalization is applied over all images in order to standardize the intensity. Finally, data augmentation process is used to overcome the overfitting problem. Still, in order to avoid irrelevant result of augmentation which is inconsistent with real MR images, reliability of medical image is taken into account. Therefore, for the augmentation techniques, we used rotation and brightness. Besides, images are scaled to a range of 0 to 1.

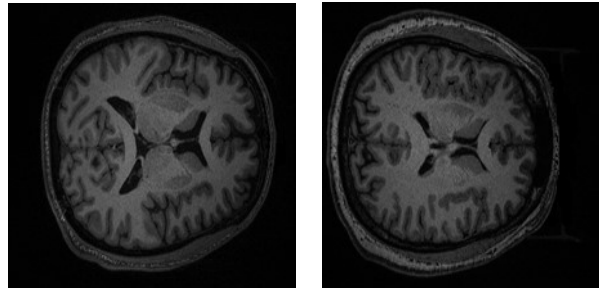


Fig 1. 224x224 resized MR images of patients bipolar (left) and schizophrenia (right).

3 Methods and Simulation Results

In this study, we employed ResNet50 deep neural network architecture in order to classify psychological disorder named bipolar and schizophrenia. The ResNet50 model, originally designed by [14], uses 50-layer deep residual block including skip connections that enhances the performance of neural networks.

For our classification task, pre-trained Resnet model, with images from ImageNet database which include more than 14 million images was used. Here, we fine-tuned the standard ResNet50 architecture both by adding a global average pooling layer followed by a dense layer and training the last 10 layers and freezing the earlier layers. Global average pooling layer computes the spatial average of the feature maps obtained from the last convolutional layer of ResNet50. By this way, the number of features is reduced in order to handle the large amount of data. Besides, dense layer consisting of 128 units with a ReLU (Rectified Linear Unit) activation function was added.

In order to optimize the cost function while training the model, Adam optimizer was employed. The learning rate of optimizer is selected as 0.001 to optimize the training process and the batch size was chosen as 64. These hyperparameters of model are found experimentally that yield the best results. The dataset was strategically partitioned, allocating 10% for testing, 5% for validation, and the remaining 85% for training purposes.

The performance of the model was assessed based on simulation results derived from the test data. The evaluation criteria utilized for this assessment included accuracy, F1-score, recall, and precision metrics. Simulation results are given in Table 1 below. As shown in Table 1, F1-Score and recall of bipolar disorder are higher than schizophrenia. On the other hand, precision of schizophrenia is higher compared to bipolar and overall accuracy of deep learning model is %73.43.

Table 1. Performance (%) of model for classification of bipolar and schizophrenia.

	Recall	Precision	F1-Score
Bipolar	0.7786	0.7286	0.7443
Schizophrenia	0.6886	0.7729	0.7114
Overall Accuracy			73.43

4 Conclusion

In this study, deep learning techniques is utilized to classify individuals with schizophrenia and bipolar disorders using MR images. The data is obtained from the UCLA Consortium of Neuropsychiatric Phenomics LA5c study and series of data prepro-

cessing operations such as resizing and normalization is applied on them. Data augmentation techniques were also implemented to prevent overfitting. Simulation results indicate that deep learning models can be used to classify psychiatric disorders from MR images of individuals diagnosed with schizophrenia and bipolar disorders. According to our results, these disorders can be classified with an accuracy of %73.43. These results show the potential to develop automated diagnostic systems that can reduce human error. Besides, it can be used to improve diagnostic accuracy.

5 References

- [1] E. Amaro Jr. and G. J. Barker, "Study design in fMRI: Basic principles," *Brain Cogn.*, vol. 60, no. 3, pp. 220–232, Apr. 2006, doi: 10.1016/j.bandc.2005.11.009.
- [2] A. Berger, "Magnetic resonance imaging," *BMJ*, vol. 324, no. 7328, p. 35, Jan. 2002.
- [3] A. C. Elias-Jones, T. Jaspan, D. H. Mellor, and B. S. Worthington, "Magnetic resonance imaging in neurological disorders.," *Arch. Dis. Child.*, vol. 65, no. 9, pp. 922–929, Sep. 1990.
- [4] "Structural Magnetic Resonance Imaging - an overview | ScienceDirect Topics." Accessed: May 16, 2024. [Online]. Available: <https://www.sciencedirect.com/topics/neuroscience/structural-magnetic-resonance-imaging>
- [5] G. Chen *et al.*, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, Apr. 2011, doi: 10.1148/radiol.10100734.
- [6] G. S. Dichter, D. Gibbs, and M. J. Smoski, "A systematic review of relations between resting-state functional-MRI and treatment response in major depressive disorder," *J. Affect. Disord.*, vol. 172, pp. 8–17, Feb. 2015, doi: 10.1016/j.jad.2014.09.028.
- [7] E. S. Paykel, "Life events, social support and depression," *Acta Psychiatr. Scand. Suppl.*, vol. 377, pp. 50–58, 1994.
- [8] E. Canu *et al.*, "Mapping the structural brain changes in Alzheimer's disease: The independent contribution of two imaging modalities," *J. Alzheimers Dis.*, vol. 26, no. Suppl 3, pp. 263–274, 2011, doi: 10.3233/JAD-2011-0040.
- [9] B. Sutubasi *et al.*, "Anatomical connectivity changes in bipolar disorder and schizophrenia investigated using whole-brain tract-based spatial statistics and machine learning approaches," *Neural Comput. Appl.*, vol. 31, no. 9, pp. 4983–4992, Sep. 2019, doi: 10.1007/s00521-018-03992-y.
- [10] P. Arrondo, Ó. Elía-Zudaire, G. Martí-Andrés, M. A. Fernández-Seara, and M. Riverol, "Grey matter changes on brain MRI in subjective cognitive decline: a systematic review," *Alzheimers Res. Ther.*, vol. 14, no. 1, p. 98, Jul. 2022, doi: 10.1186/s13195-022-01031-6.
- [11] A. Breier, R. W. Buchanan, A. Elkashef, R. C. Munson, B. Kirkpatrick, and F. Gellad, "Brain morphology and schizophrenia. A magnetic resonance imaging study of limbic, prefrontal cortex, and caudate structures," *Arch. Gen. Psychiatry*, vol. 49, no. 12, pp. 921–926, Dec. 1992, doi: 10.1001/archpsyc.1992.01820120009003.

- [12] A. C. Altamura *et al.*, "The impact of psychosis on brain anatomy in bipolar disorder: A structural MRI study," *J. Affect. Disord.*, vol. 233, pp. 100–109, Jun. 2018, doi: 10.1016/j.jad.2017.11.092.
- [13] Bilder, R and Poldrack, R and Cannon, T and London, E and Freimer, N and Congdon, E and Karlsgodt, K and Sabb, F (2018). UCLA Consortium for Neuropsychiatric Phenomics LA5c Study. OpenNeuro.
- [14] He, K., Zhang, X., Ren, S., & Sun, J., "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016, doi:10.48550/arXiv.1512.03385

Ses Dalgalarından Öfke Belirleme İçin Evrişimli Sinir Ağı Modeli

Mehmet YİĞİT¹[0000-0001-6115-0537], Mehmet Erşan KALAYCI²[0000-0002-4797-9020], Kemal TURHAN³[0000-0001-7871-3025]

¹ Ordu üniversitesi Eğitim ve Araştırma Hastanesi, Ordu, Türkiye

² Karadeniz Teknik Üniversitesi, Trabzon, Türkiye

³ Karadeniz Teknik Üniversitesi, Trabzon, Türkiye

mehmettygt@gmail.com, mehmetersankalayci@gmail.com,
kemalturhan@ktu.edu.tr

Convolutional Neural Network Model for Detecting Anger from Sound Waves

Abstract. The audio data also contains emotional content. Nowadays, it is aimed to develop a method that will help to determine the verbal and physical violence that prevents health workers from working in a safe environment. The model created to determine the verbal violence that healthcare workers will be exposed to due to the work environment. It will contribute to the prevention of possible violence by warning the security personnel at the time of violence with the help of machine learning from the audio data to be collected in the work environment. In this paper, they developed a model that detects anger from sound waves in order to provide a safe working environment for healthcare workers. The study used the CNN (Convolutional Neural Networks) method to analyze approximately 10-second-long sound fragments obtained from Turkish dubbed television series. The model was able to distinguish between angry and neutral voices with high accuracy and achieved a success rate of 75%. The results of the study show that higher success rates can be achieved when the data set is expanded, and that this model can be used for security purposes in the field of healthcare.

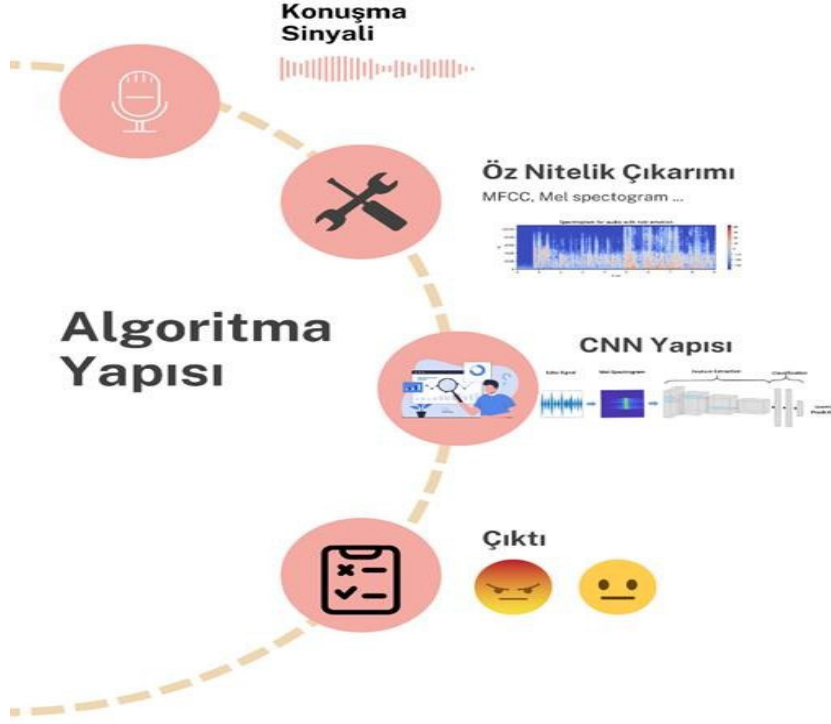
Keywords: Machine Learning, Medical Informatics, Speech Recognition Software

1 Giriş

Bireyler, birbirleri ile iletişim kurmasını sağlayan ve toplumlarca kullanılmakta olan semboller sistemini kullanmaktadır. Konuşma ise bu sözel sembolleri ifade eden nöro müküller süreçlerdendir. Dil, düşüncelerimizi başkalarıyla iletişim kurmamızı sağlar. Bu fikirlerimizi başkalarıyla paylaşmamıza ve onların fikirlerinden öğrenmemize olanak tanır [1-2]. Sağlık kuruluşlarında çalışan sağlık profesyonelleri, hastalarla veya hasta yakınlarıyla etkileşim-iletişim halinde oldukları için zaman zaman şiddete uğrama riskiyle karşı karşıya kalabilirler. Bu etkileşimin tedavi sürecinde başarılı olması durumunda, hastaların memnuniyeti ve sağlık profesyonelinin iş doyumunu artırmaktadır [3]. Bazen iş yoğunluğu ve ortamın stresinden, beklenmedik durumlar karşısında hastaların veya hasta yakınlarının duygusal tepkileri (üzüntü, öfke vb.) sağlık çalışanlarına yönelik sözel, psikolojik, fiziksel şiddet biçiminde kendini gösterebilmektedir. [4]. Tüm dünyada olduğu gibi ülkemizde çeşitli yöntemler ile şiddetin önlenmesine yönelik faaliyetler bulunmaktadır. Sağlık Bakanlığı tarafından uygulanan Beyaz Kod Çağrı Sistemi herhangi bir şiddet (sözel, psikolojik, fiziksel) durumunda sağlık profesyonelinin güveni için kullanılan bir yöntemdir. Bu yöntemde çalışan olası bir şiddet durumunda hastane içerisindeki güvenlik birimlerine bir çağrı iletmekte ve güvenlik birimlerinin olası şiddet mekanına ulaşmasını sağlamaktadır [5]. Herhangi bir şiddet esnasında öfke duygusu bireyin sesini yükseltmesine ve konuşmasının hızının artmasına neden olur. Öfke duygusu, bireyin kan basıncının artmasına da neden olabilmektedir. Öfkeli bir konuşma ortamdaki bireyleri de etkilemektedir [6]. Konuşma mesaj, konuşmacı, dil ve duygu hakkında bilgi içeren karmaşık bir sinyaldir [7]. Bu sesler makinelerce işlenmek üzere sinyallere dönüşür. İnsan kulağına benzer bir şekilde makineler, bu sinyalleri algılamaya çalışmaktadır. Genellikle makine öğrenmesi yöntemleri verileri işlemek için kullanılan yöntemlerdendir [8]. Bu çalışmada, öfkeli bir konuşma anında oluşan ses sinyalleri model tarafından saptanmaya çalışılacaktır. Mevcut yapılan çalışmalar genellikle profesyonel stüdyolarda hazırlanmış ve işaretlenmiş veri setleri (RAVDESS, SAVEE, CREMA-D, TESS, EMODB) aracılığı ile yapılmaktadır. Bu veri setleri üzerinden duygu çıkarımı yapılmaktadır. 8 adet (öfke, beklenti, üzüntü, şaşkınlık, tikslenme, güven, sevinç ve korku) duygunun çıkarımı ile ilgili birçok model vardır [9]. Bu çalışmanın amacı, sağlık profesyonellerinin çalışma ortamında oluşabilecek güvenlik sorunlarını sesli iletişim cihazları ile tespit ederek sözel ve fiziksel şiddetin önlenmesini sağlamaktır.

2 Gereç ve Yöntem

Genellikle mevcut literatürdeki çalışmalarda, küçük ses parçaları ve duyguları içeren ses kayıtları stüdyo ortamında kayıt yapılarak etiketlenmektedir. Çalışmada bu yöntem kullanılarak açık kaynak bir dizindeki sesler ~10sn'lik bir konuşma içeren metin halinde bölünerek sezgisel olarak etiketlenmiştir. 362 adet nötral ve 280 adet öfkeli şekilde toplamda 642 adet ses verisi etiketlenmiştir.



Şekil. 1. Oluşturulan Model Yapısı

3 Materyal ve Metot

Veri Kümesi Genellikle mevcut literatürdeki çalışmalar küçük ses parçaları ve duyguları içeren ses kayıtları stüdyo ortamında kayıt yapılarak etiketlenmektedir. Çalışmada mevcut yöntem kullanılarak açık kaynak bir televizyon dizisinin Türkçe dublajındaki sesler kullanılmıştır. Sesler ~10sn'lik bir konuşma içeren anlamlı haldeki parçalara bölünerek sezgisel olarak etiketlenmiştir. Toplamda 642 ses verisi, 362 nötral ve 280 öfkeli olacak şekilde etiketlenmiştir [10].

Öznitelik Çıkarımı Ses verilerinin işlenmesinde en sık kullanılan ve en başarılı yöntem olan CNN (Convolutional Neural Networks) özellikle görüntü işleme ve bilgisayarla görme görevlerinde yaygın olarak kullanılan derin öğrenme

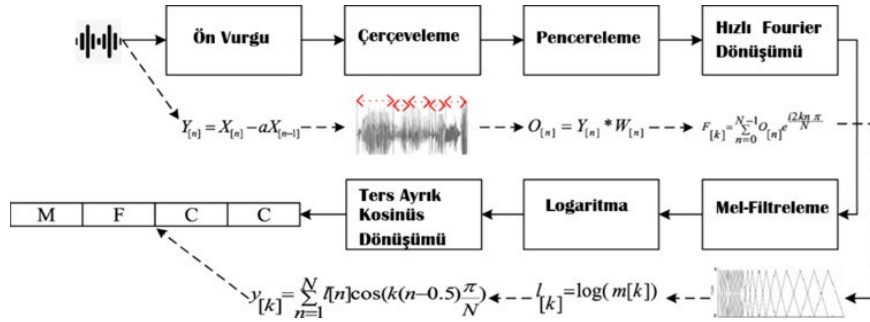
modellerindedir. CNN'ler, geleneksel yapay sinir ağlarından farklı olarak, yerel bağlantılar ve ağırlık paylaşımı gibi özellikler sayesinde, özellikle iki boyutlu verilerle (görüntüler gibi) etkili bir şekilde çalışabilmektedir. CNN yapısında eğitim modeli olarak ses verileri MFCC (Mel-Frequency Cepstral Coefficients), öz nitelik çıkarımı yönteminden yararlanılmaktadır. MFCC, ses sinyallerini insan algısına daha yakın bir şekilde temsil ettiği için, ses ve konuşma işleme görevlerinde önemli bir araçtır. Bu nedenle, özellikle makine öğrenmesi ve derin öğrenme tabanlı ses işleme uygulamalarında yaygın olarak kullanılır.

MFCC

Ses sinyallerini algoritmanın tanımlamasını ve analiz etmesini sağlamaktadır. Yaygın olarak ses verilerin tanımlamada ve duygu çıkarımı yapılmasında kullanılır. MFCC'lerin avantajı ses verilerini düşük boyutlu ve anlamlı bir temsil haline getirmektedir. Makine öğrenimi ve derin öğrenme metodlarında oldukça kolaylık sağlamaktadır.

Ses Sinyalinin Parçalanması (Framing) Ses sinyalleri, küçük analiz edilebilir parçalar (frame) haline getirilir. Bu çerçeveler 20-40 ms uzunluğundadır. Çerçeveler birbirleri örtüşür, her çerçevenin belirli kısmının önceki ve sonraki çerçevelerle ortak olmasını sağlar.

Ses Sinyalinin Parçalanması (Framing) Ses sinyalleri, küçük analiz edilebilir parçalar (frame) haline getirilir. Bu çerçeveler 20-40 ms uzunluğundadır. Çerçeveler birbirleri örtüşür, her çerçevenin belirli kısmının önceki ve sonraki çerçevelerle ortak olmasını sağlar.



Şekil. 2. MFCC oluşum şeması [11].

Pencereleme (Windowing) Her çerçeve ile pencereleme fonksiyonu (Hamming penceresi) çarpılır. Bu, çerçevenin başındaki ve sonundaki ani değişikliklerin etkisini azaltır.

Hızlı Fourier Dönüşümü (FFT) Her çerçeve üzerinde FFT uygulanarak frekans spektrumu elde edilir. Zaman alanındaki sinyali, frekans alanına dönüştürür.

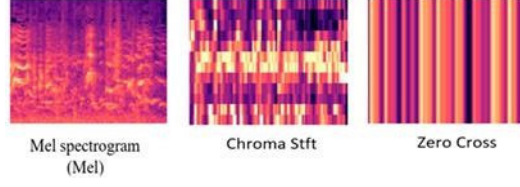
Mel-Frekans Ölçeği Frekans spektrumu insan kulağının frekansları algılama biçimine benzer şekilde çalışan Mel-Frekans Ölçeğine dönüşür.

Logaritmik Enerji Hesaplama Mel-frekans bantlarının her birinin logaritmik enerjisi hesaplanmaktadır. İnsan kulağının ses yoğunluğunu algıma biçimiyle benzerdir.

Ters Fourier Dönüşümü (DCT) Son olarak, logaritmik Mel frekans spektrumuna ters Fourier dönüşümü DCT (Discrete Cosine Transform) uygulanır. Bu, son MFCC'leri oluşturur. İlk birkaç MFCC katsayısı genellikle en anlamlı olanlardır ve çoğu zaman sadece bu katsayılar kullanılır.

Veri Artırma (Data Augmentation) Ses dosyalarında veri artırma (data augmentation) yöntemleri, makine öğrenimi modellerinin performansını artırmak ve aşırı öğrenmeyi (overfitting) azaltmak için kullanılan tekniklerdir. Bu yöntemler, mevcut ses verilerini çeşitli şekillerde dönüştürerek veri kümesini genişletir ve çeşitlendirir. Veri setimizin başarısını arttırmak için ses kayıtlarında oluşabilecek Librosa kütüphanesinden yararlanılarak 3 akustik ses artırma yöntemi kullanılmıştır. Gürültü enjeksiyonu (Additive Noise), Zaman Çarpıtma (Time Stretching), Perdeyi Ton Kaydırma (Pitch Shifting) gibi yöntemler kullanılarak doğala en yakın bir ses kaydı elde edilmektedir. Bu yöntemlerin her biri, modelin daha genel ve esnek olmasını sağlamak için ses verilerini çeşitli yollarla dönüştürür ve genişletir. Genellikle, birden fazla veri artırma yöntemi ile birlikte kullanılır ve bu eğitim veri kümesinin çeşitliliğini artırarak modelin daha iyi genelleme yapmasını sağlar. Veri Setine, veri artırma işleminden sonra öznitelik çıkarımı yapılmaktadır [11].

Modellerin Hazırlanması Çalışmada kullanılacak öznitelikler sırasıyla; Sıfır Geçiş Oranı (ZCR), Chroma Stft, MFCC, RMS (kök ortalama kare değer) ve Mel Spektrogramıdır. Her bir özelliğin ortalama değerleri bir araya getirilir ve tek bir vektör olarak haline getirilir. Ses verilerini yükler, ardından bu ses verileri üzerinden ayrıca gürültü eklenmiş bir versiyonunun da özelliklerini çıkarır. Son olarak, her iki özellik vektörünü dikey olarak (sıra bazında) birleştirir ve bir özellik matrisi olarak döndürür. Bu kod, ses dosyalarından elde edilen özellikleri bir veri kümesine hazırlamak için kullanılabilir [12].



Şekil. 3. Mel Spectrogram, Chroma Stft, Zero Cross Yapıları

Derin Öğrenme Algoritmaları 1 Boyutlu CNN yapısı

Tablo 1. 1 Boyutlu CNN yapısı Veri Seti Kümesi

Veri Seti	Veri Yüzdesi	Veri Sayısı
Öfkeli Ses (Test)	%20	53
Öfkeli Ses Eğitim)	%80	210
Nötral Ses (Test)	%20	66
Nötral Ses (Eğitim)	%80	261

Veri Kümesi, Veri Arttırma (Data Augmentation) yöntemi ile büyütülerek eğitime sokulmuştur. Bu sonuçların çıktılarını sonuçlara dahil edilmemiştir. Eğitim verisine de veri büyütme yapıldığı için sonuçları etkilediği varsayılmıştır [13]. Giriş katmanındaki filtre sayısı 256'dır. Kernel Size (Kernel Boyutu):5X5, her filtrenin genişliğini belirtir (5 zaman adımı boyunca konvolüsyon yapılır). Stride (1 zaman adımı) filtrelerin her adımda ne kadar kaydırılacağını belirtir, girdi ve çıktının aynı boyutta olmasını sağlar, sınırdaki öğeler için sıfır doldurma (zero padding) kullanır. Bu işlem için ReLU (Rectified Linear Unit) aktivasyon fonksiyonu kullanılmıştır. Girdi verisinin şekli ile zaman adımı sayısını ve her adımda bir özellik olduğu belirtilmiştir.

Havuzlama penceresinin genişliği 5 stride (5 zaman adımı), havuzlama penceresinin her adımda ne kadar kaydırılacağını 2 stride (2 zaman adımı) olarak belirlenmiştir. Sınırdaki öğeler için sıfır doldurma (zero padding) kullanılmıştır. Yine aynı yapı, ancak bu sefer girdi şekli otomatik olarak önceki katmandan alınarak oluşturulan bu katmanlar, verinin daha fazla filtre ve özellik çıkarımı yapması için kullanılmıştır. Bu katmandaki filtre sayısı 128 olarak belirlenmiştir. Daha az filtre ile daha derin özellikler çıkarılmıştır. Diğer katmandaki filtre sayısı ise 64 olarak belirlenmiştir. Daha derin ve daha az filtre ile özellik çıkarımı yapılmıştır. Flatten, çok boyutlu bir girişi tek boyutlu bir vektöre dönüştürür. Özellikle yoğun (fully connected) katmanlardan önce kullanılır. 32 nöronlu tam bağlantılı bu katman ReLU aktivasyon fonksiyonu ile %30 oranında rastgele olarak nöronları kapatıp, 2 nöronlu çıkış katmanı (ikili sınıflandırma

için) kullanılan Softmax aktivasyon fonksiyonu, çıktıları olasılık dağılımı olarak normalize etmiştir. “Adam” optimizasyon algoritması kullanılmıştır.

Çok sınıflı sınıflandırma problemleri için kategorik çapraz entropi kayıp fonksiyonu kullanılmıştır. Modelin başarısını ölçmek için doğruluk (accuracy) metriği kullanılmıştır. %20 oranında rastgele olarak nöronları kapatılarak aşırı öğrenme (overfitting) önlenmiştir.

```
model.Sequential()
model.add(ConvID(56, kernel_size=5, strides=1, padding='same', activation='relu', input_shape=(x_train.shape[1], 1)))
model.add(MaxPoolingID(pool_size=2, strides=2, padding='same'))
model.add(ConvID(56, kernel_size=5, strides=1, padding='same', activation='relu'))
model.add(MaxPoolingID(pool_size=2, strides=2, padding='same'))
model.add(ConvID(28, kernel_size=5, strides=1, padding='same', activation='relu'))
model.add(MaxPoolingID(pool_size=2, strides=2, padding='same'))
model.add(Dropout(0.2))
model.add(ConvID(4, kernel_size=5, strides=1, padding='same', activation='relu'))
model.add(MaxPoolingID(pool_size=2, strides=2, padding='same'))
model.add(Flatten())
model.add(Dense(units=32, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(units=0, activation='softmax'))
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()
Model: "sequential_14"
-----
Layer (type)                Output Shape                Param #
-----
conv1d_53 (Conv1D)          (None, 162, 256)           1536
max_pooling1d_40 (MaxPoolin (None, 81, 256)            0
gID)
conv1d_54 (Conv1D)          (None, 81, 256)           327936
max_pooling1d_41 (MaxPoolin (None, 41, 256)            0
gID)
conv1d_55 (Conv1D)          (None, 41, 128)           163968
max_pooling1d_42 (MaxPoolin (None, 21, 128)            0
gID)
dropout_26 (Dropout)        (None, 21, 128)            0
conv1d_56 (Conv1D)          (None, 21, 64)            41024
max_pooling1d_43 (MaxPoolin (None, 11, 64)            0
gID)
flatten_13 (Flatten)        (None, 704)                0
dense_24 (Dense)            (None, 32)                 22560
dropout_27 (Dropout)        (None, 32)                 0
dense_25 (Dense)            (None, 0)                  264
-----
Total params: 557,288
Trainable params: 557,288
Non-trainable params: 0
```

Şekil 4. 1 Boyutlu CNN modeli

Derin Öğrenme Algoritmaları 2 Boyutlu CNN yapısı

Tablo 2. 2 Boyutlu CNN yapısı Veri Seti Kümesi

Veri Seti	Veri Yüzdesi	Veri Sayısı
Öfkeli Ses (Test)	%20	58
Öfkeli Ses Eğitim)	%80	263
Nötral Ses (Test)	%20	60
Nötral Ses (Eğitim)	%80	327

Veri Seti, Eğitim ve Test verisi olmak üzere ikiye ayrılmıştır. Bu yapılar öz nitelik çıkarımı yapılarak 2 boyutlu CNN modelinde eğitilmiştir. Bidirectional LSTM (Long Short-Term Memory) modeli de kullanılmıştır. Bidirectional LSTM katmanı hem ileri hem de geri yönlerde işler. Giriş katmanındaki filtre sayısı 128'dir, LSTM katmanındaki hücre sayısını belirtir. Her adımda giriş birimlerinden %5'ini rastgele kapatır. Her adımda geri döngü bağlantılarının %20'sini rastgele kapatır. Her zaman tam bir dizi (sequence) döndürüleceğini belirtir (örneğin, başka bir LSTM katmanı eklemek için gereklidir). Giriş verisinin 25 stride (zaman adımına) ve 1 özelliğe sahip olduğunu belirtir.

Dense, tam bağlantılı (fully connected) katmanlardır. Aktivasyon fonksiyonunu kullanır. Bu fonksiyon, negatif girdileri sıfırlar ve pozitif girdileri aynen geçirir. Dropout katmanı, %20 oranında rastgele olarak nöronları kapatarak aşırı öğrenmeyi (overfitting) önlemeye yardımcı olur.

Flatten, çok boyutlu bir girişi tek boyutlu bir vektöre dönüştürür. Özellikle yoğun katmanlardan önce kullanışlıdır. 2 birimli bir çıkış katmanıdır. Bu, modelin ikili sınıflandırma yapacağını gösterir. Bu modelde, Sigmoid aktivasyon fonksiyonu kullanılmıştır. Bu, çıktıyı [0, 1] aralığına sıkıştırır ve her bir sınıf için olasılık döndürür. İkili sınıflandırma için uygun olan kayıp fonksiyonudur. Bu modelde, Adam optimizasyon algoritması kullanılmıştır. Modelin başarısını ölçmek için doğruluk (accuracy) metriğini kullanır.

```

model = Sequential()
model.add(Bidirectional(LSTM(128, dropout=0.05, recurrent_dropout=0.20, return_sequences=True, input_shape = (25,1))))
model.add(Dense(128, activation='relu'))
model.add(Dense(64, activation='relu'))
#model.add(Dropout(0.6))
model.add(Dense(32, activation='relu'))
#model.add(Dense(32, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(32, activation='relu'))
model.add(Flatten())

model.add(Dense(2, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='Adam', metrics=['acc'])

model.summary()

```

Layer (type)	Output Shape	Param #
bidirectional_12 (Bidirectional)	(None, 25, 256)	133120
dense_68 (Dense)	(None, 25, 128)	32896
dense_69 (Dense)	(None, 25, 64)	8256
dropout_19 (Dropout)	(None, 25, 64)	0
dense_70 (Dense)	(None, 25, 32)	2080
dense_71 (Dense)	(None, 25, 32)	1056
dropout_20 (Dropout)	(None, 25, 32)	0
dense_72 (Dense)	(None, 25, 32)	1056
flatten_12 (Flatten)	(None, 800)	0
dense_73 (Dense)	(None, 2)	1602

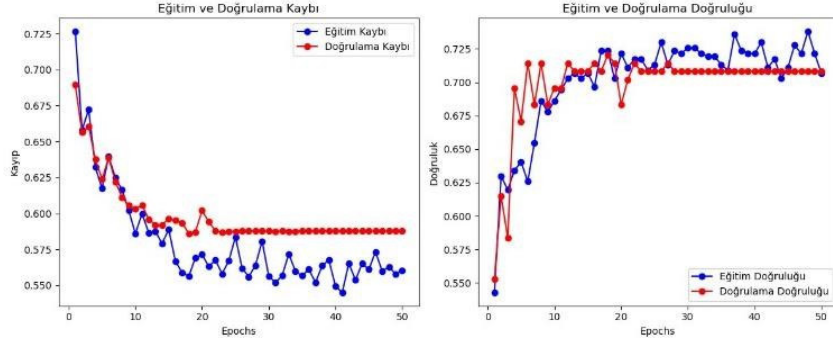
Total params: 180,866
Trainable params: 180,866
Non-trainable params: 0

Şekil 4. 2 Boyutlu CNN modeli

4 Bulgular

Modeller ve başarı oranları aşağıda verilmiştir.

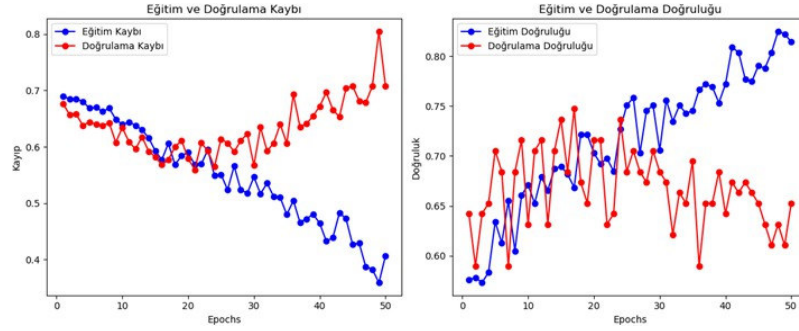
1 Boyutlu CNN modelinin başarısı



Şekil 5. Modelin Doğruluğu epochs=50, batch_size=64'da

Doğruluk: %70,8'dir. Hata oranı: 0,5877'dir.

2 Boyutlu CNN modelinin başarısı



Şekil 6. Modelin Doğruluğu epochs=100, batch size=128'da

Doğruluk: %76,4'dür. Hata oranı : 0,5754'tür.

Tablo 3. Modellerin Başarı Değerlendirme Metrikleri

	Kesinlik	Duyarlılık	F1 Puanı
1 Boyutlu CNN Modeli			
Öfke	0,79	0,47	0,59
Nötr	0,68	0,90	0,77
Doğruluk			0,71
Ağırlıklı Ortalama	0,73	0,71	0,69
2 Boyutlu CNN Modeli			
Öfke	0,80	0,62	0,70
Nötr	0,70	0,87	0,78
Doğruluk			0,74
Ağırlıklı Ortalama	0,75	0,75	0,74

5 Sonuç ve Tartışma

Derin öğrenme yöntemlerinden birisi olan CNN'in en çok tercih edilen sınıflandırma modeli olan 1D CNN ve 2D CNN gibi yöntemler çalışmamızda kullanılmıştır. Ses verilerinden öfke tahmininde en başarılı model '2D CNN' modeli olmuştur. Literatürde, 3D CNN-LSTM modeli kullanılarak yapılan çalışmada kızgın (öfkeli) seslerde % 75 doğruluk bulunmuştur [13]. MFCC yöntemi kullanılarak yapılan çalışmada ise mutlu, üzgün ve kızgın duygularının doğrulamasında % 80 tahmin başarısına ulaştığı görülmüştür (Hakdağlı ve Demirtaş). İngilizce konuşma dilinde uyarlanmış ses kütüphanelerinden yapılan bu çalışmalarda veri seti örneklem sayısı oldukça geniştir [14-15]. Bu çalışma da veri seti ile sınırlıdır.

Verilerin işaretlenmesi için birbirinden bağımsız iki hakem tarafından değerlendirilmesi, uyumsuzluk durumunda konsensus kararı ile işaretleme yapılması durumunda yanlışlığın azabileceği öngörülmüştür. Bu ses verilerinin, kelimeleri de anlam bilgisi ile ilişkilendirildiğinde başarının artacağı düşünülmektedir. Çalışmada veri seti büyütülerek daha fazla örneklem kullanılması durumunda daha iyi bir model ve daha fazla doğruluk elde edileceği varsayılmaktadır. İleriye dönük yapılacak bir çalışma ile sağlık çalışanlarına çeşitli ses kayıt görevleri verilebilir. Elde edilen ses kayıtları, yanlışlık olmaması amacıyla bağımsız bir grup tarafından etiketlenilerek ses

verileri toplanabilir ve bu seslerden veri tabanı oluşturulabilir [16]. Bu çalışma sağlıkta şiddetin önlenmesi kapsamında, toplumda şiddetin engellenmesi ve müdahale edilmesine yönelik öncü bir çalışma olması amacıyla literatüre sunulmuştur.

Kaynaklar

1. Duru, H., Akgün, E.G., Maviş, İ.: Dil ve Konuşma Terapisi Mesleğine Yönelik Farkındalığın Belirlenmesi. *Dil Konuşma ve Yutma Araştırmaları Dergisi*, 1(3), 257-280 (2018).
2. Clark HM. Neuromuscular treatments for speech and swallowing: a tutorial. *Am J Speech Lang Pathol*. November 12(4):400-415 (2003). doi: 10.1044/1058-0360(2003/086). PMID: 14658992.
3. Vermeir, Peter, et al. Job Vermeir, P., Degroote, S., Vandijck, D., Mariman, A., Deveugele, M., Peleman, R., Verhaeghe, R., Cambré, B., & Vogelaers, D.: Satisfaction in relation to communication in health care among nurses: A narrative review and practical recommendations. *Sage Open* 7(2), 1-11 (2017).
4. Yeşilbaş, H.: Sağlıkta şiddete genel bakış. *Sağlık ve Hemşirelik Yönetimi Dergisi* 1(3), 44-54 (2016).
5. Öz, P. Çırak, M.: Sağlıkta Şiddetin Beyaz Kod Verileri ile Değerlendirilmesi. *Medical Journal of Bakirkoy* 15(4) :393-398 (2019).
6. Siegman, Aron Wolfe, Robert A. Anderson, and Tal Berger.: The angry voice: its effects on the experience of anger and cardiovascular reactivity. *Psychosomatic Medicine* 52(6), 631-643 (1990).
7. Kwon, O. W., Chan, K., Hao, J., Lee, T. W.: Emotion recognition by speech signals. In *Interspeech* pp. 125-128, San Diego, USA (2003).
8. Koolagudi, S. G., Rao, K. S.: Emotion recognition from speech: a review. *International journal of speech technology*, 15, 99-117(2012).
9. Zielonka, M.: Recognition of emotions in speech using convolutional neural networks on different datasets. *Electronics* 11(22) 3831(2022).
10. Öztürk, Ömer Faruk, and Elham Pashaie. Konuşmalardaki duygunun evrimsel LSTM modeli ile tespiti. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi* 12(4) 581-589 (2021).
11. Bingol, M.C., Aydogmus, O. Implementation of speech recognition for robot control using support vector machine. In *International Eurasian Conference on Science, Engineering and Technology Eurasian Sci En Tech*, pp. 814-819, Ankara, Türkiye (2018).
12. Schlüter, J., Grill, T.: Exploring data augmentation for improved singing voice detection with neural networks. In *ISMIR* pp. 121-126, Vienna, Austria (2015).
13. Atila, O., Şengür A.: Attention guided 3D CNN-LSTM model for accurate speech-based emotion recognition. *Applied Acoustics*, 182, pp. 108260 (2021).
14. Demirtaş S. C., Hakdağlı Ö.: CNN Modeli ile Türkçe Konuşma Verisi Üzerinde Duygu Tanıma Emotion Recognition on Turkish Speech Data with Transformer-CNN Eleco Elektrik-Elektronik ve Biyomedikal Mühendisliği Konferansı Bursa, 93, Bursa, Türkiye (2022).
15. Michael G. R., Kandali A.B.: Emotion Recognition from Manipuri Language Using MFCC and Convolution Neural Network. *International Journal for Research in Applied Science & Engineering Technology- IJRASET*, 10(9), pp. 861-865 (2022).
16. Bost, X., Labatut, V., Linares, G.: Serial speakers: a dataset of tv series. arXiv preprint . 12th International Conference on Language Resources and Evaluation (LREC 2020), pp. 4256-4264, Marseille, France (2020).

An Example Application For Machine Learning-Assisted Retrosynthesis

Güven ŞAHİN¹[0000-0003-3901-5819], Mehmet Erşan KALAYCI²[0000-0002-4797-9020], Kemal TURHAN²[0000-0001-7871-3025]

¹ Recep Tayyip Erdoğan University, Rize, Türkiye,

² Karadeniz Technical University, Trabzon, Türkiye

guven.sahin@erdogan.edu.tr

mehmetersankalayci@ktu.edu.tr

kemalturhan@ktu.edu.tr

Literature Research: Products with different ingredients are needed for the treatment of various diseases. One of the most commonly used methods to obtain these products is organic synthesis. Although organic synthesis has a history of 190 years, it is not fast enough to discover new drugs and materials. Expert knowledge is relied upon in the process of determining synthetic routes that will result in the target product. This situation causes to prolong the discovery process and increase the incurred expenses in this direction. To overcome this problem, the discovery of synthetic routes has become different with the developing machine learning technologies. One of the important points in the discovery of new molecules lies in well-designed and applicable retrosynthetic routes. The aim of retrosynthesis work is to design synthetic routes from the target molecule to the starting materials. When the literature was examined within the scope of the research, it was seen that machine learning-supported retrosynthesis studies were still in their early stages. It is thought that the study conducted in this direction will help chemists find better synthesis routes that will reach the target product more quickly.

Purpose of the study: The purpose of this study is to perform both synthesis and retrosynthesis applications of the determined target molecules.

Method: The study started with the identification of target molecules. The molecules were selected as atropine, lidocaine, loratadine, acetaminophen (Paracetamol) and acetylsalicylic acid (Aspirin) from the World Health Organization's list of essential drugs. Synthesis plans were obtained through literature review of the identified products. After each reactant and product in the synthesis plan was obtained from the PubChem platform, files with .sdf extension were downloaded. Using these files, the values of the products and reactants in SMARTS format were obtained with the PubChem Sketcher v2.4 tool. Synthesis formulas were obtained by combining the obtained formulas in SMARTS format. Using these SMARTS format synthesis formulas obtained with RDKit, an open source toolkit for chemoinformatics, a new molecule was tried to be obtained by adding different reactants to the target molecules. IBM's web-based RXN platform was used for the retrosynthesis process of target molecules. Reactants were tried to be predicted by giving the values of the products in SMILES format to the system.

Result and Evaluation: No new molecule was found in the synthesis study performed with RDKit. It has been observed that the retrosynthesis process using the IBM RXN platform can predict the reactants in the synthesis plan for atropine, lidocaine, loratadine, acetaminophen (Paracetamol) and acetylsalicylic acid (Aspirin). This study did not train a machine learning model, so there is no data set used or any metrics measuring model performance. Retrosynthesis of the identified target molecules was carried out with the web-based RXN tool. As a result, it has been shown that the tool can accurately predict the synthesis plans of target molecules. In future studies, it is planned to develop a machine learning-supported model that predicts retrosynthesis similar to this tool.

Keywords: Machine Learning, Synthesis Plan, Retrosynthesis

Pulmonary Nodule Decision Support Model Using Convolutional Neural Network-Based Transfer Learning Approach

Hilal TİRYAKİ¹[0000-0002-3303-5192], Burçin KURT²[0000-0001-5781-2382], Polat KOŞUCU³[0000-0002-8381-7600], Selçuk AKKAYA⁴[0000-0002-6680-1156]

¹ School of Medical Informatics, Karadeniz Technical University, Department of Computer Programming, Avrasya University
61080 Trabzon, TURKEY
hilaltiryaki@avrasya.edu.tr

² Faculty of Medicine, Basic Medical Sciences, Biostatistics and Medical Informatics, Karadeniz Technical University
61080 Trabzon, TURKEY
burcinkurt@ktu.edu.tr

³ Faculty of Medicine, Internal Medicine, Radiology, Karadeniz Technical University 61080 Trabzon, TURKEY
polatkosucu@ktu.edu.tr

³ Faculty of Medicine, Internal Medicine, Radiology, Karadeniz Technical University 61080 Trabzon, TURKEY
selcukakkaya@ktu.edu.tr

Abstract

In lung cancer, as with other types of cancer, early diagnosis is crucial for improving patient survival. This study aims to develop a decision support model using Convolutional Neural Network (CNN) to classify pulmonary nodules as benign or malignant for lung cancer diagnosis. A dataset of 1111 Thoracic CT (computed tomography) images from 600 patients, obtained from the Radiology Department of Karadeniz Technical University Farabi Hospital, was utilized. The developed decision support model automatically classified pulmonary nodules for lung cancer diagnosis using Thoracic CT images, achieving the highest accuracy of 84% through five-fold cross-validation. To enhance classification performance, the transfer learning approach commonly used in deep networks was employed. Subsequently, a CNN-based VGG16 model was trained using images obtained from the Radiology department of Farabi Hospital. To prevent overfitting, 5-fold cross-validation was performed on the model, and the highest classification performance was calculated as 70% accuracy, 80% sensitivity, and 64% specificity.

Keywords: Pulmonary Nodule, Convolutional Neural Network, Transfer Learning, Computed Tomography

Radyoloji Uzmanlık Eğitiminde Yapay Zekanın Yeri: Ulusal Anket Çalışması

Emre Emekli¹[0000-0001-5989-1897] Özlem Coşkun²[0000-0001-8716-1584] Işıl İrem Budakoğlu
2[0000-0003-1517-3169]

¹ Eskişehir Osmangazi Üniversitesi Tıp Fakültesi, Radyoloji Anabilim Dalı, Eskişehir, Türkiye
² Gazi Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı, Ankara, Türkiye

Abstract. The aim of this study was to investigate the knowledge levels and attitudes of radiology resident physicians in our country regarding artificial intelligence (AI). 80.6% of the participants stated that they were familiar with AI terminologies, however, it was observed that their knowledge levels were low. Majority (90.3%) did not possess programming skills, yet 84.5% of the participants supported the inclusion of AI education in medical residency programs. 84.5% of the participants also stated that no resources were provided for AI education during their residency, with only 1.9% receiving less than 1 hour and 4.5% receiving 1-5 hours of education. Moreover, the majority of participants (76.82%) were not aware of online AI resources. These findings indicate a low awareness level and a lack of education on AI among radiology resident physicians. Therefore, it is necessary to provide more comprehensive education on AI for radiology residents and increase their awareness.

Keywords: Artificial Intelligence, Radiology, Medical Education

Özet. Ülkemizdeki radyoloji asistan hekimlerinin yapay zeka (YZ) ile ilgili bilgi düzeyleri ve tutumları araştırılması amaçlanmıştır. Katılımcıların %80,6'sı YZ terimlerinden haberdar olduğunu belirtmiş, ancak bilgi düzeylerinin düşük olduğu gözlemlenmiştir. Çoğunluk (%90,3) programlama bilgisine sahip değilken, katılımcılar YZ eğitiminin tıpta uzmanlık programlarında yer alması gerekliliğini %84,5 oranında desteklemiştir. YZ ile ilgili 131 kişi (%84,5) uzmanlık eğitiminde herhangi bir kaynak sunulmadığını, eğitim aldığını ifade edenler 3 (%1,9) kişi 1 saatten az, 7 (%4,5) kişi 1-5 saat arası eğitim aldığını ifade etmiştir. Bununla birlikte, katılımcıların çoğunluğu (%76,82) online YZ kaynaklarından haberdar değildir. Bu bulgular, radyoloji asistanlarının YZ konusunda farkındalık düzeyinin düşük olduğunu ve eğitim eksikliği olduğunu göstermektedir. Bu nedenle, radyoloji asistanlarının YZ konusunda daha kapsamlı eğitim alması ve farkındalıklarının artırılması gerekmektedir.

Keywords: Yapay Zeka, Radyoloji, Tıp Eğitimi

1 Giriş

Yapay zeka (YZ) ve derin öğrenme (DÖ) algoritmaları, tıbbi görüntüleme yöntemlerine uygulandığında umut verici performans göstermektedir [1,2] YZ radyologlar için iş akışını daha verimli kullanmaları, daha hızlı ve tekrarlanabilir segmentasyon yapımları, lezyonları daha kolay saptamaları noktasında fırsatlar sunmaktadır [3]. Bu gelişmeler günümüzde daha çok kongreler ve literatür düzeyinde belirgin olarak hissedilmektedir. Klinik olarak ise henüz pratik işleyişte YZ araçlarının benimsenmesinin erken aşamalarında olduğu söylenebilir [4]. YZ alanındaki bu değişimlerin bazen literatür ve çoğunluklu ana akım medyada olumsuz yönlerinin ortaya koyulması ise tıp öğrencileri, stajyerler bazı durumlarda radyologlar arasında da, bilgisayar veya diğer disiplinler tarafından yerlerinden edilebilecekleri korkusunu geliştirebilmektedir [5]. Diğer bir taraftan ise YZ, DÖ ve makine öğrenmesinin (MÖ) her alanda olduğu gibi radyoloji alanında da bu şekilde aratarak etkisinin görülmesi ve entegrasyonun artmasıyla, radyoloji eğitiminde de YZ eğitiminin müfredatta yer almasına yönelik ilgi de artmaktadır. Bu nedenle ulusal ve uluslararası çeşitli kaynaklar eğitim için sunulmaktadır [6].

Bu çalışmanın amacı ülkemizdeki radyoloji asistan hekimlerinin YZ'ya yönelik mevcut bilgi ve genel tutumlarını araştırmaktır. İkinci olarak radyoloji tıpta uzmanlık eğitimindeki mevcut YZ / MÖ / DÖ eğitiminin mevcut durumunu tespit etmek, sunulan eğitim kaynaklarının asistanlar tarafından farkındalığını ve kullanımını değerlendirmek ve asistanların YZ eğitiminin genişliği ile tercih ettikleri öğrenme yöntemleri/araçları hakkındaki bakış açısını anlamak amaçlanmaktadır.

2 Yöntem

Gazi Üniversitesi Etik Komisyonundan onay alınmıştır. Çalışma Google Formlar aracılığı ile oluşturulan online anket formu ile kesitsel olarak gerçekleştirilmiştir. Anketin veri toplama aşaması Eylül 2023-Şubat 2024 tarihleri arasında gerçekleştirilmiştir. Anket formunda çalışma hakkında kısa bir bilgilendirme içeren, çalışmanın amacını açıklayan bir onam formu oluşturulmuştur. Anket demografik veriler, YZ bilgisi, yapay zekaya karşı tutum, tıpta uzmanlık eğitiminde yapay zekanın yeri olacak şekilde dört ana başlıktan oluşmaktadır.

Ankette YZ'nin radyoloji asistanlık eğitiminde yer alması ve YZ/MÖ/DÖ konularında katılımcıların bilgi düzeyleri 7'li Likert ölçeği şeklinde sorgulanmıştır (1=çok kötü / kesinlikle olmamalı, 5= çok iyi / kesinlikle olmalı). Anket Türk Radyoloji Derneği aracılığıyla üyelere e-posta olarak gönderilmiştir.

3 Bulgular

Çalışmaya katılmayı onaylayan 155 kişi çalışmaya dahil edilmiştir. Çalışmaya katılanların yaş ortalaması $28,81 \pm 4,77$ yıl, 94 (%60,6) 'i erkek, 60 (%38,7) 'i kadındır, bir kişi

(%0,6) cinsiyet belirtmemiştir. Çalışmaya katılan asistan hekimlerin tıpta uzmanlık seneleri ve kurumları Tablo 1'de ifade edilmiştir.

Tablo 1. Katılımcıların tıpta uzmanlık seneleri ve kurumları

Tıpta uzmanlık senesi	Sayı(n)	Yüzde (%)
1	43	27,7
2	49	31,6
3	25	16,1
4	18	11,6
5	20	12,9
Kurumu	Sayı(n)	Yüzde (%)
Üniversite	112	72,3
EAH	35	22,6
Afiliye	8	5,2
Vakıf	0	0

Katılımcılardan 125'i (%80,6) YZ/MÖ/DÖ terimlerinden haberdar olduğunu ifade etmiştir. Bu katılımcıların bilgi düzeylerini 7'li Likert ile değerlendirmesi istendiğinde $3,02 \pm 1,39$ ortalama skor elde edilmiştir. Katılımcıların 140 'ı (%90,3) programlama bilgisine sahip olmadığını, 15 kişi (%9,7) temel düzeyde bilgi sahibi olduğunu belirtmiştir. Asistan hekimlerin 120'si (%77,4) YZ/MÖ/DÖ araçlarını hiç kullanmadığını, 7'si (%4,5) birkaç kez, 2'si (%1,3) günlük, 7'si (%4,5) haftalık, 19'u (%12,3) ise aylık kullandığını ifade etmiştir.

YZ'nin radyolojinin geleceğini değiştirip değiştirmeyeceği sorusuna 114 (%73,4) katılımcı evet, 5 (%3,2) katılımcı hayır, 36 (%23,2) katılımcı belki cevabını vermiştir. Önceki soruya evet / belki cevabını veren katılımcılardan 12 kişi (%7,7) beş yıldan az, 78 kişi (%50,3) 5-10 yıl, 43 kişi (%27,7) 10-20 yıl ve 17 kişi (%11) 20 yıldan fazla bu değişimin gerçekleşeceğini düşünmektedir. Radyologların işinin tehlikede olduğunu düşünüp düşünmedikleri katılımcılara sorulduğunda 16 (%10,3) evet, 94 (%60,6) hayır, 45 (%29) belki cevabı alınmıştır. Asistan hekimler neden radyologların işini tehlike de görüp görmediğine ilişkin veriler Tablo 2 de belirtilmiştir.

Tablo 2. Katılımcıların radyolojinin geleceğine YZ etkisi hakkında görüşleri

Neden Tehlikede (n=87)	Sayı (n)	Yüzde (%)	Neden Tehlikede değil (n=120)	Sayı (n)	Yüzde (%)
Yapay zekâ ile birlikte ihtiyaç duyulan radyolog sayısı azalacaktır.	23	26,44	Radyologların rolü yapay zekâ ile değişmez hatta işleri kolaylaştırabilir.	60	50
Yapay zekâ araçları, görüntü yorumlama / karşılaştırma gibi görevleri daha verimli bir şekilde yerine getirecek, radyologlara duyulan ihtiyacı azalacaktır.	40	45,98	Radyologların rolü değişebilir, ancak yapay zekâ onların yerini alamaz.	60	50
Yapay zekanın maliyet etkinliği, bunu kurumlar için cazip hale getirecek ve radyologların yerini alacaktır.	24	27,59			

Kurumlarda YZ ile ilgili öğretim üyesi varlığı ve herhangi bir araştırmada yer alma durumu Tablo 3'te özetlenmiştir. YZ'nın radyoloji eğitiminin bir parçası olmalı mı sorusuna (7'li Likert) verilen cevapların ortalaması $5,09 \pm 1,72$ 'ydi.

Tablo 3. YZ ile ilgili akademik görev alma durumu ve YZ alt disiplin olması hakkında görüşler

	Evet n (%)	Hayır n (%)	Bilmiyorum / Emin değilim n (%)
Bölümünüzde radyolojide YZ ile ilgilenen öğretim görevlileri var mı?	51 (32,9)	48 (31)	56 (36,1)
YZ/MÖ/DÖ ile ilgili herhangi bir araştırmada yer aldınız mı?	19 (12,3)	136 (87,7)	
Görüntüleme bilişimi / yapay zekanın radyolojide bir alt uzmanlık alanı olması gerektiğini düşünüyor musunuz?	57 (36,8)	33 (21,3)	65 (41,9)

YZ ile ilgili 131 kişi (%84,5) herhangi bir kaynak sunulmadığını, 3 kişi (%1,9) amfi dersi, 4 (%2,6) kişi online materyal, 1 (%0,6) kişi ulusal düzeyde ders, 7 (%4,5) kişi ulusal kongre / sempozyum, 2 kişi (%1,3) kurum eğitimi aldığını ifade etmiştir. Bu eğitimi süresini 3 (%1,9) kişi 1 saatten az, 7 (%4,5) kişi 1-5 saat arası olarak ifade etmiştir. 142 (%91,6) katılımcı kurumları dışında eğitim almadığını ifade etmiştir. ESR AI blog, ESR "Masterclass in AI" dersleri, RSNA tarafından radyolojide yapay zeka videoları, Radiology: Artificial intelligence Journal, "ACR-Resident and Fellow" grubu tarafından yönetilen "AI Journal club", ACR Data Bilim Enstitüsü /AI Lab seçeneklerinden hangilerinden haberdar olduğu katılımcılara sorulduğunda 119 (%76,8) katılımcı hiçbirinden haberdar olmadığını belirtmiştir.

4 Tartışma ve Sonuç

Katılımcıların YZ bilgi ve tutumları değerlendirildiğinde %80,6'sı YZ terimlerinden haberdar olduğunu ifade etmiştir. Literatürde yapılan çalışmalarda da benzer oranlar saptanmıştır [7]. Bilgi düzeyleri sorulduğunda ortalama 3,02/7 olarak hesaplanmıştır. Literatürde ise bilgi düzeylerinin radyologlarda temel ve orta düzeyde olduğu ifade edilmektedir [8]. Programlama bilgisine hiç sahip olmayanlar %90,3 olarak belirlenmiştir, Salastekar, ve ark yaptığı çalışmada bu oran %75 olarak bulunmuştur. Literatürde YZ korkusunun yapılan bir çalışmada radyologlar arasında % 39 olduğu ifade edilmektedir. Aynı çalışmada YZ bilgisi az olanlarda bu korkunun daha fazla olduğu ifade edilmiştir [8]. Bu çalışmada ise katılımcıların %60,6'ı YZ nedeniyle radyologların işinin tehlikede olmadığını, %29'u belki tehlikede olduğunu ifade etmiştir.

Bu çalışmada katılımcılar YZ eğitiminin uzmanlık eğitiminde yer alması gerekliliğini 5,09/7 olarak belirlemişlerdir. Amerika'da (ABD) yapılan çalışmalarda Salastekar katılımcıların %83'ünün Huisman ise %79'unun radyoloji asistanlık eğitiminde YZ müfredatının olması gerektiğini düşündüğünü ifade etmekteydiler [9,10]. Singapur (%84.8) ve İspanya (%92.9) da yapılan çalışmalarda da benzer oranlar bildirilmiştir [11,12]. ABD'de yapılan çalışmalarda %25 ile %50 arasında değişen oranlarda YZ müfredatının tıpta uzmanlık müfredatında yer aldığı ifade edilmiştir [11,13]. Bu çalışmada %84,5 oranında katılımcı YZ eğitimi için herhangi bir kaynak sunulmadığını ifade etmektedir. Ülkemizde literatüre göre eğitim verilme oranı çok düşük olarak saptanmıştır. Bu çalışmada katılımcıların %76,82'si online YZ kaynaklarından haberdar olmadığını belirtmiştir. Literatürde de benzer şekilde online kaynaklar açısından radyoloji asistanlarının farkındalığının az olduğu ifade edilmektedir [9].

Sonuç olarak literatür ile kıyaslandığında bu çalışma popülasyonunda da benzer şekilde katılımcıların YZ'dan haberdar oldukları, ancak bilgi düzeylerinin ve temel bilgilerinin az olduğu saptanmıştır. Bu bulgular literatür verileri ile örtüşmektedir. Ancak tıpta uzmanlık eğitimi değerlendirildiğinde çoğu kurumda YZ ile ilgili eğitim verilmediği, birçok katılımcının da online kaynaklar hakkında bilgi sahibi olmadığını saptanmıştır.

Kaynaklar

1. Liu, X., Faes, L., Kale, AU. et al: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health* 1:e271–e297 (2019).
2. Jha, S., Topol, EJ.: Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA* 316:2353–2354 (2016).
3. Wichmann, J., Willeminck, M., De Cecco, C.: Artificial intelligence and machine learning in radiology: current state and considerations for routine clinical implementation. *Investigative Radiology* 55:619–627 (2020).
4. Lincoln, C., Chatterjee, R., Willis, M.: Augmented radiology: looking over the horizon. *Radiology: Artificial Intelligence* 1:e180039 (2019).

5. Collado-Mesa, F., Alvarez, E., Arheart, K.: The role of artificial intelligence in diagnostic radiology: a survey at a single radiology residency training program. *Journal of the American College of Radiology* 15:1753–1757 (2018).
6. Gorospe-Sarasua L., Muñoz-Olmedo, J.,M., Sendra-Portero, F., et al.: Challenges of radiology education in the era of artificial intelligence. *Radiologia* 64(1):54–59 (2022).
7. Gray, K., Slavotinek, J., Dimaguila, G.L., Choo, D., Artificial Intelligence Education for the Health Workforce: Expert Survey of Approaches and Needs. *JMIR Medical Education* 4;8(2):e35223 (2022).
8. Salastekar, N.V., Maxfield, C., et al.: Artificial Intelligence/Machine Learning Education in Radiology: Multi-institutional Survey of Radiology Residents in the United States. *Academic Radiology* 30(7):1481-1487 (2023)
9. Huisman, M., Ranschaert, E., Parker, W., et al.: An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. *European Radiology* 31(9):7058-7066 (2021).
10. Gallix, B., Chong, J.: Artificial intelligence in radiology: who's afraid of the big bad wolf? *European Radiology* 29:1637–1639 (2019).
11. Gong, B., Nugent, J.,P., Guest, W., et al.: Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: a national survey study. *Academic Radiology* 26:566–577 (2019).
12. Waymel, Q., Badr, S., Demondion, X., et al.: Impact of the rise of artificial intelligence in radiology: what do radiologists think? *Diagnostic and Interventional Imaging* 100:327–336 (2019).
13. Pinto Dos Santos, D., Giese, D., Brodehl, S. et al.: Medical students' attitude towards artificial intelligence: a multicentre survey. *European Radiology* 29:1640–1646 (2019).

Tümör Tedavi Alanları (Ttfields) Terapisinin Glioblastomada (Gbm) Hayatta Kalma ile İlişkisi: Sistemik Bir İnceleme ve Meta-Analiz

Mehmet KIVRAK¹, Hamit YILMAZ²

¹ Recep Tayyip Erdoğan Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi

Orcid id: 0000-0002-2405-8552

² Recep Tayyip Erdoğan Üniversitesi Tıp Fakültesi, Biyofizik

Orcid id: 0000-0002-8324-1891

Giriş: Tumor-treating fields (TTFields) terapi yöntemi; kanser hücreleri üzerinde seçici sitotoksik etkiye dayanan yeni bir anti-kanser modalitedir. TTFields terapisinde elektrik alanının bölünebilen hücrelere etkisi esas olarak mitotik süreçte kritik rol oynayan yüklü makro molekül ve organellerde meydana getirdiği biyofiziksel etkiyle açıklanmaktadır. Bu hücresel elemanların disfonksiyonu mitotik süreci baskılayabilir. Dipol hizalanma ve dielektroforez bu etkileşimin altında yatan temel iki fiziksel mekanizmadır. Mitoz ve sitokinezden sorumlu makromoleküllerin polaritesi yüksek olduğundan lokalize bir elektrik alan uygulamasından kolaylıkla etkilenir ve bölünme sürecinde bozulmalar meydana gelebilir. Glioblastoma multiform (GBM), merkezi sinir sisteminde ortaya çıkan son derece agresif, invaziv ve yüksek insidanslı (~3,2- 4,5/100.00) vasküler tümörlerin oluşturduğu heterojen bir hastalık grubudur. TTFields, randomize kontrollü EF-14 çalışmasına dayanarak 2015 yılında FDA onayı alınıp GBM tedavi protokollerinde yer almaya başlamıştır. Ancak geçen süreye rağmen klinisyenler arasında, ortalama sağ kalıma etkisi konusunda, hala tam bir fikir birliği oluşmamıştır.

Metot: Çalışmamızda 2010'dan 2023'e kadar sistemik bir literatür taraması yapılmış ve TTFields ile GBM arasındaki sağkalım oranlarını inceleyen çalışmalar dahil edilmiştir. Literatür analizinde "glioblastoma", "TTFields", "survival", "hazard" ve "etki" gibi anahtar kelimeler seçilmiştir. Daha sonra PubMed (MEDLINE) ve Web of Science veritabanı arama motorları seçilmiştir. Boole operatörlerini kullanan birleştirilmiş anahtar kelimeler de ise: ("Glioblastoma" VEYA "GBM") VE ("TTFields" VEYA "Tumor Therapy Field") VE ("sağkalım" VEYA "hazard") VE ("etki" VEYA "etkin") belirlenmiştir. Çalışmamızda dahil edilme kriteri olarak: orijinal araştırma olması, eşik altı GBM semptomlarının operasyonel

tanımını kullanılması, hastaların TTFields uygulaması örnekleri ile sınırlandırılması (>18 yaş), fonksiyonel sonuçlar değerlendirilmesi ve ingilizce makaleler taranması belirlenmiştir. Hariç tutma kriteri olarak: vaka çalışması ve derlemeler, çalışma ingilizce olarak mevcut değilse veya PubMed ve WOS aracılığıyla mevcut değilse belirlenmiştir. Etki büyüklüğünü değerlendirmek için bir meta-analiz gerçekleştirilmiştir. Bir orman planı oluşturulup, yayın yanlılığını incelenip ve homojenlik değerlendirilmiştir. GBM ile TTFields arasındaki sağkalıma ait ilişkiyi araştırılmıştır.

Bulgular: Oluşturulan PRISMA şemasına göre literatür taramasında ilgili olduğu düşünülen 1160 makale belirlenmiş ve detaylı analiz sonucunda 39 makaleye indirgenmiştir. Bu 39 makaleden dahil etme ve hariç tutma kriterlerimizin tamamını karşılayan yalnızca 8 nitelikli çalışma değerlendirilmiştir. Kesişme değeri 0,331 (pozitif) olup, çalışmalar arasında orta derecede heterojenlik olduğunu göstermektedir. Bu, çalışmaların sonuçlarının birbirinden farklılık gösterdiği ve bu farklılıkların bir kısmının muhtemelen rastgele varyasyondan kaynaklandığı anlamına gelmektedir. Standart hata (0,083), etki büyüklüğü tahmininin çalışmalar arasında daha kesin ve güvenilir olduğunu göstermektedir. Rastgele etkiler modeline dayalı tahmini ortalama log olasılık oranı 0,331'dir (%95 GA: 0,17 ila 0,49). Yani glioblastoma hastalarında TTFields uygulaması sağkalım üzerinde artırıcı bir etkisi görülmektedir. Non-TTFields'ye göre (log olasılık oranı: 0.139) ancak yüksek bir düzey gözlemlenmedi ($z = 3,99$, $p = <0,001$). Kendall's Tau değeri 0,051 ve p değeri $<0,001$ olup, TTFields'nin glioblastoma riskini baskılama potansiyeline sahip olduğunu ancak kesin bir sonuca varmak için daha fazla araştırmaya ihtiyaç olduğunu göstermektedir.

Sonuç: Meta analiz çalışmamızda yer alan klinik karşılaştırılmalı TTFields çalışmalarına dayanarak Konvansiyonel GBM tedavisine (maksimum rezeksiyon, eşzamanlı adjuvan temozolomid ve radyoterapi) dahil edilen TTFields modalitesinin ortalama sağ kalıma pozitif etkisinin olabileceği söylenebilir. Bunun yanında TTFields'nin etkinliğini araştıran çalışmalar arttıkça, bu tedavinin altın standart olarak kabul gören modalitelerle kıyaslanmasının kolay yapılması ve klinik kullanım yaygınlığının artması beklenmektedir.

Anahtar Kelimeler: Glioblastoma, Tümör Tedavi Alanları, Meta-Analiz

Kaynaklar

- 1- Urbanska K, Sokolowska J, Szmidt M, Sysa P. Glioblastoma multiforme - An overview. *Wspolczesna Onkologia*. 2014;18(5):307–12.
- 2- Brodbelt A, Greenberg D, Winters T, Williams M, Vernon S, Collins VP. Glioblastoma in England: 2007-2011. *European Journal of Cancer*. 2015;51(4):533–42.
- 3- Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *The Lancet Oncology* [Internet]. 2009;10(5):459–66. Available from: [http://dx.doi.org/10.1016/S1470-2045\(09\)70025-7](http://dx.doi.org/10.1016/S1470-2045(09)70025-7).
- 4- Gilbert MR, Wang M, Aldape KD, Stupp R, Hegi ME, Jaeckle KA, et al. Dose-dense temozolomide for newly diagnosed glioblastoma: A randomized phase III clinical trial. *Journal of Clinical Oncology*. 2013;31(32):4085–91.
- 5- Campos B, Olsen LR, Urup T, Poulsen HS. A comprehensive profile of recurrent glioblastoma. *Oncogene*. 2016;35(45):5819–25.
- 6- Wang Y, Pandey M, Ballo MT. Integration of Tumor-Treating Fields into the Multidisciplinary Management of Patients with Solid Malignancies. *The Oncologist*. 2019;24(12):e1426–36.
- 7- Kirson ED, Schneiderman RS, Dbal V, Tovary F, Vymazal J, Itzhaki A, et al. Chemotherapeutic treatment efficacy and sensitivity are increased by adjuvant alternating electric fields (TTFields). *BMC Medical Physics*. 2009;9(1):1–13.
- 8- Pless M, Weinberg U. Tumor treating fields: Concept, evidence and future. *Expert Opinion on Investigational Drugs*. 2011;20(8):1099–106.
- 9- Davies AM, Weinberg U, Palti Y. Tumor treating fields: A new frontier in cancer therapy. *Annals of the New York Academy of Sciences*. 2013;1291(1):86–95.
- 10- Stupp R, Taillibert S, Kanner A, Read W, Steinberg DM, Lhermitte B, et al. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma a randomized clinical trial. *JAMA - Journal of the American Medical Association*. 2017;318(23):2306–16.
- 11- Vymazal J, Kazda T, Novak T, Slanina P, Sroubek J, Klener J, et al. Eighteen years' experience with tumor treating fields in the treatment of newly diagnosed glioblastoma. *Frontiers in Oncology*. 2023;12(January):1–9.

Analysis of Findings in Radiology Reports Using Non-Negative Matrix Factorization

Emir KARAYAĞIZ¹[0009-0005-1673-621X] and Tolga BERBER¹[0000-0002-6487-5581]

¹ Department of Computer Sciences, Karadeniz Technical University,
Trabzon 61080, TURKEY
emirkyzmain@gmail.com
tolga.berber@fen.ktu.edu.tr

Abstract. Medical progress has allowed for the collection of large quantities of data, which can be utilized to discover connections between individual patient results and offer useful insights for analyzing and interpreting future cases. Topic modelling methods can be utilized to classify text data based on the words it includes. An issue when dealing with text data is its huge magnitude, which complicates performing calculations and generating useful insights. This study employs text vectorization techniques and Non-Negative Matrix Factorization, an algorithm for matrix decomposition, to analyze the finding section of radiology reports. The goal is to group words that frequently appear together in reports into clusters and analyze findings to assign them to the appropriate cluster. The primary problem in topic modelling is ascertaining the optimal number of clusters to extract. Word cluster comparisons are conducted by assessing the coherence and relevance of the terms within each cluster. Only one data set is utilized for this research. The dataset is examined with various cluster sizes, and their coherence scores are computed. Optimally selecting the number of clusters leads to a more significant distinction between words for analysis.

Keywords: Topic Modelling, Text Vectorization, Non-Negative Matrix Factorization, Coherence Score.

1 Introduction

Medical advances have enabled the generation of vast amounts of data, which can be used to identify correlations between individual findings of patients and provide valuable information for analysis and interpreting future patients. By using topic modeling methods, it is possible to categorize text data according to the words it contains. One of the problems working with text data comes from the fact that text data is on a large scale, making it hard to do calculations and create a meaningful analysis. The text vectorization methods, which are data pre-processing methods, and, for ease of interpretation of the vectorized data, matrix decomposition algorithms are used to eliminate this problem. Text vectorization is the preparation of text data for numerical calculations. By using text vectorization, non-calculable text data is converted to matrix form.

Bag-of-Words (BoW) (1) and Term frequency-inverse document frequency (TF-IDF) (2) are methods used to convert text data to numerical data. The output of these methods is in matrix form, which contains numerical data.

Non-Negative Matrix Factorization (NMF) (3) algorithm is widely used in matrix decompositions.

In this study, TF-IDF method, which holds more information compared to BoW method, was applied to radiology findings data. For the classification of the matrix data, the NMF algorithm is used to create meaningful and easily interpretable outcomes.

2 Literature Review

Topic modelling is widely used in different fields of medicine. For example, Zhao, Juan, et al (4) used the NMF algorithm for identifying associations between disease phenotypes and genetic variants. Different variants of NMF were developed by other researchers to meet their needs. Shah, Adnan Muhammad, et al (4) applied a customized text mining technique, NMF topic modeling to identify the major topics on two platforms. More recent study from Grotheer, Rachel, et al (6) made use of hierarchical nonnegative matrix factorization to organize literature related to the novel coronavirus into a tree structure that allows researchers to search for relevant literature based on detected topics.

3 Material and Methods

3.1 Dataset

Radiology reports in Indiana University chest x-ray collections available from the National Library of Medicine's Open-I service dataset (7) is used in this study. The dataset contains 3,425 radiology reports collected from various hospital systems. For our analysis, we used the "findings" field as our text data. After creating dictionary from the data, we got 1428 unique words found. Some examples from the dataset can be seen below.

Table 1. Samples from Dataset

<i>Sample Nr</i>	<i>Report Sample</i>
1	The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.
2	The trachea is midline. The cardiomeastinal silhouette is normal. The lungs are clear, without evidence of focal infiltrate or effusion. There is no pneumothorax. The visualized bony structures reveal no acute abnormalities.

<i>Sample Nr</i>	<i>Report Sample</i>
3	Normal heart and mediastinum. Clear lungs. Trachea is midline. No pneumothorax. No pleural effusion. Radiopaque foreign body overlying left chest.
4	The lungs are clear. There is hyperinflation of the lungs. There is no pleural effusion or pneumothorax. The heart and mediastinum are normal. The skeletal structures are normal.
5	Calcified granulomata are present in the right middle lobe and right upper lobe. These are unchanged since the previous examination. Both lungs are free of active infiltrates. Heart size normal.

3.2 Preprocessing

Before trying to convert our text data into numerical matrix form, we need to eliminate any data that may affect our analysis. This includes words like “_” or “XXXX” which are common replacements for patient names or any other information that is not suitable to be shared publicly. We need to remove these to get better results. Additionally, we need to convert all letters to lower case to prevent multiple copies of same word, so in a way, we normalize the data. After converting the letters, we remove all symbols using Unicode symbol categories. In radiology reports, most of the times symbols do not affect the meaning of the finding, so it is removed to not affect the results.

3.3 Text Vectorization

Topic modelling is made of two main methods. Text vectorization and matrix decomposition. Some of the most known vectorization methods will be discussed in this section.

Bag-of-Words: The bag of words method vectorizes texts using each words’ frequency values. Firstly, all texts are split into vectors where, each value consists of one word.

While texts are being split, at the same time Stop Words, words that have no meaning of their own, might be removed depending on the data being worked on. These include words like “to”, “for” and “be”. After word splitting, a dictionary is made that contains of each word once. Then, for each sentence, each word’s frequency value is calculated after that these values are placed in their respective places according to word dictionary.

This process gets repeated for each sentence, and results in creating one vector for each sentence. Only after these vectors combined vertically, a matrix is formed. This matrix’s row count is how many documents or sentences you have, and the column count is how many unique words you have in your word dictionary.

It can be seen that no interpretation can be made about whether a word is important for a sentence to identify or if it is just a commonly used word. Because of this reason other methods, which have different steps added in calculations, were proposed through the years.

TF-IDF: The TF-IDF method in its core works like BoW method, but TF-IDF holds more knowledge about words and their impact for a sentence. It is modelled as follows by calculating two separate values and multiplying them.

$$TF(x, s) = \frac{\text{number of times } x \text{ appears in } s}{\text{total number of terms in } s} \quad (1)$$

$$IDF(x) = \log \left(\frac{\text{total count of documents}}{1 + \text{count of documents that contains } x} \right) \quad (2)$$

$$TFIDF(x, s) = TF * IDF \quad (3)$$

Values calculated this way have the effect of highlighting words that are distinct in each document.

After applications of these formulas, it can be found that words that are most common have TF-IDF values close to zero. This means that they hold low value for their respective sentences. In some cases, values can be zero. Because of the matrix calculations that will be done in the next section, we add value one to these values, this keeps the information that they have the least value among any other words. In this study, TF-IDF method is used to get better understanding of the dataset.

3.4 Matrix Factorization

Matrix factorization is basically the process of factorizing a matrix. Some of the methods used for this are Principal Component Analysis, Singular Value Decomposition and Non-Negative Matrix Factorization. In this study, NMF is used because of its ability to perform better on non-negative data, which includes text, and data that contains small length sentences. NMF was first suggested in 1990s by Shen and Israel in 1989 (8), followed by Paatero and Tapper in 1994 (9). NMF gained its popularity through research that made it easier to calculate by Lee & Seung in 1999 (3). With that paper, it was observed that NMF was very effective at factorizing matrices using iterative calculation rules. These are some of the reasons why NMF was chosen for his study. NMF can have different calculation rules depending on factorization level.

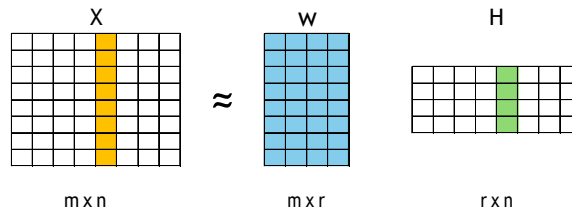


Figure 1. Visualization of NMF for factorization into two matrices

In the Figure 1 basic version of NMF can be seen. In this version r value represents how many clusters or topics we want at the end of our analysis. For this reason, it is crucial to determine optimal “ r ” value before starting calculations.

$$r < \frac{m*n}{m+n} \quad (4)$$

$$r < \frac{\text{count of non-zero values of } X}{m+n} \quad (5)$$

In equation 4 the general upper limit for r value can be calculated. For sparse matrices, equation 5 is the more suitable option to calculate upper limit for topic count. Sparse matrices mostly made of value 0. Having a high amount of 0 values leads us to reduce dimensions or factorize the matrix into smaller and more informative one. Because of the way we vectorized our texts, we end up with a big sparse matrix as well. For this using latter method for choosing upper limit is used in this study.

The “W” matrix’s row number is the total and number of documents and the columns are the number of topics. Each value in matrix represents the score for given document and the most possible topic it may be in. The maximum of this values can give the information about which topic does a document belongs to.

The “H” matrix’s row number is the total amount of topics, and the columns are the total number of words that exists in our word dictionary. Each value in this matrix represents the score for given word and its most possible topic. These values are sorted and desired number of words for each topic can be extracted and analysis can be made for those topics.

The iterative update rules of NMF are made of two calculations each iteration. First both W and H matrices can be initialized using uniform distribution. After that, iterative calculative equations can be seen below.

$$\tilde{W}^{t+1} = \tilde{W}^t \otimes (X\tilde{H}'^t \oslash \tilde{W}^t\tilde{H}^t\tilde{H}'^t) \quad (6)$$

$$\tilde{H}^{t+1} = \tilde{H}^t \otimes (\tilde{W}^{t+1}X \oslash \tilde{W}'^{t+1}\tilde{W}^{t+1}\tilde{H}^t) \quad (7)$$

These calculations can be repeated until mean square error between starting matrix X and the result matrix \tilde{X} reaches a small threshold or number of iterations done by program reaches a certain number.

To further increase the liability of iterative calculations and eliminate randomness between different tests, an algorithm called Non-Negative Single Value Decomposition (NNDSVD) algorithm (10) can be implemented to initialize W and H matrices. Pseudo code implementation for this algorithm can be seen below. This algorithm eliminates the random initialization at the start and provides a more suitable starting point according to our X matrix. The biggest benefit of using NNDSVD is that it enables us to re-analyze data in the same initialization condition for different methods.

Algorithm 1. NNDSVD initialization of non-negative matrix

Input: $X \in \mathbb{R}^{m \times n}$, $r < \min(m, n) \in \mathbb{N}^+$

Output: $\tilde{W}^0 \in \mathbb{R}^{m \times r}_+$, $\tilde{H}^0 \in \mathbb{R}^{r \times n}_+$

1 Compute the largest r singular triplets of X : $U, S, V = tda(X, r)$

2 Initialize $\tilde{W}_{:,1} = \sqrt{s_{1,1}} \times U_{:,1}$ and $\tilde{H}_{1,:} = \sqrt{s_{1,1}} \times V'_{:,1}$

3 **for** $\forall j \in [2, r]$ **do**

3.1 Define $x = U_{:,j}$ and $y = V'_{:,j}$

$$x_+ = \begin{cases} x_i, & x_i > 0 \\ 0, & \text{else} \end{cases}, \quad x_- = \begin{cases} -x_i, & x_i < 0 \\ 0, & \text{else} \end{cases}, \quad y_+ = \begin{cases} y_i, & y_i > 0 \\ 0, & \text{else} \end{cases} \quad \text{ve} \quad y_- = \begin{cases} -y_i, & y_i < 0 \\ 0, & \text{else} \end{cases}$$

3.2 $\mu_+ = \|x_+\| \|y_+\|$ ve $\mu_- = \|x_-\| \|y_-\|$

3.3 $\mu_+ > \mu_-$ ise $u = \frac{x_+}{\|x_+\|}, \mu_+ > \mu_-$ ise $v = \frac{y_+}{\|y_+\|}$ ve $\sigma = \max(\mu_+, \mu_-)$

3.4 $\mu_- > \mu_+$ ise $u = \frac{x_-}{\|x_-\|}, \mu_- > \mu_+$ ise $v = \frac{y_-}{\|y_-\|}$ ve $\sigma = \max(\mu_+, \mu_-)$

3.5 $\tilde{W}_{:,j} = \sqrt{s_{j,j}} \times \sigma \times u$ ve $\tilde{H}_{j,:} = \sqrt{s_{j,j}} \times \sigma \times v'$

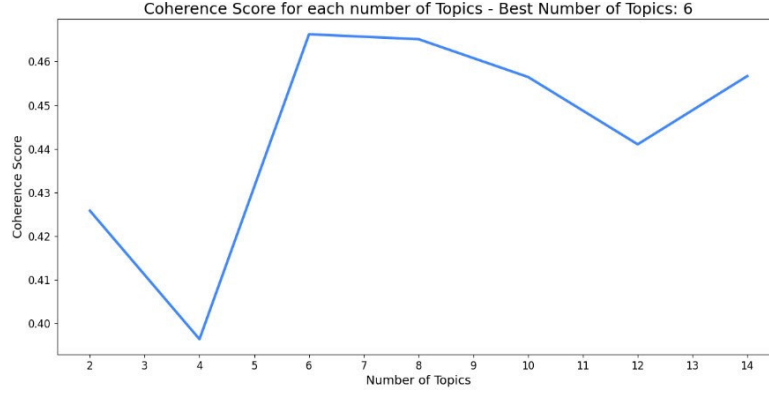
4 Results

When TF-IDF is applied to “findings” field in radiology reports, we end up with matrix which has 3425 rows and 1428 columns. This shows us that we have 1428 unique words in our word dictionary. After vectorization algorithm has ended, the matrix’s 98.60021% consisted of value 0. When calculated, only 68.462 values of matrix consisted data different than 0 out of total $3425 \times 1428 = 4890900$ values. When we applied the optimum topic clusters rule from equation 5, we ended up result below:

$$r = \frac{68462}{4853} = 14.1071 \quad (8)$$

We ended up with a value of 14.1071. This result means that our maximum number of topics should not exceed 14. After that, Cv coherence score calculation algorithm from Gensim (11), a python framework, is applied to our text data. Figure 2 includes coherence scores of topic numbers between 2 and 14.

Figure 2. Cv Coherence Scores for each value with 14 as maximum



It is apparent from the graph that 6 is the best number of topics for our dataset with coherence value of 0.46627. After that, the NMF algorithm is applied with the decided number of topics. The words with highest values for each topic cluster can be seen in Table 2. The table also includes topic titles generated by ChatGPT. Also, Figure 3 includes word clouds generated using topic keyword weights.

Topic Nr	Keywords	Topic Title
Topic #1	<i>spine, change, degenerative, thoracic, mild, aorta, disease, tortuous, airspace, atherosclerotic, focal, pulmonary, stable, limit, area</i>	<i>Thoracic Spine and Vascular Conditions</i>
Topic #2	<i>abnormality, without, evidence, structure, osseous, acute, visualized, thorax, bilaterally, specifically, unremarkable, cardio, silhouette, bony, consolidation</i>	<i>Radiological Assessment of Thoracic Structures</i>
Topic #3	<i>free, visible, mass, nodule, air, intraperitoneal, diaphragm, fluid, infiltrates, appear, grossly, vascularity, within, mediastinal, limit</i>	<i>Intraperitoneal and Mediastinal Abnormalities on Imaging</i>
Topic #4	<i>lateral, chest, frontal, view, examination, radiograph, show, consists, configuration, obtained, abdomen, airway, cardiomeastinal, cardiac, silhouette</i>	<i>Radiographic Imaging Techniques and Anatomy Evaluation</i>
Topic #5	<i>left, right, lobe, opacity, atelectasis, lower, base, upper, stable, calcified, scarring, small, granuloma, bilateral, tip</i>	<i>Lobar and Bilateral Opacities: Patterns and Characteristics</i>
Topic #6	<i>finding, alveolar, definite, typical, edema, seen, pulmonary, hilar, previous, process, consolidation, stable, granulomatous, mediastinal, limit</i>	<i>Radiological Assessment of Pulmonary Edema and Granulomatous Processes</i>

Table 2. First 15 words for each Topic and Their Titles Generated by ChatGPT

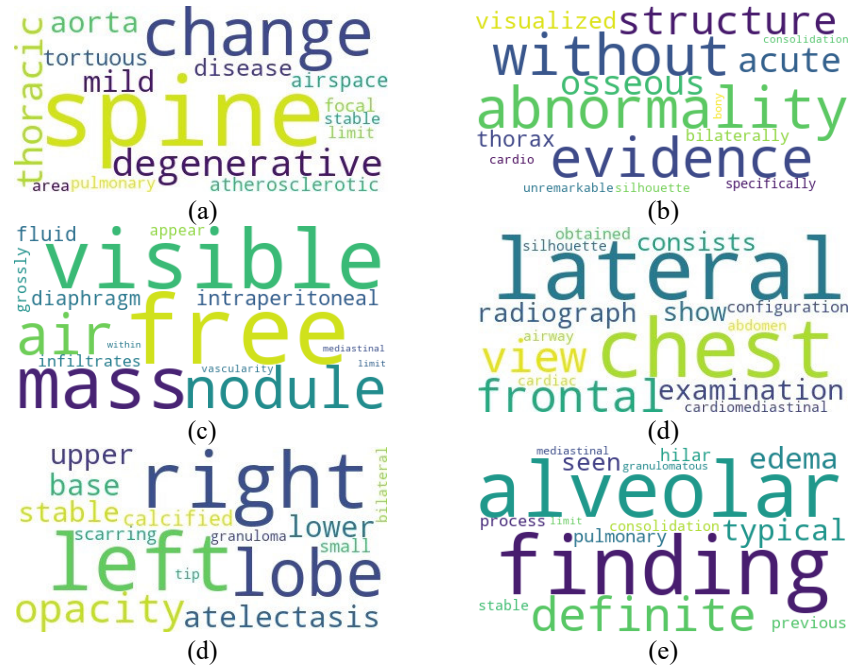


Figure 3. Word Clouds for Each Topic 1-6 given in (a)-(e) respectively

5 Conclusion

According to the results, it can be interpreted that our findings are mostly about general structure of chest area and Thoracic related abnormalities. The likelihood of the topics is caused by not having large dataset. This analysis can be vastly improved by gathering extra reports about different parts of the body. By providing a wide range of reports, the assessment of the analyze can be improved and can be used to categorize new reports right after they are recorded thus helping medical personnel.

6 References

1. Qader, Wisam & M. Ameen, Musa & Ahmed, Bilal. (2019). An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges. 200-204. 10.1109/IEC47844.2019.8950616.
2. (2011). TF-IDF. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_832
3. Lee, D., Seung, H. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999).

4. Zhao, Juan, et al. "Using topic modeling via non-negative matrix factorization to identify relationships between genetic variants and disease phenotypes: A case study of Lipoprotein (a)(LPA)." *PloS one* 14.2 (2019): e0212112.
5. Shah, Adnan Muhammad, et al. "Use of sentiment mining and online NMF for topic modeling through the analysis of patients online unstructured comments." *Smart Health: International Conference, ICSH 2018, Wuhan, China, July 1–3, 2018, Proceedings* 6. Springer International Publishing, 2018.
6. Grotheer, Rachel, et al. "COVID-19 literature topic-based search via hierarchical NMF." *arXiv preprint arXiv:2009.09074* (2020).
7. Demner-Fushman, Dina, et al. "Design and development of a multimodal biomedical information retrieval system." *Journal of Computing Science and Engineering* 6.2 (2012): 168-177.
8. Shen, J., and G. W. Israël. "A receptor model using a specific non-negative transformation technique for ambient aerosol." *Atmospheric Environment* (1967) 23.10 (1989): 2289-2298.
9. Paatero, Pentti, and Unto Tapper. "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values." *Environmetrics* 5.2 (1994): 111-126.
10. Boutsidis, Christos, and Efstratios Gallopoulos. "SVD based initialization: A head start for nonnegative matrix factorization." *Pattern recognition* 41.4 (2008): 1350-1362.
11. Rehurek, R., & Sojka, P. (2011). *Gensim–python framework for vector space modelling*. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2)

In Silico Screening of the Interaction of Factor Xa and *Melilotus Officinalis* Phytochemicals

Işık ÇAKMAK¹[0000-0003-4758-4077], Ayşe Banu PAK^{1,2}[0000-0002-9490-3504] and Mustafa Emre ERCİN³[0000-0002-7340-8045]

¹ Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey

² Distance Education Application and Research Center, Trabzon University, Trabzon, Turkey

³ Ankara Atatürk Sanatorium Training and Research Hospital, Department of Medical Pathology, 06290, Ankara, Turkey
isikozdinc@gmail.com

Abstract. Disseminated intravascular coagulation (DIC) causes both microvascular and macrovascular clotting. As a result, blood flow is impaired. It is defined as widespread uncontrolled clot formation that can lead to multi-organ dysfunction syndrome. The main approach in the treatment of DIC is to eliminate the underlying pathologic condition causing DIC, supportive therapy and correction of the coagulation mechanism. Factor Xa inhibitors are among these treatments. It is known that there are many plants known as anticoagulants in the literature. In this study, the interaction between Factor Xa (PDB:2W26) and phytochemicals of *Melilotus officinalis* was investigated using molecular docking method. For the analysis, the ligand in the 2W26 receptor was downloaded from the Pubchem database. The -7.5 kcal/mol value found as a result of the docking analysis was determined as the reference value. Docking analysis was performed with 70 phytochemicals of *Melilotus Officinalis* plant downloaded from IMPPAT database. When filtered according to the reference ligand, 10 compounds were considered to be suitable molecules. In this study, Dicumarol (-9.4 kcal/mol), Rosmarinic acid (-8.5 kcal/mol), Myricetin (-8.0 kcal/mol), Quercetin (-7.9 kcal/mol), Pimara-8(14)15-diene (-7.8 kcal/mol), Honokiol (-7.8 kcal/mol), Kaempferol (-7.6 kcal/mol), Liquiritigenin (-7.5 kcal/mol), Magnolol (-7.5 kcal/mol), Rubiadin (-7.5 kcal/mol) were suggested as potential anticoagulant molecules for Factor Xa. However, this is a basic study and the proposed molecules need to be further investigated with in vitro and in vivo tests.

Keywords: Disseminated Intravascular Coagulation (DIC), *Melilotus officinalis*, Factor Xa, Molecular docking.

1 Introduction

Disseminated intravascular coagulation is not a disease but a syndrome that causes disseminated uncontrolled clot formation with over activation of the coagulation cascade as a result of various pathological conditions such as sepsis, severe infection, trauma and organ injury, solid tumors, and Covid-19 [1][2]. The coagulation cascade has two

initial pathways called the intrinsic pathway and the extrinsic pathway [3]. Various factors are involved in these pathways. After the successive activation of these factors, factor Xa is the last common point. With the activation of factor Xa, prothrombin is converted into thrombin. Thrombin converts fibrinogen into fibrin and a clot is formed [3] [4]. The formed clot adversely affects human life. The main approach in the treatment of DIC is treatment of the underlying disease, supportive therapy and correction of the coagulation mechanism [5]. Treatments in DIC are subdivided according to bleeding type and symptomatic status. Factor Xa inhibitors are among these treatments. Factor Xa inhibitors strongly and selectively inhibit thrombin formation [6]. It is known that plants are used in the treatment of many diseases. Due to the side effects of existing anticoagulants, herbs are referred to as alternative therapeutic agents. Complementary and alternative medicine (CAM) is widely used all over the world. There are many plants known as anticoagulants in the literature. Plants that can enhance or inhibit the effect of anticoagulant therapy have an important place in secondary treatment for cardiovascular diseases [7]. In this context, *Melilotus officinalis* has the potential to have anticoagulant effect. *Melilotus officinalis* is known to be in the Leguminosae family. *Melilotus Officinalis* plant is shown in Figure 1 (Figure 1).



Figure 1. Melilotus Officinalis Plant (<https://kocaelibitkileri.com/>)

An ideal clinical anticoagulant should inhibit thrombin activity without causing bleeding. In this context, new plant components that can be an alternative to the drugs used as factor Xa inhibitors are being investigated. In this context, the molecular mechanism of *Melilotus Officinalis* on coagulation was investigated in our study.

Molecular docking analysis is used in the study of plant phytochemicals. The focus of this analysis is on the interaction between receptor(s) (target(s)) and small molecules (ligands).

2 Materials And Methods

The crystal structure of Factor Xa (PDB ID: 2W26) identified as the receptor was downloaded from the Protein Data Bank (PDB) (<https://www.rcsb.org/>). PDB ID: 2W26 consists of two chains, A and B. It is complexed with BAY59-7939 (RIV). The protein was prepared using Chimera 1.16 program before docking analysis and saved in pdb format

(Figure 2). In the preparation phase, ions, water, ligand structures in the protein were removed and H and Gasteiger charges were added. The prepared protein was saved in pdbqt format using AutoDockTools-1.5.6. Phytochemicals of *Melilotus officinalis* plant were identified with the keyword 'Anticoagulants' and obtained from IMPPAT: Indian Medicinal Plants, Phytochemistry And Therapeutics database and 70 components were downloaded in sdf format. Here, the components of the selected plant (*Melilotus officinalis*) were used as ligand and factor Xa (PDB ID: 2W26) was used as receptor. Ligand structures were optimized with the General Amber Force Field (GAFF) algorithm in the command interface of Ubuntu 20.04.4 operating system and the energy for the molecule was minimized. The optimized structures were then converted to the pdbqt format of Open Babel (<https://openbabel.org/>). A grid box was used to determine the binding site for docking. For the PDB: 2W26 protein, the grid size was $10\text{\AA}\times 10\text{\AA}\times 16\text{\AA}$ with a 1.0\AA step, covering the ligand site located at 7.884\AA (along the x-axis), 5.131\AA (along the y-axis) and 21.259\AA (along the z-axis) to the center. RIV complexed with Factor Xa was used as the reference ligand in this study. Molecular docking analysis was performed using AutoDock Vina software. Chimera 1.16, Discovery Studio Visualizer software was used to visualize the analysis result. For a new molecule discovery to become a drug, it must have a desired biological activity. In addition, it must be safe and effective and have an appropriate pharmacokinetic and physicochemical profile [8]. These pharmacokinetic properties are absorption, distribution, metabolism and excretion (ADME) parameters. It has been reported that estimation of these parameters at the early stage of drug discovery greatly reduces the rate of pharmacokinetic failure in clinical trials [9]. The drug similarity and ADME properties of the molecules determined according to the reference as a result of molecular docking analysis were examined with the swissADME (<http://www.swissadme.ch/>) tool. Lipinski's rule of five (Ro5) was used to assess drug similarity [10]. According to Lipinski's rule; MW: Molecular weight less than 500, HBA: Hydrogen Bond Acceptors less than 10, HBD: Hydrogen Bond Donors less than 5, Log P less than 5 [11].

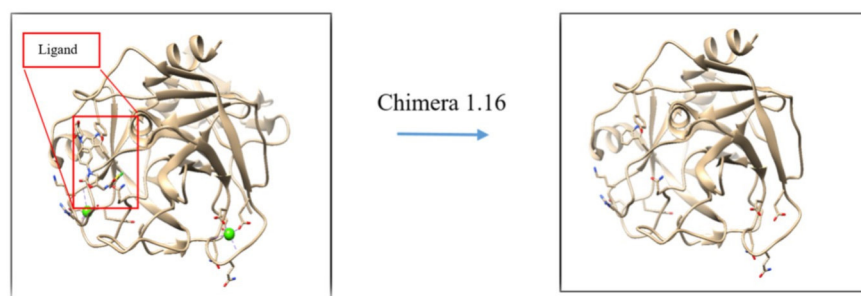


Figure 2. Cleaning of the protein from the structures inside using Chimera 1.16 software before the docking process

3 Conclusion and Discussion

It is known that the development of new drugs is a very expensive and time-consuming process. Computer-aided drug design (CADD) methods are used to demonstrate interactions between a specific molecule and a target. These methods have become a widely used technique because they require significantly less cost and time and the number of potential medicinal compounds obtained can be significantly reduced [12]. In our study, the interaction of phytochemicals from the *Melilotus Officinalis* plant with the 2W26 target structure determined as the Factor Xa receptor and their binding properties to the determined active site of the receptor were examined by the molecular docking method. The binding energy of the reference ligand used in the study was found to be -7.5 kcal/mol. As a result of the docking analysis, components with energy equal to or lower than the reference value were determined (Table 1).

Table 1. Binding Energies Determined as a Result of Molecular Docking Analysis

Component	Binding Energy	SMILES
Reference	-7.5	C1COCC(=O)N1C2=CC=C(C=C2)N3CC(OC3=O)CNC(=O)C4=CC=C(S4)Cl
Dicumarol	-9.4	O=c1oc2ccccc2c(c1Cc1c(=O)oc2c(c1O)cccc2)O
Rosmarinic acid	-8.5	O=C(O[C@@H](C(=O)O)Cc1ccc(c(c1)O)O)/C=C/c1ccc(c(c1)O)O
Myricetin	-8	Oc1cc(O)c2c(c1)oc(c(c2=O)O)c1cc(O)c(c(c1)O)O
Quercetin	-7.9	Oc1cc(O)c2c(c1)oc(c(c2=O)O)c1ccc(c(c1)O)O
Pimara-8(14),15-diene	-7.8	C=C[C@]1(C)CC[C@H]2C(=C1)CC[C@@H]1[C@]2(C)CCCC1(C)C
Honokiol	-7.8	C=CCc1ccc(c(c1)c1ccc(c(c1)CC=C)O)O
Kaempferol	-7.6	Oc1ccc(cc1)c1oc2cc(O)cc(c2c(=O)c1O)O
Liquiritigenin	-7.5	Oc1ccc(cc1)[C@@H]1CC(=O)c2c(O1)cc(cc2)O
Magnolol	-7.5	C=CCc1ccc(c(c1)c1cc(CC=C)ccc1O)O
Rubiadin	-7.5	O=C1c2ccccc2C(=O)c2c1c(O)c(c(c2)O)C

In this context, when the binding energies of the obtained phytochemicals were compared with the binding energy of the reference ligand, components with lower energy and higher binding affinity were included in the analysis. The binding regions of the reference ligand in the 2W26 protein include the Gly219, Gln192, Cys220, Lys96, Trp215, Tyr99, Phe174, Cys191, Gln192, Gly216, Ala190, Val213, Tyr228 regions, which were visualized using Discovery Studio. Gly219, Gln192, Cys220, Lys96 contain hydrogen bonds. Others include hydrophobic interactions. The 2D structures of the reference ligand and the phytochemicals showing the best binding affinity are given in Figure 3. When the binding regions of Liquiritigenin, Pimara-8(14),15-diene and Magnolol molecules are examined with the reference ligand, it is seen that they form hydrophobic interactions.


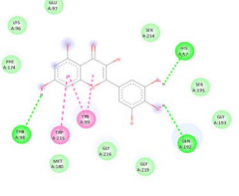
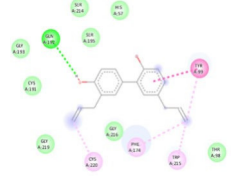
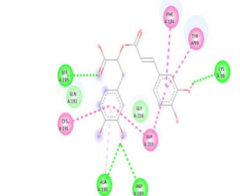
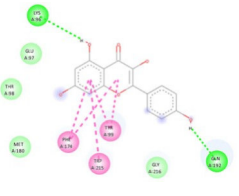
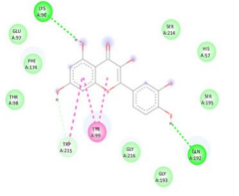
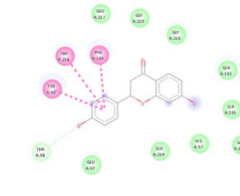

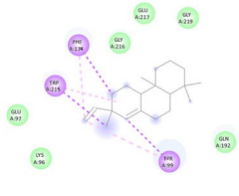
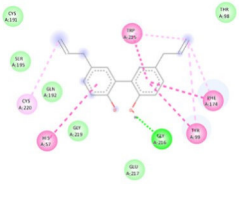
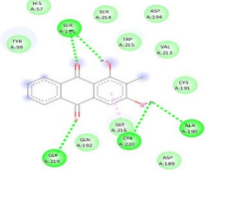
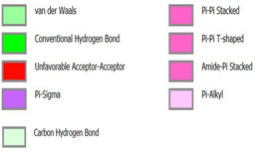
		
Rivaroxaban	Myricetin	Honokiol
		
Rosmarinic acid	Kaempferol	Quercetin
		
Liquiritigenin	Dicumarol	Pimara-8(14),15-diene
		 <ul style="list-style-type: none"> ■ van der Waals ■ Conventional Hydrogen Bond ■ Unfavorable Acceptor-Acceptor ■ Pi-Sigma ■ Carbon Hydrogen Bond ■ Pi-Pi Stacked ■ Pi-Pi T-shaped ■ Amide-Pi Stacked ■ Pi-Alkyl
Magnolol	Rubiadin	Interaction colors

Figure 3. 2D representation of the top 10 compounds and reference compounds showing affinity for the 2W26 (Factor Xa) receptor structure.

In the study, molecules with lower binding energy than the binding energy of -7.5 kcal/mol reference value, that is, molecules that showed better binding, were examined. In this context, it is thought that the phytochemicals identified may be drug candidate molecules to be used as Factor Xa inhibitors in the treatment of DIC. The web tool SwissADME (<http://www.swissadme.ch/>) was used for ADME evaluation of the proposed molecules. The ADME analysis results are shown in table 2. Physicochemical, pharmacokinetic and drug-like properties of compounds are very important parameters in the drug development phase [9]. As a result of molecular docking, the suitability of phytochemicals was determined using the Lipinski rules of the physicochemical and pharmacokinetic properties of the proposed molecules. It was determined that among the 10 suggested molecules, 2 molecules complied with Lipinski's rules with 1 violation, while the others complied with Lipinski's rules with no violation. Among the molecules that are excluded from this violation when compared to the reference molecule, there are molecules with low gastrointestinal absorption. It was also determined that some molecules could cross the blood-brain barrier. It was determined that the reference ligand Pgp substrate was positive (+) and the other

Table2. SwissADME results

Molecule	MW (g/mol)	HBA	HBD	RB	iLOGP	MLOGP	logKp (cm/s)	TPSA(Å ²)	GIA	BBBP	LV
Mol 1	435.88	5	1	6	2.95	1.41	-7.19	116.42	High	No	0
Mol 2	336.29	6	2	2	2.27	2.04	-6.88	100.88	High	No	0
Mol 3	360.31	8	5	7	1.48	0.9	-6.82	144.52	Low	No	0
Mol 4	318.24	8	6	1	1.08	-1.08	-7.4	151.59	Low	No	1
Mol 5	302.24	7	5	1	1.63	-0.56	-7.05	131.36	High	No	0
Mol 6	272.47	0	0	1	3.89	5.82	-2.97	0	Low	No	1
Mol 7	266.33	2	2	5	2.96	3.78	-4.39	40.46	High	Yes	0
Mol 8	286.24	6	4	1	1.7	-0.03	-6.7	111.13	High	No	0
Mol 9	256.25	4	2	1	1.73	1.27	-6.23	66.76	High	Yes	0
Mol 10	266.33	2	2	5	3.28	3.78	-4.39	40.46	High	Yes	0
Mol 11	254.24	4	2	0	1.94	0.92	-5.67	4.6	High	Yes	0

components except molecule 9 were negative (-). The analysis results are given in the table below (Table 2). Mol 1 is the reference molecule.

MW: Molecular weight, HBA: Hydrogen Bond Acceptors, HBD: Hydrogen Bond Donors, RB: Number of Rotatable Bonds, TPSA: Topological Polar Surface Area, GIA: Gastrointestinal Absorption, BBBP: Blood Brain Barrier Permeability, LV: Lipinski Violations [13].

The 10 molecules identified in our study were compared with reference compounds as potential anticoagulant drug candidates (Dicumarol (-9.4 kcal/mol), Rosmarinic acid (-8.5 kcal/mol), Myricetin (-8.0 kcal/mol), Quercetin (-7.9 kcal/mol), Pimara-8(14), 15-

diene (-7.8 kcal/mol), Honokiol (-7.8 kcal/mol), Kaempferol (-7.6 kcal/mol), Liquiritigenin (-7.5 kcal/mol), Magnolol (-7.5 kcal/mol), Rubiadin (-7.5 kcal/mol)) were determined. The molecules determined as a result of the docking analysis were determined according to their energy values. Figure 3 shows that some compounds have hydrophobic interactions.

It has been stated that dicoumarol and its derivatives, which are stated to have anticoagulant effects, also show anticancer, antimicrobial and antiviral activities [14]. In another study, it was reported that Rosmarinic acid inhibited enzymatic thrombin activity [15]. It has been stated that the inhibitory effect of Quercetin on Factor X has been confirmed in studies [16].

Finally; Since our study is an *in silico* study and is based only on data analysis, it is thought that it should be examined in more detail with *in vitro* and *in vivo* experiments.

4 References

1. Palta S, Saroa R, Palta A (2014). Overview of the coagulation system. *Indian Journal of Anaesthesia* 58: 515–523.
2. Cate H (2000). Pathophysiology of disseminated intravascular coagulation in sepsis 28: 9–11.5.
3. M, Cengiz MS, Çeray Y, Ramazanoğlu A (2011). Trauma and Coagulation. *Türk Yoğun Bakım Derneği Dergisi* 9.
4. Atalan N (2013). Hemostaz. *Gogus-Kalp-Damar Anestezi ve Yoğun Bakım Derneği Dergisi* 19: 109–112.
5. Çelebi H. *Türk Hematoloji Derneği, Temel Hemostaz Tromboz Kursu. Yaygın damar içi pıhtılaşması* 60–63.
6. Rupprecht HJ, Blank R (2010). Clinical pharmacology of direct and indirect factor xa inhibitors. *Drugs* 70: 2153–2170.9.
7. Saw JT, Bahari MB, Ang HH, Lim YH (2006). Potential drug–herb interaction with antiplatelet/anticoagulant drugs. *Complementary Therapies in Clinical Practice* 12: 236–241.
8. Dulsat J, López-Nieto B, Estrada-Tejedor R, Borrell JI (2023). Evaluation of Free Online ADMET Tools for Academic or Small Biotech Environments. *Molecules*.Jan 12;28(2):776. doi: 10.3390/molecules28020776. PMID: 36677832; PMCID: PMC9864198
9. Daina, A., Michielin, O. & Zoete, V(2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7, 42717 .<https://doi.org/10.1038/srep42717>
10. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature chemistry*, 4(2), 90–98.
11. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery re-views*, 64, 4–17.

12. Baig MH, Ahmad K, Roy S, Ashraf JM, Adil M, Siddiqui MH, Khan S, Kamal MA, Provazník I, Choi I (2016). Computer Aided Drug Design: Success and Limitations. *Curr Pharm Des.*22(5):572-81. doi: 10.2174/1381612822666151125000550. PMID: 26601966.
13. Evranos Aksöz B, Kaynak Onurdağ FK, Aksöz E, Özgen S. (2021). In Silico ADME Screening and Evaluation of Antimicrobial and Antimycobacterial Activities of 3,5-Diphenyl Pyrazoline Derivatives. *Süleyman Demirel Üniver- sitesi Sağlık Bilimleri Dergisi.*;12(2):184-91.
14. Sun, C., Zhao, W., Wang, X., Sun, Y., & Chen, X. (2020). A pharmacological review of dicoumarol: An old natural anticoagulant agent. *Pharmacological Research*, 160, 105193.
15. Chapado, L., Linares-Palomino, P. J., Salido, S., Altarejos, J., Rosado, J. A., & Salido, G. M. (2010). Synthesis and evaluation of the platelet antiaggregant properties of phenolic antioxidants structurally related to rosmarinic acid. *Bioorganic chemistry*, 38(3), 108-114.
16. Sabbagh, G. M., Al-Beik, L. M., & Hadid, I. (2023). In silico and in vitro anticoagulant activity detection of quercetin, rutin, and troxerutin as new potential inhibitors of factor Xa. *Egyptian Journal of Chemistry*, 66(2), 151-166.

Parkinson Hastalığının Teşhisinde EEG-F1 Kanal Aktivitesinin Rolü

Neslihan BAKI¹[0000-0002-7850-7333] ve Nurhan GÜRSEL ÖZMEN²[0000-0002-7016-5201]

¹ Karadeniz Teknik Üniversitesi, Biyoistatistik ve Tıp Bilişimi, Trabzon

² Karadeniz Teknik Üniversitesi, Makine Mühendisliği, Trabzon
neslihanbaki520@gmail.com, gnurhan@ktu.edu.tr

Özet. Parkinson hastalığı (PH), tüm dünyada giderek önemi artan, hem bireye hem de ülkelerin sağlık sistemlerine maliyeti çok yüksek olan, Alzheimer hastalığından sonra en sık görülen ikinci kronik ve ilerleyici nörodejeneratif hastalıktır. Parkinson hastalığına karşı etkili tedavi tam olarak sağlanamamıştır. Hastalığın erken tanısı tedavi seçimini etkilediğinden oldukça önem kazanmaktadır. Bu çalışmada PH'nin erken tanısı için tek kanallı EEG veri analizi ile spektral öznitelikler çıkarılarak bir analiz gerçekleştirilmiştir. Çalışma, 20'si PH'li ve 20'si kontrol olmak üzere 40 kişiden toplanan açık kaynaklı Turku EEG verileri üzerinde gerçekleştirilmiştir. Veriler yaş ve cinsiyete göre demografik olarak eşleştirilmiş ve dinlenme durumunda gözleri açık olarak kaydedilmiştir. Literatürden farklı olarak, EEG verilerinin beyin ön bölgesinde yer alan F1 kanalına ait, Delta, Theta, Alpha, Beta ve Gamma bantlarının Güç Spektral Yoğunlukları (PSD) çıkarılarak analiz edilmiştir. Daha sonra bunlardan Kolmogorov Algoritmik Karmaşıklık (KAC) özniteliği çıkarılmıştır. Sınıflandırma yeni bir topluluk öğrenme algoritması olan CatBoost Algoritması ile gerçekleştirilmiştir. Çalışmanın sonucunda tek kanal olan F1 kanal verileriyle % 80 doğruluk, % 79 kesinlik, % 80 duyarlılık ve % 77 F1 puanı elde edilmiştir. Öte yandan, tüm kanallar olan 64 kanal verisi ile % 70 doğruluk, % 73 kesinlik, % 70 duyarlılık ve % 69 F1 puanı, 18 merkezi kanal verisi ile % 70 doğruluk, % 71 kesinlik, % 70 duyarlılık ve % 68 F1 puanı elde edilmiştir. Bu sonuçlara dayanarak, KAC özniteliği kullanılarak, tek EEG kanalı F1 kullanımının, aynı veri setini kullanan mevcut çalışmalara kıyasla başarılı bir sınıflandırma gerçekleştirdiği sonucuna varılabilir. Bu çalışmanın sonuçları gelecekteki PH tanı araştırmaları için umut verici olacaktır.

Anahtar Sözcükler: Parkinson, EEG Sinyal İşleme, Kolmogorov Algoritmik Karmaşıklık (KAC), Güç Spektral Yoğunluğu, CatBoost Algoritması.

Role of EEG - F1 Channel Activity for Parkinson's Disease Detection

Neslihan BAKI¹[0000-0002-7850-7333] and Nurhan GÜRSEL ÖZMEN²[0000-0002-7016-5201]

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 the second most common chronic and progressive neurodegenerative disease after Alzheimer's disease, which is becoming increasingly important all over the world, and has a very high cost both to the individual and to the health systems of countries. Effective treatment against Parkinson's disease has not been fully achieved. Early diagnosis of the disease becomes very important as it affects the choice of treatment. In this study, a single channel EEG data analysis by extracting spectral features is tested for early diagnosis of PD. The study is conducted on open-source Turku EEG data collected from 40 individuals, 20 with PD and 20 controls. The data is demographically matched for age and gender and recorded in a resting state with eyes open. Apart from the literature, the F1 channel, located in the frontal region of the brain, was analyzed by extracting the Power Spectral Densities (PSDs) of Delta, Theta, Alpha, Beta and Gamma bands of EEG data. Then the Kolmogorov Algorithmic Complexity (KAC) feature was extracted from them. A new ensemble learning method CatBoost algorithm is used for classification. The results of the study achieved an 80% accuracy, 79% precision, 80% sensitivity, and 77% F1 score with single F1 channel data. However, 70% accuracy, 73% precision, 70% sensitivity, and 69% F1 score were obtained with whole 64 channels data and 70% accuracy, 71% precision, 70% sensitivity, and 68% F1 score with 18 central channels data. Based on these results, it can be concluded that the use of single EEG channel F1 with KAC feature perform a successful classification compared to the existing studies using the same dataset. The results of this study would be promising for future PD diagnosis research.

Keywords: Parkinson, EEG Signal Processing, Kolmogorov Algorithmic Complexity (KAC), Power Spectral Density, CatBoost Algorithm.

1 Introduction

PD was first introduced by James Parkinson in 1817 with his essay "An essay on swaying paralysis" [1]. The prevalence of PD varies between 1.5 and 22 per 100 thousand people per year for all age groups. These results may vary according to the demographic characteristics of the study population, diagnostic criteria and case definition source. The incidence of PD increases with age [2]. PD is a progressive neurodegenerative disease caused by the loss of nerve cells in the substantia nigra, the brain region responsible for the production of the hormone dopamine. The hormone dopamine, provides inter-nerve transmission. The loss and damage of these nerve cells in the substantia nigra leads to a decrease in dopamine production. A decrease in dopamine below its normal level causes problems in the body's ability to fulfill its normal functions [3, 4].

Four main symptoms have been identified in PD: hypokinesia, which is the lack of movement, or bradykinesia, which is the slowing down of movement; stiffness in areas such as the wrist, shoulder and neck; tremor at rest due to imbalance of dopamine and neurotransmitters; and postural instability, which is the inability to maintain normal body posture [5]. In addition to these motor disorders, loss of physiospatial abilities such as spatial perception and orientation has been reported to be seen from the early stage of the disease [6]. The diagnosis of PD is usually based on the presence of bradykinesia, rigidity and tremor at rest [7-9].

In the current state of the art, PD cannot be diagnosed at an early stage. Since the symptoms are non-specific, the disease is very difficult to diagnose. There is no specific medical test in the diagnostic process. The patient's medical history and clinical examination are important in the diagnosis of PD [10]. Early diagnosis of disease can have error rates of up to 30% [11]. Early diagnosis of the disease affects the choice of treatment. Early initiation of neuroprotective treatment to prevent excessive loss of dopaminergic neurons in the substantia nigra and to protect these neurons may delay the progression of the disease and the development of dementia.

Recent research on diagnosing the disease using magnetoencephalography (MEG) or electroencephalography (EEG) is also pioneering PD research. Since PD is caused by subcortical dysregulation, the use of EEG to evaluate the electrical activity of the brain has been a very appropriate idea [12]. It is easy to apply, low cost and has no harm to the individual. Studies using EEG in PD are available in the literature [13, 14]. It has been found that EEG amplitudes at theta and alpha frequencies are significantly different in PD. These studies suggest that EEG recordings may be useful in identifying changes in PD. Therefore, there is still a need for the use of new analysis methods in research with EEG electrical signals in PD. However, such research will also contribute to revealing other features of the brain. Thus, the findings that cause brain dysfunctions can be revealed.

For all these reasons, in this study, Kolmogorov Algorithmic Complexity (KAC) feature is extracted from EEG data, and CatBoost algorithm is used for distinguishing PD and HCs for early diagnosis of PD.

2 Method

2.1 Data Set and Pre-processing

An opensource dataset which was recorded by the University of Turku, Finland was used [15]. The Turku dataset contains resting state EEG recordings of 40 patients, 20 PD and 20 SC groups. These recordings include both eyes open and eyes closed data in the medicated state. Only the eyes open data set was used for this study. A 64-channel NeurOne Tesla device was used for recordings. It has a sampling frequency of 500 Hz. The data were recorded for two minutes and referenced to the mean reference. The electrode channels of the Turku dataset are Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz, Iz, FC1, FC2, CP1, CP2, FC5, FC6, CP5, CP6, FT9, FT10, TP9, TP10 consists of F1, F2, C1, C2, P1, P2, AF3, AF4, FC3, FC4, CP3, CP4, PO3, PO4, F5, F6, C5, C6, P5, P6, AF7, AF8, FT7, FT8, TP7, TP8, PO7, PO8, Fpz, CPz, POz, Oz.

The analysis scheme is shown in Figure 1. Here, first classification was performed using 64 channels from 64-channel EEG data. Then, classification was performed using 18 central channels in the sensorimotor cortex. Finally, in order to evaluate single channel performance, only the F1 channel was analyzed. The electrode layout of the NeurOne Tesla EEG system is given in Figure 2. The raw EEG image of the F1 channel of the control group and individuals with PD is shown in Figure 3.

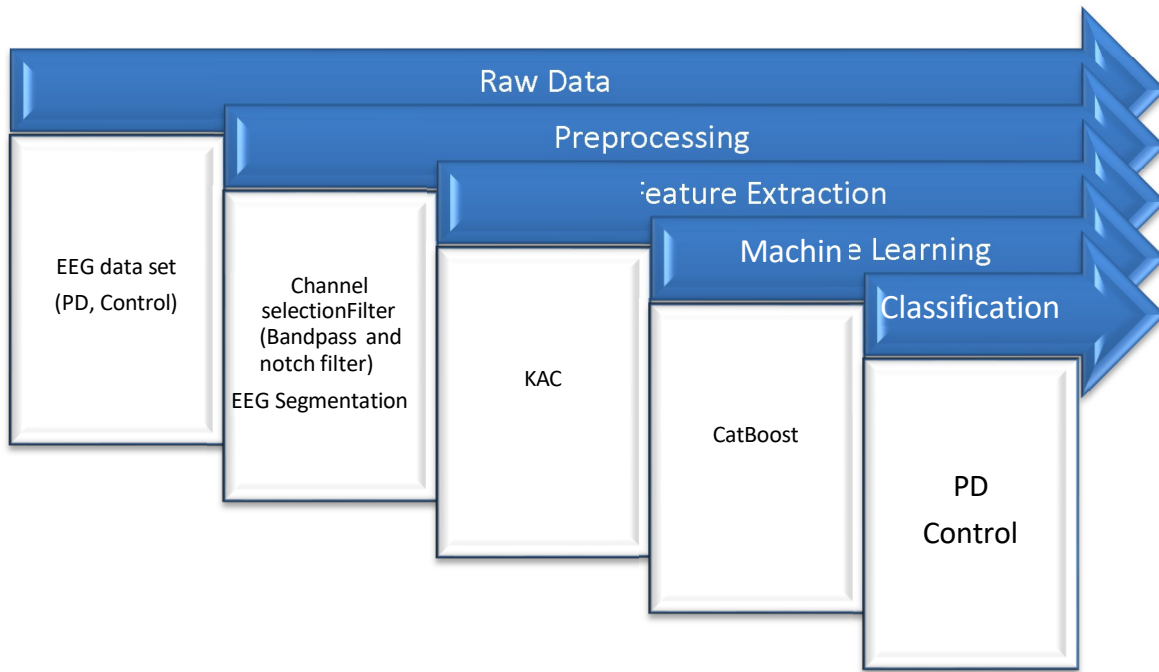


Figure 1. Analysis Scheme of EEG data

The EEG data was segmented for 40 seconds and band-pass filtered between 0.1 and 120 Hz and notch filter was used at 50 Hz and 100 Hz.

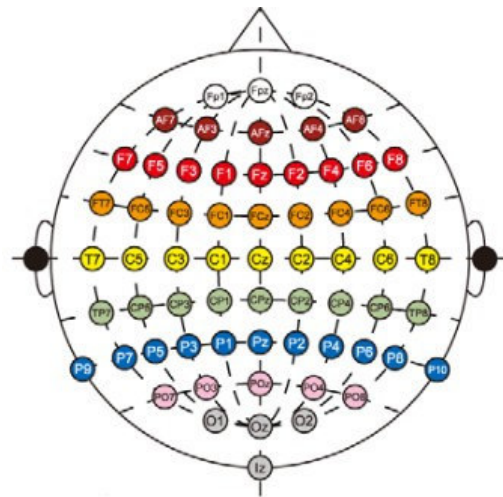


Figure 2. 64 channels electrode locations of NeuroOne Tesla 10-20 System EEG System

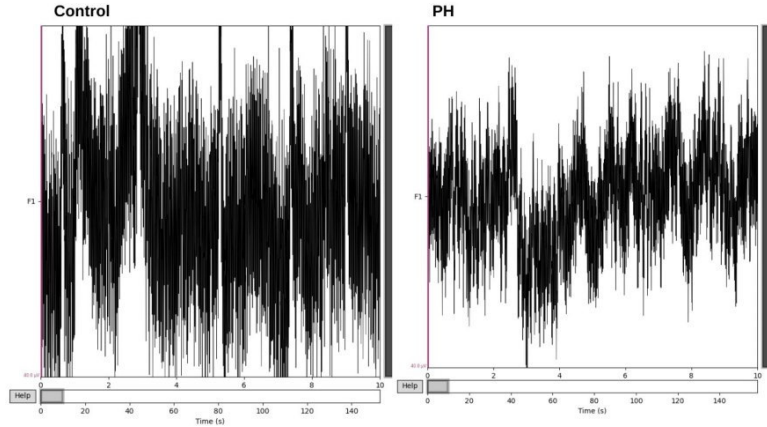


Figure 3. F1 channel raw EEG image of control and PD subjects from Turku dataset

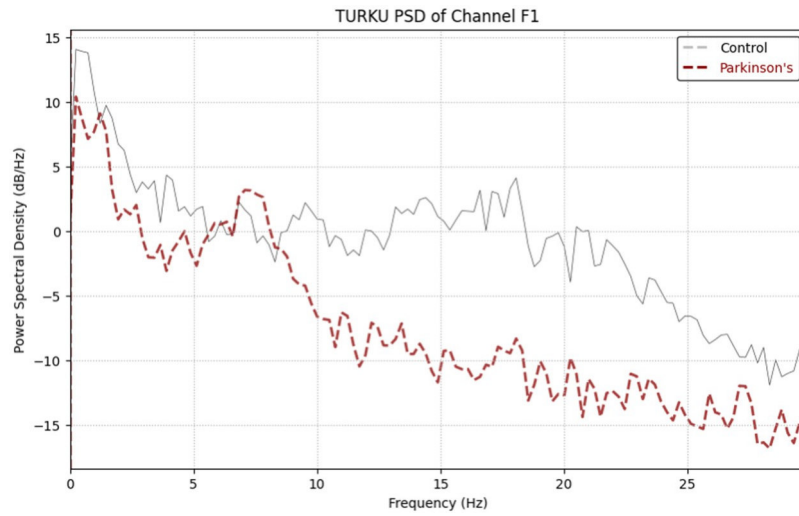


Figure 4. PSD image of the F1 channel in the range of 0-30 Hz for one person each in the control and PD groups

2.2 Feature Extraction

In this study, Algorithmic "Kolmogorov" Complexity (KAC) is extracted from spectral features. The Algorithmic "Kolmogorov" Complexity of a sequence s is the shortest program length that computes or outputs s . It formulates the concepts of simplicity and complexity. If $d(s)$ of a sequence s with the fewest bits is of minimum length, this is defined as the minimum description of s . The length $d(s)$ found here is called the Kolmogorov Complexity of s . Kolmogorov Complexity is denoted by $K(s)$ [16]. The length of the shortest description is given by equation (1) below.

$$K(s) = |d(s)| \quad (1)$$

Here K will be defined as a function on the natural numbers N consisting of finite binary sequences of arbitrary length. Therefore, $K: \{0,1\}^* \rightarrow N$ is a function defined by 'objects' represented by binary arrays. Here, the data for each EEG channel is represented by binary numbers.

2.3 Classification

Classification was done with the CatBoost algorithm. CatBoost is a new ensemble learning method proposed by Yandex and derived from the Gradient Boosting and Decision Tree algorithms, which have been shown to have the ability to deal well with heterogeneous data [17]. CatBoost also uses forgetful trees that are balanced as base predictors. Thus, it overcomes the problems of overlearning. In general, it exhibits superior classification

performance compared to other ensemble learning algorithms. The Python programming language and the Google Colab environment were used in the analysis of this study. The raw data was first preprocessed, then the Kolmogorov Algorithmic Complexity feature was extracted and binary classification was performed with the CatBoost Algorithm.

3 Experiments and Results

The EEG dataset contains a total of 40 data belonging to 20 PD and 20 control groups. The data were classified by 10-fold cross-validation with the CatBoost algorithm. The KAC feature was selected from the spectral features of the frequency domain and the model success was evaluated by first classifying the data consisting of PD and control group with 64 channels. Then 18 central channels data and finally single channel F1 channel data were analysed with the CatBoost algorithm. Table 1 shows the ten-fold cross-validation results obtained with the CatBoost algorithm for the three different cases of the Turku dataset. The table shows the accuracy, sensitivity, precision, precision, F1 score value of the model. In the study, 80% accuracy, 79% precision, 80% sensitivity and 77% F1 score were obtained from F1 channel, and 70% accuracy, 73% precision, 70% sensitivity and 69% F1 score from 64 channels were obtained. Finally 70% accuracy, 71% precision, 70% sensitivity and 68% F1 score with 18 central channels were obtained. According to the table, the highest classification accuracy was achieved using the single F1 channel (80.00%).

Table1. Classification Results

Channels	Accuracy	Precision	Recall	F1 score
64 channels	70	73	70	69
18 channels	70	71	70	68
F1 channel	80	79	80	77

Table 2. Turku dataset literature results

Made in Study	Data set (eyes open)	Feature analysis	Machine learning techniques	Classification	Accuracy
Henry Railo et al. [9].	Turku	Budget-based search algorithm and 10 channels	Logistic Regression	PD and Control	% 77.80
Mehmet Fatih Karakas et al. [18].	Turku	Gray-level co-occurrence matrix (GLCM) properties and 64 channels	Support Vector Machine	PD and Control	% 63.25
Ilkka Suuronen et al. [19].	A total of 3 datasets with Turku	Custom budget-based search algorithm and 5 channels	Logistic Regression	PD and Control	% 76.00
Proposed study	Turku	KAC Feature (single feature) and single channel (F1)	CatBoost	PD and Control	% 80.00

4 Discussion and Conclusions

PD is the second most common neurodegenerative disease worldwide. There is no known cure. In recent years, researchers have been focusing on developing automated methods to detect PD in its early stages in order to prevent as much as possible the disease symptoms that reduce patients' quality of life. Therefore, this study proposes a novel approach to characterize early-stage PD activity over single EEG channel. The proposed method selects the KAC feature from the frequency domain features and performs classification over 64, 18 and single channels. Experiments show that choosing a single channel is a good choice. Gradient summation based CatBoost algorithm is used in the study. CatBoost facilitates the analysis by making it easier to focus on the contribution of a single feature to the model, rather than a particular value of a particular feature.

As can be seen in Table 2, among the EEG-based methods, the proposed method stands out by achieving the best classification result, outperforming Logistic Regression by 2.20%. As intended, the channel analysis provides an overview of PD activity in the scalp. As can be seen, the F1 channel, located in the frontal region, achieved high success compared to the other multiple channels. Oscillatory responses in the delta frequency range were examined in PD, and the responses of delta sensory evoked oscillations and event-related oscillations were significantly lower in PD compared to controls. Event-related oscillation responses were found to decrease in all frontal, central and parietal regions [20]. When we evaluate the studies in the literature, classification was performed with traditional machine learning methods using multiple channels and multiple features. Compared to the traditional machine learning methods used in the literature, the CatBoost algorithm, a state-of-the-art ensemble learning method, was used to perform successful classification over a single feature (KAC) and a single channel (F1 channel). The use of a single channel and a single feature significantly reduces the analysis time of the study. In summary, the results of this pilot study with a single electrode channel, seems promising for future PD research, since a high classification rate is obtained compared to the literature.

In the future, with advances in Telemedicine, Cloud and Internet of Things (IoT) computing technology, it can be developed as a real-time smart health monitoring method for the prevention and detection of PD. With an advanced EEG device, the study can detect EEG signals through a single electrode and transmit these recordings to remote neurologists, enabling remote health monitoring of patients. However, it can also be evaluated to detect other diseases including Alzheimer, Epilepsy, Autism, Schizophrenia, etc.

References

1. Parkinson, James. "An essay on the shaking palsy." *The Journal of neuropsychiatry and clinical neurosciences* 14.2 (2002): 223-236.
2. Wirdefeldt, Karin, et al. "Epidemiology and etiology of Parkinson's disease: a review of the evidence." *European journal of epidemiology* 26 (2011): 1-58.
3. Parkinson's UK, What is Parkinson's. <http://www.parkinsons.org.uk/about-parkinsons/what-is-parkinsons.aspx>. last accessed 2024/05/10.
4. Khan, Naheed L., and Thomas Britton. "Clinical features, pathophysiology and genetics." *HOSPITAL PHARMACIST-LONDON* 11.1 (2004): 9-17.
5. Jankovic, J. "Pathophysiology and assessment of parkinsonian symptoms and signs." *Handbook of Parkinson's disease*. New York: Taylor and Francis Group LLC (2007): 79-104.
6. 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.
7. Kalia, Lorraine V., and Anthony E. Lang. "Parkinson's disease." *The Lancet* 386.9996 (2015): 896-912.
8. Jankovic, Joseph. "Progression of Parkinson disease: are we making progress in charting the course?." *Archives of neurology* 62.3 (2005): 351-352.
9. Beitz, Janice M. "Parkinson's disease: a review." *Front Biosci (Schol Ed)* 6.1 (2014): 65-74.
10. Parkinson's UK Homepage, <http://www.parkinsons.org.uk/about-parkinsons/what-is-parkinsons.aspx>, last accessed 2024/05/10.
11. 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.
12. 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.
13. 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.
14. 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.
15. Railo, Henry, et al. "Resting state EEG as a biomarker of Parkinson's disease: Influence of measurement conditions." *BioRxiv* (2020): 2020-05.
16. Kolmogorov Karmaşıklığı Homepage, https://en.wikipedia.org/wiki/Kolmogorov_complexity, last accessed 2024/05/10.
17. Prokhorenkova, Liudmila, et al. "CatBoost: unbiased boosting with categorical features." *Advances in neural information processing systems* 31 (2018).
18. Karakaş, Mehmet Fatih, and Latifoğlu Fatma. "Distinguishing Parkinson's disease with GLCM features from the hankelization of EEG signals." *Diagnostics* 13.10 (2023): 1769.
19. Suuronen, Ilkka, et al. "Budget-based classification of Parkinson's disease from resting state EEG." *IEEE Journal of Biomedical and Health Informatics* (2023).
20. Emek-Savaş, Durusu Derya, et al. "Decrease of delta oscillatory responses in cognitively normal Parkinson's disease." *Clinical EEG and neuroscience* 48.5 (2017): 355-364.

A Bibliometric Analysis of the Use of Machine Learning Methods in Variant Effect Prediction

Glbahar Merve ŐILBIR ¹ [0000-0003-0321-7259], Burin KURT ¹ [0000-0001-5781-2382]

¹ Department of Biostatistics and Medical Informatics, Karadeniz Technical University,
Trabzon, Turkey
gmerve.cakmak@gmail.com
burcinnkurt@gmail.com
il.com

Abstract. Studies using computational methods such as artificial intelligence, machine learning and deep learning to predict the effect of variants occurring in the human genome on the phenotype have increased recently. The aim of this study is to provide an overview of scientific research using machine learning methods in variant effect prediction using the bibliometric method. For this purpose, the Web of Science Core Collection (WoSCC) database was used to access the relevant literature in the study. Countries, institutions, authors, journals, quotations and keywords were analyzed using the "bibliometrix" library in the R-Studio program. As a result of the analysis, it has been seen that the popularity of scientific publications on the use of machine learning methods in variant effect prediction has increased in recent years, and the largest share of this increase is due to joint research by institutions in the United States with China, Germany, England and Australia. In the publication production in this field, it can be seen that the publications made by researchers Majid Masso and Yuedong Yang have spread over a long period of time, and when looking at the publications made in the last few years, researchers Yongguo Liu, Yun Zhang, Haicang Zhang and Jiajing Zhu come to the fore. It was observed that the most cited author was researcher Jian Zhou (1.116). Although there has been an increasing trend in publications in this field in recent years, it has been determined that older publications are still cited more. Therefore, it has become clear that there is still a need to conduct further research in this field, to strengthen international cooperation and communication, and to increase the quality of the literature by gaining experience.

Keywords: Bibliometric Analysis, Bibliometrix, Variant Effect Prediction, Machine Learning

Transforming Human-Computer Interaction: Gesture Recognition Technologies for Accessible Smart Home

Ufuk Beşenk¹, Ahmet Özkurt^{1,2}, Ömür Öcal³, Tuğçe Toprak⁴, Pembe Keskinoglu⁵, M. Alper Selver^{1,2}

¹ Department of Electrical and Electronics Engineering, Dokuz Eylul University, Izmir, Turkey

² Izmir Health Technologies Development and Accelerator (BioIzmir), Dokuz Eylul University, Izmir 35160, Turkey

³ Core Bina Otomasyon Teknolojileri San ve Tic Ltd Şti, 35410, Izmir/ Turkey

⁴ Institute of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey

⁵ Faculty of Medicine Basic Medical Sciences Department of Biostatistics And Informatics, Dokuz Eylul University, Izmir 35160, Turkey

Abstract. This study presents a novel smart home control system utilizing gesture-based interactions to enhance accessibility and usability, particularly for users with physical disabilities. As smart homes become increasingly common, traditional control mechanisms like touchscreens and remotes often fail to meet the needs of all users. This research addresses this gap by developing an intuitive, non-verbal communication system that interprets static human hand gestures as commands using advanced image processing and machine learning techniques.

The core technology integrates a camera module with software that captures and processes hand movements in real time. The dataset is organized into 10 distinct gestures containing 316.777 images for training and 100.000 for testing. Convolutional Neural Networks are employed to recognize and classify these gestures into commands for controlling various home devices, such as lights and thermostats. To optimize computational efficiency, MobileNet models, designed for edge computing, are utilized. These models provide high accuracy while being computationally lightweight, making them ideal for gesture recognition in smart homes. Their streamlined architecture ensures rapid processing and responsiveness, crucial for real-time applications in accessibility-focused environments.

An extensive and careful testing and validation framework demonstrates the system's high accuracy and efficiency, even in varied environmental conditions, making it a viable solution for improving smart home interactions for users with limited mobility. It is worth to point that this study, focuses on static hand gesture recognition, with future research to explore dynamic recognition in depth. The findings indicate that dynamic gesture recognition, although more computationally intensive, offers richer interaction by interpreting continuous movements and subtle nuances. Conversely, static gesture recognition is less resource-intensive and faster, suitable for simpler command executions.

This research advances the fields of IoT and human-computer interaction by pushing the boundaries of traditional home automation interfaces. It offers a more inclusive, user-friendly approach, laying the groundwork for future innovations in smart home technologies and emphasizing the importance of accessibility and ease of use in consumer electronics.

Keywords: Smart Home Control System, Gesture-Based Interactions, Convolutional Neural Networks (CNNs), Digital Image Processing, Machine Learning.

1 INTRODUCTION

1.1 Scope of the Research

The integration of smart home technology, significantly driven by advancements in the Internet of Things (IoT), has revolutionized residential spaces. These technologies enhance convenience, security, and energy efficiency by enabling seamless interactions between devices. However, traditional control mechanisms like touchscreens, remote controls, and mobile applications often fail to meet the accessibility needs of all users, particularly those with physical disabilities. This gap underscores the need for innovative solutions to enhance the inclusiveness and usability of smart home systems.[1]

Gesture recognition technology is a pivotal advancement in human-computer interaction, offering a more intuitive and natural communication method by mirroring human gestures. Adapting gesture-based controls within smart homes is particularly beneficial, providing a non-verbal and accessible interface that accommodates users with varying levels of mobility and dexterity.

Hand gesture recognition involves analyzing a user's hand movements through computational methods to control devices. This process has two phases: detection and classification. The detection phase identifies the presence and position of a hand within digital images obtained via cameras or sensors, using techniques like skin detection, background subtraction, and noise reduction to enhance localization accuracy. The classification phase then interprets specific gestures using advanced machine learning models, such as Convolutional Neural Networks (CNNs), which excel at recognizing complex visual patterns.[2]

Studies have shown that Convolutional Neural Networks (CNNs) are highly effective in recognizing and interpreting complex patterns in visual data, making them well-suited for tasks involving gesture identification [3]. However, challenges remain in reliably building systems that depend on hand gestures as input, with historical accuracy rates averaging 78%. Incorporating deep learning algorithms has significantly improved this accuracy, allowing systems to exceed 90% [4].

The primary goal of this project is to develop a smart home control system that employs advanced digital image processing and machine learning techniques to interpret static hand gestures as commands for home device control. This initiative aims to enhance the accessibility of smart home technologies, presenting an intuitive user interface that does not depend on physical touch or fine motor skills, and effectively mitigates environmental variables such as lighting variations or complex backgrounds.

1.2 Convolutional Neural Network Techniques

Convolutional Neural Networks (CNNs) are deep learning models adept at processing visual information, inspired by the human visual system. They learn spatial hierarchies of features from images through their multi-layer architecture. Convolutional layers use filters to extract spatial features like edges and textures, creating feature maps for further analysis. Pooling layers reduce the spatial dimensions of these maps, enhancing

computational efficiency and reducing sensitivity to feature positioning. Fully connected layers handle high-level reasoning, transforming spatial features into outputs such as class labels. Activation functions like ReLU introduce non-linearity, enabling the CNN to model complex patterns. Dropout layers prevent overfitting by randomly deactivating neurons during training, increasing network robustness.[5,6,7,8,9]

1.3 Convolutional Neural Network Applications

CNNs have become a cornerstone in the field of computer vision, supporting a wide array of tasks such as image classification, object detection, semantic segmentation, and gesture recognition. A pivotal moment in the history of CNNs was the introduction of AlexNet by Krizhevsky et al., a revolutionary architecture that significantly outperformed competitors in the ImageNet competition. This success not only demonstrated the efficacy of CNNs in image recognition but also popularized their use in deep learning applications across various domains [3].

Following AlexNet, later models like VGGNet, ResNet, and MobileNet have continued to advance the capabilities of CNNs. These models have introduced refinements in architecture that enhance processing efficiency and adaptability, enabling them to manage more complex and varied tasks effectively. Such developments underscore the ongoing evolution and importance of CNNs in tackling real-world challenges through advanced pattern recognition and image processing techniques [11,12,13].

Overall, the impact of CNNs extends beyond theoretical applications, proving essential in practical machine learning scenarios. Their unparalleled ability in identifying and learning from visual data makes them an indispensable tool in the continuous advancement of artificial intelligence technologies.

1.4 Convolutional Neural Network Models

VGG16 (2014), is a convolutional neural network model known for its architectural simplicity and depth, widely used in image recognition tasks. Developed by the Visual Graphics Group at the University of Oxford, VGG16 features 16 trainable layers (13 convolutional and 3 fully connected) aimed at enhancing feature extraction. Its use of small (3x3) convolution filters across all layers simplifies the structure and aids in learning complex features. Max pooling layers manage spatial dimensions and reduce computation in deeper layers. However, VGG16 is computationally demanding, requiring significant resources for training and inference, making it less suitable for resource-constrained environments like mobile or edge devices, where more efficient models like MobileNet are preferred. [12]

MobileNetV2 (2018), developed by Mark Sandler and his team at Google in 2018, is a convolutional neural network designed for mobile and edge computing. It emphasizes computational efficiency and minimal memory usage while delivering robust performance in image classification and object detection. The model innovates with inverted residual connections, using a narrow bottleneck layer connected to a wider expansion layer, which allows for efficient, compact feature representations. It employs depthwise

separable convolutions that separate the filtering and combining processes into two stages, significantly reducing the computational load. Linear bottlenecks within each inverted residual block reduce dimensionality and compress information effectively, minimizing information loss while maintaining high accuracy. The ReLU6 activation function enhances numerical stability, especially beneficial for quantized models deployed on edge devices. MobileNetV2 is extensively adopted for developing deep learning models on mobile devices due to its excellent balance between efficiency and accuracy, making it particularly advantageous for applications requiring low power consumption and limited computational resources, such as image classification, object detection, and semantic segmentation. Empirical evaluations demonstrate that MobileNetV2 consistently achieves high accuracy and computational efficiency, often surpassing other models in its category, establishing it as a significant contribution to lightweight, high-performance computing in deep learning applications.[13,14]

MobileNetV3 (2019), developed by Andrew Howard and his team at Google in 2019, is the latest addition to the MobileNet series. It builds on MobileNetV1 and V2, introducing innovative design principles to further optimize performance and reduce computational demands for mobile and edge computing. The design process used Network Architecture Search (NAS) to optimize the balance between latency and accuracy. Squeeze-and-Excitation modules enhance feature relevance by adjusting channel-wise responses. Continuing from MobileNetV2, it employs inverted residuals and linear bottlenecks for efficiency. The new Hard-Swish activation function increases non-linearity and improves stability during model quantization. MobileNetV3 is available in two variants: MobileNetV3-Large, which prioritizes higher accuracy with a larger architecture, and MobileNetV3-Small, designed for environments with strict resource constraints, offering reduced latency. MobileNetV3 excels in mobile and embedded vision tasks like image classification, object detection, and semantic segmentation, achieving an outstanding balance between performance and resource utilization, making it ideal for real-world edge computing applications. [15,16]

EfficientNet (2019). EfficientNet, introduced by Mingxing Tan and Quoc V. Le in 2019, advances CNN architectures through optimized model scaling, achieving high accuracy with reduced computational costs. Its compound scaling method balances network width, depth, and resolution, enhancing performance without excessive computational demands. The EfficientNet series starts with EfficientNet-B0, established through neural architecture search, and scales up to EfficientNet-B7 using this method. It incorporates inverted residuals, linear bottlenecks, and squeeze-and-excitation blocks to efficiently handle high-dimensional data and improve feature extraction. EfficientNet models outperform older architectures like ResNet and Inception in both accuracy and computational efficiency, excelling in benchmarks such as ImageNet. EfficientNet-Lite variants are optimized for mobile and edge devices, making them suitable for tasks like object detection, semantic segmentation, and image classification.

Implementing a gesture-based control system in smart homes enhances accessibility and offers intuitive interaction through natural hand movements. This reduces cognitive

load and increases task efficiency. This project will explore static and dynamic gesture recognition to determine the most effective method for smart home interaction. [17]

1.5 Performance Metrics

Precision, recall, and F1 score are crucial metrics for evaluating the performance of machine learning models, especially in classification tasks. These metrics help provide a more nuanced understanding of how a model handles both true and false predictions.

Precision. Precision measures the proportion of true positive predictions out of all the positive predictions made by the model. It evaluates the accuracy of the model in identifying relevant instances among those predicted as positive.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positives} \quad (1)$$

High precision indicates that the model is specific and minimizes false positives [18].

Recall. Recall, also called sensitivity or the true positive rate, measures the proportion of true positives out of all actual positive instances. It assesses how well the model identifies positive examples.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatives} \quad (2)$$

High recall means that the model is sensitive and can identify most of the positive examples [18].

F1 Score. The F1 score is the harmonic mean of precision and recall. It is particularly useful in evaluating models on imbalanced datasets, as it balances the trade-offs between precision and recall.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

The F1 score provides a comprehensive view of a model's performance when dealing with imbalanced datasets. It is a more informative evaluation metric compared to the ROC plot in such cases, as it effectively highlights the model's trade-offs between precision and recall [18].

1.6 Literature Analysis

Prior to initiating the project, a thorough review of existing datasets relevant to gesture recognition was conducted. The literature references two types of datasets: one designed for sign language interpretation and the other for recognizing hand movements akin to gestures used in this project. A detailed analysis revealed that while sign language datasets are systematically studied, the datasets involving general hand gestures

lack comprehensive, standardized evaluations. Most studies in this category tend to report findings based on their bespoke datasets and models, which highlights a gap in universal dataset standards and methodologies.

Table 1. Sign Language Literature Analyzes

Sign Language Dataset	Dataset features	Model	Accuracy	Year	Ref
Turkish Sign Language (TSL) Alphabet	22,000 images	3D CNN with score level fusion	95%	2020	[19]
Indian Sign Language (ISL) Digits	5,000 images	k-NN and Naïve Bayes Classifier	98%	2021	[20]
American Sign Language (ASL) Alphabet	87,000 images	MobileNet V2	99%	2022	[21]
Assamese gestures	30,000 images	Mediapipe and sign classification with Deep Learning	99%	2023	[22]
American Sign Language (ASL) Alphabet	166,000 images	3D ConvNet and Mediapipe Hand Tracking Module	100%	2023	[23]

Table 2. Hand Gesture Literature Analyzes

Hand Gesture Dataset	Dataset features	Model	Accuracy	Year	Ref
Hand Gesture	20,100 images	VGG16	89%	2023	[2]
Hand Gesture	20,100 images	MobileNetV2	93%	2023	[2]
Hand Gesture	4,150 images	MobileNet	86.54%	2023	[24]
Hand Gesture	4,150 images	G-MobNet	%96.19	2023	[24]
HaGRID	554,800 images	MobileNetV3 Large	91.9%	2024	[25]
HaGRID	554,800 images	ResNeXt-50	98.3%	2024	[25]

2 METHODOLOGY

2.1 System Overview

The proposed system includes several key components: a camera module captures real-time video of hand gestures, an image processing unit optimizes image quality for recognition, and a gesture recognition module uses CNNs to accurately identify gestures. The control system converts these recognized gestures into commands for home automation devices, and a feedback mechanism provides immediate visual or auditory confirmation to the user.

2.2 Dataset Preparation:

The dataset is organized into 10 distinct folders representing different gesture classes. Each folder contains images that belong to a specific class. Separate folders are prepared for training and testing. There are 406,777 images for training and 10,000 images for testing.



Fig. 1 Left to right: Index, Peace, Three, Four, Palm

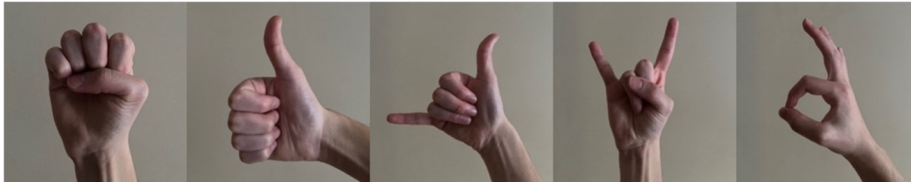


Fig. 2. Left to right: Fist, Like, Call, Rock, Okey

Data Augmentation. An ImageDataGenerator is initialized with various augmentation parameters, including rescaling pixel values to $[0, 1]$, randomly rotating images up to 20 degrees, shifting images horizontally and vertically by 20%, adding shearing and zooming effects, performing horizontal flipping, and using nearest-neighbor filling for new pixels generated.

Data Loading. The generator loads training and validation data separately from the main dataset directory. Training data is loaded with the training subset option, while validation data uses the validation subset. Both are resized to 224×224 pixels, categorized with one-hot encoding, and shuffled for randomized batches.

2.3 Model Architecture:

The model is compiled using the Adam optimizer and categorical cross-entropy loss for multi-class classification. A custom Learning Rate Finder dynamically adjusts the learning rate during training, starting at $1e-5$ and increasing to $1e-1$, with initial training running for two epochs to find the best rate. Once identified, the model is trained using this optimal rate.

Base models used include VGG16, MobileNetV2, MobileNetV3 Small, MobileNetV3 Large, and EfficientNetB7, all pre-trained on ImageNet and configured to accept 224x224 RGB input images. Initial layers are frozen to retain pre-trained weights, while the final four layers remain trainable.

The custom architecture, built using Sequential, includes GlobalAveragePooling2D for aggregating feature maps, a Dense layer with 512 units and ReLU activation, a Dropout layer to reduce overfitting, and a final Dense layer that outputs class probabilities with softmax activation. The ReduceLROnPlateau callback adjusts the learning rate based on validation loss to avoid overfitting and enhance convergence.

2.4 Model Test

After training, the model's accuracy and loss are plotted to evaluate performance. Testing data is then loaded using the same ImageDataGenerator settings. The test directory is structured similarly to the training dataset. The trained model is evaluated on the test set to measure its generalization capability across unseen data.

This methodology allows efficient training of the deep learning model on gesture recognition tasks, leveraging data augmentation, transfer learning, and learning rate optimization to build a robust classifier.

2.5 Mobile Algorithms for Smart Home Integration

Incorporating mobile-friendly algorithms in smart home systems offers significant advantages, primarily by reducing computational load. Models like MobileNet, designed for edge computing, provide high accuracy while maintaining a lightweight footprint, making them ideal for gesture recognition in smart homes.[14]

Mobile algorithms such as MobileNet and EfficientNet use depthwise separable convolutions to reduce computational demands without sacrificing performance. Optimized for edge devices, these models enable local processing on smart cameras or embedded systems, eliminating the need for powerful central servers and improving response times.[14,15]

These models are also power-efficient, which is crucial for battery-operated smart home sensors, reducing overall power consumption. The reduced computational load allows data processing across multiple devices simultaneously, supporting the scalability of gesture-based controls throughout a broader network of smart home devices.

On-device processing enhances privacy and security by ensuring that sensitive data does not leave the home network, thereby significantly improving data security and protecting user privacy.[10,14]

2.6 Model Development and Training

The development of the model began with the selection of suitable machine learning frameworks and libraries. TensorFlow and Keras were chosen for their extensive support and ease of use in implementing CNNs. Multiple models, including VGG16, MobileNetV2, and EfficientNetB7, were evaluated for their performance in handling real-time image data.

Training the models involved preparing a dataset of hand gesture images that has 407,000 images, which were augmented to enhance the diversity and volume of training data. This process helped in improving the model's ability to generalize from the training data to real-world scenarios. Initial training results have shown promising accuracy rates.

2.7 Training Results

Table 3. Performance Metrics of model that trained with VGG16 (Left) and EfficientNetB7 (Right)

Classification Report:									
	precision	recall	f1-score	support		precision	recall	f1-score	support
Call_Test	0.94	0.93	0.94	1000	Call_Test	0.89	0.95	0.92	1000
Fist_Test	0.90	0.95	0.93	1000	Fist_Test	0.97	0.92	0.94	1000
Four_Test	0.93	0.86	0.89	1000	Four_Test	0.97	0.88	0.92	1000
Index_Test	0.92	0.88	0.90	1000	Index_Test	0.84	0.96	0.90	1000
Like_Test	0.91	0.94	0.92	1000	Like_Test	0.92	0.93	0.93	1000
Ok_Test	0.99	0.84	0.91	1000	Ok_Test	0.98	0.79	0.88	1000
Palm_Test	0.94	0.94	0.94	1000	Palm_Test	0.95	0.95	0.95	1000
Peace_Test	0.73	0.94	0.82	1000	Peace_Test	0.85	0.94	0.89	1000
Rock_Test	0.94	0.88	0.91	1000	Rock_Test	0.97	0.95	0.96	1000
Three_Test	0.86	0.84	0.85	1000	Three_Test	0.86	0.91	0.88	1000
accuracy			0.90	10000	accuracy			0.92	10000
macro avg	0.91	0.90	0.90	10000	macro avg	0.92	0.92	0.92	10000
weighted avg	0.91	0.90	0.90	10000	weighted avg	0.92	0.92	0.92	10000

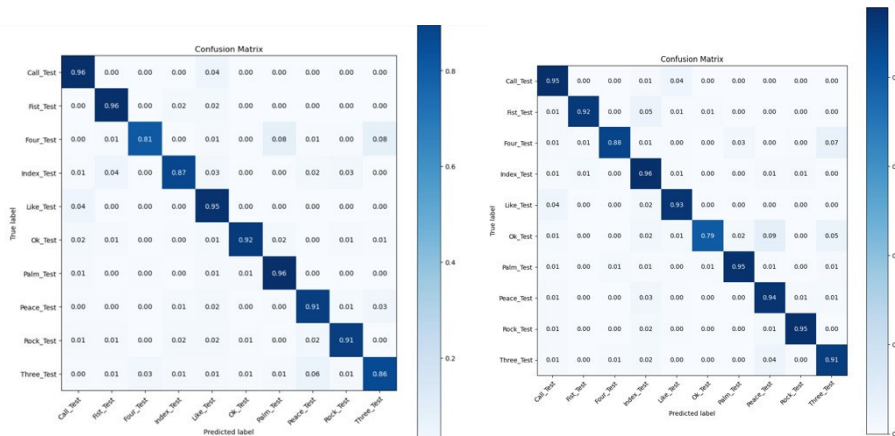


Fig.3. Confusion Matrix of model that trained with VGG16 (Left) and EfficientNetB7 (Right)

4. KEY FINDINGS

This research project has made significant contributions to human-computer interaction, smart home technologies, and gesture recognition systems. The findings demonstrate the feasibility and effectiveness of the proposed solutions and suggest future research and development directions.

One key outcome was developing a high-accuracy gesture recognition model using Convolutional Neural Networks (CNNs) and integrating MobileNet and EfficientNet architectures. This system accurately classifies various hand gestures in real-time, performing well in varying lighting and background conditions and improving recognition rates from initial benchmarks.

The use of lightweight, mobile-oriented architectures like MobileNetV2 and MobileNetV3 significantly enhanced computational efficiency. These models are optimized for edge devices, crucial for real-time smart home applications, allowing faster processing times and lower power consumption, making them suitable for widespread deployment in consumer-grade devices.

The project demonstrated that gesture-based interfaces could significantly enhance usability and accessibility in smart homes. Users with limited mobility or dexterity found the system particularly beneficial, highlighting the potential of gesture recognition technologies to make smart home systems more inclusive and user-friendly.

The research addressed environmental variability challenges, such as different lighting conditions and backgrounds. The enhanced models maintained high accuracy despite these variabilities, ensuring the system's reliability and functionality in diverse real-world settings, making it practical for everyday use.

Finally, the study found the system scalable and easily integrable with existing and future smart home technologies. Its modular design allows for easy adaptation and scaling for various applications, from simple residential setups to complex systems in commercial or healthcare settings.

5 CONCLUSION

This research significantly advances smart home technologies by integrating advanced machine learning algorithms into IoT devices. Through meticulous planning, design, implementation, and optimization, it provides critical insights into developing efficient, user-friendly smart home interfaces. It also highlights the transformative impact of machine learning and digital image processing on enhancing human-computer interaction within the IoT ecosystem.

Evolving hand gesture recognition technologies offer more natural and engaging user experiences in applications like interactive gaming, virtual reality, and assistive technologies. In smart homes, the developed gesture recognition system allows users to control their homes effortlessly without physical touch or voice commands. This touch-free control is particularly beneficial for individuals with limited mobility, enabling easier management of living spaces.

Incorporating AI and machine learning enhances hand gesture recognition accuracy and deepens the understanding of user intentions. This makes smart home systems more intelligent and responsive, significantly improving the user experience and potentially accelerating smart home technology adoption.

The gesture recognition system improves the accessibility and usability of smart home controls, offering an inclusive interface for individuals with varying physical abilities. The project sets a robust methodological foundation for future research, highlighting effective strategies for integrating deep learning techniques within IoT applications. Future iterations will focus on incorporating dynamic gesture recognition to enrich user interaction with the smart home system, requiring sophisticated computational strategies and real-time data processing. Enhancements in system robustness through hyperparameter tuning and additional pre-trained models are essential for improving performance and reliability.

In conclusion, this project demonstrates the practical implementation of a smart home control system using hand gesture recognition and sets a precedent for future innovations. By integrating deep learning technologies with IoT, it highlights the potential for creating smarter, more accessible living environments, making daily life more efficient and enjoyable for a broader user base. The project's success invites further exploration and development, promising significant advancements in smart home interactions and accessibility.

REFERENCES

1. M. R. Alam, M. B. I. Reaz and M. A. M. Ali, "A Review of Smart Homes—Past, Present, and Future," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1190-1203, Nov. 2012, doi: 10.1109/TSMCC.2012.2189204.
2. Oudah, M.; Al-Naji, A.; Chahl, J. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *J. Imaging* 2020, 6, 73. <https://doi.org/10.3390/jimaging6080073>
3. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. Curran Associates Inc., Red Hook, NY, USA, 1097–1105.
4. H. Ahuja, K. Varadarajan and M. Jammal, "Towards Smart Interaction: Hand Gesture Recognition Using Machine Learning in IoT Scenarios," 2023 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Dubai, United Arab Emirates, 2023, pp. 48-52, doi: 10.1109/GCAIoT61060.2023.10385100.
5. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
6. Sermanet, Pierre, Soumith Chintala, and Yann LeCun. "Convolutional neural networks applied to house numbers digit classification." *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE, 2012.
7. Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biological Cybernetics* 36 (1980): 193-202.

8. Nair, Vinod and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines." International Conference on Machine Learning (2010).
9. Srivastava, Nitish et al. "Dropout: a simple way to prevent neural networks from overfitting." J. Mach. Learn. Res. 15 (2014): 1929-1958.
10. Lane, ND; Georgiev, P; (2015) Can deep learning revolutionize mobile sensing? In: Manweiler, J and Choudhury, RR and Rozner, E and Lane, N, (eds.) Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications. (pp. pp. 117-122). Association for Computing Machinery: Santa Fe, New Mexico, USA.
11. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
12. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
13. Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).
14. Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
15. Howard, Andrew, et al. "Searching for mobilenetv3." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
16. Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.
17. Saito, Takaya and Marc Rehmsmeier. "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets." PLoS ONE 10(2015): n. pag.
18. Gökçe, Çağrı, et al. "Score-level multi cue fusion for sign language recognition." Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020.
19. Ashok Kumar Sahoo, "Indian Sign Language Recognition Using Machine Learning Techniques." Volume 397, Issue 1 Special Issue: National Conference on Functional Materials — NCFM 2020 June 2021 <https://doi.org/10.1002/masy.202000241>
20. Elsayed, Nelly, Zag ElSayed, and Anthony S. Maida. "Vision-Based American Sign Language Classification Approach via Deep Learning." arXiv preprint arXiv:2204.04235 (2022).
21. Jyotishman Bora, Saine Dehingia, Abhijit Boruah, Anuraag Anuj Chetia, Dikhit Gogoi, Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning, Procedia Computer Science, Volume 218, 2023, Pages 1384-1393, ISSN 1877 0509, <https://doi.org/10.1016/j.procs.2023.01.117>.
22. Kumar, Rupesh, Ashutosh Bajpai, and Ayush Sinha. "Mediapipe and cnns for real-time asl gesture recognition." arXiv preprint arXiv:2305.05296 (2023).
23. Agyemang, Isaac Osei, et al. "Gesture Control of Micro-drone: A Lightweight-Net with Domain Randomization and Trajectory Generators." arXiv preprint arXiv:2301.12470 (2023).
24. Kapitanov, Alexander, et al. "HaGRID--HAnd Gesture Recognition Image Dataset." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024.

Türkiye’de Tıp Bilişimi Alanında Üretilen Tezlerin Analizi

Hazal TAŞ¹ ve Neşe ZAYİM¹

¹ Akdeniz Üniversitesi Biyoistatistik ve Tıbbi Bilişim AD

Özet. Tıp Bilişimi dünyada 1950’li, Türkiye’de 1990’lı yıllardan itibaren çalışmaların yürütüldüğü ve gün geçtikçe popülerliği artan bir çalışma alanıdır. Tıp ve sağlık hizmetlerinde veri, enformasyon ve bilginin sistematik olarak işlenmesiyle ilgilenen bir disiplin olarak tanımlanabilen tıp bilişimi alanında gerek akademik gerekse dernekler bünyesinde pek çok çalışma yürütülmektedir. Akademik olarak Tıp Bilişimi, dünyada da ülkemizde de tek başına anıldığı gibi bazen de Biyoistatistik ve Tıp Eğitimi gibi başka birimlerle anılmaktadır. Bu çalışmada ülkemizdeki bu birimlerde Tıp Bilişimi ile ilgili yapılan tezlerin incelenmesi ve bu tezlerdeki konu eğilimlerinin çıkartılması amaçlanmıştır. Bu kapsamda Yükseköğretim Kurulu Ulusal Tez Merkezi bünyesinde taranan tezler Tıp Bilişimi, Biyoistatistik ve Tıp Bilişimi, Biyoistatistik ve Tıbbi Bilişim, Tıp Eğitimi ve Bilişim, Sağlık Bilişimi ve Sağlık Bilişim Sistemleri olmak üzere Anabilim Dallarında taranmış ve alanda yapılan tezlere erişilmiştir. Erişim kısıtı olmayan tüm tezler incelendikten sonra dışlama kriterlerine göre bazıları elenmiş ve nihai olarak 217 tez çalışmaya alınmıştır. Bu tezlerin anahtar kelimeleri çıkartılmış ve beş yıllık periyodlar halinde incelenmiştir. Sonuç olarak dünyadaki teknolojik gelişmelere paralel olarak ülkemizde Tıp Bilişimi alanındaki tez çalışma konularının da değiştiği görülmüştür. Ancak büyük veri analizi, doğal dil işleme, kişiselleştirilmiş tıp, dijital sağlık gibi alanlarda yapılan tezlerin sayıca azlığı dikkat çekmektedir.

Anahtar Kelimeler: Tıp Bilişimi, Tıp Bilişimi Eğitimi, Yüksek Öğrenim, Tez

Abstract. Medical Informatics is a field of study where studies have been carried out in the world since the 1950s and in Turkey since the 1990s and its popularity is increasing day by day. Defined as a discipline dealing with the systematic processing of data, information, and knowledge in medicine and healthcare services, Medical Informatics is the focus of numerous studies conducted both academically and within associations. Academically, Medical Informatics is sometimes referred to as a standalone discipline both globally and in our country, and at times it is also associated with other units such as Biostatistics and Medical Education. This study aims to examine the theses conducted in units related to Medical Informatics in our country and to extract the topic trends in these theses. Within this scope, the theses scanned in the National Thesis Center of the Council of Higher Education were searched in Departments such as Medical Informatics, Biostatistics and Medical Informatics, Medical Education and Informatics, Health Informatics and Health Informatics Systems. After examining all the theses with no access restriction, some were eliminated based on exclusion criteria, and ultimately, 217 theses were included in the study. The keywords of these theses were extracted, and they were examined in five-year periods. As a result, it has been observed that the topics of thesis studies in the field of Medical Informatics in our country have changed parallel to technological developments worldwide. However, it is noteworthy that the number of theses in areas such as big data analysis, natural language processing, personalized medicine and digital health is low in number.

Key Words: Medical Informatics, Education of Medical Informatics, Higher Education, Thesis

1 Giriş

Haux, tıp bilişimini “tıp ve sağlık hizmetlerinde veri, enformasyon ve bilginin sistematik olarak işlenmesiyle ilgilenen disiplin” olarak tanımlamıştır [1]. Tıp bilişimi bilgisayar bilimi, bilgi bilimi, bilişsel bilimler ve sosyal bilimler dahil olmak üzere birçok disiplinden yararlanan çok disiplinli bir alandır [2]. Her ne kadar Morris F Collen tıp bilişimini "bilgisayar teknolojisinin tıbbın tüm alanlarına uygulanması" olarak tanımlamış olsa da [3] Greenes ve Shortliffe tıp bilişiminin bilgisayarların tıpta uygulanmasından daha fazlasını içerdiğini ileri sürmüşlerdir [4].

Tıp bilişimi ile ilgili ilk makale 1959 yılında Ledley ve Lusted tarafından yayınlanmıştır. Hekimlerin daha kesin tanı koymalarına ve tedaviyi bilimsel bir şekilde belirlemelerine yardımcı olmak için tıbbi karar vermenin modellenmesine yönelik hesaplama tekniklerini ortaya koymuşlardır [5]. 1950’li yıllarda ilk uygulamaları ve makaleleri yayımlandıktan sonra, tıp bilişimi ile ilgili yayınların sayısı her otuz yılda hızla artmıştır [6]. 1970’li yıllarda tıp bilişimi yeni bir uzmanlık alanı olarak tanımlanmıştır [7].

ABD’de bir tıp fakültesinde Tıbbi Bilişim bölümü, ilk olarak Utah Üniversitesi’nde 1964 yılında “Department of Biomedical Informatics” adıyla kurulmuştur [8]. ABD’nin tanınmış üniversiteleri olan Harvard Üniversitesi Tıp Fakültesi [9] ve Stanford Üniversitesi Tıp Fakültelerinde [10] “Biomedical Informatics” bilim dalları bulunmaktadır. Washington Üniversitesi’nde olduğu gibi, Tıp Eğitimi Anabilim dalı ile birlikte (Department of Biomedical Informatics and Medical Education), Wisconsin Madison Üniversitesi’nde olduğu gibi Biyoistatistik Anabilim dalı ile birlikte (Department of Biostatistics and Medical Informatics) kurulmuş Anabilim dalları da

vardır. Tıp bilişimi eğitimi ABD’de genellikle mezuniyet sonrası yüksek lisans veya doktora programı içerisinde verilmekle birlikte son zamanlarda mezuniyet öncesi eğitim olarak da verilmektedir. Avrupa’da 1970’li yılların başından itibaren çeşitli tıp fakültelerinde Tıp bilişimi ile ilgili bölümler kurulmuştur.

Türkiye’de de dünyada olduğu gibi pek çok üniversitede Tıp Bilişimi bölümü yapılanması, Biyoistatistik anabilim dalları ile başlamış, daha sonra ise 2005-2006 yıllarında Akdeniz Üniversitesi ve Ege Üniversitesinde Biyoistatistik anabilim dallarının adı “Biyoistatistik ve Tıp Bilişimi” olarak değiştirilmiştir. YÖK, 22.12.2008 tarihli kararı ile bunlardan başka 11 üniversitede tıp fakültelerinde Biyoistatistik Anabilim Dallarının adının “Biyoistatistik ve Tıp Bilişimi” olarak değiştirilmesini istemiştir. Bu kararla aynı zamanda tıp fakültesi bünyelerinde yer alan "tıp eğitimi anabilim dallarının" kapatılarak, Sağlık Bilimleri Enstitüsü bünyesinde "tıp eğitimi anabilim dalı" açılmasına, tıp eğitimi ve bilişimi anabilim dalının birlikte kurulduğu tıp fakültelerinden biyoistatistik anabilim dalı bulunanlarda "biyoistatistik ve tıp bilişimi anabilim dallarının" birleştirilmesine karar verilmiştir. Danıştayın 25.5.2019 tarihinde bu kararın yürütmesini durdurmasıyla birlikte Tıp Fakültelerinde “biyoistatistik ve tıp bilişimi” veya “tıp eğitimi ve tıp bilişimi” anabilim dalları kurulması konusunda bir kısıtlama yapılmamıştır [11]. Mevcut durumda, Türkiye’deki Tıp/Sağlık Bilişimi Anabilim Dallarının yapılanması Tablo 1’de verilmiştir [12].

Tablo 1. Tıp/Sağlık Bilişimi Anabilim Dalları

Fakülte	Anabilim Dalı	Anabilim Dalı Sayısı
Tıp Fakültesi/Temel Tıp	Biyoistatistik ve Tıp Bilişimi/Tıbbi Bilişim	49
Tıp Fakültesi/Temel Tıp	Tıp Bilişimi Anabilim Dalı	3
Tıp Fakültesi/Temel Tıp	Tıp Eğitimi ve Bilişimi	
Sağlık Bilimleri Fakültesi/Sağlık Yönetimi	Sağlık Bilişimi ve Teknolojileri	8
Sağlık Bilimleri Fakültesi/Sağlık Yönetimi Bölümü	Sağlık Bilişimi	3
İktisadi ve İdari Bilimler Fakültesi/Sağlık Yönetimi	Sağlık Bilişimi ve Teknolojileri	3
İktisadi ve İdari Bilimler Fakültesi/Sağlık Yönetimi Bölümü	Sağlık Bilişimi	1

Ülkemizde Tıp Bilişimi alanında lisans programı bulunmamakta, eğitim lisansüstü düzeyde yürütülmektedir. 1990’da Akdeniz Üniversitesinde “Tıbbi Bilgisayar Uygulamaları” adı altında ilk yüksek lisans programı açılmış, sonraki yıllarda bu programın ismi Tıp Bilişimi olarak değiştirilmiş, bunu 2003 yılında ODTÜ’de yüksek lisans ve doktora programlarının açılması izlemiştir. Daha sonraki yıllarda Akdeniz Üniversitesinde (2006), ve Karadeniz Teknik Üniversitesinde (2015) doktora programı açılmıştır, günümüzde ise birçok üniversitede de lisansüstü eğitim programları açılmıştır. YÖK, Yükseköğretim İstatistikleri’ ne göre 2012-2013 /2022-2023 öğrenim yılları arasında Tıp/Sağlık Bilişimi alanında lisansüstü mezun öğrenci sayıları Tablo 2’de verilmektedir [13]. Tabloya göre bu yıllar arasında tezli Yüksek lisans (164) ve doktora (90) programından olmak üzere toplam 254 öğrenci mezun olmuştur.

Tablo 2. Tıp/Sağlık Bilişimi Tezli Lisansüstü Programlardan Mezun Öğrenci Sayıları

DÖNEM	PROGRAM ADI	YÜKSEK LİSANS	DOKTORA	TOPLAM
2022-2023	BİYOİSTATİSTİK VE TIP BİLİŞİMİ	15	12	27
	TIP BİLİŞİMİ/SAĞLIK BİLİŞİMİ	13	12	25
2021-2022	BİYOİSTATİSTİK VE TIP BİLİŞİMİ	16	7	23
	TIP BİLİŞİMİ /SAĞLIK BİLİŞİMİ	7	6	13
2020-2021	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM	7	7	14
	TIP BİLİŞİMİ/BİYOSAĞLIK BİLİŞİMİ	2	4	6
2019-2020	TIP BİLİŞİMİ/SAĞLIK BİLİŞİMİ		2	2
	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM	10	6	16

2018-2019	TIP BİLİŞİMİ	3	2	5
	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM	26		26
2017-2018	TIP BİLİŞİMİ	2	2	4
	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM	9	5	14
2016-2017	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM /BİYOİSTATİSTİK VE TIP BİLİŞİM/BİYOİSTATİSTİK VE TIP BİLİŞİMİ	16	2	18
	TIP BİLİŞİMİ /TIBBİ BİLİŞİM	6	3	9
2015-2016	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM /BİYOİSTATİSTİK VE TIP BİLİŞİM/BİYOİSTATİSTİK VE TIP BİLİŞİMİ	7	2	9
	TIP BİLİŞİMİ	2	4	6
2014-2015	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM/BİYOİSTATİSTİK VE TIP BİLİŞİM	2	1	3
	TIP BİLİŞİMİ	5	1	6
2013-2014	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM	5	1	6
	TIBBİ BİLİŞİM/TIP BİLİŞİMİ	8	4	12
2012-2013	BİYOİSTATİSTİK VE TIBBİ BİLİŞİM /BİYOİSTATİSTİK VE TIP BİLİŞİM	3	4	7
	TIP BİLİŞİMİ		3	3
	TOPLAM	164	90	254

"Sağlık Bilişimi" temel alanı, Üniversitelerarası Kurul'un Sağlık Bilimleri Temel Alanı Doçentlik başvuruları için Ekim 2019 döneminden itibaren bilim dalları listesinde, "Sağlıkla İlgili Çok Disiplinli Bilimler" başlığı altında ilk defa kabul edilmiştir [14]. Bu listede "Biyostatistik" ve "Tıp Eğitimi" ayrı temel alan olarak verilirken, "Biyostatistik ve Tıp Bilişimi", "Tıp eğitimi ve Tıp Bilişimi" adı verilmiş doçentlik başvuru alanı yer almamaktadır.

2 Yöntem

Bu çalışmada Türkiye'deki çeşitli üniversite ve anabilim dallarında yapılmış Tıp Bilişimi alanında yüksek lisans ve doktora tezleri incelenmiştir. Ulusal Tez Merkezinde [15] tez taramada lisansüstü program adı seçeneği bulunmaması nedeniyle taranan tezler için detaylı tarama seçeneği kullanılarak Ana Bilim Dalı tarama seçeneğinden Tıp Bilişimi, Sağlık Bilişimi, Biyoistatistik ve Tıp Bilişimi, Biyoistatistik ve Tıbbi Bilişim, Sağlık Bilişim Sistemleri Ana Bilim Dallarına olarak taranmıştır. Tarama sonucunda ilki 1993 yılına ait olmak üzere günümüze kadar yapılan 259 teze ulaşılmıştır (Tablo 3).

Çalışmadan dışlanma kriterleri:

- Biyoistatistik ve Tıp Bilişimi/ Biyoistatistik ve Tıbbi Bilişim Anabilim dallarında Biyoistatistik alanında lisansüstü programlar yer alabileceğinden bu bölümlerde yer alan programlar ilgili program web sitesinden program adı ve Türkiye Yeterlikler Çerçevesi veri tabanı taranarak diploma program adı "Biyoistatistik ve Tıp Bilişim/Bilişimi Yüksek Lisans/Doktora" programı veya diploma programı olmayan bölümlere ait tezler,
 - Biyoistatistik ve Tıp Bilişim/Bilişimi Yüksek Lisans/Doktora programı altında olup konuları tıp/sağlık bilişimi kapsamında olmayan tezler,
- Aynı Anabilim Dalı altında Sağlık/Tıp Bilişimi ve biyoinformatik lisansüstü programı bulunan (O.D.T. Ü.) anabilim dallarına ait, tıp/sağlık bilişimi konuları kapsamında bulunmayan tezler,
 - Anahtar kelimesi olmayanlar ve erişim engeli bulunan tezler çalışmaya dahil edilmemiştir.

Ulaşılan tez ve çalışmaya dahil edilen tezlerin bölümlere göre dağılımı aşağıdaki tablo 3'te sunulmuştur. Çalışmaya dahil edilen 217 tezin 145'i yüksek lisans; 72'si doktora tezidir.

Tablo 3. Ulaşılan tez ve çalışmaya dahil edilen tezlerin lisansüstü programlara göre dağılımı

Program Adı	Erişilen Bütün Tezler			Çalışmaya Dahil Edilen Tezler		
	Yüksek Lisans	Doktora	Toplam	Yüksek Lisans	Doktora	Toplam
Sağlık Bilişimi	4	0	4	4	0	4
Tıp Bilişimi	22	15	37	22	15	37
Tıp Bilişimi	36	37	73	35	34	69
Tıbbi Bilişim YL	9	0	9	9	0	9
Sağlık Bilişimi	6	0	6	6	0	6
Tıp Bilişimi	23	11	34	23	11	34
Tıbbi Bilgisayar Uygulamaları (Medical Informatics)	15	0	15	13	0	13
Sağlık Bilişimi YL	1	0	1	1	0	1
Biyoistatistik ve Tıbbi Bilişim YL	16	1	17	5	0	5
Biyoistatistik ve Tıp Bilişimi YL	17	0	17	2	0	2
Biyoistatistik ve Tıp Bilişimi	14	4	18	14	4	18
Biyoistatistik ve Tıp Bilişimi	15	11	26	9	8	17
Biyoistatistik ve Tıp Bilişimi	2	0	2	2	0	2
TOPLAM	180	79	259	145	72	217

Şekil 1’de yıllara göre dağılımı verilen çalışmaya dahil edilen bu tezlerin anahtar kelimeleri çıkartılmış ve analiz edilmiştir. Beş yıllık periyodlar halinde analiz edilen anahtar kelimeler ile Türkiye’de Tıp Bilişimi alanında yapılan tezlerin konu eğilimleri tespit edilmeye çalışılmıştır.



Şekil 1. Tez sayılarının yıllara göre dağılımı

3 Bulgular ve Sonuç

Ulusal Tez Merkezinde taranan tezlerden dışlama kriterleri uygulandıktan sonra kalan tezlerin anahtar kelimeleri çıkartılmış ve incelenmiştir. Yıllara göre tez dağılımlarına bakıldığında 1993-2003 yılları arasında az sayıda tez olduğu görülmektedir. Bu sebeple ilk on yıl bir periyod, diğer tezler ise beşer yıllık periyodlar halinde değerlendirilmiştir.

1993-2003 yılları arasında yapılan tezler incelendiğinde hastane bilişim sistemleri, veritabanları en çok çalışılan konulardan olmuştur. Bu yıllarda yapılan tezlerin Akdeniz Üniversitesinde yapıldığı bilinmektedir. Tez başlıkları da incelendiğinde Akdeniz Üniversitesi Hastanesi bilgi sistemleriyle ilgili tezlerin sayıca fazlalığı dikkat çekmektedir.

Tablo 4. 1993-2003 yılları arasında yapılan tezlerde sık kullanılan anahtar kelimeler

kelime	sıklık
sistem	5
veritabanı	4
hastane	3
bilişim	3
analiz	3

2004-2008 yılları arasında yapılan tezler incelendiğinde hastalar, hasta yakınları ve hekimler için web tabanlı eğitim portalları ile yine sağlık bilgi sistemlerinin sıklıkla çalışıldığı görülmüştür. Tablo 5’de bu yıllar içerisinde yapılan tezlerde en sık kullanılan anahtar kelimeler sunulmuştur.

Tablo 5. 2004-2008 yılları arasında yapılan tezlerde sık kullanılan anahtar kelimeler

kelime	sıklık
eğitim	6
portal	3
sistem	2
ilaç	2
sağlık	2

2009-2013 yıllarında ise bir önceki periyotta sık karşılaştığımız eğitim ve web kavramlarının yerini aşağıdaki tabloda verilen kelimeler almıştır (Tablo 6). Bu periyotta sağlık bilgi sistemleri, karar destek sistemleri ve veri madenciliğinin çalışma konusu olarak öne çıktığı görülmektedir.

Tablo 6. 2009-2013 yılları arasında yapılan tezlerde sık kullanılan anahtar kelimeler

kelime	sıklık
sistem	11
bilgi	8
Karar, destek	8
sağlık	7
veri	5
madenciliği	4

2014-2018 yıllarında belirgin bir şekilde destek vektör makineleri, karar destek sistemleri, yapay sinir ağları, makine öğrenmesi gibi konuların öne çıktığı görülmektedir. Bunun yanı sıra bu beş yıllık periyotta veri madenciliği ve genetik alanı da çalışma konusu olarak tercih edilmiştir.

Tablo 7. 2014-2018 yılları arasında yapılan tezlerde sık kullanılan anahtar kelimeler

kelime	sıklık
makine	13
analiz	11
destek	11
veri	9
vektör	8
ağları	7
gen	6
madenciliği	6
yapay	6
karar	6

2019 yılından günümüze kadar yapılan tezler incelendiğinde makine öğrenmesinin halen en güncel çalışma alanı olduğu görülmekle birlikte derin öğrenme de popüler çalışma konuları arasına girmiştir. Yapay zeka kavramının anahtar kelime olarak kullanıldığı tezlerin hemen hemen hepsinin makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak yapılan tezler olması dikkat çekmektedir. İçinde bulunduğumuz bu son periyotta, yapay sinir ağları, bilgisayar destekli tanı sistemi, ilaç-ilaç, ilaç-protein etkileşimi, genetik konuları da öne çıkan çalışma alanlarından olmuştur.

Tablo 8. 2019 yılından günümüze yapılan tezlerde sık kullanılan anahtar kelimeler

kelime	sıklık
öğrenme	50
makine	25
veri	18
ağ	18
sistemi	17
analiz	14
derin	12
yapay	12
karar	11
sinir	11
bilgi	11

Sonuç olarak Türkiye’de Tıp Bilişimi alanında yapılan tezler beş yıllık periyotlar halinde incelendiğinde tez çalışma konularının dünyada yaşanan teknolojik gelişmelere ve konu trendlerine paralel olarak değiştiği görülmektedir. Ancak günümüzde dünyada son zamanlarda popüler olan [16] büyük veri analizi, blockchain, doğal dil işleme, kişiselleştirilmiş tıp, dijital sağlık gibi alanlarda yapılan tezlerin sayıca azlığı dikkat çekmektedir. Bunun nedeninin bu alana ayrılmış proje bütçelerinin yetersizliği ve veri erişimindeki kısıtlar olduğu düşünülmektedir.

Kaynaklar

1. Haux, R.: Aims and tasks of medical informatics, International journal of medical informatics, cilt 44, no. 1, pp. 9-20, 1997.
2. Patel, V. L., Arocha, J. F. ve Kaufman, D. R.: A primer on aspects of cognition for medical informatics,» Journal of the American Medical Informatics Association, cilt 8, no. 4, pp. 324-343, 2001.
3. Masic, I.: The history and new trends of medical informatics, Donald School J Ultrasound Obstet Gynecol, cilt 7, no. 3, p. 72-83, 2013.
4. Greenes, R. A. ve Shortliffe, E. H.: Medical informatics: an emerging academic discipline and institutional priority, Jama, cilt 263, no. 8, pp. 1114-1120, 1990.
5. Ledley, R. S. ve Lusted, L. B.: Reasoning foundations of medical diagnosis, Science, cilt 130, no. 3366, pp. 9-21, 1959.
6. Collen, M. F. ve Ball, M. J.: The history of medical informatics in the United States, Springer, 2015.
7. Collen, M. F.: Origins of medical informatics, Western Journal of Medicine, cilt 145, no. 6, p. 778, 1986.
8. <https://medicine.utah.edu/dbmi>, son erişim 10.05.2024
9. <https://dbmi.hms.harvard.edu>, son erişim 10.05.2024
10. <http://med.stanford.edu/bmi.html>, son erişim 10.05.2024
11. <https://tip.baskent.edu.tr/kw/genel.php?birim=600&id=140197>, son erişim 10.05.2024
12. Yüksek Öğrenim Kurumu. 2023-2024 Öğretim Yılı Yükseköğretim İstatistikleri. <https://istatistik.yok.gov.tr/>
13. Yüksek Öğrenim Bilgi Yönetim Sistemi, Yüksek Öğrenim İstatistikleri, Öğrenim Düzeyleri ve Birimlere Göre Mezun Sayıları. <https://istatistik.yok.gov.tr/>
14. Üniversitelerarası Kurul Başkanlığı, 2019 EKİM DÖNEMİ DOÇENTLİK BAŞVURU ŞARTLARI <https://www.uak.gov.tr/Sayfalar/docentlik/basvuru-sartlari/2019-ekim-donemi-docentlik-basvuru-sartlari.aspx>
15. <https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp>, son erişim 10.05.2024
16. Damar, M., Yüksel, İ., Çetinkol, A.E. et al. Advancements and integration: a comprehensive review of health informatics and its diverse subdomains with a focus on technological trends. Health Technol., 2024. <https://doi.org/10.1007/s12553-024-00872-5>

The Inclusion of Medical Informatics Topics in the Curricula of Medical Faculties in Turkey: A Document Analysis Study

Ayşenur Duman Dilbaz¹, Yasemin Güner¹, Melek Şahin², Barış Sezer³

¹ Karadeniz Technical University Faculty of Medicine, Department of Medical Education

² Karadeniz Technical University Faculty of Medicine

³ Hacettepe University Faculty of Medicine, Department of Medical Education and Health Informatics

Introduction

The role of technology in the health sciences is becoming increasingly significant. Information technologies are being used in various areas, including patient care, health record management, diagnosis and treatment processes. These technologies are becoming an important tool for improving the quality of healthcare [1]. The advent of the global pandemic has led to a significant increase in the adoption of tele-medicine, e-health applications, wearable health technologies and patient tracking devices in health service delivery [2]. The use of e-prescription platforms, hospital information management systems, laboratory and imaging information management systems are now an integral part of the daily practice of health professionals [1, 3].

Although numerous studies have demonstrated that the utilisation of health informatics has a beneficial impact on patient outcomes and patient safety, increases satisfaction and improves efficiency by facilitating daily procedures, it persists as a challenge for physicians [3, 4]. It is becoming increasingly crucial that medical education curricula adapt to these technological developments and include health informatics topics. Considering this necessity, the national core education programme of our country (NCCP 2020) states that graduates of medical faculties should be able to utilise information technologies to enhance the efficacy of their studies in the fields of health service, research and education. Furthermore, it emphasises that the interaction of technology and health is an important topic that should be included in the curriculum [5]. The International Medical Informatics Association (IMIA) provides and publishes recommendations to support educators in the development of curricula in biomedical and health sciences. These recommendations include the essential knowledge domains, skills and competencies required for medical students at different levels and fields (6).

Our literature review revealed a lack of data on the extent to which medical informatics topics are included in the curriculum of medical faculties in Turkey. The objective of this study was to examine the current situation regarding the inclusion of medical informatics in medical education curricula.

Methods:

A document analysis study was conducted.

The list of medical education programmes in Turkey was accessed on 1 May 2024 via the website of the Higher Education Institution. The contents of the programmes in this list were accessed through the course programmes and ECTS (European Credit Accumulation and Transfer System) information packages published publicly on the faculties' websites. It was noted that six of the 144 faculties listed here offer joint education with other faculties, with students from different faculties sharing a single education programme. In order to ensure the accuracy of the data, these faculties were excluded from the analyses. The present research comprised a total of 138 education programmes.

Analyses were conducted on the information provided in the information packages and course programmes. Information about the departments in the faculties was also obtained through their own websites.

The quantitative data obtained was analysed using MS Excel software. The researchers determined which courses were compatible with the medical informatics context within the faculty programmes by qualitatively analysing the course titles, accessible course contents and learning objectives. The researchers conducted the analyses based on the medical informatics topics recommended by the Medical Informatics Association (IMIA) for inclusion in biomedical and health sciences curricula (6). Once the researchers had completed their individual document analyses, they reported the results by coming together as many times as needed to reach a consensus.

Results:

A total of 138 training programmes were examined, of which 22 were in English and 116 were in Turkish. Two programmes could not be accessed due to unavailability of content or course titles. Only two years of course programmes from a newly established university, which has been teaching for two years, could be accessed.

Upon examination of the departments within faculties, it was observed that 64 faculties (%45) lacked a department of medical informatics, either in isolation or within another department.

The field of medical informatics was situated within the departments of biostatistics and medical informatics in 43 faculties (%31), while in 24 faculties (%17) it was located within the departments of medical education and medical informatics. In 7 faculties (%5), both departments were present, while in 3 faculties (%2), the department of medical informatics existed in isolation.

A total of 62 programmes (%45) included medical informatics-related courses. While 31 of the programmes included compulsory courses in medical informatics, 31 of them included only elective courses. 26 (%19) of the programmes included a compulsory course only in the preclinical phase. Only 2 programmes (%1) included a compulsory course at the fourth grade.

Discussion

Although 55% of faculties have departments that include medical informatics, only 26% of programmes have medical informatics courses embedded in the main curriculum. It was observed that the majority of such courses occur in the first year, yet the inclusion of medical informatics courses in the clinical phase is insufficient.

A study conducted in 2012 examined the coverage of health informatics topics in health and medical informatics programmes in Turkey (7). It was observed that most of the competencies recommended by IMIA were not included in the programmes. The study found that the lack of trainers and the intensity of the curricula were effective factors in this situation (7). We found that medical informatics-related courses were not included in a significant part of the educational programmes. Furthermore, we revealed that 45% of faculties lack a department of medical informatics. A systematic review conducted by Damar et al. in 2024 found that our country ranked 27th among 30 countries in terms of conducting research in the field of medical informatics (8). This situation indicates that the number of qualified people working in the field is low, similar to the situation in 2012. To equip future physicians with the necessary competencies to thrive in a digitally transforming healthcare landscape, there is an urgent need to expand the pool of professionals trained in medical informatics. This can be achieved by developing robust and comprehensive medical informatics training programs and fostering career development opportunities for graduates of such programs.

It is the responsibility of medical faculties to ensure that the physicians they train possess the requisite qualifications to serve the society. The ability to utilise information technologies, which is a necessity of our age, has become a compulsory competence to be gained by medical students in the context of healthcare. In Turkey, there is a clear need for a health informatics curriculum that extends to the clinical years, in order to ensure that graduates are able to comprehend the relationship between information technologies and the clinic, as well as the impact of these technologies on patient outcomes.

Limitations

In our study, we were able to obtain as much information about the faculties' educational programmes as was available on their websites and information packs. However, we were unable to reach information about the depth in which health informatics topics are covered in the curriculum. Topics in health informatics that are not clearly stated in the course principles but are included in the course may have been overlooked. Since the course titles, contents and learning objectives were analysed qualitatively by the researchers, subjective evaluations emerged, which are within the limitations of our research.

References

1. Patrício, L., Sangiorgi, D., Mahr, D., Čaić, M., Kalantari, S., & Sundar, S. Leveraging service design for healthcare transformation: Toward people-centered, integrated, and technology-enabled healthcare systems. *Journal of Service Management*, 31(5), 889-909 (2020).
2. Shah, S., Diwan, S., Kohan, L., Rosenblum, D., Gharibo, C., Soin, A., ... & Provenzano, D. A. The technological impact of COVID-19 on the future of education and health care delivery. *Pain physician*, S367-S380 (2020).
3. Masters, K., Ellaway, R. H., Topps, D., Archibald, D., & Hogue, R. J. Mobile technologies in medical education: AMEE Guide No. 105. *Medical teacher*, 38(6), 537-549 (2016).
4. Knaup, P. Necessity and potential of educating medical students, physicians and other health care professionals in medical informatics. *Medical teacher*, 21(1), 73-76 (1999).
5. Ulusal Cep-2020 UCG, Ulusal Cep-2020 UYVYCG, Ulusal Cep-2020 DSBBCG. *Medical Faculty - National Core Curriculum 2020*. TED. 19(57 - 1):1-146 (2020).
6. Bichel-Findlay, J., Koch, S., Mantas, J., Abdul, S. S., Al-Shorbaji, N., Ammenwerth, E., ... & Wright, G. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics: second revision. *International Journal of Medical Informatics*, 170, 104908 (2023).
7. N. ZAYİM, "Mezuniyet Öncesi Sağlık/Tıp Eğitiminde Sağlık Bilişimi Eğitimi," IX. Ulusal Tıp Bilişimi Kongresi , Antalya, Turkey, pp.39 (2012) . <https://turkmia.net/kongre2012/cd/bildiriler2012.htm>. Last accessed 24/05/2024.
8. Damar, M., Küme, T., Yüksel, İ., Çetinkol, A. E., Pal, J. K., & Erenay, F. S. Medical Informatics as a Concept and Field-Based Medical Informatics Research: The Case of Turkey. *Duzce Medical Journal*, 26(1), 44-55 (2024).
9. Roberts, T. Tomorrow's curriculum for tomorrow's doctors. *Medical Teacher*, 33(7), 517-517 (2011).

An Ethical Examination of ChatGPT Use in the Field of Medicine

Belkıs Birden¹[0000-0002-6144-4292]

¹ Başkent University Medical Faculty The Department of Medical Ethics and History,
Ankara, TR06790 Türkiye
blksbrdn@gmail.com

Abstract

Chat Generative Pre-Trained Transformer, or ChatGPT, is a language model developed by OpenAI and released on 30 November 2022. The use of the system, which is a chatbot and uses artificial intelligence (AI) technologies, is critical in the field of medicine as well as other sectors. In the studies, it is stated that ChatGPT contributes to people making informed decisions about their health and in the field of medicine for healthcare professionals identifying and designing scientific research topics, determining clinical and laboratory diagnostic methods and predicting disease outcomes. There are also studies showing the use of ChatGPT in various fields related to medical education. The main purpose of this study is to examine the ethical problem areas that occur or may occur without ignoring the use of ChatGPT in various fields of medicine. It is necessary to examine the benefits, risks, and ethical issues arising from using ChatGPT, which has become a world agenda in a short time, in clinical applications, academic research and education processes in medicine. In this context, in this study, solutions to ethical problems and answers to how the correct use can be examined. When it comes to artificial intelligence applications such as ChatGPT, it is emphasized that researchers and health professionals in medicine should act according to professional ethical rules.

Keywords: ChatGPT, Medicine, Artificial Intelligence (AI), Ethics, Ethical Issues

Tıp Alanında ChatGPT Kullanımının Etik Açısından İncelenmesi

Özet

Chat Generative Pre-Trained Transformer veya ChatGPT, OpenAI tarafından geliştirilen ve 30 Kasım 2022 tarihinde piyasaya sürülen bir dil modelidir. Bir chatbot olan ve yapay zekâ (YZ) teknolojilerini içeren sistemin, diğer sektörlerin yanı sıra tıp alanında kullanımı kritik önem taşımaktadır. Yapılan çalışmalarda ChatGPT'nin, insanların sağlıkları hakkında bilinçli karar vermelerine, tıp alanında bilimsel araştırma konularının belirlenmesi ve tasarlanması, klinik ve

laboratuvar tanı yöntemlerinin saptanması ve hastalık sonuçlarının öngörülmesi gibi konularda sağlık profesyonellerine katkı sunduğu ifade edilmektedir. ChatGPT'nin tıp eğitimiyle bağlantılı çeşitli alanlarda kullanımını gösteren çalışmalar da mevcuttur. Bu çalışmanın temel amacı ise, ChatGPT'nin tıbbın çeşitli alanlarında kullanımını göz ardı etmeden oluşan veya oluşabilecek etik sorun alanlarını incelemektir. Kısa sürede dünya gündemine oturan ChatGPT'nin tıpta klinik uygulamalarda, akademik araştırmalarda ve eğitim süreçlerinde kullanımından doğan faydaların, risklerin ve etik sorunların incelenmesinin gerekli olduğu düşünülmektedir. Bu bağlamda bu çalışmada etik sorunlara yönelik çözüm önerileri ve doğru kullanımın nasıl olabileceğine dair yanıtlar incelenmiştir. ChatGPT gibi yapay zekâ uygulamaları söz konusu olduğunda tıp alanındaki araştırmacıların ve sağlık profesyonellerinin mesleki etik kurallara göre hareket etmelerinin gereği vurgulanmaktadır.

Anahtar Kelimeler: ChatGPT, Tıp, Yapay Zekâ (YZ), Etik, Etik Sorunlar

Giriş

Son yıllarda yapay zekâ (YZ) alanı oldukça ilerlemiş, özellikle ChatGPT gibi üretici dil modelleri tıp da dâhil olmak üzere bütün sektörleri temelden etkilemiştir (Wong ve ark. 2023). ChatGPT resmi olarak Şubat 2023'te piyasaya sürülmüş olmakla birlikte, bu teknolojiye ilişkin alanyazın yeni uygulamalarla hızla gelişmektedir (Roumeliotis ve Tselikas 2023). ChatGPT, tıp ve sağlık bilimlerinde tanı, hastalık yönetimi, tedavi rejimleri, akademik değerlendirme ve yazım desteği gibi geniş bir alanda uygulama potansiyeli sunmaktadır (Sumbal ve ark. 2024; Salvagno 2023). ChatGPT'nin öğrenmeyi ve fikir üretimini arttırdığı, bunun yanında üretilen bilginin doğruluğu ve gerçekliği konusunda ChatGPT'ye aşırı güvenin, bazı sorun alanlarının doğmasına neden olabileceği yapılan çalışmalarda vurgulanmaktadır (Wong ve ark. 2023; Malik ve ark. 2023).

Sağlık hizmetlerinin sunumunda ve bilimsel araştırma sürecinde yaşanan dijital dönüşüm, ChatGPT kullanımıyla birlikte, gizlilik, güvenlik ve şeffaflık gibi bir takım etik konuları ön plana çıkarmıştır (Altamimi ve ark. 2023). Bu durum bilimsel araştırmalarda onam ve mahremiyet kavramlarının işleyişini karmaşık bir hale getirmiş, bazı durumlarda araştırma etiğine ilişkin temel ilkelerin yürütülememesi neticesinde belirli etik zorlukları ortaya çıkarmıştır (Samuel ve ark. 2021).

Alanyazında ChatGPT'nin tıp alanında kullanımı konusunu etik bağlamda ele alan çalışmaların sınırlı olduğu görülmektedir. Bu bağlamda bu çalışmada, ChatGPT'nin tıp alanında

kullanılması sonucu ortaya çıkan etik sorunlar ve sorun alanlarına yönelik çözüm yolları alanyazın incelemesine dayanarak geleneksel bir derleme şeklinde ele alınmaktadır.

Etik Alan

Ahlaki değerler felsefesi şeklinde tanımlanan etik, felsefenin bir alt disiplini. İyi-kötü değerlendirmelerini yapabilmek için dikkate alınanları inceleyen ve gerekçelendiren bir alandır. Etik, yalnızca bir eylem sınırlamasını tanımlamak ve gerekçelendirmek için değil, aynı zamanda kişinin ve kişiliğinin ortaya çıkarılması için de vardır (Bendel 2023). Tıbbın uygulanması sürecinde ortaya çıkan değer sorunları ise tıp etiğinin temel konusunu oluşturmaktadır (Oğuz ve ark. 2005). Tıp alanında karşılaşılan etik sorunlara yönelik Beauchamp ve Childress tarafından geliştirilen etkin etik yaklaşımlardan birisi, temel etik ilkelerdir (Varkey 2021).). Bu bakımdan dört temel ilke *zarar vermeme (nonmaleficence)*, *yarar sağlama (beneficence)*, *özerkliğe saygı (respect for autonomy)* ve *adalet*'tir. (*justice*) Sağlık profesyonelleri bu etik yaklaşımı kullanarak değer çatışmalarını (çatışan ilkeleri) belirleyebilir, neyin üstün gelmesi gerektiğini saptayabilir, gerekirse etik danışmanlık alabilirler(Varkey 2021).

Fırsatlar, Sınırlılıklar ve Etik Sorunlar

ChatGPT'nin, akademik ve eğitim sürecine sağladığı katkılar yanında, (Örneğin bir çalışmanın ChatGPT desteğiyle kısa sürede düzenlenmesi sonucu artan zamanın diğer çalışmalar için kullanılması) klinik tıpta hasta eğitiminde, mental sorunların teşhisinde ve sağlık hizmetlerinin sunumunda giderek dönüştürücü bir rol oynadığı ifade edilmektedir (Wong ve ark. 2023; Altamimi ve ark. 2023; Wen ve Wang 2023). ChatGPT'nin derlediği bilgilerin her zaman doğru olmaması, tıbbın gerektirdiği karmaşık görevleri yerine getirme yeteneğinden yoksun olması, insanlar için adil olmayan sonuçlara yol açabilen belirli grupları ya da fikirleri kayırbilen önyargılı-yanlış verileri üretebilmesi tıp alanında kullanımındaki sınırlılıklardan (Wong ve ark. 2023; Wen ve Wang 2023). Bunun yanında ChatGPT desteğiyle hasta verilerine aydınlatılmış onam alınmadan erişilebilmesi ve klinik karar verme sürecinde yanlış kararlar alınabilmesi, sağlık profesyonellerinin göz önünde bulundurmaları gereken etik konulardır (Wong ve ark. 2023). Sağlık hizmetleri sunumunda hastalara ait hassas verilerin toplanması, saklanması ve işlenmesi sürecinde kişisel verilerin korunması, tıpta mahremiyet ilkesinin karşılığı olarak konuyla ilgili birtakım sorunları gündeme getirmektedir (Wang ve ark. 2023).

ChatGPT'nin tıp eğitiminde çeşitli alanlarda kullanıldığı, sınav katılımcısı olarak girdiği sınavların çoğunu geçtiği belirtilmektedir (Sumbal ve ark. 2024). Bir çalışmada United States

Medical Licensing Examination (USMLE) sınavını geçtiği, diğer bir çalışmada ise 3. sınıf tıp öğrencisi için eşdeğer geçme puanı elde ettiği gösterilmektedir (Wong ve ark. 2023; Mbakwe ve ark. 2023; Gilson ve ark. 2023).

Son yıllarda ChatGPT'nin farklı alanlar için kullanılabilir hale gelmesi, bilimsel yayınlarda etik sorun alanlarının doğmasına yol açmıştır (Rahimi ve Abadi 2023). Bu bakımdan yayın etiği kapsamında ön plana çıkan konular *plagiarism (intihal)* ve *yazarlık hakkı* sorunlarıdır (Salvagno ve ark. 2023). ChatGPT, başkalarına ait eserleri uygun şekilde atıfta bulunmadan sunabilir veya yeniden ifade ederek başkalarını kopyalamaktan kaçınabilir. Bununla birlikte, intihal yüzdesini azaltmak için kişisel bir katkı sunmadan yazıları yeniden formüle edici programlar kullanmanın akademik dürüstlük ihlali olduğunu vurgulamak gereklidir (Salvagnove ark. 2023).

Tıp alanında bilimsel yayın sürecinde ChatGPT'nin desteğini alan ve araştırmalarda ortak yazar olarak yer almasını kabul eden bilim insanları ve bazı bilimsel yayın platformlarının olduğu görülmektedir (Rahimi ve Abadi 2023; ChatGPT ve Zhavoronkov 2022; O'Connor ve ChatGPT 2023; Stokel-Walker 2023). Diğer taraftan ChatGPT'nin bilimsel bir araştırmada ortak yazar olarak yer almasının bilimsel ve etik açıdan doğru olmadığını belirten COPE (Committee on Publication Ethics) ve ICMJE (International Committee of Medical Journal Editors) gibi uluslararası kuruluşlar da bulunmaktadır (Rahimi ve Abadi 2023; Stokel-Walker 2023). ICMJE'nin yazarlıkla ilgili yönergeleri bu bakımdan önem taşımaktadır. Bu kriterler, bir makaleye önemli ölçüde katkıda bulunmak, içeriği eleştirel bir şekilde gözden geçirmek, yayından önceki son halini onaylamak, yayınlanan içeriğin bütünlüğü ve doğruluğu için sorumluluk almak şeklinde açıklanmaktadır (ICMJE 2024).

Bu çalışmanın yazarı da bilimsel yayınlarda sorumluluk ve hak kavramları açısından ChatGPT'nin önemini görmekle birlikte, yazarlık kriterlerini karşılayabilecek niteliklerden yoksun olması gerekçesiyle, destek alındığı takdirde ChatGPT'nin ortak yazar olmaması gerektiğini düşünmektedir. Hak kavramı tarihsel süreçte insan merkezli bir bakış açısıyla insan için kullanılagelmiştir (Aşar 2018). Bunun yanında kavramın hayvanlar gibi insan dışı canlılar ya da çevre açısından ele alınabildiğini hatırlamak gerekir. Bu noktada konuya insanın diğer canlılarla ve cansız çevreyle ahlaki ilişkisini konu edinen çevre etiği perspektifinden bakıldığında, insan dışı canlıların, örneğin hayvanların hakları olduğu fikri, canlı merkezli etik yaklaşımın savunucularından Peter Singer'in görüşleriyle farklı bir yorum kazanmaktadır (Aşar 2018; Birden 2016). Singer, hayvanlar gibi haz ve acı duyabilen varlıkların ahlaki yönden ilgi görme ehliyetlerinin olduğunu düşünmektedir (Aşar 2018). Bu durumda bir varlığın hak sahibi

olabilmesi, acı ve haz duyma yetisine bağlanmıştır (Aşar 2018). Konuya bu açıdan bakıldığında, insan olmayan bir yapay zekâ uygulamasının acı ve haz duyma gibi yetilerinin olmaması dolayısıyla ahlaki yönden çıkar sahibi olmadığı görülmektedir. Bu konuda ahlaki değerlere sahip bir varlık olduğu düşünülen insanın sorumluluğu kilit rolü oynamaktadır. Çünkü sorumluluk ancak insana yüklenebilen bir kavramdır (Oğuz ve ark. 2005). Bu noktada araştırmacıların bilimsel yayın sürecinde akademik sorumluluk bilinciyle ChatGPT’den yardım aldıkları zaman, çalışmanın *Teşekkür* ya da *Gereç ve Yöntemler* bölümlerine ChatGPT’den destek aldıklarını belirten bir metin eklemeleri izlenebilecek yollardan birisi olarak kabul edilebilir.

Sonuç

Tıp alanında ChatGPT kullanımının diğer teknolojik gelişmelerde olduğu gibi mesleki uygulamalara olumlu ve olumsuz yönde etkileri bulunmaktadır. Sağlık hizmetlerinin kalitesinin yükseltilmesi, hekim-hasta ilişkisinin güvene dayalı bir şekilde sürdürülebilmesi ve bilimsel çalışmaların devam edebilmesi için sağlık profesyonelleri ve araştırmacıların temel etik ilkelere (*zarar vermeme, yarar sağlama, özerkliğe saygı, adalet, mahremiyet, sır saklama ilkesi*) bağlı kalmaları esastır. Tıp alanında akademik dürüstlük ve mesleki standartların sürdürülebilmesine yönelik atılabilecek önemli adımlardan biri ChatGPT’nin sorumlu-doğru kullanımının bilinmesi ve mesleki uygulamalara yansıtılabilmesidir. Bunun yolu temelde dürüstlük ve şeffaflık ilkelerinden geçmektedir. Yanlış kullanımı önlemek için klinik uygulamalarda kapsamlı etik yönergeler hazırlanması, sağlık profesyonellerinin ChatGPT kullanımı konusunda doğru davranış pratikleri geliştirmelerine yardımcı olabilir.

Bu çalışmanın tıp ve sağlık bilimleri alanında çalışan araştırmacılara, sağlık profesyonellerine, tıp fakültesi öğrencilerine ve konuya ilgi duyan okurlara etik farkındalık katacağı düşünülmektedir.

Kaynaklar

- 1- Wong, R.S., Ming, L.,C., Raja, Ali R.A. 2023. The intersection of ChatGPT, clinical medicine, and medical education. *JMIR Med Educ* 9:e47274.
- 2- Roumeliotis, K.I. ve Tselikas, N.D. 2023. ChatGPT and open-AI models: A preliminary review. *Future Internet* 15(6): 192.
- 3- Sumbal, A., Sumbal, R., Amir, A. 2024. Can ChatGPT-3.5 Pass a Medical Exam? A Systematic Review of ChatGPT’s Performance in Academic Testing. *J Med Educ Curric Dev* 11:23821205241238640.

- 4- Salvagno, M., Taccone, F.S., Gerli, A.G. 2023. Can artificial intelligence help for scientific writing? *Critical Care* 27:75.
- 5- Malik, A., Khan, M. L., Hussain, K. 2023. How is ChatGPT Transforming Academia? Examining its Impact on Teaching, Research, Assessment, and Learning. *SSRN Electronic Journal*
- 6- Altamimi, I., Altamimi, A., Alhumimidi, A.S., Altamimi, A., Temsah, M.H. 2023. Artificial Intelligence (AI) Chatbots in Medicine: A Supplement, Not a Substitute. *Cureus* 15(6):e40922.
- 7- Samuel, G., Chubb, J., Derrick, G. 2021. Boundaries Between Research Ethics and Ethical Research Use in Artificial Intelligence Health Research. *Journal of Empirical Research on Human Research Ethics* 16(3):325-337.
- 8- Bendel, O. Image Synthesis from an Ethical Perspective. 2023. *AI & Society*.
- 9- Oğuz, N.Y., Tepe, H., Büken, N.Ö., Kucur, D.K. 2005. *Biyetik Terimleri Sözlüğü* Türkiye Felsefe Kurumu. Ankara.
- 10- Varkey, B. 2021. Principles of Clinical Ethics and Their Application to Practice. *Med Princ Pract* 30(1):17-28.
- 11- Wen, J. ve Wang, W. 2023. The Future of Chatgpt in Academic Research and Publishing: A Commentary for Clinical and Translational Medicine. *Clin Transl Med* 13(3):e1207.
- 12- Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., Liu, J. 2023. Ethical Considerations of Using ChatGPT in Health Care. *J Med Internet Res* 25:e48009.
- 13- Mbakwe, A.B., Lourentzou, I., Celi, L.A., Mechanic, O.J., Dagan, A. 2023. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digit Health* 2(2):e0000205
- 14- Gilson, A., Safranek, C.W., Huang, T., Socrates, V., Chi, L., Taylor, R.A., et al. 2023. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ* 9:e45312.
- 15- Rahimi, F. ve Abadi, A.T.B. 2023. ChatGPT and Publication Ethics. *Arch Med Res* 54(3):272-274.
- 16- ChatGPT., ve Zhavoronkov, A. 2022. Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. *Oncoscience* 9:82-84
- 17- O'Connor, S. ve ChatGPT. 2023. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice* 66.
- 18- Stokel-Walker, C. 2023. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613(7945):620–621.
- 19- International Committee of Medical Journal Editors, ICMJE. 2024. Defining the Role of Authors and Contributors.
Erişim tarihi: 03.04.2024 Erişim Adresi:
<https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
- 20- Aşar, H. 2018. Hayvan haklarına yönelik temel görüşler ve yanılgıları. *Kaygı* 30, 239-251.
- 21- Birden, B. 2016. Çevre Etiğinde Bireyin Ahlaki Sorumluluğuna Kısa Bir Bakış. *Türkiye Biyoetik Dergisi* 3(1): 4-14.

Statistical Analysis of Physiological Stress Signals and Exam Performance

Ayşegül K. Kasap¹ [0000-0002-8560-5759], Burçin KURT² [0000-0001-5781-2382]

¹School of Medical Physiology Karadeniz Technical University,
Trabzon, Turkey, Department of Therapy and Rehabilitation, Gumushane University,
Gumushane, Turkey

²Department of Biostatistics and Medical Informatics, Karadeniz Technical University,
Trabzon, Turkey
aysegulkurtkasap@gumushane.edu.tr
burcinkurt@gmail.com

Abstract. The availability of datasets capturing real-world cognitive stressors over extended durations is indeed limited compared to those focused on artificial stimuli in controlled laboratory settings. This limitation poses a significant challenge for researchers seeking to comprehend the intricacies of stress responses in real-life scenarios. “A Wearable Exam Stress Dataset” aims to bridge the divide between the established knowledge of the human stress response primarily derived from laboratory studies and the real-world fluctuations in people's skin conductance during stressful exam experiences. The dataset includes electrodermal activity (EDA), heart rate (HR), and skin surface temperature (TEMP) data collected for 10 students during three exam sessions (midterm 1, midterm 2, and finals), along with corresponding grades. In this study, we explore the impact of stress on exam performance by analyzing datasets comprised of physiological signals and exam scores. The findings indicate significant differences in grade values between midterm 1 and the final exam ($p < 0.05$), as well as between midterm 2 and final exams ($p < 0.05$). Furthermore, electrodermal activity and heart rate displayed a moderate, negative correlation ($r = -0.504$; $p < 0.01$). Additionally, electrodermal activity demonstrated a weak, positive, but significant correlation with skin surface temperature ($r = 0.417$; $p < 0.05$).

Keywords: Physiological Stress Signals, Student Performance, Statistical Analyze

OFTALMOLOJİDE YAPAY GÖRME (RETİNA İMPLANTI) ÜZERİNE YAPILAN ÇALIŞMALAR

Evren Sezgin¹[0000-0003-1270-581X], Ahmet Yardımcı¹[0000-0001-7241-4913]

¹ Akdeniz Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, 07070, Antalya, Türkiye
esezgin@akdeniz.edu.tr, yardimci@akdeniz.edu.tr

Özet Yapay görme, doğal olarak görme yeteneğini kaybetmiş veya sınırlı görme kapasitesine sahip bireyler için geliştirilen teknolojik sistemler veya cihazlar aracılığıyla görme sağlama sürecidir. Yapay görme sistemleri, dışarıdan gelen ışık ve görüntü bilgisini algılamak, işlemek ve beyne iletmek için çeşitli yöntemler kullanır.

Yapay görme teknolojileri 3 kategoride ele alınmaktadır:

1-Retinal Protez Sistemleri: Retina üzerine yerleştirilen elektrot dizileri aracılığıyla ışık ve görüntü bilgisini algılayan ve beyne ileten sistemlerdir.

2-Görsel İmplantlar: Göz içine yerleştirilen mikroçipler veya implantlar aracılığıyla görüntü bilgisini işleyen ve beyne ileten sistemlerdir. Görsel implantlar, retina veya optik sinir üzerine yerleştirilebilir.

3-Görsel Dönüştürücüler: Dışarıdan gelen görüntüyü ışık veya desen şeklinde algılayan ve bunu işleyip beyne ileten sistemlerdir.

Retinal protezler ise göz içinde yerleştirilme yerlerine göre; Retinanın üzerinde olanlar (Epiretinal), retina ile pigment epitel tabakası arası olanlar (Subretinal) ve pigment epitel tabakası ile damar ağı tabakası arasına (Suprakoroidal transretinal) yerleştirilenler olarak 3'e ayrılır.

Özellikle son 20 yılda yapılan çalışmalar ile Retinitis pigmentosa gibi retinanın dış tabakasında olan hastalıkların tedavisi için çeşitli yöntemler uygulanmaktadır. Yapay görme teknolojileri, genellikle kişinin doğal görme yeteneğini tam olarak geri kazandırmazlar, ancak kısmi bir görme sağlayabilirler. Bu sistemler sürekli olarak geliştirilmekte ve iyileştirilmektedir. Bu çalışmada yapay görmede retina implantı ile ilgili yapılan çalışmalar ve kullanılan yöntemlerin karşılaştırılması yapılmaktadır.

Anahtar Kelimeler : Yapay Görme, Retina İmplantı, Biyonik Göz.

STUDIES ON ARTIFICIAL VISION (RETINA IMPLANT) IN OPHTHALMOLOGY

Abstract: Artificial vision is the process of providing technological systems or devices for individuals who have naturally lost their ability to see or have limited parts. Artificial transport systems use a variety of storage to detect, process and transmit external light and image information to the brain.

Machine vision technologies consist of 3 categories:

1-Retinal Prosthesis Systems: These are systems that store light and image information and transmit it to the brain through electrode arrays placed on the retina.

2-Visual Implants: These are systems that store image information and transmit it to the brain through microchips or implants placed inside the eye. Visual implants can be placed on the retina or optic nerve.

3-Visual Converters: These are systems that perceive external light or patterns, process it and transmit it to the brain.

Retinal prostheses, depending on their placement in the eye; They are divided into 3 groups: those located above the retina (Epiretinal), those located between the retina and the pigment epithelial layer (Subretinal), and those located between the pigment epithelial layer and the vascular network layers (Suprachoroidal transretinal).

It is distributed at various intervals, especially for the treatment of diseases in the outer layer of the retina, such as Retinitis pigmentosa, based on studies conducted over the last 20 years. Artificial vision technologies generally do not fully restore a person's natural ability to function, but they cannot preserve some of it. These systems are constantly being developed and improved. Studies on retinal implants in this possible artificial vision and comparison with the methods used are made.

Key Words: Artificial Vision, Retinal Implant, Bionic Eye.

1 Giriş

Görme kaybı, dünya genelinde milyonlarca insanı etkileyen yaygın bir sağlık sorunudur. Geleneksel tedavi yöntemleri, bazı durumlarda etkili olmasına rağmen, kimi zaman görme kaybının kalıcı olduğu durumlarda sınırlıdır. Ancak, son yıllarda yapay görme (biyonik göz) teknolojilerindeki hızlı ilerlemeler, görme kaybı yaşayan bireyler için umut verici yeni tedavi seçenekleri sunmaktadır.

Biyonik gözler, yapay retina implantları ve optik sinir uyarıcıları gibi cihazlar aracılığıyla görme yeteneğini geri kazandırmayı amaçlayan yenilikçi tıbbi cihazlardır. Bu cihazlar, görme sinirlerine doğrudan elektriksel uyarılar göndererek veya retina üzerindeki foto sensör hücrelerini simüle ederek görme yeteneğini geri kazandırmayı hedeflemektedir. Bu teknolojiler, retina dejenerasyonu, retinitis pigmentosa, yaşa bağlı makula dejenerasyonu gibi nedenlerle görme kaybı yaşayan bireyler için özellikle umut vericidir.

Bu makalede, oftalmolojideki yapay görme (biyonik göz) çalışmalarını inceleyeceğiz. İlk olarak, yapay görmenin temellerini ve işleyişini açıklayacağız. Daha sonra, farklı biyonik göz cihazları ile yapılmış çalışmaların sonuçlarını karşılaştıracacağız. Ayrıca, biyonik göz teknolojilerinin mevcut durumu ve gelecekteki potansiyel etkileri yer almaktadır.

1.1. Biyonik Göz Nasıl Çalışır?

Sağlıklı görme, gözbebeğine gelen ışığın, mercekle tarafından odaklanarak vitreous humour'dan geçmesi ve retinaya ulaşması ile başlar. Buradan retinanın yüzeyindeki ganglion hücrelerini içeren bir grup şeffaf hücreden geçerek ışığa duyarlı 'rod' ve 'konlar'a gelir. Işık tarafından uyarılan rod ve konlar, retinanın üzerindeki diğer hücreleri kullanarak ganglion hücrelerine elektrik uyarıları gönderir. Ganglion sinir hücreleri de optik sinir ağınını kullanarak, beynin görme korteksine sinyaller göndererek görme gerçekleşir [1].

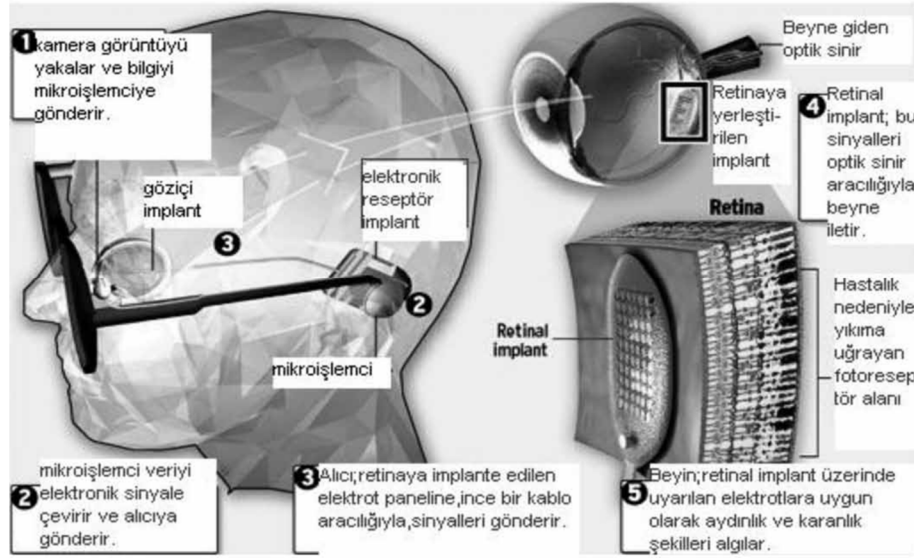
Doğal olarak görme yeteneğini kaybetmiş veya sınırlı görme kapasitesine sahip bireyler için geliştirilen teknolojik sistemler veya cihazlar aracılığıyla görme sağlama sürecidir. Teknolojik gelişmeler ile görme kaybının giderilmesi ile ilgili çalışmaların sayısı artmaktadır. Temel amaç, görme yollarındaki sağlam kalmış hücreleri tekrar aktiveleştirmektir. Bununla ilgili çok sayıda tedavi yöntemi uygulanmaktadır. Tablo 1'de Yapay görme ile ilgili uygulanan çalışma ve hatalıklar gösterilmektedir.

Tablo 1- Görme ile ilgili hastalıklar ve uygulanan yöntemler [2]

Uygulama	Hastalık
Biyonik göz (vizüel protez)	Retinitis pigmentosa
Retinal	Yaşa bağlı Makula Dejenerasyonu
Kortikal	
Optik sinir	
Gen tedavisi	Retinitis pigmentosa Leber'in konjenital amorozu (retinal distrofi) Leber'in herediter optik nöropatisi
Kök hücre tedavisi	Retinitis pigmentosa Yaşa bağlı makula dejenerasyonu
Optogenetik	Kornea hastalıkları Retinitis pigmentosa

Biyonik göz sistemleri, gözlüğe yerleştirilmiş bir kamera görüntüyü yakalayıp kablolu ya da kablosuz bir teknoloji ile genellikle bele takılan bir mikroişlemci ya da video işleyiciye görüntüyü aktarır. Burada görüntü çözünürlüğü düşürülerek elektrik sinyallerine dönüştürülür. Oluşan görüntüyü tekrar gözlüğe takılı alıcıya aktarır, buradan retinada bulunan implantın mikro elektrot paneline aktarılır. Elektrotlar retinayı uyarak bilgiyi görme sinirinde oluşturur ve oluşan görüntü beyne aktarılır (Şekil 1).

Retinis pigmentosa ve yaşa bağlı makula dejenerasyonu, ışık algılayan fotoreseptör hücrelerin dejenerasyonuna bağlı görme kaybıyla sonuçlanan, etkin ve kesin tedavisi olmayan retina hastalıklarıdır [3].



Şekil 1- Biyonik göz nasıl çalışır ((www.sciencecodex.com/get_ready_for_a_bionic_eye))

Retinal protezler yerleşim yerine göre isimlendirilmektedir. Subretinal protez, retina dokusu ile pigment epitel tabakası arasında yerleştirilir. Epiretinal protez retinanın üzerine yerleştirilir. Suprakoroidal protez ise pigment epitel ve damar ağı tabakası arasında yerleştirilmektedir [4].

Tablo 2- Retinal Protez Çalışmaları

YIL	AYGIT	TAKIM	TİP	ONAY
1974	Cortical	Dobelle	Cortical	
2000	ASR	Optobionics	Subretinal	
2002	Argus I	Second Sight	Epiretinal	
2006	ARGUS II	Second Sight	Epiretinal	FDA
2010	MDPA	RIAG	Subretinal	CE
2014	IRIS I	Pixium	Epiretinal	
2016	IRIS II	Pixium	Epiretinal	CE

Retina implantları, retinitis pigmentosa ve yaşa bağlı makula dejenerasyonu gibi retina dejeneratif hastalıkları olan bireylere bir dereceye kadar görme fonksiyonunu geri kazandırabilir [4].

YÖNTEM

Yapay görme ile ilgili alanında yapılmış ilk çalışmaları tespit için, PubMed, Science-direct veri tabanlarında literatür taraması yapıldı. İncelemeye, insan veya hayvan deneyimlerinde ilerleme gösteren hakemli makaleleri seçilmiştir. Her yeni cihaz ile yapılmış ilk makalelerin çözüm özellikleri ve klinik sonuçları sunulmuştur.

SONUÇ

Retina implantları, görme engelli bireyler için umut verici bir tedavi seçeneği olarak ön plana çıkmaktadır. Bu implantlar, görsel bilgiyi doğrudan retina veya optik sinir sistemine aktararak, kullanıcıların sınırlı veya kaybolmuş görme yeteneklerini geri kazanmalarına yardımcı olmaktadır. Son yıllarda yapılan araştırmalar, retina implantlarının etkinliği, güvenilirliği ve kullanıcı deneyimi üzerine önemli bilgiler sunmaktadır. Bu çalışmalar, farklı implant türleri ve teknolojileri arasında karşılaştırmalar yaparak, en etkili ve uygun çözümleri belirlemeyi amaçlamaktadır. Retina implantları üzerine yapılan çalışmaların sonuçları ve karşılaştırmaları, bu alandaki bilimsel ve klinik

ilerlemeleri anlamamıza ve görme engelli bireylerin yaşamlarını iyileştirmek için daha etkili tedavi stratejileri geliştirmemize olanak tanımaktadır.

Dobelle ve ekibinin yapmış olduğu bu çalışmada [5], yedi gönüllünün hiçbirinde epileptik semptomlar veya implantla ilgili başka sistemik sorunlar görülmemiştir. Sistemin test aşamasında 1,5 metre yükseklikteki 15 cm kare şeklinde "E" harfini tanıdıkları ve parmak sayabildikleri belirtilmiştir. Bu makalede istatistiksel olarak herhangi bir bilgi verilmemiştir.

Chow ve arkadaşları [6], ASR ismini verdikleri mikroçipi kullanarak bir sistem oluşturdu. ASR mikroçipi, yaklaşık 5000 mikro elektrot uçlu mikro fotodiyot içeren ve gelen ışıkla çalışan, 2 mm çapında silikon bazlı bir cihazdır. Retinitis pigmentozalı 6 hastanın sağ gözüne ASR mikroçipi yerleştirilmiş, sol gözü ise kontrol olarak kullanılmıştır. 6 18 aylık takipten sonra tüm ASR'ler elektriksel olarak çalıştı ve hiçbir hastada implant reddi, enfeksiyon, iltihaplanma, erozyon, neovaskülarizasyon, retina dekolmanı veya migrasyon belirtileri görülmedi. Parlaklık, kontrast, renk, hareket, şekil, çözünürlük ve görsel alan boyutunun algılanmasında iyileşmeler gözlemlendiği belirtilmiştir. Bu makalede istatistiksel bir bilgi yer almamıştır.

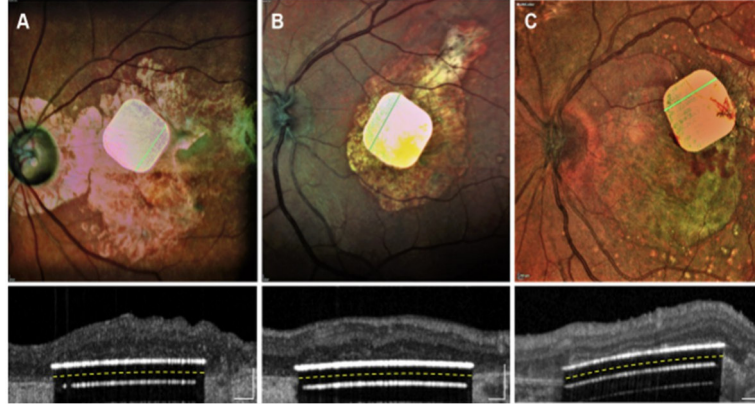
Humayun ve arkadaşları [7,8] (2002 ve 2004) tarafından yapılan çalışma, altı deneye 16 elektrotlu bir retina implantının test edilmesini içeriyordu. Bu çalışmanın sonuçlarına göre, deneyin eğitimi sonucunda:

1. 120 cm uzaklıktaki bir duvardaki bir ışık noktasının konumunu tespit edebildiler.
2. Karanlık bir odada, 200 cm uzaklıkta bulunan bir kişinin taşıdığı flaş ışığının konumunu gösterebildiler.
3. Eğitim sürecinin artmasıyla, denekler normal oda aydınlatması altında karanlık bir nesnenin yerini tespit edebildiler ve hareket yönünü gösterebildiler.

Bu sonuçlar, retina implantının deneklerin görsel algılarını iyileştirmede potansiyel bir etkiye sahip olduğunu göstermektedir. Araştırma, yapay görme teknolojisinin gelişiminde önemli bir adımdır ve görme kaybı olan bireyler için umut verici bir ilerleme sağlamaktadır.

Geruschat ve arkadaşları [9], 60 elektrot bulunan Argus II Retinal Protez Sistemini 26 hastaya implante etmişlerdir. Ortalama 36 aylık takip sonrasında yayınladıkları çalışmalarında, günlük hayat ile ilgili 35 görevi gözlemciler tarafından derecelendirerek puanlama yapmışlardır. Hastanın bir görevi tamamlama kolaylığı, kolaydan (1 puan) Zora (4 puan) kadar değişen dört puanlık bir ölçek kullanılarak puanlamışlardır. 35 görevin 24 'ünde deneklerin günlük işleri istatistiksel olarak kolaylaştırdığı sonucuna varmışlardır. Cihazlarının aktif durumuna karşı, pasif durumdayken hastaların performansında yüzde 19 ile 38 arasında bir iyileşme tespit etmişlerdir ($p<0,05$).

Palanker ve arkadaşları [10] kablosuz 378 elektrotlu mikroçip geliştirmişlerdir. 2 x 2 milimetre karelik implant 30 mikron kalınlığındadır; yani insan saçı kalınlığının yaklaşık üçte biri kadardır. Cerrahi olarak retinanın altına veya subretinal boşluğa implante edilmektedir (Şekil 2). Pixium'un mikroçipi, görsel sahnedeki bilgileri kodlayan elektriksel uyarıyı kullanarak bu iletişimdeki eksik bağlantıyı sağlıyor. 5 hasta üzerinden yapılan çalışmalarda, 3 hastada görme keskinliği, belirlenen hedefe %90, 2 hastada ise %70 olarak ölçülmüştür. Bunun nedeni olarak da implantın, koroidin içinde ve merkezin 2 mm dışında olmasından kaynaklı olduğunu belirtmişlerdir. Gelecekteki implantlarda daha küçük piksellerin kullanılması görme keskinliğini daha da artırabileceği sonucuna ulaşmışlar.



Şekil 2-Pixium Mikroçipinin Retinaya yerleşimi

Biyonik göz teknolojilerinin bazı sınırlamaları vardır. Örneğin, görüntü çözünürlüğü hala sınırlı olabilir ve görsel algının doğallığı tam olarak sağlanamayabilir. Şu ana kadar yapılan çalışmalara oluşan görüntülerin siyah-beyaz ya da gri formlarda olduğunu göstermektedir. Ayrıca, bu cihazların maliyetleri çok yüksek olduğu için erişilebilirlik sıkıntıları yer almaktadır. Retina implantlarının yerleşimi uzmanlık gerektirdiği için her zaman cerrahi risk içerebilmektedir [11].

Gelişen nöroprotezler ve yapay zekâ algoritmaları, biyonik gözlerin işlevselliğini ve güvenilirliğini artırarak, kullanıcıların daha doğal ve etkili bir görme deneyimi yaşamalarını sağlayabilir. Gelecekte, bu teknolojilerin yaygınlaşmasıyla, görme engelli bireylerin yaşam kalitesinin önemli ölçüde iyileşmesi beklenmektedir. Bununla birlikte, bu teknolojilerin yapılacak çalışmalarda, maliyet, erişilebilirlik ve cerrahi riskler gibi faktörlerin ele alınması önemlidir. Biyonik gözlerin ilerlemesi, gelecekte daha fazla insanın görsel dünyaya erişimini sağlayabilir.

Kaynakça

1. Niyazoglu, G., Ataseven, G., & Eltutar, K. (2011). Bionic Eye and Advanced Technology in Ophthalmology. *Istanbul Medical Journal*, 12(2), 79–84. <https://doi.org/10.5505/1304.8503.2011.46338>.
2. Yapay Görme Hastalıkları, <https://www.herkesebilimteknoloji.com/haberler/saglik/yapay-gorme-ve-biyonik-goz-ile-korluklerin-onlenmesinde-onemli-adimlar-atildi>, Erişim Tarihi: 10.04.2024.
3. Liu, W., & Humayun, M. (2004). Retinal prosthesis. Paper presented at the 2004 IEEE International Solid-State Circuits Conference, San Francisco (IEEE Cat. No. 04CH37519).
4. Güven, D., Weiland, J. Mark, Öz.G. (2005). Retinal Protezler. *Ret-Vit* 2005;13: 247-256.
5. Dobbelle, Wm. H.. Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex. *ASAIO Journal* 46(1):p 3-9, January 2000.
6. Chow AY, Chow VY, Packo KH, Pollack JS, Peyman GA, Schuchard R. The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa. *Arch Ophthalmol*. 2004 Apr;122(4):460-9. doi: 10.1001/archoph.122.4.460. PMID: 15078662
7. Humayun, M.S., de Juan Jr, E., Weiland, J.D., Dagnelie, G., Katona, S., Greenberg, R., ... & Hinton, D.R. (1999). Pattern electrical stimulation of the human retina. *Vision Research*, 39(15), 2569-2576.
8. Humayun M. S., Weiland J. D., Fujii G. Y., Greenberg R., Williamson R., Little R. , Brian Mech, V. Cimmarrusti, G. Van Boemel, G. Dagnelie, E. Juan, Visual perception in a blind subject with a chronic microelectronic retinal prosthesis, *Vision Research*, Volume 43, Issue 24, 2003, Pages 2573-2581, ISSN 0042-6989, [https://doi.org/10.1016/S0042-6989\(03\)00457-7](https://doi.org/10.1016/S0042-6989(03)00457-7).
9. Geraschat DR, Richards TP, Arditi A, da Cruz L, Dagnelie G, Dorn JD, Duncan JL, Ho AC, Olmos de Koo LC, Sahel JA, Stanga PE, Thumann G, Wang V, Greenberg RJ. An analysis of observer-rated functional vision in patients implanted with the Argus II Retinal Prosthesis System at three years. *Clin Exp Optom*. 2016 May;99(3):227-32. doi: 10.1111/cxo.12359. Epub 2016 Jan 24. PMID: 26804484; PMCID: PMC5066820.
10. Palanker D., Mer Y., Said S.M., Muqit M., Sahel J., Photovoltaic Restoration of Central Vision in Atrophic Age-Related Macular Degeneration, *Ophthalmology*, Volume 127, Issue 8, 2020, Pages 1097-1104, ISSN 0161-6420, <https://doi.org/10.1016/j.ophtha.2020.02.024>.
11. Da Cruz, L., Coley, B.F., Dorn, J., Merlini, F., Filley, E., Christopher, P., ... & Sahel, J.A. (2016). The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. *British Journal of Ophthalmology*, 100(2), 1-5.

Derin Öğrenme Öznelik Çıkarım Algoritmalarının Makine Öğrenmesi Modelleri ile Kullanılması: Diyabetik Retinopati Örneği

Fatih Serdar AYDEMİR¹[0009-0005-0197-5138] and Uğur ŞEVİK²[0000-0002-2056-9988]

^{1,2} Karadeniz Technical University Department of Computer Science, Trabzon, Türkiye

Using Deep Learning Feature Inference Algorithms with Machine Learning Models: Diabetic Retinopathy Example

Abstract. Diabetic Retinopathy (DR) is caused by complications of diabetes in the eye. Diabetic Retinopathy affected approximately 127 million people worldwide in 2010 and is expected to affect approximately 191 million people by 2030. Regular screenings using highly efficient computer-based systems are of great importance for the early detection of DR cases. As a result of the literature studies, it has been observed that various DR classification studies have been carried out on different datasets. As a result of the studies, studies using deep learning algorithms have shown higher classification success than studies using machine learning algorithms. Due to the high physical cost required by deep learning models, the techniques used in some studies were tested with machine learning models. In this study, VGG19, ResNet50 and MobileNetV3 Large models with feature extraction feature were run on the open dataset shared by the Asia Pacific Tele-Ophthalmological Society (APTOS). The resulting feature matrices were used to train Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF), Naïve Bayes (NB) and AdaBoost algorithms. For each machine learning algorithm, feature extraction algorithms were tested and the best results were obtained as F1-Score 67% and success rate 74% with VGG19 + SVM cooperation, F1-Score 67% and success rate 70% with MobileNetV3 Large + RF cooperation, and F1-Score 67% and success rate 74% with ResNet50 + SVM cooperation.

Keywords: Diabetic Retinopathy, SVM, KNN, AdaBoost, Random Forest, Naïve Bayes, Feature Extraction Algorithms

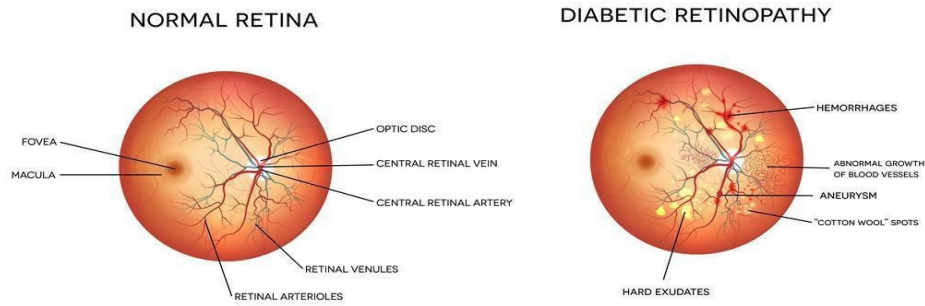
Özet. Diyabetik Retinopati (DR), gözdeki diyabet komplikasyonlarından kaynaklanır. Dünya genelinde 2010 yılında yaklaşık 127 milyon insanda görülen Diyabetik Retinopati hastalığının 2030 yılına kadar yaklaşık 191 milyon insanı etkilemesi beklenmektedir. Yüksek verimli bilgisayar tabanlı sistemler kullanan düzenli taramalar, DR vakalarının erken tespiti için büyük önem taşımaktadır. Yapılan literatür çalışmaları sonucunda farklı veri kümeleri üzerinde çeşitli DR sınıflandırma çalışmalarının yapıldığı görülmüştür. Yapılan araştırmalar

neticesinde derin öğrenme algoritmalarının kullanıldığı çalışmalar, makine öğrenmesi algoritmalarının kullanıldığı çalışmalara ziyade daha yüksek sınıflandırma başarısı göstermiştir. Derin öğrenme modellerinin gerektirdiği fiziksel maliyetinin yüksek olması nedeniyle bazı çalışmalarda kullanılan teknikler makine öğrenmesi modelleri ile test edilmiştir. Bu çalışmada öznelik çıkarma özelliğine sahip VGG19, ResNet50 ve MobileNetV3 Large modelleri Asya Pasifik Tele-Oftalmoloji Derneği (APTOS) tarafından paylaşılan açık veri seti üzerinde çalıştırılmıştır. Elde edilen öznelik matrisleri Support Vector Machine(SVM), K-Nearest Neighbor(KNN), Random Forest(RF), Naïve Bayes(NB) ve AdaBoost algoritmalarını eğitmek için kullanılmıştır. Her bir makine öğrenmesi algoritmasına karşılık öznelik çıkarma algoritmaları test edilerek en iyi sonuçlar; VGG19 + SVM işbirliği ile F1-Score %67, başarı oranı %74, MobileNetV3 Large+ RF işbirliği ile F1-Score %67, başarı oranı %70 ve ResNet50+SVM işbirliği ile F1-Score %67, başarı oranı %74 olarak elde edilmiştir.

Anahtar Kelimeler: Diyabetik Retinopati, SVM, KNN, AdaBoost, Random Forest, Naïve Bayes, Öznelik Çıkarma Algoritmaları

1 Giriş

Diyabetik Retinopati (DR), gözdeki diyabet komplikasyonlarından kaynaklanır. DR gözde bulunan kan damarlarına zarar vererek görme duyusunda sorunlara yol açar. İnsanların günlük yaşantısını etkileyen DR hastalığında erken ve doğru tanı oldukça önemlidir. Dünya genelinde 2010 yılında yaklaşık 127 milyon insanda görülen Diyabetik Retinopati hastalığının 2030 yılına kadar yaklaşık 191 milyon insanı etkilemesi beklenmektedir [1]. DR hastalığının dört adet evresi vardır. Bu evreler sırası ile hafif, orta, şiddetli ve proliferatif retinopati (retinada yeni ve sağlıklı damar oluşumu) evreleridir. Normal bir retina ve DR Hastası göz retinası aşağıda **Şekil 1** ile gösterilmiştir.



Şekil 1. Normal Göz Retinası ve DR Hastası Göz Retinası

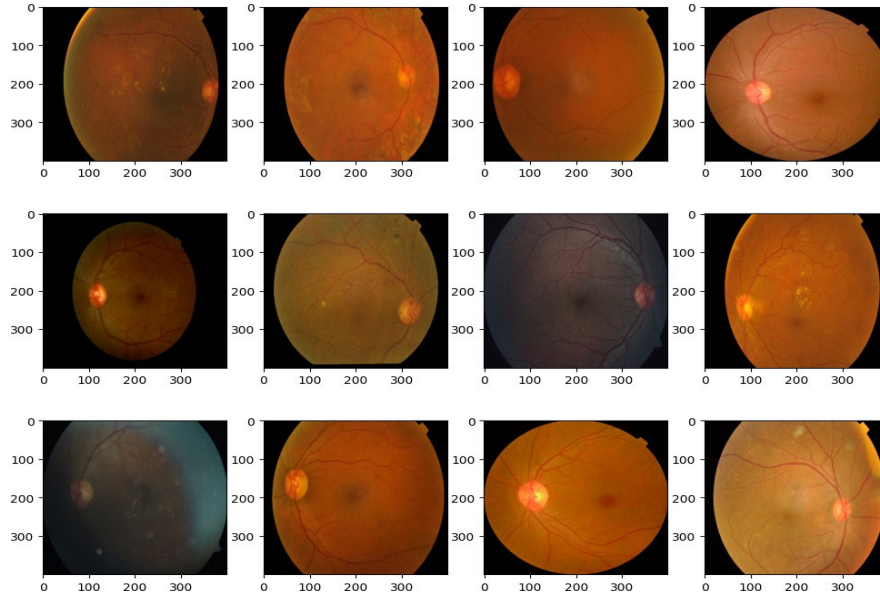
DR hastalığı ile alakalı literatürde derin öğrenme ve makine öğrenmesi yaklaşımları birçok araştırmada kullanıldığı görülmektedir. Örneğin, APTOS veri seti ile yapılan bir çalışmada ResNet50 öznelik çıkarma algoritması ile kullanılan CNN modelinin başarı oranı %77, F1-Score değeri ise %72 olarak bulunmuştur[2].İncelenen bir diğer çalışmada, APTOS veri seti ile eğitilen CNN modeli F1-Score olarak %70 performans ölçüsüne ulaşmıştır[3].

Yapılan bu çalışmada derin öğrenme modellerinin işlem maliyetinin yüksek olduğu göz önüne alındığında öznelik çıkarma algoritmalarının makine öğrenmesi algoritmaları ile ne kadar başarılı olabileceği araştırılmıştır.

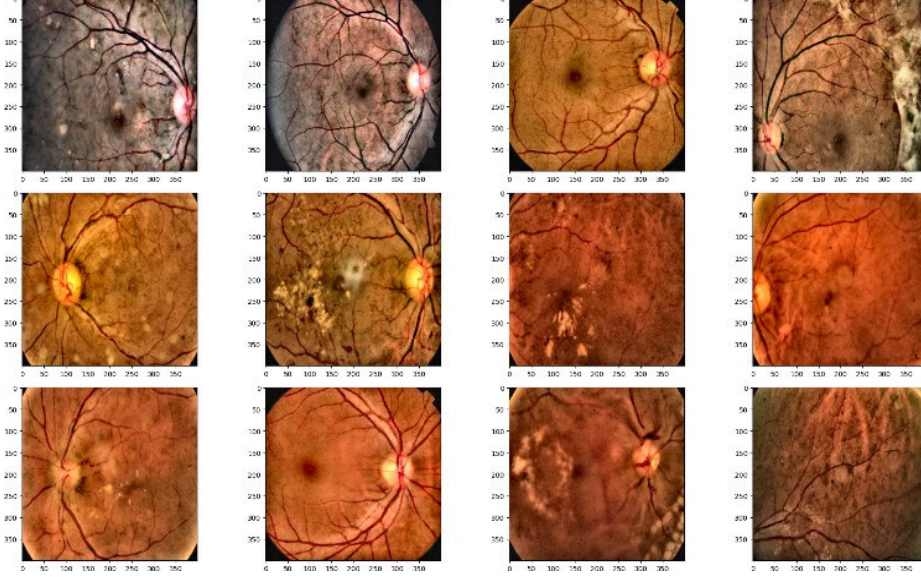
2 Yöntem

2.1 Veri Seti

Yapılan çalışmada kullanılan veri seti Asya Pasifik Tele-Oftalmoloji Derneği (APTOS) tarafından açık erişimli olarak paylaşılan gerçek hasta verilerinden oluşmaktadır[4]. Veri setinde toplamda sağlıklı retina görüntülerinden 1805 adet, hafif evre retina görüntülerinden 370 adet, orta evre retina görüntülerinden 999 adet, şiddetli evre görüntülerinden 193 adet, proliferatif evre retina görüntülerinden 295 adet içermektedir. Öznelik çıkarma algoritmaları eğitilmeden önce veriler ön işlem adımlarına tabi tutulmuştur. Boyutları 400*400 olarak ayarlanan görsellere kontrast iyileştirme yaklaşımlarından Contrast Limited Adaptive Histogram Equalization (CLAHE) işlemi uygulanmıştır. Uygulama öncesinde mevcut retina görüntüleri ile işlem sonucunda oluşan görsellere ait örnekler Şekil 2 ve Şekil 3 ile aşağıda verilmiştir.



Şekil 2. Ön İşleme Öncesi Mevcut Retina Görüntüleri Örneği



Şekil 3. Ön İşleme Sonucunda Oluşan Retina Görüntüleri Örneği

2.2 Kullanılan Modeller

Yapılan çalışmada VGG19[5], ResNet50[6] ve MobileNetV3[7] öznitelik çıkarım algoritmaları ile SVM[8], KNN[9], Random Forest[10], AdaBoost[11] ve Naïve Bayes[12] makine öğrenmesi algoritmaları kullanılmıştır.

VGG19 algoritmasının 19 katmandan oluşan basit ve tutarlı yapısı öznitelik çıkarımında oldukça iyi sonuçlar vermektedir. Modelin sadeliği ve önceden eğitilmiş ağırlıklarla çalışabilmesi modelin bu tarz projelerde tercih edilebilirliğini arttırmaktadır. ResNet50 modeli büyük ve karmaşık veri kümelerinde tercih edilmektedir. VGG19 modeline göre daha derin bir ağ yapısına sahip olduğundan daha karmaşık öznitelik çıkarımlarında başarılı olabileceğinden tercih edilmiştir. MobileNetV3 modeli düşük gecikme süreleri, gerçek zamanlı uygulamalarda kullanılabilirliği ve hesaplama gücü açısından verimi yüksek olduğundan bu çalışmada tercih edilmiştir.

Çalışmada kullanılan makine öğrenmesi modelleri arasında SVM, Random Forest ve AdaBoost algoritmaları aşırı öğrenme (overfitting) durumuna karşı dayanıklı olabildiği için tercih edilmiştir. Naïve Bayes algoritması ise hızlı ve verimli bir algoritma olduğu için tercih edilmiştir. KNN algoritması uygulanabilirliği kolay, verimli ve hızlı bir algoritma olduğu için yapılan çalışmada bir diğer karşılaştırma algoritması olarak seçilmiştir.

3 Bulgular

Yapılan çalışmada makine öğrenmesi modeli parametreleri her bir öznelik çıkarıcı algoritma denenirken sabit bırakılmıştır. VGG19 ve ResNet50 öznelik çıkarım algoritmaları için SVM modelinin parametreleri standart ayarlarda bırakılmıştır. MobilNetV3 + RF modeli için ise RF modelinin parametreleri “n_estimators=250, max_depth=8, min_samples_split=5, class_weight='balanced' ” olarak kullanılmıştır. Deneyler sonucunda en iyi performans sonuçları belirlenen bu parametrelerle elde edilmiştir.

Yapılan çalışmada en iyi değerleri veren öznelik-model sonuçları **Tablo 1.** ile aşağıda gösterilmiştir.

MODEL / METRİK	Precision	Recall	F1-Score	Accuracy
VGG19-SVM	%70	%74	%67	%74
ResNet50-SVM	%71	%74	%67	%74
MobileNetV3-RF	%70	%70	%67	%70

Tablo 1. En İyi Sonuçları Veren Modellerin Karşılaştırmalı Tablosu

Çalışmada kullanılan KNN, AdaBoost ve Naïve Bayes algoritmalarının diğer modellere göre daha düşük başarı oranı göstermesinin nedenleri olarak veri setinde bulunan görsellerin farklı hastanelerde farklı görüntüleme cihazları ile elde edilmesinden kaynaklı olarak kalite düşüklüğü ve öznelik çıkarım algoritmaların elde ettiği matrislere uygun olmadığı düşünülmektedir.

İleride yapılması planlanan akademik çalışmalarda, araştırmacılara fikir vermesi için en iyi sonuçları veren algoritmaları haricinde daha düşük başarı oranı gösteren KNN, AdaBoost ve Naïve Bayes algoritmalarına ait sonuçları ve parametreleri **Tablo 2**, **Tablo 3** ve **Tablo 4** ile aşağıda gösterilmiştir.

KNN / METRİK	Precision	Recall	F1-Score	Accuracy
KNN-VGG19	%64	%67	%63	%67
KNN-ResNet50	%65	%68	%66	%68
KNN-MobileNetV3	%64	%68	%66	%69

Tablo 2. KNN Algoritması : n_neighbors=5,weights='distance'

ADABOOST / METRİK	Precision	Recall	F1-Score	Accuracy
AdaBoost-VGG19	%57	%69	%61	%68
AdaBoost -ResNet50	%58	%70	%63	%71
AdaBoost -MobileNetV3	%56	%70	%61	%69

Tablo 3. AdaBoost Algoritması : $n_estimators = 100$, $learning_rate = 0.1$, $algorithm = 'SAMME'$

NB / METRİK	Precision	Recall	F1-Score	Accuracy
NB-VGG19	%56	%58	%53	%58
NB-ResNet50	%58	%61	%55	%61
NB-MobileNetV3	%58	%61	%59	%62

Tablo 4. Naïve Bayes Algoritması : Standart parametreler ile çalıştırılmıştır.

4 Sonuç

Derin öğrenme modellerinin işlem maliyeti göz önüne alındığında, makine öğrenmesi algoritmaları ile kurulan sistemlerin sağlık çalışanlarına gerekli ön bilgiyi yukarıda kullanılan modellerin %67'lik F1-Skor performansı ile verebileceği öngörülmüştür.

References

- Alwakid, G., Gouda, W., Humayun, M., & Jhanjhi, N. (2023). Enhancing diabetic retinopathy classification using deep learning. *DIGITAL HEALTH*, 9. <https://doi.org/10.1177/20552076231203676>
- Chaturvedi, S. S., Gupta, K., Ninawe, V., & Prasad, P. S. (n.d.). *Automated Diabetic Retinopathy Grading using Deep Convolutional Neural Network*.
- Chengsheng, T., Huacheng, L., & Bing, X. (2017). AdaBoost typical Algorithm and its application research. *MATEC Web of Conferences*, 139, 00222. <https://doi.org/10.1051/mateconf/201713900222>
- APTOS Homepage, <http://2019.asiateleophth.org/> last accessed 2024/05/06
- Evgeniou, T., & Pontil, M. (2001). *Support Vector Machines: Theory and Applications* (pp. 249–257). https://doi.org/10.1007/3-540-44673-7_12
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). *Searching for MobileNetV3*.
- Mansour, Y., & Schain, M. (2001). Random Forests. *Machine Learning*, 45(2), 123–145. <https://doi.org/10.1023/A:1010950718922>
- Oulhadj, M., Riffi, J., Chaimae, K., Mahraz, A. M., Ahmed, B., Yahyaouy, A., Fouad, C., Meriem, A., Idriss, B. A., & Tairi, H. (2022). Diabetic retinopathy prediction based on deep learning and deformable registration. *Multimedia Tools and Applications*, 81(20), 28709–28727. <https://doi.org/10.1007/s11042-022-12968-z>

9. Rewaskar, V., Chaudhari, S., Patil, Y. V., Agrawal, S. A., Rewaskar, V. D., Agrawal, R. A., Chaudhari, S. S., Patil, Y., & Agrawal, N. S. (2023). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Advancements in NSFW Content Detection: A Comprehensive Review of ResNet-50 Based Approaches. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2023, Issue 4). www.ijisae.org
10. Rish, I., & Rish, I. (n.d.). *An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier*. <https://www.researchgate.net/publication/228845263>
11. Sudha, V., & T. R. Ganeshbabu, Dr. (2020). A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Computers, Materials & Continua*, 66(1), 827–842. <https://doi.org/10.32604/cmc.2020.012008>
12. Sun, J., Du, W., & Shi, N. (2018). A Survey of kNN Algorithm. *Information Engineering and Applied Computing*, 1(1). <https://doi.org/10.18063/ieac.v1i1.770>

Utilizing Pretrained GPT-2 Model for Analyzing Drug Reviews

Sinem ÖZDEMİR¹[0000-0002-6332-0891], Kemal TURHAN¹[0000-0001-7871-3025]

¹ Karadeniz Teknik Üniversitesi, Trabzon, Türkiye
sinemozdemir@ktu.edu.tr, kemalturhan@ktu.edu.tr

Abstract. Accurately predicting drug side effects is critical to ensuring patient safety and optimizing health outcomes, while drug efficacy prediction plays a crucial role in directing clinicians and patients to treatments that provide maximum benefit with minimal side effects. This study used the pre-trained GPT-2 model by fine tuning to analyze drug reviews. Our aim is to estimate side effects in three categories (mild, severe and no side effects) and to classify effectiveness (ineffective, moderately effective and highly effective) into three categories.

In conclusion, drug effectiveness and drug side effects prediction based on drug reviews represents a promising paradigm in precision medicine. By integrating advanced computational methods with clinical data, It can improve our understanding of drug responses and provide personalized health care by tailoring treatments to individual patient needs.

Keywords: Drug Effectiveness Prediction, Drug Side Effect Prediction, GPT-2, Sentiment Analysis

Automatic Article Reader and Article Section Classifier Application with Web Scraping

Nihat Burak Zihni^[0000-0002-3757-9154]

¹ Karadeniz Technical University, Vocational School of Health Sciences
nbzihni@ktu.edu.tr

Abstract. In this study, a search was carried out on a specified topic in one of the openly accessible articles scanning databases, using a Python code script written using the web scraping technique. In the first stage, a classification method was applied only based on article section headings. In addition, a cluster analysis was applied that included the contents of the sections along with their headings. As a result of these analyses, contents covering more than 50% of the body text of articles that met certain conditions were obtained in one of the cluster sets used with a F_1 ratio of approximately 96%.

Keywords: web-scraping, natural language processing, text mining.

1 Introduction

Portable Document Format (PDF), standardized as ISO 32000, is a file format developed by Adobe in 1992 to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems.[1][2] Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it.[3]

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.[4] There are many libraries in the literature that convert PDF documents into text, such as “*pytesseract*”, “*pdfminer*”, “*pdftotext*”, “*pdf2image*”, and “*OpenCV*”.

However, most of them cannot successfully convert all the text content in a PDF document. The main reason for this situation is that in an actual PDF file, text portions might be split into several chunks in the middle of its running, depending on the authoring software. Therefore, text extraction needs to splice text chunks. In the figure below, two text chunks whose distance is closer than the “char_margin” (shown as M) is considered continuous and get grouped into one. Also, two lines whose distance is closer than the “line_margin” (L) is grouped as a text box, which is a rectangular area that contains a “cluster” of text portions. Furthermore, it may be required to insert blank characters (spaces) as necessary if the distance between two words is greater than the “word_margin” (W), as a blank between words might not be represented as a space but indicated by the positioning of each word. [5]

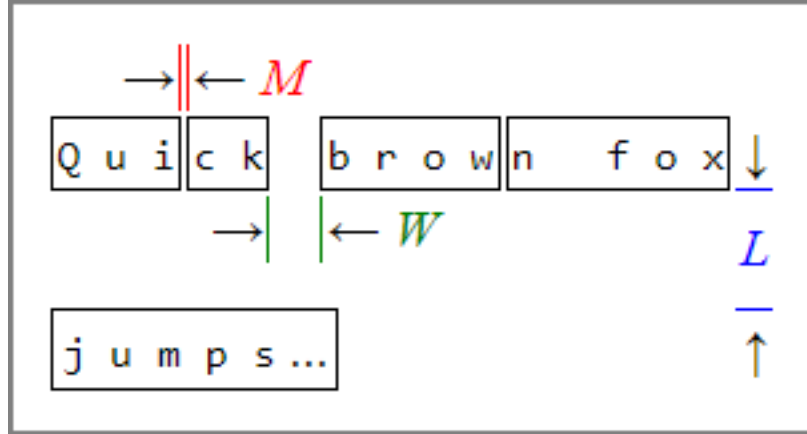


Fig. 1. Each value is specified not as an actual length, but as a proportion of the length to the size of each character in question.

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites.[6] Web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. Nowadays, most article databases (e.g. sciencedirect.com) share the content of the article on their web page, as well as downloadable pdf documents of the articles.

There is a similar study in the literature. This study was conducted on randomly selected PDF documents on various topics at “*arxiv.org*” and has an F_1 rate of 0,99 in sentence extraction and 0,96 in paragraph extraction, respectively [7].



Fig. 2. A graphical illustration of web scraping

2 Material and Methods

In this study, articles compatible with a sample search word (“covid”) selected from the “sciencedirect.com” database were searched using the web scraping technique. Among the search results, articles that also presented the article content as a web page and were “free to access” were used. The contents of these articles were captured with the “crawler” which we designed, and all text contents of each article were successfully obtained. A folder name was created with the “pii” number of each article, and the article titles and the text content of these titles were written to the disk as a “txt” file. In the next step, text mining analyzes, and natural language processing techniques were used to remove sections from the article titles that are not directly related to the main text of the article. After applying these analyses, the downloaded articles were divided into two main categories: “regular” and “irregular”.



Fig. 3a. A graphical illustration of our “crawler” build under web scraping technique
3b. A graphical illustration of our “crawler” pulled and organized data from web source

It has been observed that there is a pattern that can be easily discovered by the “human eye” in the articles whose category is called “regular”. This pattern corresponds to the use of a simple numbered bullet point. Namely, the sections of each “regular” article that we are interested in and need to read are stated as a numbered list (For example, 1. Introduction, 2. Materials and Methods, etc.). However, this simple pattern does not appear in “irregular” articles. For this reason, “the human eye method” does not work for these articles, and the IMRaD rule also returns incomplete results for such articles. “IMRaD” format refers to a paper that is structured by four main sections: Introduction, Methods, Results, and Discussion. After it was observed that the IMRaD rule did not yield results in dealing with “irregular” articles, the words forming the section titles of all articles were subjected to clustering process after going through the “bag-of-words” text mining process. The bag-of-words model is a model of text which uses a representation of text that is based on an unordered collection (or “bag”) of words.

“*Wordcloud*” graphics of the cluster elements obtained because of the clustering process were produced and the number of elements was examined. In addition, the similarities of the words representing the most common elements in each cluster and the words in the IMRaD rule were calculated. When the similarity findings were examined, it was observed that there was no standard in naming the article sections (e.g., method, material and methods, methodology). Additionally, some section names can be used together. To overcome such problems and to make our article reader application work more efficiently, it was considered to use a “*computer-based classifier*” that includes not only the article section titles but also the text content within the section title in the analysis. A classifier that works with natural language processing techniques and *K-means* algorithms was created using the “*sci-kit*” library. This classifier obtained the readable sections of all articles with a certain success rate. However, only “regular” articles were used for comparison. Because the “human eye method”, which we will consider as the golden ratio for such articles, has been a definite success. Finally, a comparison was made between the classifier we developed and the human eye method, which we accept as the golden rule.

3 Results

The first finding obtained in this study is that the most repeated words among the section titles of all articles (n = 716) which are obtained with the web scraping technique (See Fig 4).

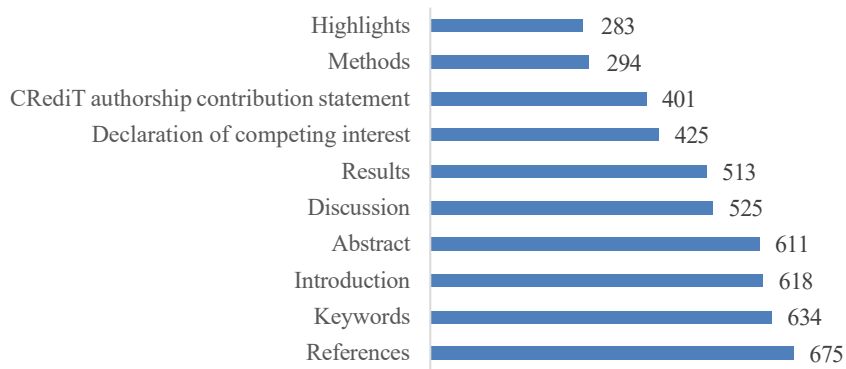


Fig. 4. The top ten words that appear most frequently among the section titles in all articles.

The findings obtained in the second step show us that the contents of all articles cannot be read successfully by using only section headings. Except for the first cluster, all other clusters match a section name. However, the headings of most sections that make up the body text of an article are collected in the first group (See Fig 5a).

The reason why the number of clusters was determined as “5” is related to the IMRaD rule [8,9]. However, it was also observed that among the most recurring section headings were some words that were not included in the IMRaD rule (See Fig 4).

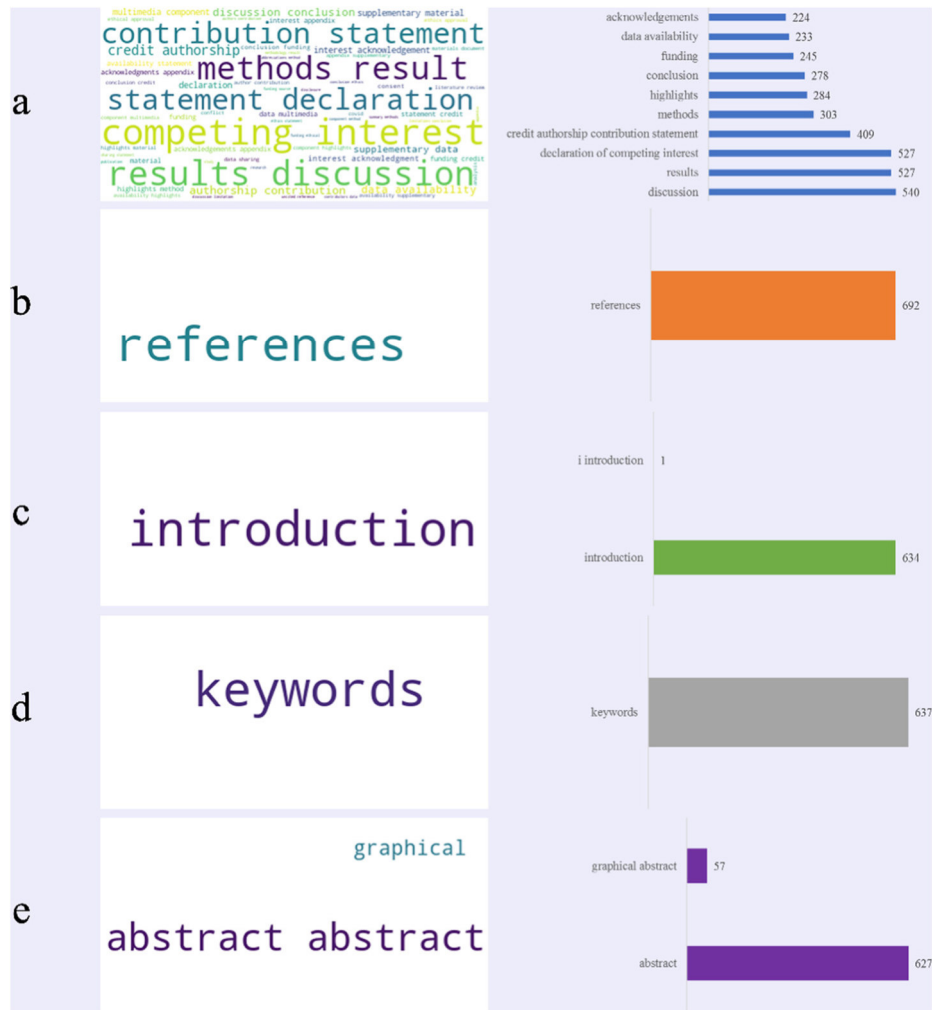


Fig. 5. Text mining analysis results applied to cluster elements obtained because of the clustering method.

Table 1. Words like the word "*Introduction*" in the IMRaD rule

Words	Similarity value	Count (n)
<i>Introduction</i>	1,00	639
<i>Introductions</i>	0,93	1
<i>Implementation</i>	0,84	3
<i>Presentation</i>	0,83	13
<i>Formulation</i>	0,82	2
<i>Interpretation</i>	0,80	3
<i>Revision</i>	0,80	1
<i>Specification</i>	0,80	1
<i>Adaptation</i>	0,80	1
<i>Publication</i>	0,79	33
Total word count		697
Section word ratio in all articles		697/756=92,20%

Table 2. Words like the word "*Methods*" in the IMRaD rule

Words	Similarity value	Count (n)
<i>Methods</i>	1,00	499
<i>Techniques</i>	0,91	1
<i>Method</i>	0,89	45
<i>Methodology</i>	0,84	66
<i>Technique</i>	0,83	1
<i>Approaches</i>	0,80	1
<i>Strategies</i>	0,79	2
<i>Processes</i>	0,77	1
<i>Algorithms</i>	0,75	1
<i>Mechanisms</i>	0,74	2
Total word count		619
Section word ratio in all articles		619/756=81,88%

Table 3. Words like the word "Results" in the IMRaD rule

Words	Similarity value	Count (n)
<i>Results</i>	1,00	587
<i>Outcomes</i>	0,77	5
<i>Outcome</i>	0,74	1
<i>Result</i>	0,73	11
<i>Findings</i>	0,73	25
<i>Resultados</i>	0,71	3
<i>Conclusions</i>	0,66	162
<i>Coefficients</i>	0,66	1
<i>Estimations</i>	0,66	1
<i>Statistics</i>	0,65	7
Total word count		803
Section word ratio in all articles		803/756=106,22%

Table 4. Words like the word "Conclusion" in the IMRaD rule

Words	Similarity value	Count (n)
<i>Conclusion</i>	1,00	314
<i>Conclusions</i>	0,91	162
<i>Conclusive</i>	0,85	1
<i>Interpretation</i>	0,73	3
<i>Implication</i>	0,71	2
<i>Objective</i>	0,71	6
<i>Findings</i>	0,71	25
<i>Intervention</i>	0,71	1
<i>Concluding</i>	0,71	5
<i>Acknowledgement</i>	0,70	25
Total word count		544
Section word ratio in all articles		544/756=71,96%

Table 5. Words like the word "Discussion" in the IMRaD rule

Words	Similarity value	Count (n)
<i>Discussion</i>	1,00	604
<i>Discussions</i>	0,91	15
<i>Suggestions</i>	0,85	5
<i>Presentation</i>	0,73	13
<i>Presentations</i>	0,71	2
<i>Question</i>	0,71	3
<i>Discourse</i>	0,71	1
<i>Consultations</i>	0,71	1
<i>Interdisciplinary</i>	0,71	2
<i>Interventions</i>	0,70	1
Total word count		647
Section word ratio in all articles		647/756=85,58%

Table 6. Distribution of "regular" and "irregular" article counts distribution.

Article Type	Count (Percentage)
<i>Regular</i>	462 (61,11%)
<i>Irregular</i>	254 (38,89%)
<i>Total</i>	756 (100%)

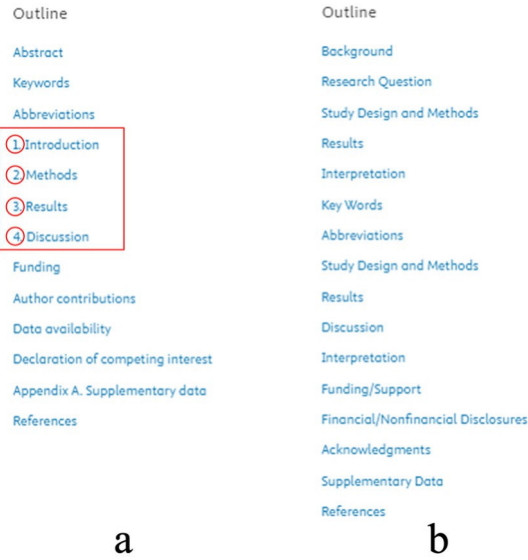


Fig. 6. Outline structure of a "regular" (6a) and "irregular" (6b) article definition in our study.

Table 7. Distribution of each classifier “readable section” word counts.

Cluster ID	Word Count (n)
0	8
1	6
2	586
3	48
4	6
5	5
6	3
7	617
8	14
9	13

Classifiers two and seven were included in further analysis (See Table 7).

Table 8. Number 2 classifier text coverage ratio analysis results.

Coverage type	Count (Coverage ratio) *
>100%	10 (2%)
100%	117 (25%)
75-<100%	165 (36%)
50-<75%	162 (35%)
25-<50%	8 (2%)
0-25%	0 (0%)
Total	462 (100%)

* The value defined as “coverage ratio” can be considered and interpreted as F₁ ratio

Table 9. Number 7 classifier text coverage ratio analysis results.

Coverage type	Count (Coverage ratio) *
>100%	297 (64%)
100%	72 (16%)
75-<100%	41 (9%)
50-<75%	44 (10%)
25-<50%	8 (2%)
0-25%	0 (0%)
Total	462 (100%)

* The value defined as “coverage ratio” can be considered and interpreted as F₁ ratio

4 Conclusion

When the findings in Table 1-5 are examined, it can be easily understood that not all articles within the scope of the study fully comply with the IMRaD rule. 92% of the articles contain an “*introduction*” or a section title with similar meaning (see Table 1). 82% of the articles contain a “*methods*” or a section title with similar meaning (see Table 2). 72% of the articles contain a “*conclusion*” or a section title with similar meaning (see Table 4). 86% of the articles contain a “*discussion*” or a section title with similar meaning (see Table 5). These findings show that even if similar words are included with the section names, the body texts of all articles cannot be read successfully by only considering the IMRaD rule. Researchers can examine Table 6 for the distribution of the articles examined within the scope of the study according to whether they have a “*regular*” or “*irregular*” structure and see Figure 6 for a clearer understanding of what we mean by “*regular*” or “*irregular*” structure. In Tables 8 and 9, article reading success rates (F_1 rate as used in the literature) using the department names of the two most successful clusters are given. It was deemed more appropriate to analyze these findings by making them categorical. When Tables 8 and 9 are examined, it is seen that the computerized reading method includes parts of some articles other than the body text within the scope of reading. The meaning of the >100% ratio in both tables expresses the situation described in the previous sentence.

This study is the preliminary step of a project that is being developed. When this study is completed, it is aimed to create a tool like the contents available at <https://www.ncbi.nlm.nih.gov/research/bionlp/Tools/>. Our current application can successfully read and obtain the text content of every accessible article on the sciencedirect.com website that has been converted to HTML format with the Web scraping technique. Our current clustering analysis among article topics, which is the next step, currently works with 96% F_1 ratio in the 50%-100% success range. We aim to develop an application that can respond to different analyzes by improving the success rate of this clustering analysis. For example, the application can read an article you want for you and tell you what its main idea is, and it can do this by analyzing the words in the article with natural language processing techniques to write a sentence that best summarizes the article.

References

1. Adobe Systems Incorporated (November 2006). "PDF Reference" (PDF). 1.7 (6th ed.). Archived from the original (PDF) on October 1, 2008. Retrieved January 12, 2023.
2. Warnock, J. (October 14, 2004) [Original date 5 May 1995]. "The Camelot Project" (PDF). Archived (PDF) from the original on July 18, 2011.
3. "What is a PDF? Portable Document Format | Adobe Acrobat DC". Adobe Systems Inc. Archived from the original on January 30, 2023. Retrieved January 12, 2023.
4. Kuhlman, Dave. "A Python Book: Beginning Python, Advanced Python, and Python Exercises". Section 1.1. Archived from the original (PDF) on 23 June 2012.
5. <https://www.unixuser.org/~euske/python/pdfminer/index.html>.

6. Thapelo, Tsaone Swaabow; Namoshe, Molaletsa; Matsebe, Oduetse; Motshegwa, Tshiamo; Bopape, Mary-Jane Morongwa (2021-07-28). "SASSCAL WebSAPI: A Web Scraping Application Programming Interface to Support Access to SASSCAL's Weather Data". *Data Science Journal*. 20: 24. doi:10.5334/dsj-2021-024. ISSN 1683-1470. S2CID 237719804.
7. Yu, C., Zhang, C., & Wang, J. (2020). Extracting Body Text from Academic PDF Documents for Text Mining. *arXiv [Cs.IR]*. Retrieved from <http://arxiv.org/abs/2010.12647>
8. P. K. R. Nair and V. D. Nair (2014). *Scientific Writing and Communication in Agriculture and Natural Resources*. Springer. p. 13.
9. Sollaci LB, Pereira MG (July 2004). "The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey". *Journal of the Medical Library Association*. 92 (3): 364–7. PMC 442179. PMID 15243643.

Nesnelerin İnterneti ve Giyilebilir Sensörlerle Yaşlı Bireylerin Sağlık Durum Değerlendirilmesi

Mustafa Yazıcı 1, Sedat Görmüş 1, Mustafa Ulutaş 1, Mukadder Okuyan 2

1 KTÜ Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

2 KTÜ Tıp Fakültesi, Fizyoloji Bölümü

Özet

Giyilebilir sensör teknolojilerinin gelişmesiyle birlikte bireylerden fizyolojik sinyallerin alınması kolaylaşmıştır. Bu sensörlerle alınan EKG (Elektrokardiyografi), EEG (Elektroensefalografi) ve diğer fizyolojik verilerin işlenmesi ve anlamlı sonuçların elde edilmesi, sağlık kurumları uzağında bulunan bireylerin sağlık durumları hakkında hızlı bilgi edinilmesini sağlar. Ölçülen verilerin analizi yapılarak kısa sürede yaşamsal önemi olan bilgiler elde edilebilir. Bu ölçümlerin nesnelerin interneti (IoT) ile birleştirilmesi gerçek zamanlı ve sürekli olarak izlenmesine olanak sağlayacaktır. Büyük yer kaplayan laboratuvar cihazları yerine giyilebilir cihazların kullanılması, bu alanda büyük kolaylık sağlayacak ve uzun süreler boyunca veri toplanması imkânı kazandıracaktır. Nesnelerin İnterneti cihazlarına kablosuz veri aktarımı, kullanıcılara mobilite kazandırırken, toplanan verilerin uzak makinelerde işlenmesine de olanak tanır. Bu çalışmada böyle bir sistem tasarımı için Bluetooth 5.2 teknolojisi önerilmiştir. İlk olarak, güvenli ölçüm yapabilen sensörler temin edilmeli ve ardından bu sensörlerden elde edilen veriler, IoT cihazlarına aktarılmayı amaçlamaktadır. Toplanan verilerin merkezi birimde veya bulut sistemlerinde saklanması, verilerin güvenli ve organize bir şekilde depolanmasını ve analiz edilmesini mümkün kılacaktır. Verilerin hangi kişiye ait olduğu bilgisi anonim tutulmalıdır. Veriler üzerinde özellik çıkarma ve sınıflandırma işlemleri yapılarak sağlık durumuna makine öğrenmesi ile karar verilecektir. Bu çalışmada EKG ve EEG ölçümlerinin nesnelerin interneti ile birleşmesiyle sağlık durumunun, bireyleri yormadan, giyilebilir sensörler ile ölçülmesi için protokol önerilmiştir. İlerleyen süreçte gerçek ölçümlerin alınması planlanmaktadır. Yaşlı ve mobilite sorunlu bireylerde sağlık durumunun izlenmesi için yenilikçi yaklaşım önerilmiştir.

Giriş

Yaşlı insan nüfusunun, 2015'te %8,5 iken 2030'lu yıllarda %12 civarında ve 2050 yılında %16 olması beklenmektedir [1]. Yaşlılıkla birlikte mobilite sınırlı bireyler için ortaya çıkan sağlık sorunlarının izlenmesi, ilerlemeden önlem alınabilmesi ve de sağlık kurumlarında öncelikli hizmet bulması kıymetlidir. Bu amaçla nesnelerin interneti ile giyilebilen pek çok ürün ve teknoloji geliştirilmiştir [2]. Bu çalışmada sağlık kurumlarına ulaşabilme zorluğu ve fırsatı bulamayan hastaların uzaktan izlenmesi amacıyla, nesnelerin interneti mimarisi önerilmesi amaçlanmıştır.

Nesnelerin internetinin, akıllı evler ve akıllı şehirler gibi alanlarda kolaylaştırıcı uygulamaları bulunmaktadır. Nesnelerin interneti sayesinde sağlık problemi yaşayan bireylere evinden dışarı

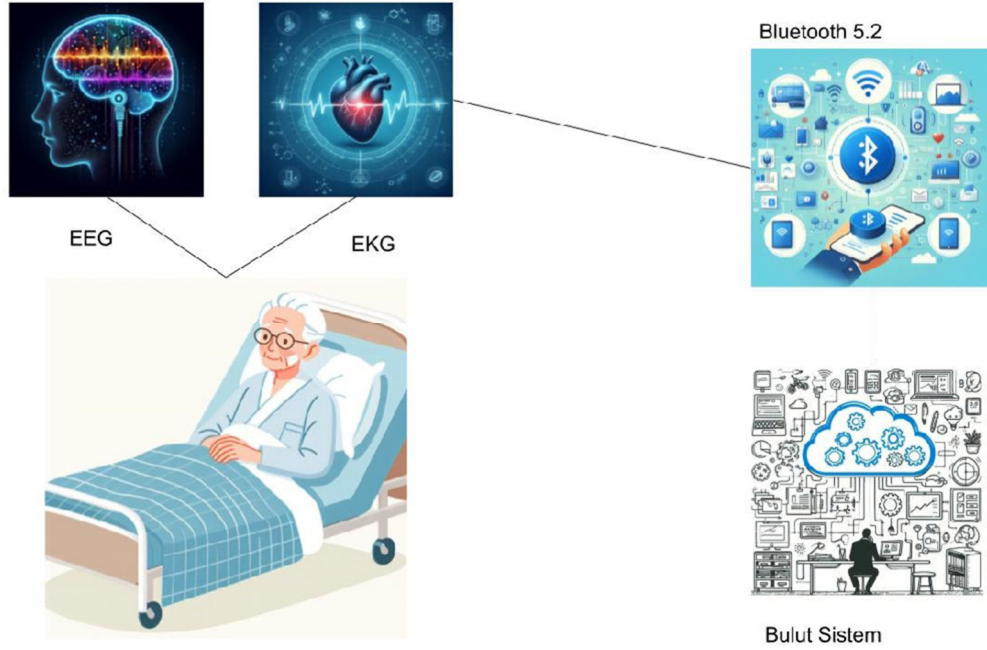
çıkmadan, temel sağlık verileri alınabilmesi mümkün kılınacaktır. Bu verilerin doktorlarla veya bireyin yakın çevresi ile paylaşılması mümkün olacaktır. Özellikle immobil hastaların izlenmesi için oldukça avantajlı bir yöntemdir ve de IoT teknolojileri bu alanda önemli katkılar sunmaktadır. Giyilebilir teknoloji, hastane ortamı dışında hasta sağlığını izlemeye yönelik bir araç olarak kabul edilmektedir [3]. Özellikle son yıllarda yaygınlaşan akıllı saatler ve giyilebilir teknolojiler ile sağlık ölçümlerinin anlık olarak izlenebilmesi ve kaydedilebilmesi hizmeti sunulmaktadır. Kalp atış hızı ve EKG sinyalleri ile hastanın izlenmesi önemli bilgiler sunmaktadır. Sağlık hizmetlerinde nesnelerin internetinin kullanımı günümüzde kullanılan teknolojileri iyileştirebilir ve sağlık hizmetlerinin her alanında maliyetleri düşürebilir [4].

EKG sinyallerinin uzaktan, sürekli ve gerçek zamanlı olarak alınarak işlenmesi, hastane içi kayıtlamaya göre hastaların durumlarının daha sıkı bir şekilde gözlemlenmesine olanak tanımaktadır [5]. Literatürde EEG, EKG veya her iki sinyalin birlikte ölçüldüğü nesnelerin interneti ile yapılan çalışmalar vardır [6, 7, 8, 9, 10]. Yaşlı bireyin sağlık durumunun anlaşılması ile yakınlarının kendisiyle ilgilenmesi sağlanabilir. Bireyin yardıma ihtiyacı olması durumunda gerekli önlemler alınabilir.

IoT giyilebilir cihazları, kişilerin kendi sağlık durumlarının takibini ve acil durumlarda hekime ulaşabilmeleri ve de hekimlerin hastaları uzaktan izlemelerini sağlar. Bu teknoloji sayesinde kişiye konfor ve istediği yerde yaşama lüksü kazandırır.

Önerilen Nesnelerin İnterneti Modeli

Çalışma kapsamında gerçekleştirilen sistemin veri gereksinimi analiz edildiğinde Bluetooth 5.2 teknolojisinin giyilebilir cihazlardan internete veri aktarımı için optimum çözümü sunabileceği öngörülmüştür. Bu bağlamda giyilebilir cihazdan alınan veriler uç veri işleme birimine aktarılarak bu noktada verilerin ön işleme/anlamlandırılması sağlanacaktır. Ön işleme yapılan verilerin saklanması ve uzun vadede analiz edilmesi için ise bulut tabanlı bir sistem tasarımının uygun olduğu sonucuna varılmıştır. Bu bağlamda önerilen sisteme ait mimari Şekil 1'de verilmiştir.



Şekil 1. Önerilen sistem mimarisi

Sonuç

EKG ve EEG cihazlarının, IoT sistemine eklenmesiyle, sağlık alanında özellikle sağlık kurumuna ulaşım sorunu yaşayan bireylerin gerçek zamanlı olarak gözlemlenmesine olanak sağlayacaktır. IoT ile izlenen ve değerlendirilen bireylerin EKG ve EEG verileri, temel sağlıkdurumu anlık olarak görülecek ve gerekli görülürse bireye sağlık desteği sunulmasına karar verilecektir. Sağlıksız gözüken bireylere, çevresinden yardım talep etmesi veya hastaneden yardımcı olunması imkânı tanıyacaktır. Fizyolojik verilerin alınmasından sonra çok farklı mimariler kullanılarak gönderilebilir. Bu çalışmada optimum performans için veri gönderiminde mimari olarak Bluetooth 5.2 önerilmektedir.

Kaynakça

- 1 Dorri S, Zabolinezhad H, Sattari M. The Application of Internet of Things for the Elderly Health Safety: A Systematic Review. *Adv Biomed Res.* 2023 Apr 27;12:109. doi: 10.4103/abr.abr_197_22. PMID: 37288027; PMCID: PMC10241622.
- 2 Matayong, S., Jetwana, K.W., Choksuchat, C. et al. IoT-based systems and applications for elderly healthcare: a systematic review. *Univ Access Inf Soc* (2023).
- 3 Pantelopoulos A., Bourbakis N.G., A survey on wearable sensor-based systems for health monitoring and prognosis (2010) *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 40 (1), art. no. 5306098, pp. 1 - 12.
- 4 Nico Surantha, Prabadinata Atmaja, David, Maulana Wicaksono, A Review of Wearable Internet-of-Things Device for Healthcare, *Procedia Computer Science*, Volume 179, 2021, Pages 936-943, ISSN 1877-

0509,

5 Neri L, Oberdier MT, van Abeelen KCJ, Menghini L, Tumarkin E, Tripathi H, Jaipalli S, Orro A, Paolucci N, Gallelli I, Dall'Olio M, Beker A, Carrick RT, Borghi C, Halperin HR. Electrocardiogram Monitoring Wearable Devices and Artificial-Intelligence-Enabled Diagnostic Capabilities: A Review. *Sensors (Basel)*. 2023 May 16;23(10):4805. doi: 10.3390/s23104805. PMID: 37430719; PMCID: PMC10223364.

6 Wan, J., A. A. H. Al-awlaqi, M., Li, M. et al. Wearable IoT enabled real-time health monitoring system. *J Wireless Com Network* 2018, 298 (2018).

7 B. Lal, M. H. Conde, P. Corsonello and R. Gravina, "Secure and Energy-Efficient ECG Signal Monitoring in the IoT Healthcare using Compressive Sensing," 2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech), Abu Dhabi, United Arab Emirates, 2023, pp. 0333-0339

8 Nisha Raheja, Amit Kumar Manocha, IoT based ECG monitoring system with encryption and authentication in secure data transmission for clinical health care approach, *Biomedical Signal Processing and Control*, Volume 74, 2022, 103481, ISSN 1746-8094.

9 Mora H, Gil D, Terol RM, Azorin J, Szymanski J. An IoT-Based Computational Framework for Healthcare Monitoring in Mobile Environments. *Sensors (Basel)*. 2017;17(10):2302.

10 Wang, L.; Song, F.; Zhou, T.H.; Hao, J.; Ryu, K.H. EEG and ECG-Based Multi-Sensor Fusion Computing for Real-Time Fatigue Driving Recognition Based on Feedback Mechanism. *Sensors* 2023, 23, 8386.

Gamified Facial Exercises Software for Facial Paralysis Patients: A Fun and Effective Method

Hüsnü Canbolat¹ [0000-0001-5932-5633] and Sabiha Kaya² [0000-0002-0883-4486]

¹ Trabzon University, Center for Performance Measurement and Talent Selection in Sports,
Trabzon, Turkey

² Trabzon University, Faculty of Sports Sciences, Trabzon, Turkey
husnucanbolat@trabzon.edu.tr

Abstract

Facial paralysis, also known as Bell's palsy, is a common neurological disorder affecting millions of people worldwide [1]. Bell's palsy disrupts the function of the facial nerve (VII cranial nerve), leading to weakness or paralysis of the facial muscles [2]. Various factors such as trauma, tumors, infections, and autoimmune diseases can cause Bell's palsy [3]. Symptoms of facial paralysis include asymmetry in facial expressions, inability to close the eyelid, drooping at the corner of the mouth, and impairment in the sense of taste [4].

Treatment methods for facial paralysis include medication, physical therapy, and surgery [5]. Medication therapy involves drugs such as steroids and antiviral medications [6]. Physical therapy plays a significant role in exercising and strengthening the facial muscles [7]. Traditional physical therapy methods include treatments like exercises and massage [8]. Surgery may be used in cases of facial paralysis that do not respond to other treatments [9]. Physical therapy is a highly effective method in the treatment of facial paralysis. However, traditional physical therapy methods can be time-consuming and monotonous for patients, making adherence to the treatment challenging [10]. Gamification is an approach that uses game mechanics and design principles to make treatment more enjoyable and motivating [11]. Gamification is increasingly being used in healthcare services, and evidence suggests that it can improve patient adherence and treatment outcomes [12].

MediaPipe, developed by Google Research, is a library capable of tracking 468 points on the face in real-time using artificial intelligence. Trained with deep learning models, MediaPipe performs low-latency and high-efficiency facial tracking operations thanks to optimized neural networks and fast processing algorithms. This library offers a wide range of applications in face recognition and verification, augmented reality, expression and emotion analysis, and health and aesthetics. By integrating with artificial intelligence, MediaPipe enables precise tracking of facial expressions [13].

In this study, a software integration was developed for facial paralysis patients to control video games using facial movements for home exercises. The software was developed using the Python programming language and the MediaPipe library. By utilizing gamification principles, the software aims to make the treatment process more enjoyable and motivating. However, the effectiveness of the software needs to be evaluated through more comprehensive clinical studies. By addressing the software's limitations, providing personalized exercise plans for each patient, and enabling follow-up by physiotherapists, it can become a more effective treatment tool. The MediaPipe library is used to track facial points in real-time with artificial intelligence, utilizing 468 points on the face [13]. The movements of the facial points are assigned to control the in-game character. For example, when the eyebrows are raised, the character moves up; when the mouth is opened, the character moves to the right; when the lips are pursed, the character moves to the left; and when the eyes are closed, the character moves down.

For future studies, the contribution of the software to facial paralysis treatment processes should be measured through clinical trials. Additionally, various efforts should be made to address the software's limitations. In this context, creating personalized exercise plans for each patient, enabling follow-up by physiotherapists and updating exercise plans, adapting the software for different languages and platforms, and other improvements are hoped to make it a more effective and accessible treatment method for facial paralysis patients.

References

1. Choi, J., & Kim, S. H. (2016). A review of facial paralysis. *Journal of clinical neuroscience*, 23(12), 1826-1835.
2. Furlan, M., Marsili, S., & Piazza, F. (2012). Bell's palsy: Clinical features and pathogenesis. *Journal of the Peripheral Nervous System*, 17(1), 33-42.
3. Jung, Y. S., & Park, H. W. (2011). Bell's palsy: Diagnosis and treatment. *Korean journal of neurology*, 17(4), 249-256.
4. Jeon, H. S., & Park, C. H. (2017). Clinical characteristics and treatment outcomes of Bell's palsy: A retrospective analysis of 1,023 patients. *Yonsei Medical Journal*, 58(5), 636-642.
5. Ferri, C., & Pau, A. (2012). Bell's palsy: Diagnosis and treatment. *American family physician*, 86(11), 787-793.
6. Benzaquen, J., & Cohen, S. J. (2010). Steroids for Bell's palsy. *Cochrane database of systematic reviews*, (12), CD005842.
7. DeOliveira, A. C., & Pereira, L. M. (2015). Physical therapy for Bell's palsy: A systematic review. *Arquivos de fisioterapia*, 42(1), 10-17.
8. Lee, H. J., & Park, H. W. (2014). Effects of conventional physical therapy versus electrical muscle stimulation on facial function in patients with Bell's palsy: A randomized controlled trial. *Journal of physical therapy science*, 26(2), 237-242.
9. Kim, J. H., & Kim, S. H. (2017). Surgical treatment for facial paralysis. *Journal of the Korean Neurosurgical Society*, 50(4), 417-426.
10. Dlugosz, D. J., & Madarame, H. (2011). Patient compliance in physical therapy: A review of the literature. *Journal of the American Physical Therapy Association*, 91(12), 1509-1523.

11. Deterding, S., Dixon, D., & Khaled, R. (2011). From game mechanics to game design: A know-how guide for game designers. *ACM SIGCHI Bulletin*, 34(4), 24-48.
12. Seaborn, C., & Petrie, C. (2015). Can you play yourself better? A systematic review of the use of game-like elements in digital health interventions. *JMIR mHealth and uHealth*, 1(1), e12.
13. MediaPipe. "Face Landmark Detection with MediaPipe". Google AI. Access Date: May 10, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker

Alterations in Gene Expression Related to Neuroendocrine Functions in Long-Term Endurance and Resistance Training Elderly

Arif Kamil Salihođlu¹[0000-0003-3864-9377]

¹ Karadeniz Technical University, Faculty of Medicine, Department of Physiology,
Trabzon, Türkiye
aks@ktu.edu.tr

Abstract.

Introduction: The benefits of endurance and resistance training in elderly individuals include improved muscle strength, cardiovascular health, and overall physical fitness. However, the molecular mechanisms, particularly related to neuroendocrine functions, are not yet fully understood. This study aims to elucidate the potential impacts of these exercise modalities on gene expression levels using bioinformatics tools.

Methods: The study utilized the GSE165630 dataset¹ from the Gene Expression Omnibus database. RNA-seq gene expression levels were re-analyzed using R, encompassing sedentary control ($n=5$), full resistance exercise ($n=2$), partial resistance exercise ($n=2$), full endurance exercise ($n=2$), and partial endurance exercise ($n=3$) groups of elderly men aged 65-81. Gene set enrichment analyses were conducted using the Gene Ontology (GO) web tool. Wald test and Benjamini-Hochberg correction were applied, with $p<0.05$ considered significant.

Results: Compared to the control group, all exercise groups exhibited downregulation of genes including proenkephalin (PENK), cystatin-C (CST3), adrenomedullin (ADM), neuromedin B (NMB), prokineticin-1 (PROK1), chromosome-4 open reading frame-48 (C4orf48), periostin (POSTN), secretogranin II (SCG2), leptin (LEP), adiponectin (ADIPOQ), retinoic acid receptor responder-2 (RARRES2), and C-C motif chemokine ligand-5 (CCL5) ($p<0.05$). Conversely, neuropeptide-Y receptor Y6 (NPY6R), apelin (APLN), and clock circadian regulator (CLOCK) genes were upregulated ($p<0.05$).

Conclusion: This *in silico* study provides evidence of significant changes in the expression levels of key genes involved in neuroendocrine signaling pathways. These findings offer new insights into the complex molecular foundations following endurance and resistance training, highlighting primarily neuroendocrine downregulations in shaping physiological responses to these exercise modalities.

Keywords: Endurance exercise, resistance exercise, neuroendocrine, bioinformatics.

References

1. De Sanctis P. et al. *Int J Mol Sci.* 2021; 22(4): 1539. doi: 10.3390/ijms22041539.

Hastane Bilgi Yönetim Sistemleri Merkezinin Yeniden Yapılandırılması; Hasta, Hasta Yakını ve Çalışan Memnuniyetinin Değerlendirilmesi

İsmail YILDIZ¹[0000-0001-5505-838X]

¹ Dicle Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Diyarbakır, TÜRKİYE
iyildiz@dicle.edu.tr

Abstract.

Bu çalışma ile, Hastanemizde sunulan sağlık hizmetinin en iyi şekilde verilebilmesi oluşturulmaya çalışılan kalite ve standartların tesis edilmesi ve korunması, hastanelerde, gelir ve giderlerin izlenmesi, gelir kaçaklarının tespiti ve önlenmesi, kaynakların yerinde kullanılması ve israfın önlenmesi, hastane yönetimine verilecek önemli kararlarda bilgi desteğinin sağlanması, sağlanan bilgi desteğiyle ileriye yönelik doğru hedef belirlenmesi ve bilimsel çalışma yapan tüm paydaşlarımıza doğru, hızlı ve güvenilir ve Sağlık Bakanlığı standardizasyonlarına her an hazır bir sunucu altyapısı kurulması sağlandı.

Bu araştırmada nitel ve nicel araştırma yöntemi benimsenmiştir. Dicle Üniversitesi Hastanelerindeki mevcut Hastane Bilgi Yönetim Sistemi altyapısında kullanılan; Ana Sunucu, yedek sunucu, storage, Veritabanı yazılımı, yedekleme sistemi, Bilgi Güvenliği, Kullanıcı profili, Hastane yazılımı vb alanları kapsayan, yaklaşık 3936 port üzerinde bağlantı yapan 474 son kullanıcıya; Sağlık Bakanlığı İdari ve Mali İşler Dairesi Başkanlığı Hastane Bilgi Yönetim Sistemleri Alım Kılavuzu ve ilgili yönetmelikleri ile ISO-9001&ISO-27001 ve Sağlık Bakanlığı Sağlık Kalite Standartları(SKS) Hastane Versiyonunun gereksinimlerine göre öntest anketi uygulanarak gerekli analizler yapıldı.

Hastanelerimizin de bu amaç doğrultusunda kaliteli ve güvenli sağlık hizmeti sunmasına katkı sağlaması için; Teknolojik ömrünü doldurmuş, günümüz uygulamalarının birçoğuna elektronik olarak cevap veremeyen mevcut sistemlerin (sunucu ve disk ünitesi) güncellenmesi, kalite standartlarını ve sağlık hizmetlerinde HBYS den beklenen ihtiyaçları karşılar düzeye getirilmesi, bu platforma uygun eksik lisansların tamamlanması sağlanarak; Dicle üniversitesi hastanelerinde geçmişten günümüze ve günümüzden geleceğe kesintisiz bilgi transferi ve bilgi akışını, lisanslı Veri Tabanı Yönetim Sistemi(VTBS) ile sağlandı.

Hastalara tam, zamanında ve doğru tanı-teşhis koyabilmek, bilimsel çalışmalarda, yönetsel faaliyetlerde ve planlamalarda kullanılmak üzere veri tabanının karmaşık varyasyonlarına girmeden sonuç alabilmek, çalışan memnuniyetini artırmak, sağlık bakanlığı ve diğer kural koyucuların istemlerini tam ve zamanında yapabilmek için bütün bu ihtiyaçları optimum düzeyde lisanslı VTBS yazılımı ve güçlü veritabanı sunucuları ve ortak harici veri depolama üniteleri ile sağlandı.

Keywords: Hastane Bilgi Yönetim sistemleri, Veri Tabanı Yönetim Sistemi,Sürdürülebilirlik

References

1. Hastane Bilgi Yönetim Sistemleri Alım Kılavuzu, T.C. Sağlık Bakanlığı İdari ve Mali İşlerDairesi Başkanlığı, Ağustos 2010
2. T.C. Sağlık Bakanlığı Bilgi İşlem Daire Başkanlığı Sağlık-Net Entegrasyonu
3. Güvenli Sistem Odaları için İhtiyaç Tanımı Ağ ve Sistem Destek Grubu
4. Hacettepe Üniversitesi / Bilgi İşlem Dairesi Başkanlığı
5. Sistem Odaları İklimlendirme Sistemleri Trakya Üniversitesi Fen Bilimleri EnstitüsüBilgisayar Mühendisliği Bölümü
6. Gazi Üniversitesi Sistem Odası Kullanım Yönergesi
7. AK,Bilal,Sağlık Kurumlarında Bilgi Yönetim Sistemleri, (Mimoğraf) Hacettepe Üniversi-tesi,Sağlık Bilimleri Enstitüsü,Ankara,2000
8. AK,Bilal,Hastane Bilişim Sistemleri (Mimoğraf),Gazi Üniversitesi,Sosyal Bilimler Ensti-tüsü,Hastane İşletme Yönetimi Programı,Ankara,2003
9. AK,Bilal,Hastane Bilgi Yönetim Sistemleri, (Mimoğraf) Hacettepe Üniversitesi,Sağlık İda-resi Yüksekokulu,Ankara,1999
10. AK,Bilal,Hastane Bilgi Sistemleri,SİD_ACC Seminer Notları,Ankara,Hilton,1997
11. T.C. Sağlık Bakanlığı, Hastane Bilgi Yönetim Sistemleri Alım Kılavuzu, Ağustos 2010
12. T.C. Sağlık Bakanlığı Bilgi İşlem Daire Başkanlığı Sağlık-Net Entegrasyonu için HastaneBilgi Sistemlerinin Temel Gereksinimleri PDF dosyası, Ankara, Ekim 2007

Convolutional Neural Networks Utilizing Image Representation of Amino Acid Sequence for Variant Effect Prediction

Gülbahar Merve ŞILBİR¹[0000-0003-0321-7259], Burçin KURT¹[0000-0001-5781-2382]

¹ Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey
gmerve.cakmak@gmail.com
burcinnkurt@gmail.com

Abstract. Proteins are one of the molecules that manage cellular processes in living organisms. Since protein-protein, DNA-protein and RNA-protein interactions play an important role in a large part of biological processes, variants occurring in the regions where these interactions occur can lead to serious consequences on the phenotype. Various supervised learning approaches are applied to determine the association of variants occurring in proteins with a particular disease. In this study, a convolutional neural network-based prediction model is proposed to predict the pathogenicity effect of variants on the phenotype by converting amino acid sequences into two-dimensional images. A protein embedding method using transfer learning (TAPE) was used to create the feature vector. The created feature vector was converted into a square-sized single-channel image and trained with a deep learning algorithm containing a convolutional neural network. In this study, binary classification (benign and pathogenic) was made using missense variants in the BRCA1 protein obtained from the open access ClinVar database as the dataset. Based on the findings, the developed prediction model demonstrated significant success in predicting the pathogenicity effects of variants within the functional regions of the BRCA1 protein on phenotype. Evaluation of the model's prediction results revealed that variants in the pathogen class can be classified with 90% accuracy (0.90 sensitivity). Additionally, the model exhibited strong performance in classifying both benign and pathogenic variants, achieving an AUC value of 0.89. The findings suggest that the developed prediction model may offer potential in classifying BRCA1 variants and assessing their potential pathogenicity.

Keywords: Protein Representation, Convolutional Neural Network, Variant Effect Prediction, Transfer Learning

Makine Öğrenimi Algoritmalarıyla Meme Kanseri Sağkalım Tahmini: Performans Karşılaştırması ve Klinik Uygulamaları

Performance Comparison and Clinical Applications of Machine Learning Algorithms for Breast Cancer Survival Prediction

Yunus Emre KAYAOĞLU¹[0009-0001-6243-706X] ve Uğur ŞEVİK²[0000-0002-2056-9988]

^{1,2}Karadeniz Teknik Üniversitesi, Bilgisayar Bilimleri Bölümü, Trabzon, Türkiye

¹yunusemre.kayaoglu@outlook.com

²usevik@ktu.edu.tr

Abstract. Breast cancer is a prevalent form of cancer that primarily affects women. Nevertheless, the precise cause of breast cancer has yet to be established. Creating a proper diagnosis and treatment strategy for breast cancer is challenging due to its complex nature. Genetic factors and several environmental variables are recognized to play a role in the onset of breast cancer. Early detection expedites the therapeutic course of breast cancer. Hence, it is imperative to assess the efficacy of methods employed in the timely detection and management of breast cancer to enhance patients' welfare and optimize their healthcare outcomes. The objective of the work is to develop a predictive model that determines the survival outcome of patients by utilizing demographic and clinical characteristics in a big dataset through decision tree-based machine learning techniques. In order to achieve this objective, the CatBoost, LightGBM, and XGBoost machine learning algorithms were evaluated using a dataset. According to the collected results, it has been observed that the performance of the approaches in question is quite similar and they have reached a success rate of over 90%.

Keywords: Breast Cancer Diagnosis, Machine Learning, XGBoost, LightGBM, CatBoost

Özet. Meme kanseri, öncelikle kadınları etkileyen yaygın bir kanser türüdür. Bununla birlikte, meme kanserinin oluşumunun kesin bir nedeni henüz belirlenmemiştir. Meme kanseri teşhisi ve tedavi stratejisi oluşturmak, hastalığın karmaşık yapısı nedeniyle zorluk yaratmaktadır. Genetik ve çeşitli çevresel faktörlerin meme kanserinin başlangıcında yol oynadığı bilinmektedir. Erken teşhis, meme kanseri tedavi sürecini hızlandırır. Bu nedenle, hastaların refahını artırmak ve sağlık sonuçlarını optimize etmek için meme kanserinin zamanında tespiti ve yönetiminde kullanılan yöntemlerin etkinliğini değerlendirmek önemlidir. Çalışmanın amacı, büyük bir veri setinde bulunan demografik ve klinik özelliklerin kullanımı sonucu hastaların sağkalım durumlarını belirleyen bir tahmin modeli geliştirmektir. Bu amaçla karar ağaçlarına dayalı makine öğrenimi yöntemleri kullanılacaktır. Sağkalım durumunu tahmin aşamasında XGBoost,

LightGBM ve CatBoost makine öğrenimi algoritmaları elde edilen veri setini kullanarak bir model oluşturmuştur. Oluşan bu modellerin sonuçlarına göre, söz konusu yaklaşımların performanslarının oldukça benzer olduğu ve %90 üzerinde bir başarı oranına ulaştıkları gözlemlenmiştir.

Keywords: Meme Kanseri, Makine Öğrenimi, XGBoost, LightGBM, CatBoost, Sağkalım Durumu

1 Giriş

Meme kanseri dünyada kadınlar arasında en sık görülen malign tümör olup, kadınlarda görülen tüm kanserlerin yaklaşık %30'unu oluşturmaktadır [1]. Ancak, meme kanserinin oluşumunda net bir etken henüz tespit edilememiştir. Meme kanserinin oluşumunda çok sayıda etiyolojik faktörün etkili olmasına karşın kesin etkenin belirlenememiş olması, tanı ve tedavide farklı yaklaşım ve tartışmaları da beraberinde getirmektedir [2]. Genetik faktörler, çeşitli çevresel etkenler ve yaşam tarzı gibi birçok faktör meme kanserinin oluşumunda rol oynar. Erken teşhis, meme kanserinin tedavi sürecini hızlandırmakta ve hastaların daha sağlıklı bir yaşam sürmesi için önemli bir adımdır. Çalışmanın amacı makine öğrenimi algoritmaları kullanılarak elde edilen veri setindeki demografik ve klinik özellikler arasındaki ilişkiyi analiz ederek etkili bir sağkalım tahmini yapılmasına yardımcı olmaktır.

2 Problemin Tanımı

Meme kanseri erken tanı ve tarama yöntemleri: kendi kendine meme muayenesi (KKMM), klinik meme muayenesi (KMM) ve mamografidir [3]. Mevcut geleneksel yöntemlerin sınırlarını aşmak ve hastaların yaşam sürelerini daha doğru bir şekilde tahmin etmek için makine öğrenimi algoritmalarının kullanılması gerekmektedir. Bu durumda, meme kanseri şikayetiyle gelen kişilerin belirli demografik özellikleri (yaş, ırk, medeni durum), tümör özellikleri (tümör evresi, boyutu, derecesi, lenf nodu durumu) ve tedavi yanıtlarını içeren kapsamlı bir veri seti kullanılacaktır. İstenilen sonuç, veri seti kullanılarak makine öğrenimi yöntemlerini eğiterek ve test ederek, kişilerin sağkalım sürelerini daha doğru şekilde tahmin etmek için model geliştirilmek istenmektedir. Bir noktada bu model, hastaların tedavi planlarını bireyselleştirmek ve sağlık hizmetlerini amaca yönelik şekilde yönlendirmek için kullanılabilir. Ayrıca, hangi özneliğin sağkalım durumuna etkisi olduğunu belirlemektir ve belirlenen özneliğin üzerinde çeşitli çalışmaların artırılması beklenmektedir.

3 Yöntem

3.1 Veri Seti

Veri seti, 16 farklı öznelikten oluşmaktadır. Bu öznelikler, meme kanseri hastalarının demografik ve klinik özelliklerini içermektedir. Hastaların yaşları, ırkları ve medeni durumları gibi demografik öznelikler, hastaların sosyo-ekonomik durumunu yansıtırken, meme kanseri tanısında kullanılan klinik öznelikler, hastalığın ciddiyetini ve tedavi seçeneklerini belirler. “Marital Status” özneliği, hastanın medeni durumunu ifade eder ve sosyal destek sistemini ve tedaviye uyumu etkileyebilir. “T Stage” ve “N Stage” öznelikleri, sırasıyla meme tümörünün boyutunu ve yayılma derecesini, bölgesel lenf düğümlerine yayılma derecesini belirtir. “Differentiate” ve “Grade” öznelikleri, kanser hücrelerinin farklılaşma derecesini ve tümörün hücresel derecesini ifade eder. “A Stage”, meme kanserinin ilerlemiş aşamasını temsil ederken, “Tumor Size” özneliği tümörün boyutunu belirtir. Hormon reseptörlerinin varlığını ve/veya aktivitesini gösteren “Estrogen Status” ve “Progesterone Status” öznelikleri, tedavi seçeneklerini etkileyebilir. “Regional Node Examined” ve “Regional Node Positive” öznelikleri, bölgesel lenf düğümlerinin incelenme durumunu ve pozitif kanser hücresi bulunup bulunmadığını belirtir. “Survival Months” özneliği, hastanın meme kanseri teşhisi aldıktan sonra hayatta kaldığı ay sayısını ifade ederken, “Status” özneliği hastanın sağkalım durumunu belirtir. Hastanın sağkalım durumu diğer bütün özneliklere bağlı olarak etiketlenen özneliği ifade etmektedir.

Tablo 1. Veri setinden örnekler.

Örnek Sayısı	Yaş	İrk	Medeni Hali	T Evresi	N Evresi	...	Yaşama Süresi	Durum
1	68	Beyaz	Evli	T1	N1	...	60	Hayatta
2	50	Beyaz	Evli	T2	N2	...	62	Hayatta
3	48	Siyah	Boşanmış	T2	N1	...	107	Hayatta
4	68	Beyaz	Dul	T1	N1	...	64	Ölü
5	40	Beyaz	Evli	T2	N1	...	14	Ölü
6	46	Diğer	Evli	T2	N3	...	31	Hayatta
7	42	Beyaz	Evli	T1	N3	...	39	Ölü

Tablo 1.’de veri setinde bulunan bazı örnekler ve buna bağlı bazı öznelikler görülmektedir.

3.2 Modeller

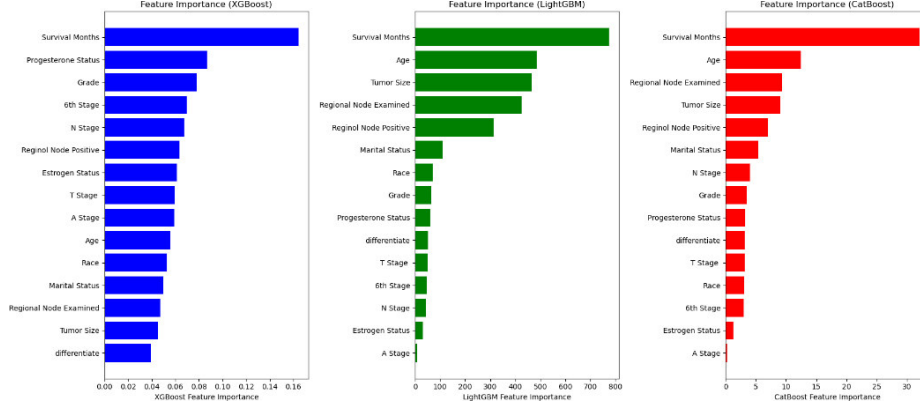
Meme kanseri sağkalım tahmininde kullanılan yöntemlerin açıklaması oldukça kritik bir süreç içermektedir. Herhangi bir makine öğrenimi algoritmasını kullanmadan önce verilerin hazırlık süreçleri vardır. Veri madenciliğinde güvenilirliğin artırılması için, veri ön işleme yapılmalıdır [4]. Kategorik verilerin sayısal verilere dönüştürmek, eksik, tutarsız ve gürültülü verileri veri setinden atmak gibi çeşitli işlemler gerçekleştirilir ve

makine öğrenimi algoritmalarının seçimi tamamlanır. Kullanılan veri setinin giriş (hastalık bulguları) ve çıkış (sağ kalım durumu) durumları bulunmasından ötürü makine öğreniminin alt alanı olan denetimli öğrenme tercih edilmektedir. Denetimli öğrenme, sistemin etiketli veriler kullanılarak eğitilmesi ile öğrenmenin sağlanmasıdır [5]. Kullanılan veri setinde etiketlediğimiz öznelik hastanın sağkalım durumudur. Gerekli ön işleme ve etiketleme adımlarından sonraki adımda denetimli öğrenme yönteminin içerdiği algoritmaların seçimi yapılır. Bu çalışmada meme kanseri sağkalım tahmininde XGBoost, LightGBM ve CatBoost algoritmaları tercih edilmiştir. Bu algoritmalar, sınıflandırma problemlerinde sıklıkla tercih edilen makine öğrenimi algoritmalarıdır.

- XGBoost algoritması karar ağaçlarını oluştururken paralel çalıştırma yaptığı için işlemler hızlı bir şekilde sonuç vermektedir [6].
- LightGBM, düşey yönde büyüme stratejisini (yapraksal büyüme) kullanırken, diğer gradyan artırma yöntemleri ise seviyesel (level-wise) büyüme stratejisini kullanmaktadır [7].
- CatBoost ise kategorik değişkenlerin otomatik olarak işlenmesini sağlayan bir diğer algoritmadır.

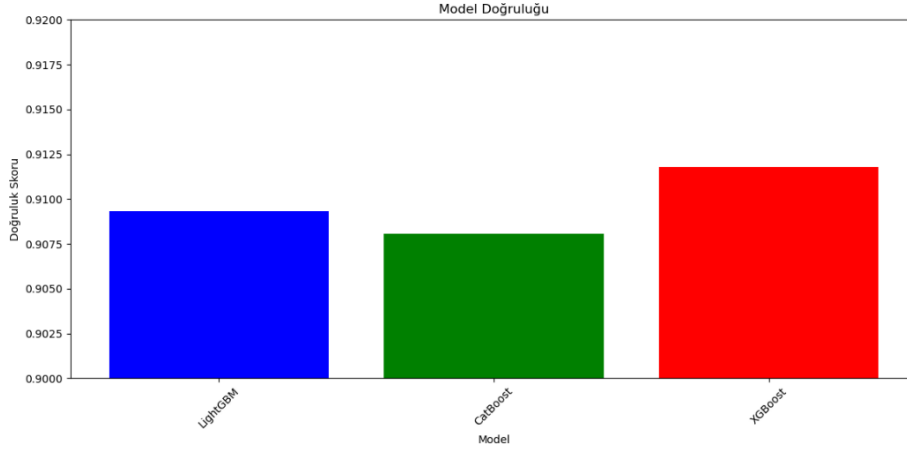
4 Sonuçlar

Meme kanseri sağkalım tahmini için kullanılan algoritmaların üzerinde gerçekleştirilen deneyler sonucunda elde edilen doğruluk değerleri LightGBM algoritması için 0.910, Catboost algoritması için 0.910 ve XGBoost algoritması için 0.911 olarak bulunmuştur. Elde edilen bu sonuçlar meme kanseri sağkalım tahmini için kullanılan yöntemlerin yüksek güvenilirliğe sahip olduğunu belirtmektedir. Bu sonuçlar, meme kanseri sağkalım tahmini alanında makine öğrenimi yöntemlerinin aktif bir şekilde rol alabileceğini göstermektedir. Ayrıca meme kanseri hastalarının kaçının çeşitli özneliklere bağlı olarak yaşayıp yaşayamayacağı konusunda bilgi vermektedir. Kullanılan makine öğrenimi yöntemleri sonucunda sağkalım durumuna büyük ölçüde etki eden bazı özellikler mevcuttur. Her üç modelde de yüksek oranda etkisi olan öznelikler, her bir hastanın teşhis konulduktan sonra hayatta kaldığı ay sayısını ifade eden “Survival Months” özneliği, hastaların yaşlarını ifade eden “Age” özneliği ve de meme kanseri hastalarında pozitif olan bölgesel lenf düğümlerinin sayısını ifade eden “Regional Node Positive” özneliğidir. Şekil 1’deki özelliklerin bilinmesi meme kanseri teşhisi konulmuş hastanın sağkalım durumunu yüksek düzeyde etkilemektedir.



Şekil 1. XGBoost, LightGBM, CatBoost algoritmalarında bulunan özniteliklerin model üzerindeki etkileri

Bunların yanı sıra kullanılan makine öğrenimi yöntemlerinin doğruluk metriğine bağlı karşılaştırılması Şekil 2.'de gösterilmektedir. Doğruluk (Accuracy) metriği, sınıflandırma modelinin doğru tahmin etme başarısını ölçmektedir. Doğru tahmin edilenlerin sayısının toplam örnek sayısına bölünmesi sonucu elde edilir.



Şekil 2. LightGBM, CatBoost, XGBoost algoritmalarının doğruluk metriği grafikleri

Modellerin karşılaştırılması sırasında, aralarındaki performansların oldukça benzer olduğu bulunmuştur. Her bir modelin doğruluk metriği altında %90 üzerinde doğruluk sağlaması ve XGBoost diğer modellere karşın daha yüksek bir doğruluk verdiği analiz edilmiştir.

5 Referanslar

1. S. Aydınтуğ, "Meme kanserinde erken tanı," *Sted*, vol. 13, no. 6, pp. 226–229, 2004.

2. N. Akyolcu, Y. Özhanlı, and D. Kandemir, "Meme kanserinde güncel gelişmeler," *Sağlık Bilimleri ve Meslekleri Dergisi*, vol. 6, no. 3, pp. 583–594, 2019.
3. A. Açıkgoz, R. Çehreli, and H. Ellidokuz, "Hastanede Çalışan Kadınların Meme Kanseri Konusunda Erken Tanı Yöntemlerine Yönelik Bilgi ve Davranışlarının Belirlenmesi, Uygulanan Planlı Eğitimin Etkinliğinin İncelenmesi.," *Meme Sağlığı Dergisi/Journal of Breast Health*, vol. 11, no. 1, 2015.
4. A. Oğuzlar, "Veri ön işleme," *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, no. 21, 2003.
5. M. Bilgin, "Gerçek veri setlerinde klasik makine öğrenmesi yöntemlerinin performans analizi," *Breast*, vol. 2, no. 9, p. 683, 2017.
6. İ. Kuş, S. B. Keser, and E. Yolaçan, "Saldırı tespit sistemlerinde topluluk öğrenme yöntemlerinin kıyaslanması," *Avrupa Bilim ve Teknoloji Dergisi*, no. 31, pp. 725–734, 2021.
7. M. Üstüner, S. Abdikan, G. Bilgin, and F. B. Şanlı, "Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması," *Türk Uzaktan Algılama ve CBS Dergisi*, vol. 1, no. 2, pp. 97–105, 2020.

The Impact of ChatGPT in Medical Informatics Education

Nur Hürsoy¹[0000-0001-5059-2268], Merve Solak¹[0000-0003-3466-7260], Ayşenur Topçu¹[0009-0002-6829-1919], Yusuf Çubukçu¹[0009-0006-5502-989X], Enver Turan¹[0009-0008-0802-0854], Esat Kaba¹[0000-0001-7464-988X]

¹ Recep Tayyip Erdoğan Üniversitesi, Rize 53100, Türkiye
nur.hursoy@erdogan.edu.tr

Abstract. Medical Informatics can be defined as the use of information technologies to solve bio-medical and health problems. Medical Informatics education is of great importance in order to be able to follow new tools, determine their usage areas, and be prepared for possible ethical and legal problems. There are current studies on the use of large language models in medical education. In this study, we aimed to evaluate the information provided by ChatGPT 3.5 and 4. Two residents (MS, ATV) prepared 60 multiple choice questions. Two other residents uploaded the questions to Chat GPT. Residents who did not know the answers to the questions were also asked to assess the reliability of the answers given by the Chat GPT. Out of a total of 60 questions, Chat GPT 3.5 was able to select the correct answer in 38 questions, while Chat GPT 4 was able to select the correct answer in 51 questions. It is seen that although CHAT GPT is useful as a training material, its limitations should be known and care should be taken in this regard.

Keywords: Generative AI, Medical Informatics, Medical Education, Chat GPT

1 Introduction

Medical Informatics can be defined as the use of information technologies to solve bi-medical and health problems. Informatics term covers "data+meaning" (1). Today, when technological developments have gained the highest momentum in history, Medical Informatics education is of great importance in order to be able to follow new tools, determine their usage areas, and be prepared for possible ethical and legal problems.

Published in 2021 by Istanbul University, renewed in 2022 and 2023 and made available in full text on the web, "Tıp Bilişimi" stands out as a comprehensive Turkish resource in this field. Consisting of 41 main chapters, the book draws attention to different aspects of the subject even with only the chapter titles(2).

Imaging Informatics is also included as a component of the program in the Specialty Training Program prepared by the Turkish Society of Radiology and updated in 2022. The learning objectives of Imaging Informatics education include having knowledge on different topics such as electronic patient records, image processing, teleradiology and artificial intelligence (3).

There are current studies on the use of large language models in medical education(4-6) . This new way of learning can be defined as learning technology with technology. In this study, we aimed to evaluate the information provided by ChatGPT 3,5 and 4, one of the widely used language models, through the questions we prepared based on the Medical Informatics book (2).

2 Material and Method

Two residents (MS, ATV) prepared 60 multiple choice questions based on the “Tıp Bilişimi” (2). The questions were evaluated by expert radiologists (NH, EK) in terms of appropriateness of wording, distraction of choices and question formats. Two other residents (YÇ, ET) uploaded the questions to Chat GPT versions 3.5 and 4. Answers were recorded with choices and explanations. Residents who did not know the answers to the questions were also asked to assess the reliability of the answers given by the Chat GPT.

3 Results

Out of a total of 60 questions, Chat GPT 3.5 was able to select the correct answer in 38 questions, while Chat GPT 4 was able to select the correct answer in 51 questions. There was a difference in the answers of 19 questions between both versions. The GPT 3.5 version gave the correct answer to 3 questions that Chat GPT 4 answered incorrectly. For 4 questions, both versions gave the same wrong answer.

Residents who uploaded the questions to ChatGPT evaluated the answers to 30 questions each. Evaluating the answers to the first 30 questions, resident (YÇ) stated that he knew the correct answer to 16 of the questions. He thought that 8 of the answers given by the Chat GPT 3.5 version were wrong and he was not sure about 1 answer. He was able to guess that three of the 7 questions were answered incorrectly, although he did not know the answer to three of them. It was determined that he thought that 3 questions that were answered correctly but in fact there were answered incorrectly by Chat GPT. Although he stated that he knew the answer to one of these questions beforehand, he evaluated it incorrectly.

For the same questions with GPT 4 version, he was not sure about one of the answers and thought that two of them were wrong. The question that he was not sure about was answered incorrectly by GPT 4. The two questions that he answered incorrectly but Chat GPT answered correctly, while he thought that the answer of Chat GPT is true for two questions but it isn't.

The second resident (ET), who evaluated the other 30 questions, stated that he knew the answers to 18 of the question. When he evaluated the answers of Chat GPT 3.5 he thought 7 of them were wrong and 11 of them were correct. Of the 12 questions that he did not know the answer to, he hesitated in answering 3 of them and guessed that

one of them was wrong and 8 of them were right. While 6 questions that he thought were wrong were answered correctly, it was determined that 3 questions that he guessed to be correct were actually answered incorrectly. Two questions that he was not sure of the answer were answered correctly and one question was answered incorrectly.

When the answers given by the ChatGPT 4 version to the same questions were evaluated, only one answer was hesitated, the other answers were guessed to be correct. The 5 wrong answers of the GPT 4 version were assumed correct by ET. The question that resident was not sure about was answered correctly.

4 Discussion

With the developments in technology, different learning opportunities have emerged, especially under the title of “adult education”. With the widespread use of Generative AI applications, which are defined as productive artificial intelligence, the usage aspects in education have also started to be investigated.

Studies on Medical Informatics Education are of particular importance in terms of keeping the subject in the minds and increasing awareness on this subject. It can be difficult to access the necessary resources to get an idea about Medical Informatics, which includes a wide range of topics. It is seen that Chat GPT 4, one of the broad language models, has a higher percentage of producing correct answers than GPT 3.5 version since its internet connection and wide data source. In addition to giving the correct answer to the questions, the explanations it offers about the subject are also broader than the 3.5 version. In this respect, large language models seem to be a helpful tool in Medical Informatics education.

The reliability of the information provided by large language models is still controversial. In our study, although at a low rate, it was observed that incorrect answers could be given. It should be noted that wrong answers can also be perceived as correct by residents. In this respect, it is especially important the validation of informations which will change our approach in the routine practice.

Other limitations of Generative AI applications can be summarized as follows(7):

- » Beginning to give similar answers over time
- » Increased margin for error when trying to give complex answers
- » Acting biased, turning to old sources that it can access and producing incorrect answers on issues when encounters for the first time
- » The risk of producing wrong answers also raises ethical issues
- » Training of systems and access to information required for training is difficult and time consuming
- » Not knowing exactly how the system works

As conclusion, efficient use of almost any tool, from simple to complex, depends on getting to know the tool in every aspect and gaining experience by using it. ChatGPT

can be considered as an educational tool that can offer different perspectives in a short time by providing quick access to various information sources, however the problem of reliability in resources is always taken into consideration.

References

1. Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, vd. AMIA Board white paper: Definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Informatics Assoc.* 2012;19(6):931–8.
2. Bozbuğa N, Gülseçen S. *Tıp Bilişimi 2021*. Istanbul University Press. Istanbul
3. Web page of TRS: <https://www.turkrad.org.tr/yeterlik-kurulu> (available at 10.05.2024)
4. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, vd. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Med Educ.* 2023;9.
5. Hasan Sapci A, Aylin Sapci H. Artificial intelligence education and tools for medical and health informatics students: Systematic review. *JMIR Med Educ.* 2020;6(1):1–14.
6. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ.* 2023;(March):1–6.
7. Kim K, Cho K, Jang R, Kyung S, Lee S, Ham S, vd. Updated Primer on Generative Artificial Intelligence and Large Language Models in Medical Imaging for Medical Professionals. *Korean J Radiol.* 2024;25(3):224–42.

Apandisit Tanısı Alan Çocuklarda Perforasyon Risk Faktörlerinin İncelenmesi

Tuğba KURT [0000-0003-0771-1289]¹, İlknur BUÇAN KIRKBİR [0000-0002-0611-0118]²

¹Faculty of Medicine, Farabi Hospital, Statistic Unit, Karadeniz Technical University, Trabzon
ktu.tugbakurt@gmail.com

²Faculty of Health Science, Department of Nursing, Karadeniz Technical University, Trabzon
ilknurbucan@gmail.com

Abstract. Akut karın sendromunun en sık rastlanan sebeplerinden biri akut apandisitir. Özellikle çocuklarda, atipik klinik seyir nedeniyle, bu durum birçok farklı karın ağrısı semptomu ile karıştırılabilir. Dolayısıyla akut apandisit klinik tanısı, hem yetişkinlerde hem de çocuklarda öncelikle klinik bulgulara dayanmaktadır. Ancak, çocuklarda tipik semptom ve bulguların seyrek olması, apandisit belirtilerinin belirlenmesini zorlaştırır ve klinik bulguların sıklıkla diğer patolojilerle örtüşmesiyle sonuçlanır. Bu da pediatrik klinisyenler için doğru teşhisi koymayı zorlaştıran bir durumdur. Perfore apandisit ise, gecikmiş tedavi veya yanlış teşhis sonucu olarak peritonit, sepsis, bağırsak tıkanıklığı, apse oluşumu ve fertilitate problemleri gibi ciddi komplikasyonlara yol açabilir. Farklı skorlama sistemlerinin mevcut olmasına rağmen, henüz standart bir karar verme testi geliştirilmemiştir. Bu bağlamda, apandisit tanısı konmuş çocuk hastalarda perforasyon risk faktörlerini belirlemek için "Permütasyon Tabanlı Özellik Önemi" yöntemi kullanılmıştır. Çalışmada kullanılan veriler, açık erişimli bir veri tabanından alınmış ve veri filtreleme ile ön işleme adımlarının ardından R programlama dili kullanılarak analiz edilmiştir. Çalışmanın sonuçlarına göre, apandisit tanısı konmuş çocuk hastalarda hastalığın şiddeti, CRP değeri ve idrarda keton bulunması gibi değişkenlerin perforasyon riskini belirlemede önemli olduğu tespit edilmiştir. Bu belirlenen değişkenler, çocuk hastalarda perforasyon riski yüksek olan hastaların tanımlanmasına ve yeni testlerin geliştirilmesine olanak sağlayabilirler.

Keywords: Apandisit, Makine öğrenmesi, Perforasyon

1 Giriş

Çocuklarda akut karın ağrısının en yaygın nedenleri arasında akut apandisitir. Ancak klinik belirtilerin belirsizliği ve spesifik olmayan bulguların varlığı nedeniyle, özellikle çocuklarda apandisit tanısı koymak oldukça zor olabilir. Apandisitli çocuklarda, başvuru öncesi ve başvuru sonrası gecikmelerin apendiks perforasyonu (apandisitlin delinmesi) ile yakından ilişkili olduğu bilinmektedir. Özellikle karın ağrısına ilk 48 saati içinde eşlik eden semptomlar tanının gecikmesine yol açabilirler ve bu da perforasyon riskini arttırabilir. Perforasyon risk faktörlerinin belirlenmesi önemlidir çünkü bu faktörler, tedavi stratejilerini etkiler ve hastanın durumunu yönlendirebilir. Bu nedenle, karın ağrısı ve ilişkili semptomları olan çocuklarda bir pediatrik cerrahla zamanında konsültasyon yapılması gerekliliği vurgulanmalıdır. Bu durum, apandisitli çocuk hastaların yönetiminde kritik bir rol oynar [1-3]. Apandisitli çocuklarda başvuru öncesinde ve sonrasındaki gecikmelerin apendiks perforasyonuna yol açtığı bilinmektedir. Özellikle ilk 48 saat içinde tanının gecikmesi, perforasyon riskini arttırabilir. [3]. Apandisitli çocuklarda perforasyon risk faktörlerinin tanımlanması, tedavi stratejisinin belirlenmesi ve sonuçların tahmin edilmesi açısından hayati önem taşır. Tedavi sürecinde yaş, tedaviye başlama süresi ve ailede apandisit öyküsü gibi faktörler, perforasyon olasılığını belirleyen önemli belirleyicilerdir. Bu nedenle, apandisit şüphesi olan çocuklarda, pediatrik bir cerrahla zamanında konsültasyon yapılması kritik öneme sahiptir. Bu, tedavi stratejilerinin doğru şekilde belirlenmesine ve hastanın sağlık durumunun yönetilmesine yardımcı olur [5,6]. Permütasyon tabanlı özellik önemi yöntemi, her değişken için ölçülen önemin dağılımını tahmin etmek amacıyla sonuç vektörünün tekrarlanan permütasyonlarını içerir ve gözlemlenen önemin p değeri, özelliğin düzeltilmiş bir ölçüsünü gösterir. Bu yöntem, değişkenlerin önemini değerlendirilmesinde makine öğrenimi modellerinin daha doğru bir şekilde yorumlanmasını sağlar. [4]. Bu amaçla yapılan bu çalışmada, apandisit tanısı alan çocuklarda perforasyonun varlığıyla ilişkili klinik ve laboratuvar bulgular makine öğrenmesi tekniklerinden biri olan permütasyon tabanlı özellik önemi yöntemi ile değerlendirilmiştir.

2 Method

2.1 Veri Seti

Bu çalışma, akut apandisit tanısı konmuş çocuk hastalardan elde edilen verilerle gerçekleştirilmiştir. Veri seti, UC Irvine açık erişimli veritabanından elde edilmiştir (<https://archive.ics.uci.edu/dataset/938/regensburg+pediatric+appendicitis>). Veri seti, demografik ve klinik bilgilerden oluşan toplamda 32 değişken ve 62 hasta verisi içermektedir. Perforasyon açısından yanlılığı önlemek için hastaların ultrason bilgileri

ve ameliyat bilgilerini içeren 9 değişken analiz dışı bırakılmış, sadece klinik ve laboratuvar bulgularına ait toplam 24 değişken analize dahil edilmiştir. Çalışmada kullanılan değişkenler Tablo 1’de verilmiştir.

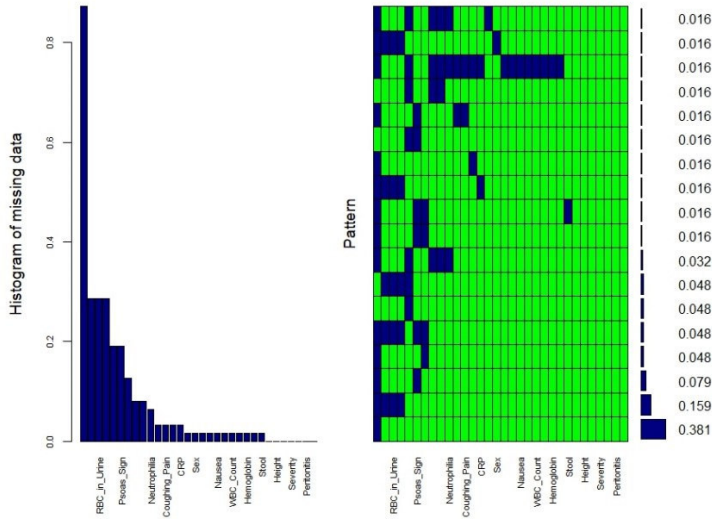
Tablo 1. Çalışmaya dahil edilen değişkenler ve açıklamaları

Değişkenler
<i>Gezici ağrı (Migratory_Pain)</i>
<i>Rebound hassasiyet testi (Contralateral_Rebound_Tenderness)</i>
<i>Apandistin şiddeti (Severity)</i>
<i>Öksürme ile ağrı (Coughing_Pain)</i>
<i>Bulantı (Nausea)</i>
<i>İştah kaybı (Loss_of_Appetite)</i>
<i>Vücut sıcaklığı (Body_Temperature)</i>
<i>Beyaz küre sayısı (WBC_Count)</i>
<i>Nötrofil yüzdesi (Neutrophil_Percentage)</i>
<i>Segmentli nötrofiller (Segmented_Neutrophils)</i>
<i>Nötrofili (Neutrophilia)</i>
<i>Kırmızı kan hücresi sayısı (RBC_count)</i>
<i>Hemoglobin</i>
<i>RDW</i>
<i>Trombosit sayısı (Thrombocyte_Count)</i>
<i>İdrarda keton (Ketones_in_Urine)</i>
<i>İdrarda kırmızı kan hücreleri (RBC_in_Urine)</i>
<i>İdrarda beyaz kan hücreleri (WBC_in_Urine)</i>
<i>CRP (C-reaktif protein)</i>
<i>İdrar yaparken ağrı-yanma (Dysuria)</i>
<i>Dışkı (Stool)</i>
<i>Peritonit (Peritonitis)</i>
<i>Psoas Belirtisi (Psoas_Sign)</i>
<i>Perforasyon (Perforation)</i>

2.2 Veri Ön İşleme ve İstatistiksel Analiz

Çalışmanın istatistiksel analiz ve model geliştirme süreci R açık kaynak programlama dili kullanılarak gerçekleştirilmiştir. Nicel verilerde normal dağılım, Kolmogorov-Smirnov testi ile değerlendirilmiştir. Sürekli değişkenler için ortalama (\pm SD), kategorik değişkenler için frekans ve yüzde kullanılmıştır. İstatistiksel anlamlılık için $p < 0.05$ kesim noktası olarak kullanılmıştır. Bağımsız gruplarda ortalama karşılaştırılması için “t-test” ve “Mann-Whitney U” testi kullanılmıştır. Kategorik değişkenler için “ki-kare” testi yapılmıştır. Veri setine filtreleme ve ön işleme adımları uygulandıktan sonra, eksik veri oranı % 40’ın altında olan değişkenler Multiple Imputation by Chained Equations (MICE) yöntemi kullanılarak tamamlanmıştır. Eksik veri oranlarına ilişkin grafik Şekil 1’de verilmiştir.

Şekil 1. Eksik verilere ilişkin satır ve sütun bazlı histogram



2.3 Permütasyon Tabanlı Özellik Önemi Yöntemi

Permütasyon tabanlı özellik önemi yöntemi, literatürde yaygın olarak kullanılan bir model inceleme tekniğidir [7]. Bu yöntem, özellikle doğrusal olmayan tahmin ediciler için etkili bir şekilde kullanılabilir ve bağımsız değişkenlerin bağımlı değişkeni tahmin etmedeki katkısını objektif bir şekilde değerlendirir [8]. Bu değerlendirme sürecinde, özelliklerin değerlerinin rastgele permütasyonu sonucunda modelin performansındaki değişiklikler ölçülür ve her bir özelliğin tahmin performansına olan etkisi belirlenir. Bu yöntem, modelin karmaşıklığını anlamak ve en önemli özellikleri belirlemek için kullanılır [9]. Ayrıca, özelliklerin katkılarını değerlendirirken, model bağımlılığını dikkate alır ve yöntemin farklı model sınıflarında kullanılabilirliğini artırır [10]. Bu nedenle permütasyon tabanlı özellik önemi yöntemi, model inceleme ve iyileştirme süreçlerinde yaygın olarak tercih edilen bir tekniktir.

3 Sonuçlar

Veri setindeki istatistiksel olarak anlamlı bulunan değişkenlere dair tanımlayıcı istatistikler ve test sonuçları Tablo 2’de verilmiştir. Veri setini oluşturan 62 çocuk hastanın 34’ünde perforasyon gelişirken, 28’inde perforasyon gelişmemiştir. Çocuk hastaların

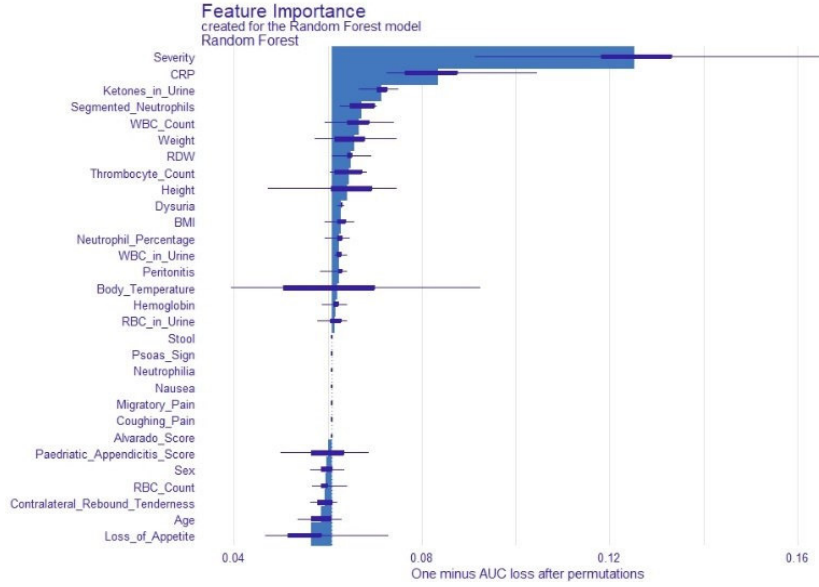
%45.2' kız, %54.8' erkektir. Tablo 1 'de perforasyon durumuna göre hipotez testlerinde anlamlı bulunan değişkenler ve demografik bilgileri sunulmuştur. Perforasyon olan vakalarda pediatrik apandisit skoru ortalaması (6.71±1.88) ile perforasyon olmayan vakaların ortalaması (5.53±1.91) arasında istatistiksel olarak anlamlı fark olduğu sonucu elde edilmiştir (p=0.019). Perforasyon olan vakalarda vücut sıcaklığı ortalaması (37.9±0.87) ile perforasyon olmayan vakaların vücut sıcaklığı ortalaması (37.33±0.83) arasında istatistiksel olarak anlamlı fark vardır (p=0.006). Benzer şekilde çocuk hastaların CRP (C-reaktif protein), yaş, RBC, WBC ve RDW ortalamaları arasında da anlamlı fark olduğu belirlenmiştir (p=0.0001). İştah kaybı, idrarda keton varlığı ve hastalığın şiddeti ile perforasyon arasında istatistiksel olarak anlamlı ilişki olduğu belirlenmiştir (p<0.05).

Tablo 2. Perforasyon durumuna göre anlamlı olan açıklayıcı değişkenler

Sayısal Değişkenler	Perforasyon (Yok=34)	Perforasyon (Var=28)	p	
	Ort±SS	Ort±SS		
Pediatrik apandisit skoru	5.53 (1.91)	6.71 (1.88)	0.019	
Vücut sıcaklığı	37.33 (0.83)	37.9 (0.87)	0.006	
RDW	12.49 (0.63)	13.17 (1.08)	0.004	
CRP	19.21 (28.56)	116.64 (98.61)	0.0001	
Yaş	11.20 (2.95)	9.71 (4.13)	0.0001	
RBC	4.75 (0.32)	4.77 (0.41)	0.0001	
WBC	13.85 (4.60)	16.89 (6.38)	0.0001	
Kategorik Değişkenler	n (%)	n (%)	p	
Apendisit şiddeti	<i>komplike</i>	19 (67.9)	4 (11.8)	0.0001
	<i>komplike değil</i>	9 (32.1)	30 (88.2)	
İştah Kaybı	<i>yüksek</i>	15 (53.6)	10 (29.4)	0.033
	<i>düşük</i>	2 (7.1)	4 (11.8)	
	<i>orta</i>	6 (21.4)	3 (8.8)	
	<i>yok</i>	5 (17.9)	17 (50.0)	
İdrarda Keton	<i>hayır</i>	6 (21.4)	18 (52.9)	0.011
	<i>evet</i>	22 (78.6)	16 (47.1)	

p<0.01

Permütasyon tabanlı özellik önemi yöntemi ile önem derecesine göre sıralanmış değişkenler Şekil 1'de verilmiştir. Çalışmanın sonuçlarına göre, apandisit tanısı konmuş çocuk hastalarda, hastalığın şiddeti, CRP değeri ve idrarda keton bulunması, perforasyon riskini belirlemede, en önemli üç değişken olarak belirlenmiştir.



Şekil 1. Permütasyon tabanlı özellik önemi yöntemi sonuçları

Çalışmanın kısıtlılıkları arasında, veri setinin boyutunun az olması (62 hasta) ve retrospektif olması sayılabilir. Gelecek dönemlerde hasta sayısı genişletilerek ve prospektif olarak veri toplanarak bu çalışma güçlendirilebilir. Elde edilen halihazırdaki bulgulara göre bu çalışma, çocuk hastalardaki apandisit vakalarında perforasyon olma riski yüksek olanların tahmin edilmesi için geliştirilecek yeni tanı araçlarına katkı sağlayabilir. Ayrıca, apandisit tanısı alan çocuklarda perforasyon gibi riskli ve komplike bir durumun için klinik yönetim stratejilerinin geliştirilmesine ve sonuçların iyileştirilmesine katkı sağlayabileceği düşünülmektedir.

Kaynaklar

1. Yang WC, Chen CY, Wu HP. Etiology of non-traumatic acute abdomen in pediatric emergency departments. *World J Clin Cases*. 2013 Dec 16;1(9):276-84. doi: 10.12998/wjcc.v1.i9.276. PMID: 24364022; PMCID: PMC3868711.
2. Lin WC, Lin CH. Multidetector computed tomography in the evaluation of pediatric acute abdominal pain in the emergency department. *Biomedicine (Taipei)*. 2016 Jun;6(2):10. doi: 10.7603/s40681-016-0010-8. Epub 2016 May 6. PMID: 27154197; PMCID: PMC4859315.
3. Ozguner IF, Buyukayavuz BI, Savas MC. The influence of delay on perforation in childhood appendicitis. A retrospective analysis of 58 cases. *Saudi Med J*. 2004 Sep;25(9):1232-6. PMID: 15448773.

4. Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10), 1340-1347.
5. Bonadio W, Shahid S, Vardi L, Buckingham C, Kornblatt A, Free C, Homel P. A pre-operative clinical scoring system to distinguish perforation risk with pediatric appendicitis. *J Pediatr Surg*. 2018 Mar;53(3):441-445. doi: 10.1016/j.jpedsurg.2017.05.017. Epub 2017 May 22. PMID: 28554816.
6. Brender JD, Marcuse EK, Koepsell TD, Hatch EI. Childhood appendicitis: factors associated with perforation. *Pediatrics*. 1985 Aug;76(2):301-6. PMID: 4022704.
7. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
8. Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC bioinformatics*, 8(1), 25.
9. Fisher, A., Rudin, C., & Dominici, F. (2018). All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(1), 571-606.
10. Cottin A, Zulian M, Pécuchet N, Guilloux A, Katsahian S. MS-CPFI: A model-agnostic Counterfactual Perturbation Feature Importance algorithm for interpreting black-box Multi-State models. *Artif Intell Med*. 2024 Jan;147:102741. doi: 10.1016/j.artmed.2023.102741. Epub 2023 Nov 29. PMID: 38184354.

Uzmanlık ve Yüksek Lisans Programlarında Görüntüleme Bilişimi Eğitimi

Nur Hürsoy¹[0000-0001-5059-2268]

¹ Recep Tayyip Erdoğan Üniversitesi, Rize, 53100

nur.hursoy@erdogan.edu.tr

Abstract. Imaging informatics covers many different areas, such as the acquisition, storage, and evaluation of images performed in the healthcare setting, the use of the information obtained, the improvement of image quality, and the facilitation of portability. In addition to image acquisition, it also plays an important role in improving the quality of healthcare services through scheduling, prioritization, and support systems. In recent years, the pace of technological development has increased. As access to information becomes easier, the amount of information to be accessed increases. In this process, education in imaging informatics is becoming increasingly important. However, it remains unclear by whom, to whom and how this education is provided. I would like to share my experience of imaging informatics education in radiology residency and biomedical engineering graduate programs by mentioned about international programs and curriculum changes in our country.

Keywords: Imaging Informatics, Radiology Residency Program, Biomedical Engineering

1 Giriş

Bilişim kelimesi 1981 yılında Prof. Dr. Aydın Köksal tarafından dilimize kazandırılmış olup, kısaca bilginin işlenmesi bilimi olarak tanımlanabilir [1]. Tıp Bilişimi alt başlıklarının büyük çoğunluğu Görüntüleme Bilişimi ile kesişmektedir. Günümüzde Tıpta Görüntüleme Bilişimi Derneği (SIIM) olarak adlandırılan ve teknolojinin gelişim sürecinde değişik isimler alan dernek 1980 yılında kurulmuştur [2]. Sağlık alanında bilişim teknolojilerinin gelişimi ve kullanımında Radyolojinin öncü rolü olmuştur.

Görüntüleme Bilişimindeki gelişmelerle birlikte yüksek teknolojiyi kullanan sağlık çalışanlarına standardizasyon getirme amacı ile 2007 yılında Amerika Görüntüleme Bilişimi Kurulu (ABII) kurulmuştur. 2023 Ağustos ayında görev tanımlaması görüntüleme bilişiminin etkili bir şekilde uygulanmasını tanımlayarak kaliteli sağlık hizmetini teşvik etmek ve bilişimcileri sertifikalandırmak şeklinde güncellenmiştir [3].

Ülkemizde de Radyoloji Uzmanlık Eğitim Programlarında Görüntüleme Bilişiminin yeri arttırılmaktadır. Üniversitelerde Bilişim Enstitüleri kurulmaktadır. Bununla birlikte söz konusu eğitimin kapsamı oldukça geniş ve uygulamada eğitim modeli tam olarak belirlenmemiştir. Kurumumuz Radyoloji Uzmanlık Eğitimi programında ve Biyomedikal Mühendisliği Yüksek Lisans Programında verdiğimiz eğitimlerle ilgili

tecrübelerimi paylaşarak, bu konudaki sorunları ve çözüm önerilerini tartışmayı amaçladım.

2 Yöntem

“Imaging Informatics” ve “Education” anahtar kelimeleri ile Google arama motorundan yapılan aramada uluslararası sertifika programlarına ulaşıldı. Üniversiteler ve Uluslararası Radyoloji Derneklerinin kursları listelendi. Örnek kursların eğitim süreleri, müfredatları, gereklilikleri ve varsa ücretleri derlendi.

Tıpta Uzmanlık Kurulu Müfredat Oluşturma ve Standart Belirleme Sistemi (TUKMOS) ve Türk Radyoloji Derneği Uzmanlık Eğitim Programlarının eski ve yeni sürümleri incelendi.

Üniversitelerimizde Bilişim Enstitüleri ve Bilişimle ilgili Yüksek Lisans programlarında Görüntüleme Bilişimi eğitiminin varlığı araştırıldı.

Kurumumuzda Görüntüleme Bilişimi hakkında hazırlanan seminer ve üç dönemdir devam eden yüksek lisans dersleri tekrar gözden geçirildi.

3 Tartışma ve Sonuç

Kuzey Amerika Radyoloji Derneği (RSNA) ve Tıpta Görüntüleme Bilişimi Derneği (SIIM)'nin ortaklaşa düzenlediği çevrimiçi kurs yurtdışında verilen eğitimlerin en kapsamlısı olarak göze çarpmaktadır [4]. Derneklerin düzenlediği kursların yanısıra Kanada'da uygulamalı sağlık bilimleri eğitimi veren tek merkez olan Michener Enstitüsünün Görüntüleme Bilişimi Sertika Programı bulunmaktadır [5]. Hollanda Maastricht Üniversitesinde ise iki yıllık Görüntüleme Mühendisliği Yüksek Lisans program dikkati çekmiştir[6].

TUKMOS'un 2024 yılı mayıs ayında öneri ve itirazlara açık olarak yayınladığı Çekirdek Eğitim Müfredatına, temel, klinik ve radyolojik bilgiler başlığı altında “Radyolojik Verilerin Yönetimi için Gerekli Bilişim Araçlarını,Çalışma Standart ve Prensiplerini ve İlgili Terminolojiyi Bilme” maddesinin eklendiği görülmektedir [7]. 2022 yılında Türk Radyoloji Derneği Yeterlik Kurulu tarafından güncellenen Eğitim Programında Görüntüleme Bilişimi öğrenim hedefleri listesinde ayrı bir başlık olarak ele alınmaktadır [8].

Eğitim, hizmet ve araştırma dengesini kurmanın oldukça zor olduğu Eğitim ve Araştırma hastanelerinde eğitim programına bilişimi dahil etmede çeşitli zorluklarla karşılaşmaktadır. Öğrenim hedeflerinin uzmanlık öğrencilerince benimsenmemesi, eğitmen kadrosunun konu hakkında yeterli bilgi ve donanıma sahip olmaması, eğitim programının güncellenmemesi, erişkin eğitimi yöntemlerinin yaygınlaşmaması bunlar arasında sayılabilir. Yüksek Lisans programlarında ise belirlenmiş öğrenim hedefleri bulunmamaktadır. Öğrenciler çok farklı bilgi ve beceri alt yapısına sahiptir, eğitimden beklentileri de belirsiz ve değişken olabilmektedir.

Bu ortamda eğitimcilerin farklı kuruluşların eğitim programlarını incelemeleri, kendi kurumlarının hedefleri ve öğrencilerin beklentileri doğrultusunda örgün eğitim dışı modelleri de kullanarak Görüntüleme Bilişimi eğitimini yaygınlaştırması önem taşımaktadır.

References

1. Türk Dil Kurumu Sözlüğü, <https://sozluk.gov.tr>, erişim tarihi 2024/05
2. The Society for Imaging Informatics in Medicine, <https://siim.org/about-siim/history>, erişim tarihi 2024/05
3. The American Board of Imaging Informatics ana sayfa, <https://www.abii.org/About.aspx>, erişim tarihi 2024/05
4. <https://www.rsna.org/education/trainee-resources/national-imaging-informatics-curriculum-and-course>, erişim tarihi 2024/05
5. https://michener.ca/ce_course/imaging-informatics-certificate-program, erişim tarihi 2024/05
6. <https://curriculum.maastrichtuniversity.nl/education/master/imaging-engineering>, erişim tarihi 2024/05
7. <https://tuk.saglik.gov.tr/TR-103291/uzmanlik-egitimi-cekirdek-mufredat-taslaklarinin-askiya-cikarilmasi.html>, erişim tarihi 2024/05
8. <https://www.turkrad.org.tr/yeterlik-kurulu>, tarihi 2024/05

15. Tıp Bilisimi Kongresi

30-31 Mayıs 2024

Ana Tema: Sağlıkta Yapay Zeka ve Büyük Dil Modelleri

15th Medical Informatics Congress

Main Theme: Artificial Intelligence and Large Language Models in Healthcare

<https://turkmia.net/kongre2024>

ISBN:978-605-7264-1-6



9 786057 256416