



JOURNAL OF INFORMATION TECHNOLOGIES

BİLİŞİM TEKNOLOJİLERİ DERGİSİ

Volume / Cilt **17** Number / Sayı **3** Year / Yıl **2024** Month / Ay **July / Temmuz**





GAZİ ÜNİVERSİTESİ (GAZİ UNIVERSITY)
BİLİŞİM ENSTİTÜSÜ (INSTITUTE OF INFORMATICS)

BİLİŞİM TEKNOLOJİLERİ DERGİSİ
(JOURNAL OF INFORMATION TECHNOLOGIES)
ISSN: 1307-9697 e-ISSN: 2147-0715



Cilt (Volume): 17

Sayı (Issue): 3

Temmuz (July) 2024

Sahibi (Owner)
Dr. Musa YILDIZ
Rektör (Rector)

Genel Yayın Yönetmeni & Baş Editör
(General Publication Director & Editor in Chief)
Dr. Ashhan TÜFEKÇİ
Bilişim Enstitüsü Müdürü
Director of Institute of Informatics

Yardımcı Editör
(Associate Editor)
Dr. Öner BARUT
Bilişim Enstitüsü Müdür Yardımcısı
Vice Director of Institute of Informatics

Editörler (Editors)

❖ Dr. Serdar KULA	Gazi Üniversitesi (Gazi University)
❖ Dr. Hüseyin POLAT	Gazi Üniversitesi (Gazi University)
❖ Dr. Resul DAŞ	Fırat Üniversitesi (Fırat University)
❖ Dr. Muzaffer KAPANOĞLU	Eskişehir Osmangazi Üniversitesi (Eskişehir Osmangazi University)
❖ Dr. Mehmet ŞİMŞEK	Milli Savunma Üniversitesi (National Defence University)
❖ Dr. Mehmet Sıraç ÖZERDEM	Dicle Üniversitesi (Dicle University)
❖ Dr. Mustafa Mahir ÜLGÜ	Sağlık Bakanlığı
❖ Dr. Murat YILMAZ	Gazi Üniversitesi (Gazi University)
❖ Dr. Oktay YILDIZ	Gazi Üniversitesi (Gazi University)
❖ Dr. Olgun DEĞİRMENÇİ	TOBB ETÜ (TOBB Economics and Technology University)
❖ Dr. Recep BENZER	Gazi Üniversitesi (Gazi University)
❖ Dr. Ö. Tolga PUSATLI	Çankaya Üniversitesi (Çankaya University)
❖ Dr. Cihangir TEZCAN	Orta Doğu Teknik Üniversitesi (Middle East Technical University)
❖ Dr. Mehmet SEVRİ	Recep Tayyip Erdoğan Üniversitesi (Recep Tayyip Erdoğan University)
❖ Dr. Muhammed Ali KOŞAN	Kahramanmaraş İstiklal Üniversitesi (Kahramanmaraş Istiklal University)
❖ Dr. Levent ÇETİNKAYA	Çanakkale Onsekiz Mart Üniversitesi (Çanakkale Onsekiz Mart University)
❖ Dr. Sinan TOKLU	Gazi Üniversitesi (Gazi University)

Yayın Danışma Kurulu (Editorial Advisory Board)

Dr. Ahmet COŞAR <i>Turkish Aeronautical Association University, Turkey</i>	Dr. Aslanbek NAZİEV <i>Ryazan State University, Russia</i>	Dr. Bogdan PATRUT <i>Alexandru Ioan Cuza University of Iasi, Romania</i>
Dr. Deepak GUPTA <i>Maharaja Agrasen Institute of Technology, India</i>	Dr. Jafar A. ALZUBİ <i>Al-Balqa Applied University, Jordan</i>	Dr. Jolanta SABAITYTĖ <i>Vilnius Gediminas Technical University, Lithuania</i>
Dr. Ilya LEVİN <i>Tel Aviv University, Israel</i>	Dr. Pınar KARAGÖZ <i>Middle East Technical University, Turkey</i>	Dr. Ufuk ÇAĞLAYAN <i>Yaşar University, Turkey</i>
Dr. Veysi İŞLER <i>Hasan Kalyoncu University, Turkey</i>	Dr. Victor Hugo Costa DE ALBUQUERQUE <i>Universidade de Fortaleza, Brazil</i>	Dr. Vijender Kumar SOLANKİ <i>CMR Institute of Technology, India</i>
Dr. Ebrahim KHOSRAVI <i>Clayton State University, United States</i>		

Dil Editörü

(Language Editor)

Dr. Çağla Gizem AKKAŞ

*Bilişim Enstitüsü
Institute of Informatics*

Teknik Sorumlu

(Technical Assistant)

Dr. Candan TÜMER

*Bilişim Enstitüsü
Institute of Informatics*

Teknik Sorumlu

(Technical Assistant)

Yasemin İÇTÜZER

*Bilişim Enstitüsü
Institute of Informatics*

Sekreterlik

(Secretary)

Bilişim Teknolojileri Dergisi

*Bilişim Enstitüsü
Institute of Informatics*

Bilişim Teknolojileri Dergisi uluslararası hakemli bir dergidir.
Journal of Information Technologies is an international refereed journal.

Yazışma Adresi (Contact Address)

Tunus Cad. No: 35 Kavaklıdere Çankaya/ANKARA

Telefon / Telephone: 0312 202 38 01

Faks / Fax: 0312 212 79 29

Çevrimiçi Değerlendirme Sistemi (Online Evaluation System)



<http://dergipark.gov.tr/gazibtd>

E-posta (e-mail): btd@gazi.edu.tr

Bilişim Teknolojileri Dergisi 3 ayda bir (Ocak, Nisan, Temmuz, Ekim) yayınlanmaktadır.
Journal of Information Technologies is published every 3 months (January, April, July, October).

Veri Madenciliği ile Hava Kalitesi Tahmini: İstanbul Örneği

Araştırma Makalesi/Research Article

 Nesibe MANAV MUTLU^{1*},  Eyüp Burak ATAHANLI²

¹İstanbul Nişantaşı Üniversitesi (Yönetim Bilişim Sistemleri (İng) Bölümü, İstanbul Nişantaşı Üniversitesi, İstanbul, Türkiye)

²İstanbul Nişantaşı Üniversitesi (Yönetim Bilişim Sistemleri Bölümü, İstanbul Nişantaşı Üniversitesi, İstanbul, Türkiye)

nesibe.manavm@gmail.com, eyupatahanli@gmail.com

(Geliş/Received:28.01.2024; Kabul/Accepted:12.05.2024)

DOI: 10.17671/gazibtd.1426942

Özet— Sanayi devrimi ile başlayan ve günümüzde tehlikeli boyutlara ulaşan hava kirliliği sorununun boyutunu ortaya çıkarmak, olası nedenleri arasındaki ilişkiyi belirlemek ve çözüm önerileri sunmak amacıyla İstanbul'un farklı ilçelerinden alınan çeşitli hava kirliliği verileri kullanılarak istatistiksel testler gerçekleştirilmiştir. Ayrıca önümüzdeki yıllarda oluşabilecek hava kirliliğinin büyüklüğünü tahmin etmek için makine öğrenmesi teknikleri de kullanılmıştır. Alınan verilere dayanarak 2023-2025 yılları için ilçeler bazında Prophet modeli kullanarak tahminler yapılmıştır. Bu tahminlere dayanarak önerilerde bulunulmuştur.(Özet)

Anahtar Kelimeler— hava kirliliği tahmini, veri madenciliği, hava kalitesi analizi

Air Quality Prediction with Data Mining: Istanbul Example

Abstract— Statistical tests were carried out using various air pollution data from different districts of Istanbul in order to reveal the extent of the air pollution problem, which started with the industrial revolution and has reached dangerous levels today, to determine the relationship between possible causes and to offer solution suggestions. In addition, machine learning techniques have been used to predict the magnitude of air pollution that may occur in the coming years. Based on the data received, predictions were made using the Prophet model on a district basis for the years 2023-2025. Recommendations have been made based on these predictions.(Abstract)

Keywords— air pollution prediction, data mining, air quality analysis.

1. GİRİŞ (INTRODUCTION)

Hava kirliliği, sadece hava kalitesini değil, su ve toprağı da etkilediği için tüm canlıların hayatını doğrudan etkileyen en önemli problemlerden biridir. Sadece iklim değişikliğine değil, ekonomiden göçe kadar pek çok sonuca da yol açar [1]. Buna karşı yıllardır alınmaya çalışılan önlemlerin yanı sıra insanların sahip olduğu farkındalık düzeyi de bu önlemlerin etkili olmasındaki önemli bir etkidir.

Hava kirliliği ile ilgili farkındalık düzeylerinin belirlenmesi amacıyla son yıllarda yerel ve uluslararası pek çok çalışma yapılmıştır [2] [3] [4][5] [6][7]. Genel anlamda temiz enerji kullanımının özendirilmesi, çeşitli tasarruf planları, doğru şehir yapılanmaları ve ağaçlandırma çalışmaları gibi global çözüm önerileri problemin çözümüne büyük katkı sağlasa da daha kapsamlı ve özel öneriler geliştirmek gereklidir [8] [9] [10] [11]. Zira hava kirliliğini etkileyen etmenler bir evin farklı odalarında bile çok değişiklik gösterebilir. Hal böyleyken gerek coğrafi koşullar gerekse sosyal yerleşim farklılıkları birbirinden çok ayrışan bölgeleri aynı havuz içinde değerlendirmek problemi uzun vadeli ve geniş bir ölçekte ele almamızı sağlasa da yerel yönetim birimleri probleme karşı yaklaşımlarını daha özel alanda yapmalı, birey olarak bizler de yaşadığımız çevrenin potansiyel risklerini daha dikkatli incelemeliyiz. Bu kapsamda büyük şehirleri mümkün olduğunca küçük birimler halinde incelemeli, hava kirliliğini arttıran etmenleri yerleşim yerine özel olarak tespit edebilmeli ve buna uygun önlemler alınabilmelidir. Çünkü hava kirliliği dünyadaki en başta gelen ölüm nedenlerinden olan kalp hastalıklarının tetikleyen en önemli etmendur [12] [13] [14]. Dünya Sağlık Örgütü 2021 verilerine göre, dünya üzerindeki her 10 kişiden 9'u hava kirliliğinden etkilenmektedir, bunun yanında her yıl hava kirliliğine bağlı olarak 4,2 milyon kişi de hayatını kaybetmektedir. Hava kirliliğinin neden olduğu hastalıklar ve ölüm oranları ülkelerin sağlık harcamalarını yani ekonomilerini da doğrudan etkilemektedir [15] [16] [17]. Sadece Türkiye'de hava kirliliği sonucu asgari 25,845 milyon TL, azami 52,492 milyon TL sağlık harcaması olduğu tespit edilmiştir [18]. UNEP(United Nations Environment Programme)'in dikkat çektiği üzere Dünya Bankası verilerine göre, hava kirliliği dünya ekonomisinde yıllık ortalama 5 trilyon dolara mal olmakta ve 225 milyar dolarlık da gelir kaybına neden olmaktadır. Bu nedenle hava kirliliği riskinin tahmin edilmesi halk sağlığı için önlemlerin önceden alınabilmesi devletlerin bütçesinde iyileştirme sağlayacağı gibi insanların da karbon ayak izi konusunda daha bilinçli hale gelmesine yardımcı olacaktır [19] [20]. UNEP bu konuda farkındalık yaratmada teknolojiyi kullanarak World Environment Situation Room (WESR) isimli bir dijital platformu 2022'de hayata geçirmiştir.

1.1. Alan Yazın (Related Works)

Literatürde hava kirliliğini makine öğrenmesi yöntemlerine dayanarak inceleyen pek çok çalışma

bulunmaktadır. Bunlardan birkaçına aşağıda yer verilmiştir.

Öklü ve Canbay tarafından yapılan çalışmada [21], City Data web sitesi üzerinden toplanan verilere hava kalitesi belirleyici değerlere göre 8 farklı sınıflandırma algoritması uygulanmıştır. Veri setleri 5- katlı çapraz doğrulama ile işleme alınarak başarı sonuçları 3 sınıflı ve 5 sınıflı veri seti için ayrı olarak sunulmuştur. Hava kalitesini sınıflandırmak için 3 ve 5 sınıflı bir sınıflandırma sistemi kullanılmıştır. Hava Kalitesi İndeksinin (AQI) hesaplanan sağlık etkilerine göre sınıfları belirlenmiştir. Makine öğrenimi teknikleri hava kalitesini 3 sınıflı modellerde %87, 5 sınıflı modellerde ise %82 doğrulukla tahmin etmiştir.

Çetin Yağmur'un çalışması [22], trafik yoğunluğu, meteorolojik koşullar ve NOX, SO₂, PM₁₀ hava kirlitcilerinin, çapı 2,5 mikrondan küçük olan PM_{2.5} parçacıklarını nasıl etkilediği incelemiştir. İstanbul Büyükşehir Belediyesi'nin veri paylaştığı açık veri portalı, Beşiktaş hava kalitesi izleme istasyonunun verilerinden yararlanılan çalışmada, Çoklu Doğrusal Regresyon (ÇDR), Rassal Orman (RO), Destek Vektör Makineleri (DVM) ve Yapay Sinir Ağları (YSA) modelleri atmosferik parçacık maddesini tahmin etmek için kullanılmıştır. Karşılaştırıldığında, algoritmaların performans kriterleri modele göre farklılık göstermekle birlikte ancak RO en iyi performans ortalamasına, ÇDR ise en kötü performans ortalamasına sahip bulunmuştur. Sonuçta, R² değeri 0,76 olan iyi bir tahmin modeli geliştirilmiştir.

Oğuz ve Pekin [23], PM₁₀ kirliliğini tahmin etmek için makine öğrenme teknikleri ve meteorolojik faktörler kullandıkları çalışmada, 39.9727 enlem, 32.8637 boylam ve 891 metre yükseklikte yer alan Meteoroloji Genel Müdürlüğü Ankara Bölge İstasyonundan elde edilen meteorolojik verilerden yararlanılmışlardır. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, Ankara Keçiören-Sanatoryum hava kalitesi istasyonunda PM₁₀ kirliliğine ilişkin veri toplamışlardır. Makine öğreniminin çalışma aşamasında, karar ağacı regresyonu, destek vektör regresyonunun, losso regresyonunun ve yapay sinir ağı uygulanmasını ile çalışmışlardır. Bu teknikler ile sıcaklık, çiğlenme noktası sıcaklığı, yağış, bağıl nem, rüzgar hızı, basınç, bulut örtüsü gibi çeşitli faktörleri ve önceki güne ait PM₁₀ ölçümlerini analiz edilmiştir. Sonuç olarak, yapay sinir ağı yöntemi önceki algoritmalarından daha iyi performans göstererek 0,6'lık bir belirleme katsayısı, 18'lik bir ortalama karekök hata ve 12'lik bir ortalama mutlak hata vermiştir.

Çelik ve Arıcı, Kovid-19 salgını öncesinde ve sırasında elde edilen kirlitici konsantrasyonlarının çoklu ölçümlerini hesaba katarak hava kalitesini tahmin etmek için makine öğrenimi tekniklerini kullandıkları çalışmada [24], gelişmiş sanayileri ve yüksek hava kirliliği seviyeleriyle tanınan Zonguldak ilindeki kirlitici konsantrasyonlarını içeren bir veri seti kullanmışlardır. Veriler Çevre ve Şehircilik Bakanlığı Hava İzleme İstasyonları internet sitesinden elde edilmiştir. Çalışma, tümü yüksek düzeyde

tahmin doğruluğu gösteren beş farklı makine öğrenimi tekniğini içermekle birlikte bu yöntemler arasında Karar Ağacı algoritması en olumlu sonucu vermiştir. Hastalığın yayılmasını azaltmak için uygulanan karantina döneminde kirlilik seviyelerinde gözle görülür bir azalma ve hava kalitesinde gözle görülür bir iyileşme yaşandığı tespit edilmiştir. Deneysel sonuçlar, önerilen modelin hava kalitesini etkili bir şekilde tanımlayabildiğini ve gelecekteki kirlenici miktarlarını tahmin edebildiğini göstermiştir.

Akgün ve Barlık çalışmalarında [25], hava kalitesi indeksi (AQI) ile ortalama sıcaklık ve yağış dahil olmak üzere meteorolojik veriler arasındaki ilişkiyi incelemişlerdir. Bağımsız değişkenler olarak bulutluluk, bağıl nem, basınç ve rüzgar hızı ile çoklu doğrusal regresyon analizi ve makine öğrenme tekniklerini kullanarak AQI'yi tahmin etmeyi amaçlamışlardır. Çoklu doğrusal regresyon analizi kullanılarak tahmin etme doğruluğu %41,8 olarak belirlenmiş, Naive Bayes algoritmasının doğruluğu %68,91 ve Karar Ağacı algoritması %70,29'luk bir doğruluğa ulaşırken, Derin öğrenme algoritması %73,05, k-NN algoritması %73,35 ve son olarak Rastgele Orman algoritması %74,89'luk bir AQI tahmin doğruluğuna ulaşmıştır.

Gültepe [26], meteorolojik verileri kullanarak Kastamonu'daki hava kirliliğini simüle ettiği çalışmada, hava kirliliğini tahmin etmek için meteoroloji ve çevre uygulamalarında kullanılan makine öğrenimi teknolojilerinden yararlanmıştır. Sistem performansına ilişkin normalleştirme yaklaşımlarını değerlendirmek için veri kümesi değişkenleri Minimum-Maksimum (Min-Maks) normalleştirme kullanılarak normalleştirilmiştir. Çalışma, performans değerlerini önceki çalışmalarla karşılaştırmış ve zorluk için en iyi sınıflandırma algoritmasını bulmuştur. Yapay Sinir Ağları (YSA), Rastgele Orman, K-En Yakın Komşuluk, Lojistik Regresyon, Karar Ağacı, Doğrusal Regresyon ve Naive Bayes ayrı ayrı oluşturulmuş ve incelenmiştir. Verilerin analizinde Naive Bayes kullanılmıştır. Eğitim ve test verileri veri setinin %70 ve %30'unu oluşturur. Tahmin edildiği gibi yapay sinir ağları %91 başarı oranıyla en doğru tahmin yaklaşımı olmuştur. Doğrusal Regresyon %30 ile en düşük sınıflandırma başarı oranına sahipken, Rastgele Orman ve Karar Ağacı %99 başarı oranına sahip olmuştur.

Ayrıca Wang, Pereira ve Hung [27], hava kirliliği için kontrol süreçlerinin altında yatan ilkeleri ve uygulamaları eleştirel bir şekilde araştırıp bunları uygulamacı mühendisler için bir dizi ayrıntılı tasarım örneğiyle açıklayan ve pratik tasarım hesaplama yöntemlerini çok sayıda sayısal hesaplama ile gösterdikleri bir kitap çalışması yapmışlardır.

Tsai, Zeng ve Chang'ın çalışmalarında [28], PM2.5 konsantrasyonunu tahmin etmek için Uzun-Kısa Süreli Belleğe sahip bir RNN (Recurrent Neural Network-Tekrarlı Sinir Ağları) kullanılmıştır. Python tabanlı Keras, ve Tensorflow'dan yararlanılmıştır. Ağ eğitimi için

2012'den 2016'ya kadar olan Tayvan Çevre Koruma İdaresi verileri kullanılmıştır. Tahmin edilen test verileri 2017 yılına ait verilerden seçilmiştir. PM2.5 konsantrasyonunun tahmin gücü dört saat boyunca Tayvan'daki 66 istasyonda test edilmiştir. Sonuçlar önerilen tekniğin PM2.5'i tahmin edebildiğini göstermiştir.

Rybarczyk ve Zalakeviciute çalışmalarında [29], meteorolojik faktörlerin bir kombinasyonuna dayalı olarak ince partikül madde tahminine yönelik yöntem belirlemeyi amaçlamışlardır. Ekvador'un Quito şehrinin uzun yıllara dayanan meteorolojik verileri, makine öğrenimi metodolojisini kullanarak modeller oluşturmak için kullanılmıştır. Makalede, ince parçacıkların konsantrasyonlarını etkili bir şekilde iki kategoriye ayıran bir karar ağacı tekniği sunulmuştur.

Bu çalışmada, yukarıdaki örnekler doğrultusunda İstanbul'un halka açık verileri [30] kullanılarak istatistiksel yöntemler ve makine öğrenmesi algoritmaları yardımıyla, İstanbul'un hava kalitesinin ölçüldüğü ilçelerinde geçmişten bugüne analizler yapılmış ve önümüzdeki yıllarda kirliliğin varabileceği boyutlar makine öğrenmesi modeliyle tahmin edilmiştir. Bu sayede tüm çevreyi etkileyen bu kirlilik krizinin araştırılması sayesinde potansiyel salgın hastalıklar, tarımsal salgınlar ve belki hayvanlardan insanlara bulaşabilen salgınlar bile önlenilebilecektir [31] [32] [33] [34]. Bulgular sayesinde, sürdürülebilir çevre politikalarının belirlenmesinde farkındalık geliştirilmesi, insanların bireysel günlük faaliyetlerinin bile hava kirliliği konusundaki etkisi hakkında bilinçlenmeleri, daha sağlıklı bir toplum için önleyici hizmetlerin hayati etkisi belirlenebilecektir.

Sürdürülebilirlik için hava kirliliği üzerine yapılacak bu çalışma üniversite düzeyinden başlayarak tüm öğrenim düzeylerindeki gençlerin dikkatini çekecek sürdürülebilir bir yatırıma dönüşecektir. Aynı zamanda İstanbul Belediyesinin ilçeler bazında alacağı önlemlere yahut yeni geliştireceği projelerde rapor niteliğinde olacaktır, kurumun sürdürülebilir yatırımlara öncelik vermesine katkı sağlayacaktır.

1.2. Hava kirliliği metriklerinin tanımlanması (Defining air pollution metrics)

Hava kirliliği verilerinde kullanılan temel değerler ve tanımları aşağıdaki şekilde listelenmiştir:

PM10 ve PM2.5: Partikül madde (PM) kirliliği, hava kirliliğinin ana kaynaklarından biridir. PM10, havadaki partiküllerin çapının 10 mikrometreden küçük olan bir bölümünü ifade ederken, PM2.5, çapı 2.5 mikrometreden küçük olan partiküllerin oranını ifade eder. Birçok çalışma PM10 seviyelerindeki artışın astım başta olmak üzere çeşitli hastalıkların semptomlarını arttırdığını göstermektedir [35] [36].

NO₂: Nitrojen dioksit, fosil yakıtların yakılması ve endüstriyel faaliyetler gibi insan kaynaklı etkenlerden kaynaklanan bir hava kirliliği türüdür [37].

SO₂: Kükürt dioksit, kömür yakma gibi endüstriyel faaliyetlerden kaynaklanan bir hava kirliliği türüdür [37].

CO: Karbon monoksit, fosil yakıtların yakılması gibi kaynaklardan salınan bir hava kirliliği türüdür [37].

Daha detaylı çalışmalarda partikül maddenin detaylı analizi ve havada bulunan kurşun, civa, gibi özel elementler incelenebilir. Bu verilerin yanında rüzgâr, nem, sıcaklık gibi meteorolojik veriler de kullanılabilir. Bu değerler, hava kalitesi indeksleri gibi farklı metriklerde de kullanılabilir ve hava kirliliğinin etkilerini ölçmek için birlikte değerlendirilir.

AQI (Air Quality Index) yani Hava Kalitesi İndeksi [38], hava kirliliğini ölçmek ve raporlamak için kullanılan bir ölçüttür. AQI, çeşitli hava kirleticilerinin konsantrasyonlarını ölçer ve bu kirlilik seviyelerini belli aralıklara göre altı farklı kategoriye ayırır:

- İyi (0-50 AQI)
- Orta (51-100 AQI)
- Hassas Gruplar İçin Kötü (101-150 AQI)
- Kötü (151-200 AQI)
- Çok Kötü (201-300 AQI)
- Tehlikeli (301-500 AQI)

AQI, PM_{2.5}, PM₁₀, NO₂, SO₂, O₃ ve CO gibi hava kirleticilerinin konsantrasyonlarını temel alarak hesaplanır ve hava kirliliği seviyesini belirlemek için kullanılır [21]. AQI'nin kullanımı, halk sağlığı açısından önemli bir araçtır ve hava kirliliği seviyelerinin ölçümünü ve takibini kolaylaştırır.

1.3. Veriye Erişim Yolları (Ways to Access Data)

Bu çalışmada veriye erişim sağlamak için İstanbul Büyükşehir Belediyesi (İBB) 'nin **api** araçları kullanılmıştır. Bu yöntemle İBB' nin veri havuzunda belirtilen yönergeler doğrultusunda hava kirliliği verisi de dahil birçok halka açık veriye erişim sağlanabilir [30]. Bu doğrultuda **Python** programlama dili ve **Json** request kütüphaneleri kullanılarak veriler çekilmiştir. Veriye erişim sürecinde, ölçüm istasyonu bulunan ilçeler ve semtler seçilerek ve istenilen tarih aralığı belirtilerek veriler alınmıştır. Çalışma kapsamında 2010 yılından 2023 yılına kaydedilen 430 MB'lık veri kullanılmıştır. Eğer seçilen konumda ve tarih aralığında ölçüm yapılmamışsa, içerik boş değeri temsil eden *NaN* değeri olarak dönmüştür. Veri, json formatında alınmıştır. Veri json formatında alındıktan sonra veri tipi dönüşümleri yapılmıştır. Bu dönüşüm çalışmanın içeriğindeki matematiksel işlemler için gereklidir. Sonrasında çekilen dosyaların formatı json'dan csv'e dönüştürülmüştür.

2. CRISPDM METODUYLA PROBLEMİ ÇÖZME (SOLVING THE PROBLEM WITH THE CRISP DM METHOD)

Endüstriler genelinde veri madenciliği süreçlerini standartlaştırmak için 1999 yılında yayınlanan bu yöntem, o zamandan beri veri madenciliği, analitik ve veri bilimi projeleri için en yaygın metodoloji olarak kullanılmaktadır. Bu makalenin ilgilendiği sorunun çözülmesinde her aşama modelin ilgili aşamasına göre uyarlanmış olarak aşağıda sırasıyla görülmektedir.

2.1. Problemi Anlamak (Understanding the Problem)

Hava kirliliği sorunu ele alınırken geniş coğrafyalarda ölçüm yaparak değerlendirmek yerine küçük birimler halinde incelenmek istenmiştir. Bu doğrultuda İstanbul' un halka açık verileri kullanılarak geçmiş yıllardaki hava kirliliği verileri analiz edilecek ve gelecek için birtakım tahminler ve önerilerde bulunulacaktır.

2.2. Veriyi Anlamak (Understanding Data)

İlgili kaynaktan edinilen veri gerekli düzenlemeler yapıldıktan sonra bir Pandas veri tipi olan Dataframe yapısına dönüştürülmüştür. Pandas Python programlama dilinde veri manipülasyonu ve analizi için kullanılan güçlü bir kütüphanedir. Dataframe ise veri analizi ve veri işleme için kullanılan bir veri yapısıdır. Genellikle tablo benzeri bir yapıya sahip olan dataframe, sütunlar ve satırlardan oluşur. Türkçeye veri çerçevesi olarak çevrilse de çalışma boyunca yaygın kullanım olan DataFrame olarak nitelendirilecektir. Bütün istasyonlar için ayrıca Dataframe oluşturulduktan sonra ortak yapılacak işlemler için bu veri yapıları bir sözlük yapısı içinde toplanmıştır. Sözlük yapısı ise anahtar-değer çiftlerinden oluşan bir koleksiyonu temsil eder. Burada anahtarlar benzersiz seçilmelidir, ancak anahtarların temsil ettiği değerler herhangi bir Python nesnesi olabilir.

Sonuç olarak 'dfs' sözlük yapısı elde edilmiştir. Bu sözlüğün içinde 'Arnavutköy', 'Beşiktaş', 'Kartal', 'Bağcılar', 'Göztepe', 'Sancaktepe', 'Tuzla', 'Sultangazi1', 'Sultangazi2', 'Sultangazi3', 'Üsküdar1', 'Yenibosna', 'Mobil', 'Sarıyer', 'Ümraniye1', 'Esenler', 'Aksaray', 'Alibeyköy', 'Avcılar', 'Kadıköy', 'Kağıthane1', 'Selimiye', 'Beylikdüzü' isimli anahtar değerler eklenmiştir. Görüldüğü üzere bazı ilçeler birden fazla veri setiyle temsil edilmektedir, veri kaynağı olan İBB o ilçede birden fazla ölçüm istasyonu bulundurduğu için ayrı ayrı değerler numaralandırılarak saklanmayı tercih edilmiştir. Ayrıca bir yere bağlı olmayan, sürekli konumu değişen Mobil ölçüm istasyonu da sözlük yapısının içinde yer almaktadır.

Buna göre ilçe ve semtlerin kısa-orta-uzun zaman dilimlerinde ortalama AQI değişimleri analiz edilmiştir. Uzun zaman dilimi için yıllık ortalama AQI değerleri, orta zaman dilimi için mevsimlere göre ortalama AQI değerleri, kısa zaman için ise günlük değişimler görselleştirilmiştir.

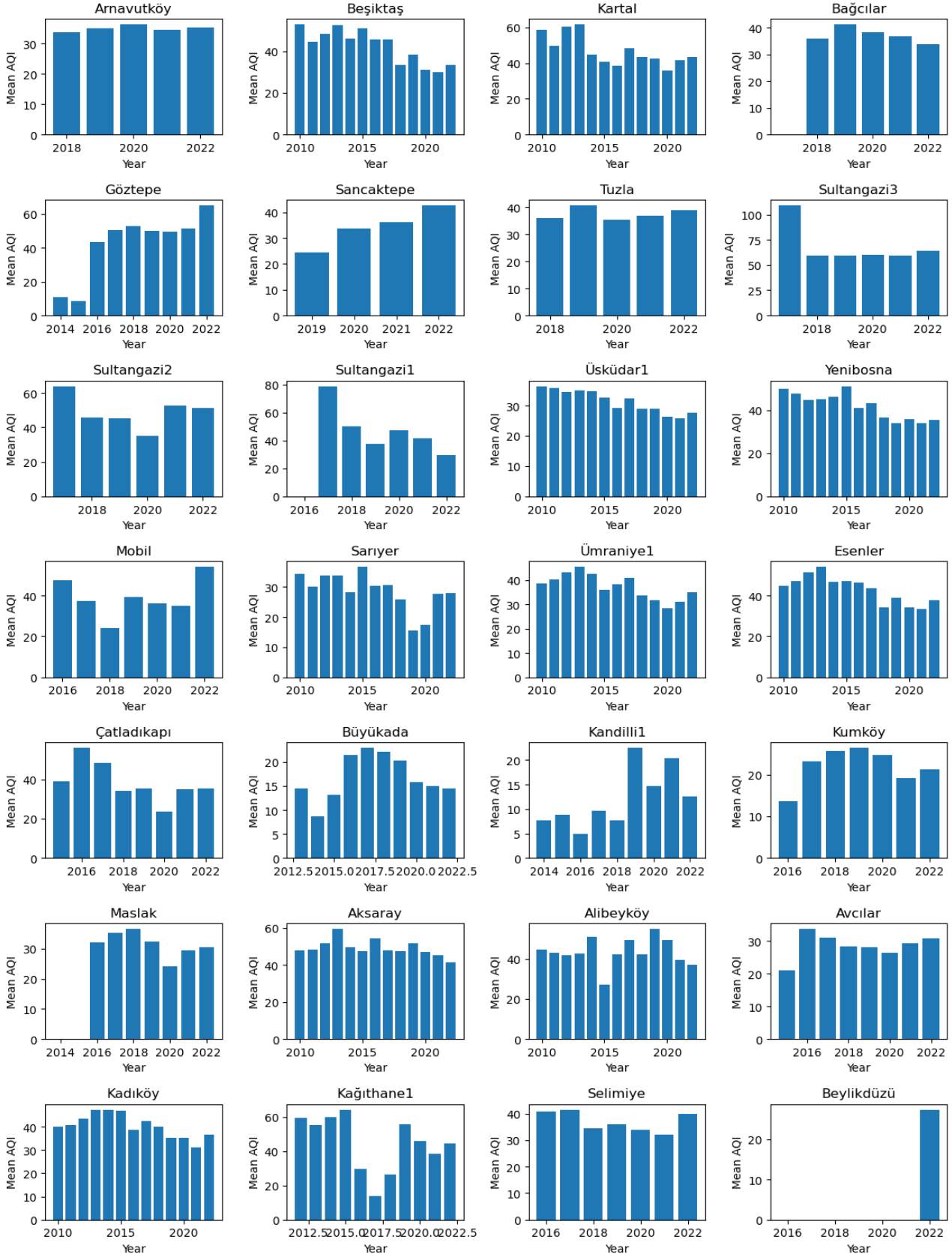
Bu görselleştirmelerden biri aşağıda yer almaktadır. Şekil 1'de yer alan grafiklere göre;

- Yıllara göre ölçüm değeri azalma eğiliminde olan yerler Beşiktaş, Kartal, Bağcılar, Üsküdar, Yenibosna, Ümraniye, Esenevler, Büyükkada, Kadıköy, Sultangazi (1.bölge) ilçe ve semtleridir. Bu yerleşim yerlerinde ölçüm yapılan yıllardan itibaren düzenli düşüş görülmesi de iyileşmeler olduğu ortadadır.
- Ölçüm değerleri artış eğiliminde olanlar ise Göztepe, Sancaktepe, Kandilli şeklindedir. Burada düzenli bir değişim görülmemekle birlikte ölçüme başlanan ilk yıllara göre hava kirliliğinde ölçülen değerlerde bir artış gözlenmektedir. Özellikle Göztepe'de son yıllarda oluşan kirliliği arttıran etmenler başka bilimsel araştırmaların konusu olabilir.
- Esenler, Beşiktaş, Üsküdar, Ümraniye, Kadıköy gibi şehrin eski yerleşim yerlerinde 2010 yılına göre ölçümlerde iyileşme görülmektedir. Bu tarih aralığında söz konusu ilçelerde bu iyileşmenin olası sebepleride araştırılıp ölçüm değerlerinde düşüğe neden olan tüm uygulamalar yaygınlaşması faydalı görülmektedir.

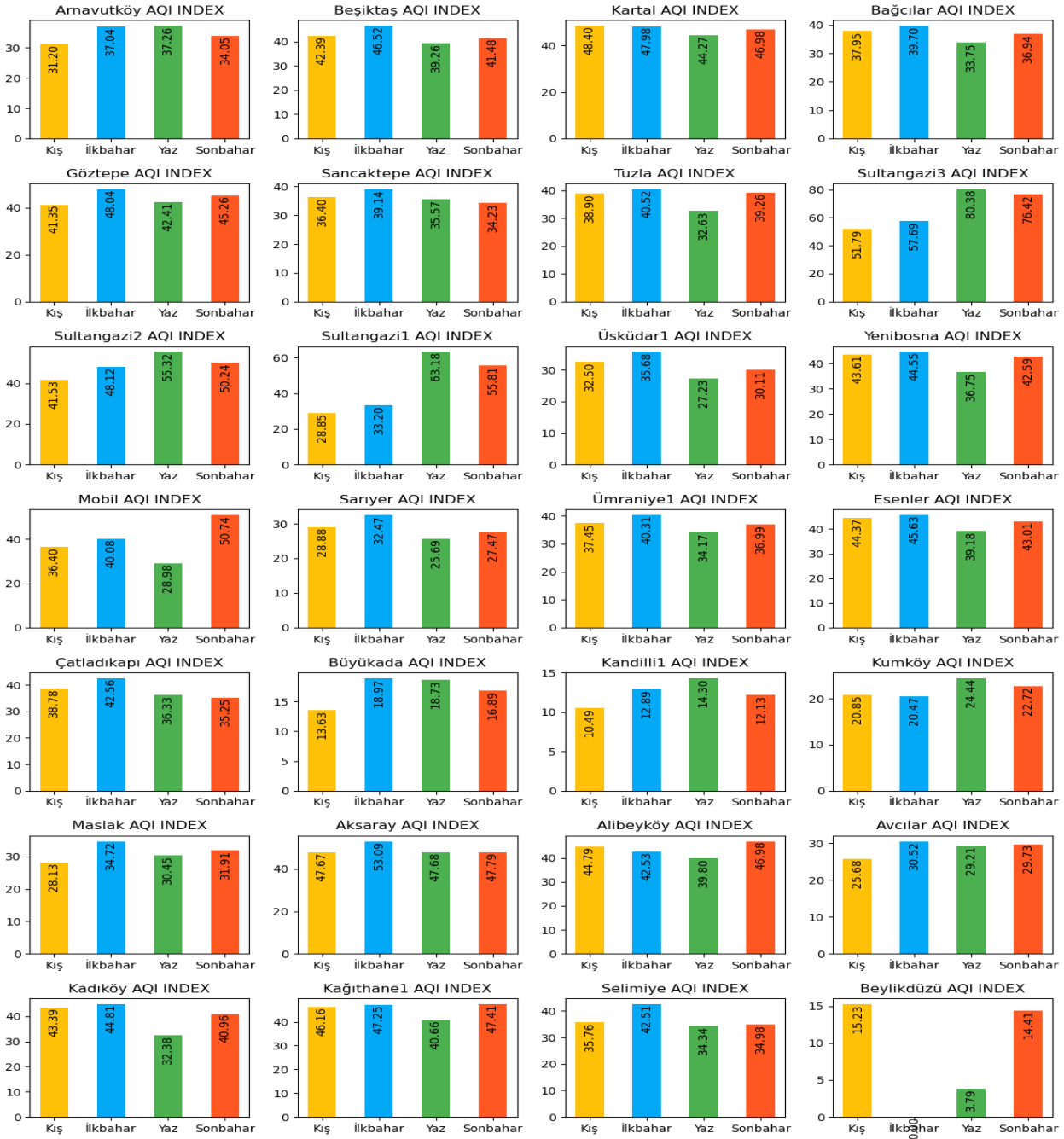
• Ayrıca, grafiklere bakıldığında bir diğer dikkat çeken durum Sarıyer istasyonunda 2019-2020 yıllarında yaşanan kayda değer düşüş olmuştur. İlk bakışta bu dönemin yaşanan Covid-19 pandemisinin etkisi söz konusu olabilir, zira bölge uzaktan çalışmaya da elverişli olan bir çok iş kolunda çalışmalar yapan kurumlara ev sahipliği yapmaktadır. Uzaktan çalışma ve esnek saatler uygulamalarının yaygınlaşmasının ardından trafik ve dolayısıyla kirliliğin azalması bu sonuçlara neden olmuş olabilir.

• Sarıyer istasyonunda olduğu gibi Kağıthane1 istasyonu da 2017'de kaydettiği dip değerler ile dikkat çekmektedir. Buradaki düşüşün %70'e varan oranda olması istasyonun ölçümlerinin incelenmesi gerekliliğine işaret etmektedir. Bu durumun, öncesi ve sonrası olarak başka yıllarda da yaşandığı göz önünde bulundurulmalıdır.

• Sancaktepe istasyonu, 2019'dan itibaren veri kaydetmeye başlamış ve takip eden yıllarda ölçüm değerleri sürekli artış göstermiştir. Bu nedenle burada artışın nedenlerinin araştırılması faydalı olacaktır.



Şekil 1. Geçmiş yıllara göre ilçe-semtlerin AQI değerleri (AQI values of districts and districts according to previous years)



Şekil 2. 2010-2022 yılları arası Mevsim kırılımında AQI ortalamaları (AQI averages by season)

Yukarıdaki Şekil 2’de istasyonların 2010-2022 yılları arasında ölçülen AQI değerlerinin mevsim kırılımında ortalamaları gösterilmiştir. Bazı istasyonlar 2010 yılından sonra veri kaydetmeye başlamıştır. Bu grafikler ilk ölçüm tarihinden 2022 yılına kadar olan sürede değerlendirilmektir.

Grafiğe göre istasyonların bulunduğu yerleşim yerlerinde ortak bir örgü bulunmamaktadır. Bu durum hava kirliliği probleminin bir çok etmene bağlı [39], küçük mesafelerde bile hızlı bir şekilde değişebilen [40], ayrıca alınan önlemlerin sonuçlarının da hızlı bir şekilde gözlemlenebilen [41] bir problem olduğunu bir kez daha

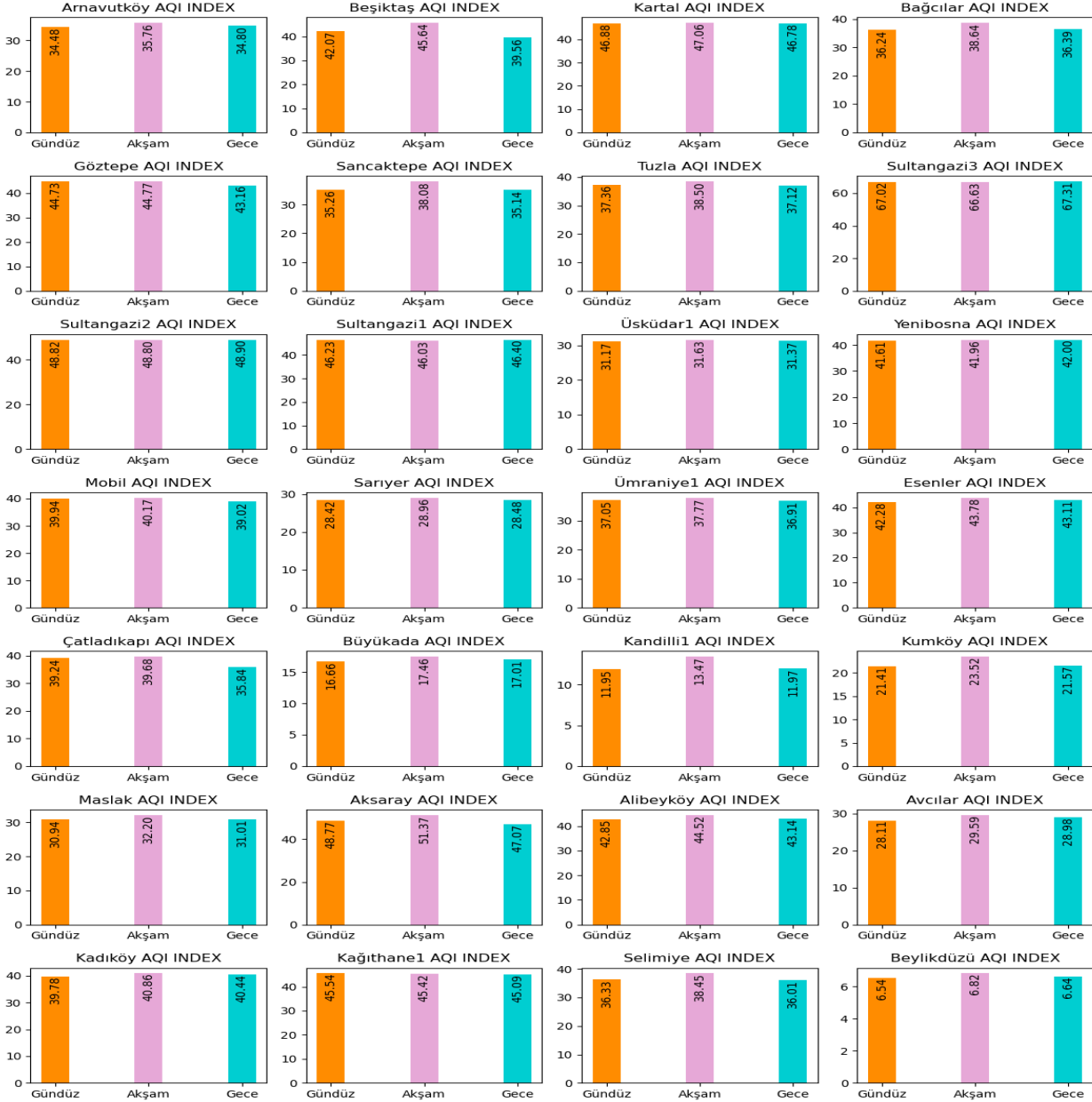
göstermiştir. Bütün bunlar bu çalışmanın önemini daha iyi bir şekilde vurgulamaktadır.

En çok dikkat çeken 3 ölçüm istasyonu da Sultangazi ilçesinde bulunmaktadır. Bu istasyonların ölçümleri mevsimler arası büyük değişim göstermektedir. İstasyonların ölçümleri yaz aylarında %110’a varan bir artış göstererek artış düzeyini gözler önüne sermektedir. Bölgedeki taş ocaklarının bu artışa sebep olup olmadığı, olası diğer etmenler başka bilimsel çalışmaların konusu olarak araştırılabilir.

Öte yandan yaz aylarında kirliliğin azaldığı, kışın %35’e varan artışların olduğu ölçüm istasyonları da

bulundurmaktadır (Beylikdüzü kısıtlı veri sebebiyle bu analize dahil edilmemiştir). Bunların ortak özelliği diğer yerleşim yerlerine göre daha eski bir tarihi olan, nüfus yoğunluğu fazla ve konutlarında kışın ısınma enerji harcamasını düşürecek yeni nesil altyapısı bulunmayan bölgeler olduğu dikkat çekmektedir.

Yaz aylarında normalde 20 binden az olan toplam Adalar nüfusunun 50-70 binlere çıkması nedeniyle [42] Büyükada'da değerlerdeki %50'ye yakın artışın turizm kaynaklı olduğu tahmin edilmektedir. Bu artışa rağmen adada taşıt trafiği olmaması, geniş ormanlık alanları sayesinde ölçüm değerleri iyidir.



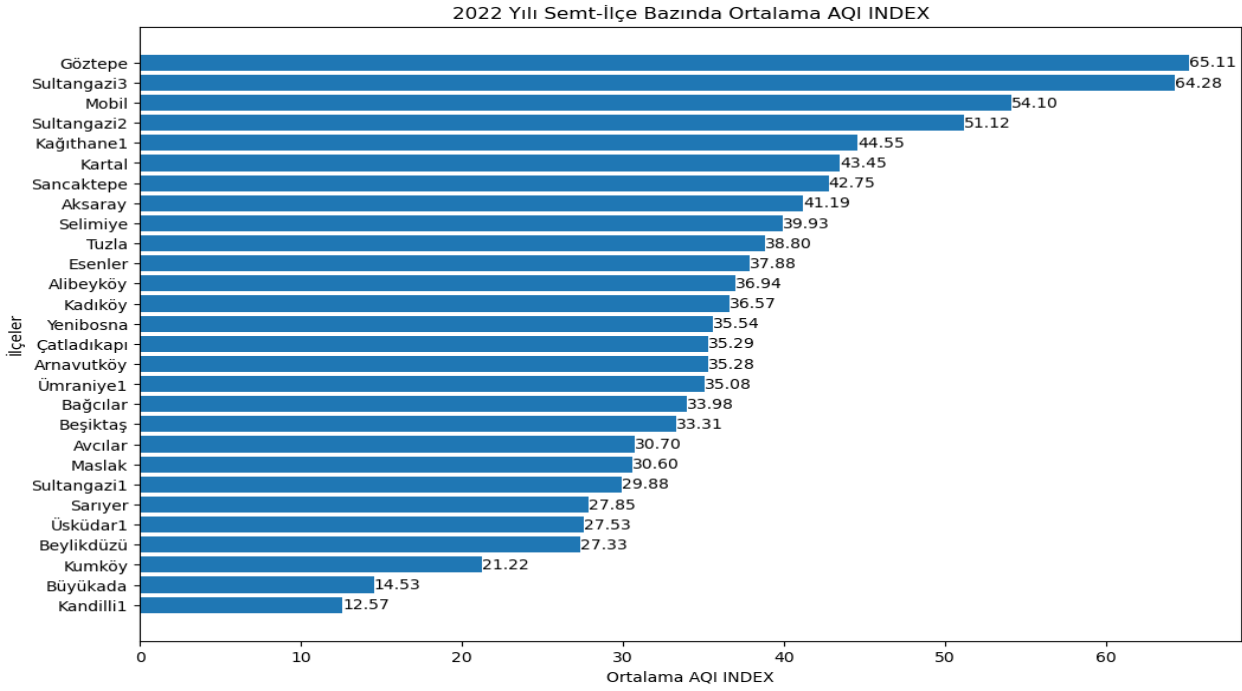
Şekil 3. Gün içi saat dilimlerinde AQI ortalama (AQI average across time zones during the day)

Yukarıdaki Şekil 3, istasyonların gün içinde gündüz (08:00 – 16:00) akşam (16:00 - 00:00) ve gece (00:00 – 08:00) olmak üzere 3 zaman diliminde saatlik AQI değerlerinin ortalamasını göstermektedir. Genel eğilime baktığımızda gün içinde değişimlerin sınırlı olduğu ortaya çıkmaktadır. Bu değişim akşam saatlerinde artan kirlilik olarak kendini göstermektedir. Özellikle ciddi trafik problemi olan Beşiktaş, Maslak, Üsküdar Kandilli bölgesinde artış daha

belirgin olarak görülmektedir [43][44]. Kirliliğin önemli unsurlarından olan partikül madde artışı trafikte dizel araçların kullanımıyla doğrudan ilişkilidir [45]. İstanbul'da dizel araçların oranı Avrupa birliği ülkelerine göre daha yüksek seviyelerdedir[46]. Söz konusu ülkelerde dizel araçların sıkı emisyon ölçüm denetimlerine tabi tutulması, bazı ülkelerde yasaklar uygulanması dizel araçlara gösterilen ilgiyi azaltmıştır. Bu durum tüm dünyada dizel

araç üretiminin azalması ve yeni nesil motor teknolojilerinin gelişmesine sebep olmuştur, öyle ki ülkemizde de trafiğe kaydı yapılan yeni araçlarda dizel araçların oranları

gittikçe düşmekte, alternatif olarak elektrikli araçların kullanımı da yaygınlaşmaktadır [47].



Şekil 4. İlçe-Semtlerin Ortalama AQI değerleri (Average AQI values of districts)

Şekil 4'e göre 2022 yılında en yüksek AQI değerlerini ölçen istasyonlar; Göztepe, Sultangazi (3.bölge), Mobil istasyonudur.

Mobil istasyon Şekil 1'e göre en yüksek değerini 2022 yılında kaydetmiştir. Bu istasyonun ölçüm yaptığı yerler kaydedilmeli ve bunlara özel analizler yapılmalıdır. Veri setinde böyle bir veri olmadığı için böyle bir yorumda bulunmak bu çalışma kapsamında mümkün olmamıştır.

Öte yandan 2022 yılı içinde en düşük AQI değerleri kaydeden istasyonlar; Kumköy, Büyükada ve Kandilli'dir.

Kumköy, İstanbul'un Sarıyer ilçesine bağlı bir sahil semtidir. Genellikle yaz aylarında yerli ve yabancı turistlerin tercih ettiği bir destinasyondur. Kilyos olarak da bilinen Kumköy, Karadeniz kıyısında yer alır ve güzel plajlarıyla ünlüdür. Kumköy daha az sanayileşmiş ve kalabalık olmayan bir bölge olduğu için, yakın çevresinde ormanlık alanlara sahip olduğu için genellikle İstanbul şehir merkezine göre daha temiz bir havaya sahiptir.

Büyükada ise İstanbul'un en popüler adalarından biridir ve Marmara Denizi'nde yer alır. Büyükada, İstanbul'un Anadolu Yakası'na bağlı olan ve İstanbul Boğazı'ndan uzakta bulunan bir adadır. Adanın toplam yüzölçümü yaklaşık 5,4 kilometrekaredir ve nüfusu oldukça sınırlıdır. Büyükada, doğal güzellikleri ve tarihi atmosferiyle ünlüdür. Ayrıca adanın doğal alanları ve ormanlık

bölgeleri, oksijen bakımından zengin bir ortam oluşturur. Büyükada'nın araç trafiğine kapalı olması ve bisiklet veya fayton gibi alternatif ulaşım araçlarının kullanılması, adanın temiz bir atmosfer sunmasını sağlar.

Kandilli ise İstanbul'un Üsküdar ilçesine bağlı bir semttir. İstasyon Boğaziçi Üniversitesi'nin Kandilli yerleşkesinde bulunmaktadır ve ormanlık alanlarla çevrilidir. Yerleşke İstanbul'un diğer bölgelerine göre daha temiz bir hava kalitesine sahiptir.

2.3. Veriyi Hazırlamak (Preparing the Data)

Verinin içinde bulunan ReadTime isimli değişken ölçüm yapılan zamanı temsil etmektedir. Temsili formatı ise; 2022-12-31 23:59:00 şeklindedir.

Bu değişkenin türü veri kaynaklarından metin türü(string) olarak gelmektedir. Ancak çalışmayı kolay yapabilmek ve gerekli istatistiksel çıkarımları görebilmek için bu string yapısını yine Python'da bulunan Datetime (zaman biçimi) türüne çevrilmiştir. Ek olarak her ilçenin hava kirliliği verilerini barından Datetime yapısına özel tarih değişkenleri eklenmiştir. Bunlar; gün, ay, yıl, saat, mevsim, günün bölümleri şeklinde olmuştur.

Bu sayede hem yapılan istatistiksel çıkarımları değerlendirmek daha kolay olmuş, hem de kurulan modele bu bağımsız değişkenleri ekleyerek bağımlı değişkenle arasındaki bağlantıyı öğrenmesi sağlanmıştır.

2.4. Modelleme (Modelling)

Bu bölümde öncelikle zaman serisi problemi modellemede kullanılan bazı modeller tanıtılmış sonrasında ise bu çalışmanın konusu olan modele yer verilmiştir.

Zaman serisi modellemede oldukça fazla alternatif mevcuttur. Bununla ilgili birkaç model aşağıda verilmiştir:

ARIMA: Zaman serilerindeki otokorelasyonu ele alan ve durağanlık olmayan bileşenleri modellemek için kullanılan bir yöntemdir. Geçmiş değerlerin kombinasyonu ile mevcut değerleri tahmin ederek gelecekteki değerleri belirleyebilir.

SARIMAX: Arima modeline benzer şekilde çalışır, ancak mevsimsel bileşenlere sahip zaman serilerini daha etkili bir şekilde modelleyebilir. Ayrıca Arima yöntemine dışsal değişkenlerin eklenmesiyle genişletilmiş bir yöntemdir. Bu model, mevsimsel örüntülerin yanı sıra dışsal faktörlerin etkisini de hesaba katarak tahminleri yapabilir.

LSTM: Derin öğrenme tabanlı bir yapay sinir ağı modelidir ve zaman serilerinde uzun vadeli bağımlılıkları ele almak için etkilidir. Geçmiş gözlemlerin hafızasını kullanarak gelecekteki değerleri tahmin etme yeteneğine sahiptir. Model başarısını birçok regresyon probleminde kanıtlamış olmasına rağmen, derin öğrenme yapısı bunun gibi birden fazla model barındıran projelerde eğitim sürelerinin uzamasına yol açmaktadır. Veri setinin saatlik olduğu düşünülürse bu dezavantaj daha da gün yüzüne çıkacaktır. Çeşitli ölçeklendirme yöntemleriyle kullanıldığında başarılı olma ihtimali yüksek bir modeldir.

Prophet: Meta (eski adıyla Facebook) tarafından geliştirilen açık kaynak kodlu bir zaman serisi tahminleme algoritmasıdır. Bu model, zaman değişkeni (ds) ve bağımlı değişken (y) olarak iki girdi alır. Prophet, bu verileri kullanarak tahmin değeri (yhat) ile birlikte tahminin üst sınırını (yhat_upper) ve alt sınırını (yhat_lower) belirleyen bir aralık tahmini yapar. Bu çalışmada, kullanım kolaylığı, hızlı ve basit yapısı ile Prophet modeli tercih edilmiştir.

Prophet modeli, zaman serilerinin yapısını analiz eder ve trendleri, mevsimsel desenleri ve diğer bileşenleri hesaba katarak gelecekteki değerleri tahmin etmeye çalışır. Tahminlerde, yhat değeri genel tahmin değerini temsil ederken, yhat_upper ve yhat_lower değerleri ise tahminin üst ve alt sınırlarını gösterir. Bu aralık tahmini, tahminin belirsizlik düzeyini ifade eder ve güven aralığı sağlar [48].

Hava kirliliği ölçümü yapan istasyonların tümünde elementlerin tamamı ölçülmediği için bazı veri setlerinde eksiklikler bulunmaktadır. Bu modelin kurulma aşamasında hata alınmasına neden olur. Bundan kaçınmak için tüm veri setinde bulunan AQI verisinin analiz edilmesi daha doğru olmuştur.

Model yapay zeka ve derin öğrenme projeleri için Google tarafından piyasaya sürülen **Google Colab** programlama ortamında Python dili ile gerçekleştirilmiştir. Colab, bulut sunucularla çalışan bilgisayar sistemleri sayesinde yalnızca internet bağlantısı ve bir tarayıcı yardımıyla

modeller kurmaya yardımcı olur. Model hızlı ve düşük bellek ihtiyacından dolayı gerekli kütüphanelerin kurulumu sonrasında giriş düzeyi donanım özellikleriyle kısa sürede sonuç alınabilmektedir. Bir sonraki bölümde bu çalışmanın çıktıları grafiklerle verilmiştir.

2.5. Değerlendirme (Evaluation)

Makine öğrenmesi modellerinin başarısını ölçmek için birden fazla hesaplama yöntemi vardır. Bu iş probleminde amacımız sürekli bir değişkeni modellemek olduğu için **MAPE** (Mean absolute percentage error) metriği seçilmiştir. Literatürde MAPE ile model doğruluğu örnekleri mevcuttur (bkz. [49] [50] [51] [52]). MAPE Türkçe karşılığı ile Ortalama Mutlak Yüzde Hatası, tahmin modellerinin performansını değerlendirmek için kullanılan bir metriktir. MAPE, tahmin edilen değerlerin gerçek değerlere göre yüzde olarak ne kadar sapma gösterdiğini ölçer [49]. Denklemi aşağıdaki eşitlikle ifade edilmiştir:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

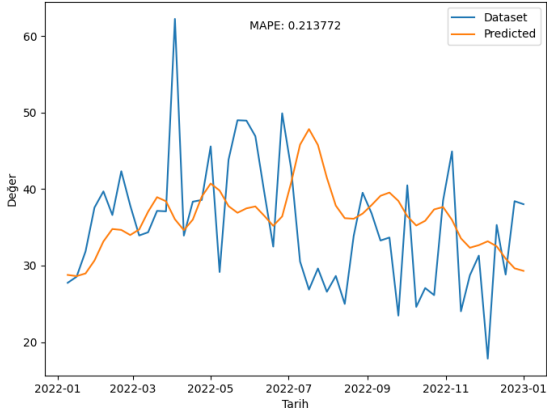
Burada Y_i gerçek değerdir ve \hat{Y}_i tahmin değeridir. Farkları Y_i 'nin gerçek değerine bölünür. Bu oranın mutlak değeri, zaman içinde tahmin edilen her nokta için toplanır ve uygun noktaların n sayısına bölünür. MAPE değeri yüzde olarak ifade edilir ve genellikle 0-1 arasında bir değer alır. MAPE'nin düşük olması, model tahminlerinin gerçek değerlerden daha az sapmış olduğunu gösterir. MAPE, temel içsel parametreleri kullanarak zaman serilerini büyük bir doğrulukla tahmin etmemizi sağlar [52].

2.6. Bulgular ve tartışma (Findings and Discussion)

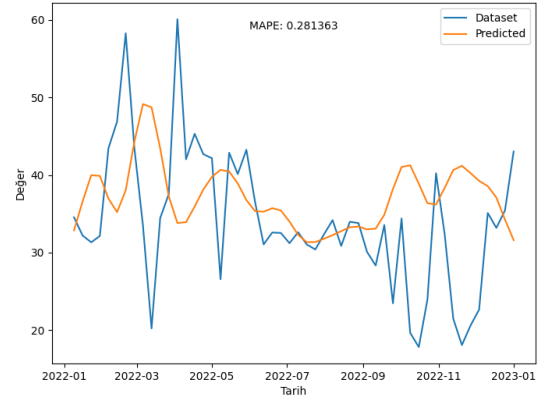
Modelleme sonucu elde edilen bulgular devamında verilmiştir.

Şekil 5'te istasyonlar için kurulan modellerin 2022 yılı içinde yaptığı tahminler ve ölçülen gerçek değerler karşılaştırılmıştır. Grafik üzerindeki MAPE değeri de model başarısını sayısal olarak ölçmek için eklenmiştir. Kurulan modellerin eğitim setinde 2022 yılına ait verilerin çıkarıldığı unutulmamalıdır. Model başarısı ancak ve ancak daha önce görmediği verileri tahmin ederek objektif olarak değerlendirilebilir.

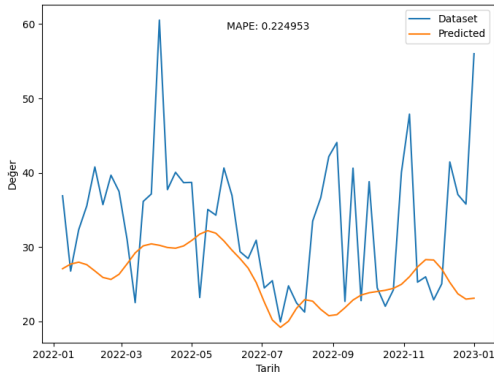
Aşağıdaki Şekillerde 5'ten 26'ya kadar olan grafiklerde mavi çizgiler ölçülen AQI değerlerini, turuncular ise modelin geçmiş verilerden yola çıkarak tahminlediği AQI değerlerini temsil etmektedir. Eğitim ve tahminler aylık ortalamalar üzerinden yapılmış, her ölçüm istasyonunun 2010 yılına kadar ölçüm yaptığı tüm zamanlardan 2021 aralık ayına kadarki kaydettiği değerler eğitim seti olarak modele sunulmuş ve sonrasında 2022 yılı için tahmin yapılması istenmiştir.



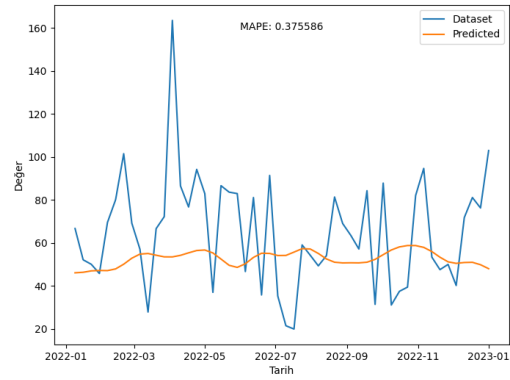
Şekil 5. Arnautköy istasyonu model değerlendirme (Arnautköy station model evaluation)



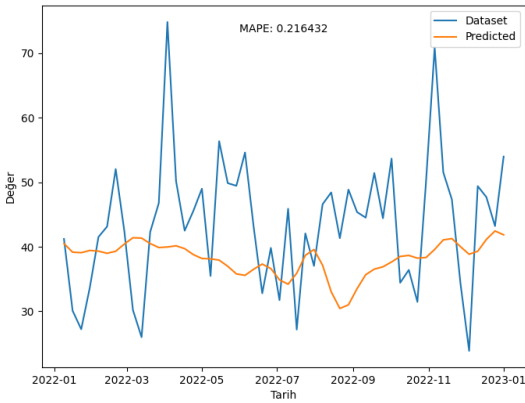
Şekil 8. Bağcılar istasyonu model değerlendirme (Bağcılar station model evaluation)



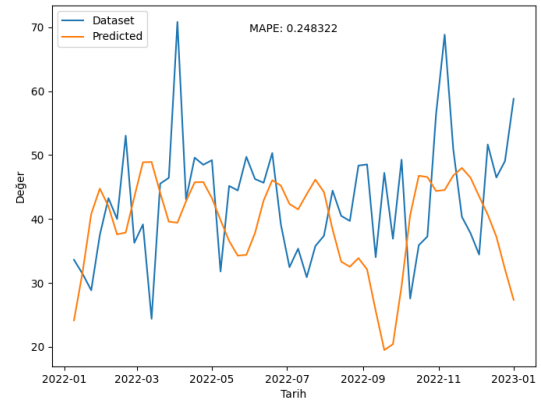
Şekil 6. Beşiktaş istasyonu model değerlendirme (Beşiktaş station model evaluation)



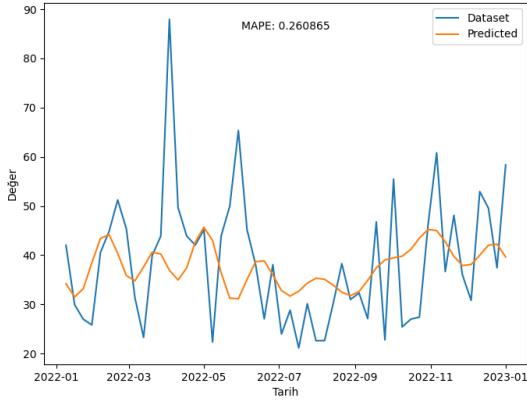
Şekil 9. Göztepe istasyonu model değerlendirme (Göztepe station model evaluation)



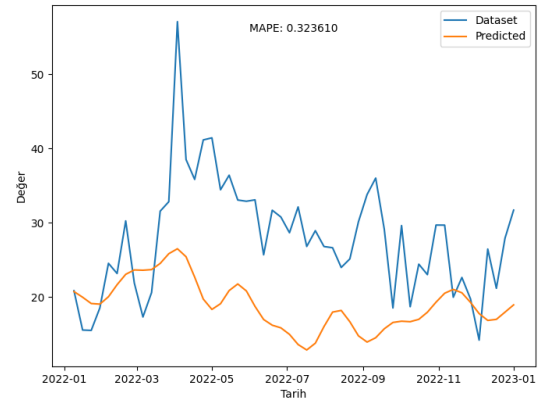
Şekil 7. Kartal istasyonu model değerlendirme (Kartal station model evaluation)



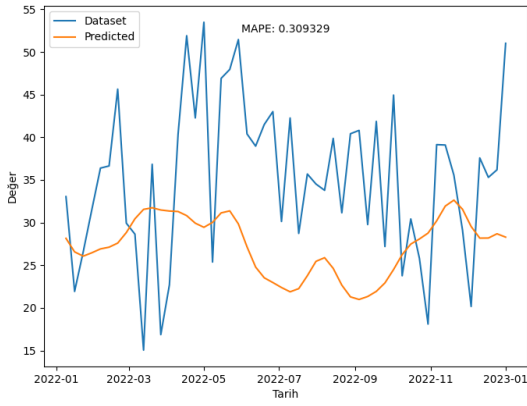
Şekil 10. Sancaktepe istasyonu model değerlendirme (Sancaktepe station model evaluation)



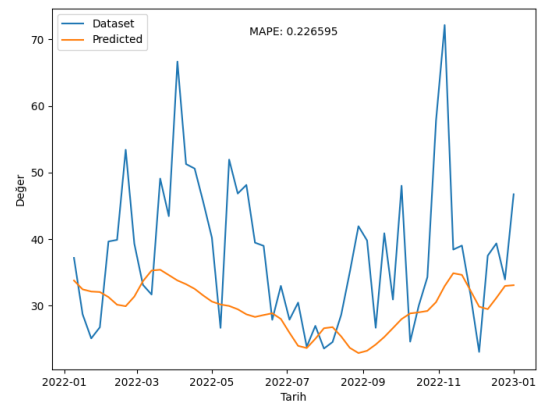
Şekil 11. Tuzla istasyonu model değerlendirme (Tuzla station model evaluation)



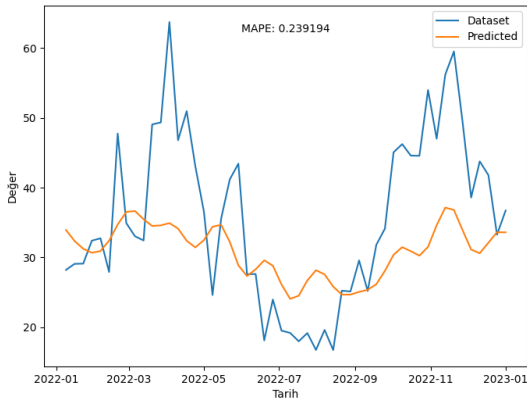
Şekil 14. Sarıyer istasyonu model değerlendirme (Sarıyer station model evaluation)



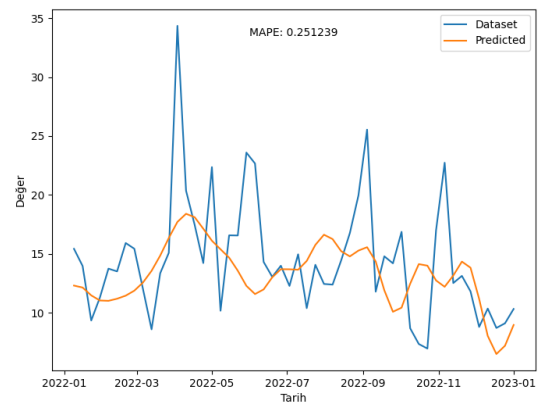
Şekil 12. Ümraniye istasyonu model değerlendirme (Ümraniye station model evaluation)



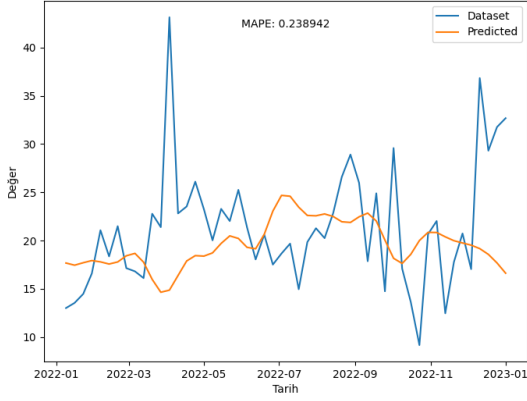
Şekil 15. Esenler istasyonu model değerlendirme (Esenler station model evaluation)



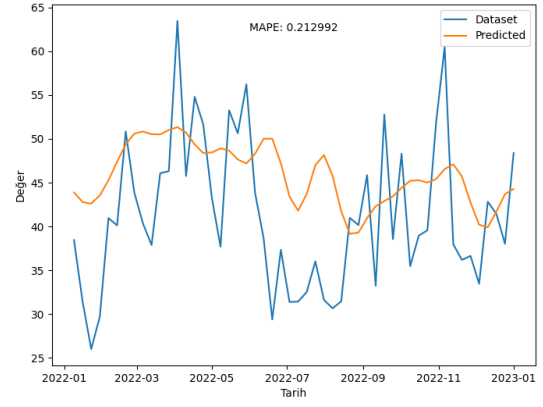
Şekil 13. Yenibosna istasyonu model değerlendirme (Yenibosna station model evaluation)



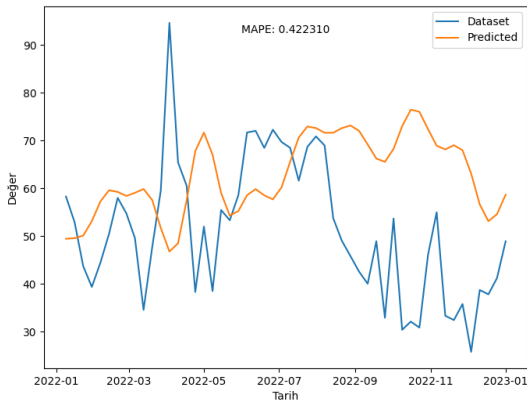
Şekil 16. Büyükada istasyonu model değerlendirme (Büyükada station model evaluation)



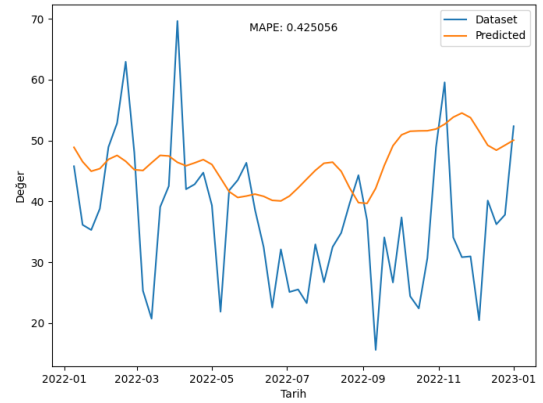
Şekil 17. Kumköy istasyonu model değerlendirme (Kumköy station model evaluation)



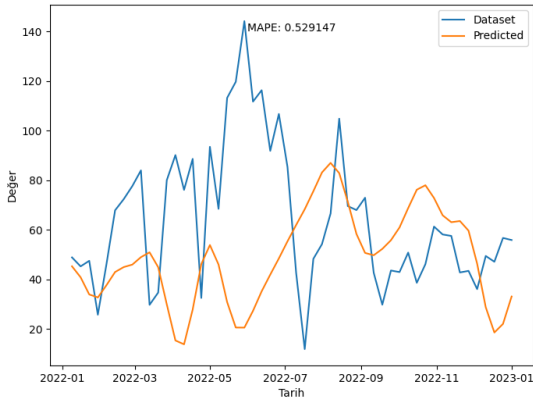
Şekil 20. Aksaray istasyonu model değerlendirme (Aksaray station model evaluation)



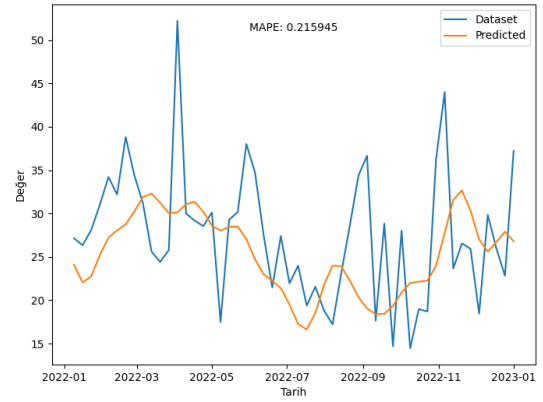
Şekil 18. Sultangazi2 istasyonu model değerlendirme (Sultangazi2 station model evaluation)



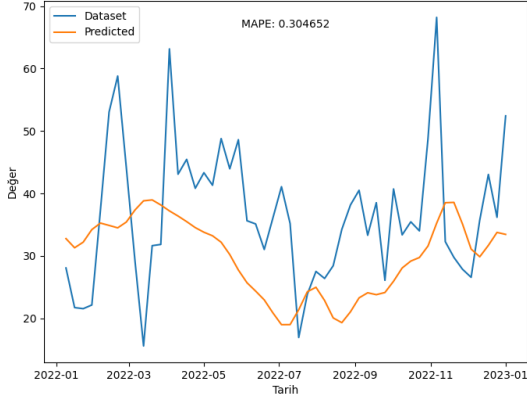
Şekil 21. Alibeyköy istasyonu model değerlendirme (Alibeyköy station model evaluation)



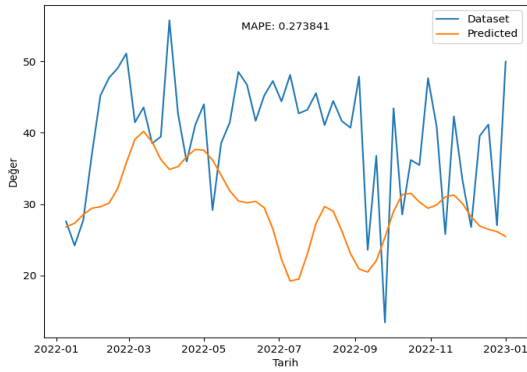
Şekil 19. Sultangazi3 istasyonu model değerlendirme (Sultangazi3 station model evaluation)



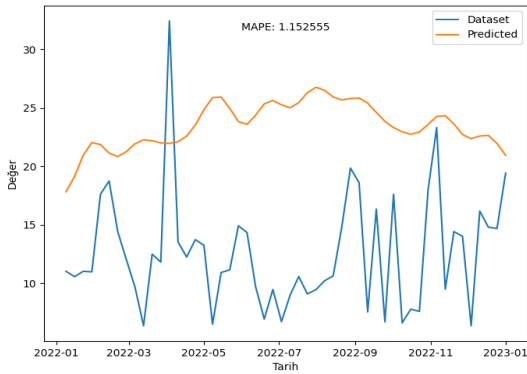
Şekil 22. Üsküdar1 istasyonu model değerlendirme (Üsküdar1 station model evaluation)



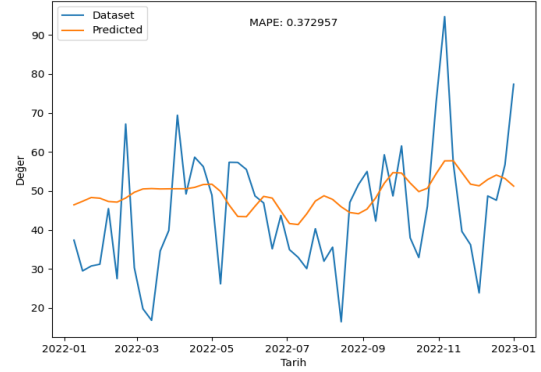
Şekil 23. Kadıköy istasyonu model değerlendirme (Kadıköy station model evaluation)



Şekil 24. Selimiye istasyonu model değerlendirme (Selimiye station model evaluation)

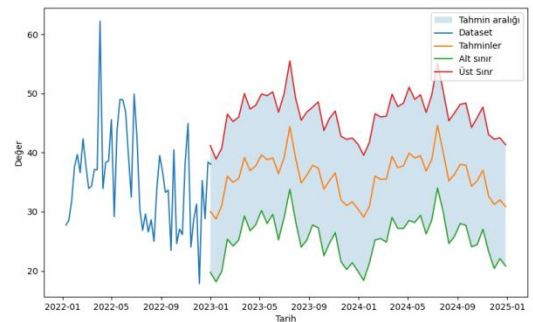


Şekil 25. Kandilli istasyonu model değerlendirme (Kandilli station model evaluation)

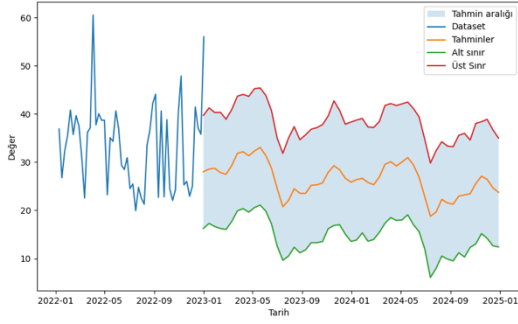


Şekil 26. Kağıthane1 istasyonu model değerlendirme (Kağıthane1 station model evaluation)

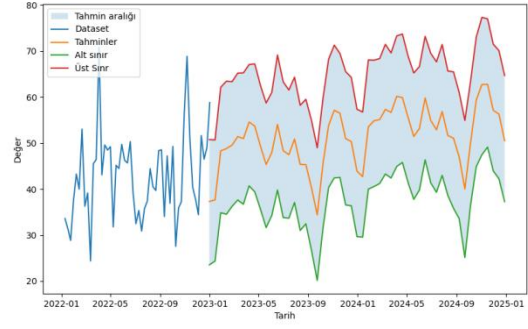
Aynı zamanda her model için MAPE değeri hesaplanmış, grafiklerin üzerinde eklenmiştir. Bazı istasyonlarda yaklaşık %20 hata oranıyla tahmin yapılırken, %100'ü aşan hata oranları da görülebilmektedir. Örneğin, sadece Şekil 25'te MAPE değeri 1.152555 olarak gelmiştir. Bunun anlamı, tahmin edilen değerlerin gerçek değerlere göre ortalama yüzde olarak %115.2555 hata ile sapma göstermesidir. Yani, tahminler gerçek değerlerden ortalama olarak yaklaşık %15 daha fazla çıkmıştır. Bu, tahmin modelinin Kandilli1 veri seti üzerinde iyi performans sergilemediğini göstermektedir. Bu fark birçok etmene bağlı olarak ortaya çıkabilir. Çalışmanın bir amacı da hem veri analizi hem modelleme aşamasında ortaya çıkan bu farka dikkat çekmektir. Nispeten düşük hata oranıyla tahmin yapan model için tamamiyle başarılı demek ya da aksi durumda başarısız demek doğru olmayacaktır. Birçok semtte veriler 5 yıldan az olurken 2010 yılından itibaren aralıksız kayıt alan istasyonlar da vardır ve bilindiği gibi veri miktarı makine öğrenmesi projelerinde başarıyı etkileyen önemli bir unsurdur. Ek olarak tahminlenen 2022 yılında ilgili istasyonda yaşanabilecek özel durumları modelin tahmin etmesi oldukça güçtür zira yalnızca bağımlı değişken ve zaman değişkeni ile eğitim ve test aşamaları tamamlanmıştır. Hem model başarısını daha doğru ölçmenin hem de çeşitli optimizasyonlarla başarıyı artırmanın yolları oldukça fazladır. Çalışmanın devamında eğitim setine 2022 yılı dahil edilmiş ve her istasyon için yeniden model kurulmuş, 2023-2025 yılı için tahminlerde bulunulmuştur. Aşağıda bu modeller görülebilir:



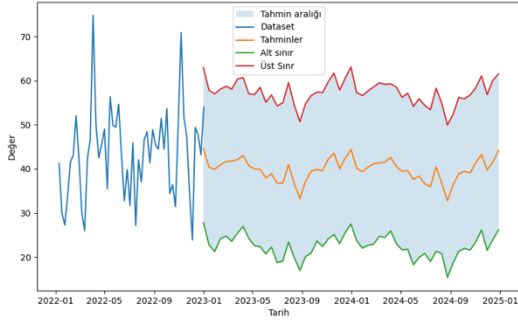
Şekil 27. Arnavutköy istasyonu 2023-2025 tahmini (Arnavutköy station 2023-2025 forecast)



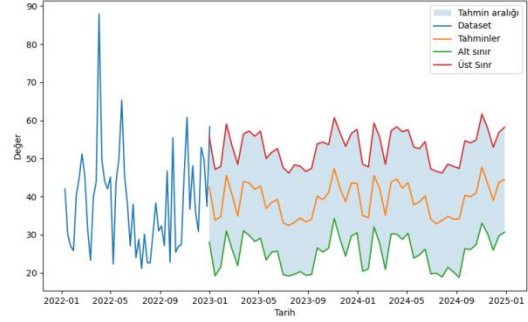
Şekil 28. Beşiktaş istasyonu 2023-2025 tahmini (Beşiktaş station 2023-2025 forecast)



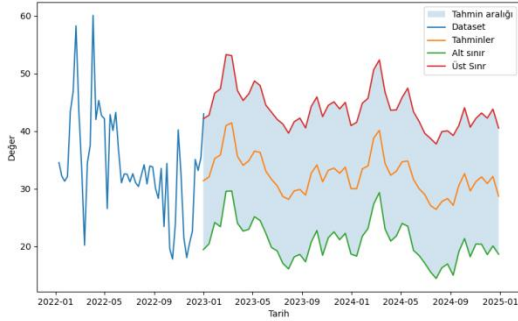
Şekil 32. Sancaktepe istasyonu 2023-2025 tahmini (Sancaktepe station 2023-2025 forecast)



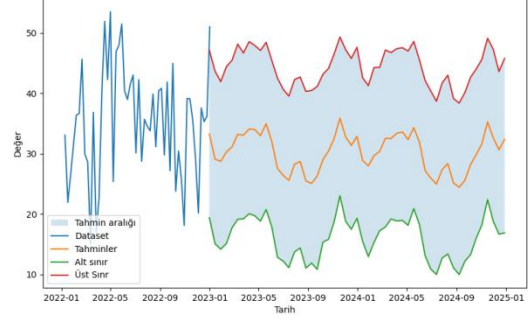
Şekil 29. Kartal istasyonu 2023-2025 tahmini (Kartal station 2023-2025 forecast)



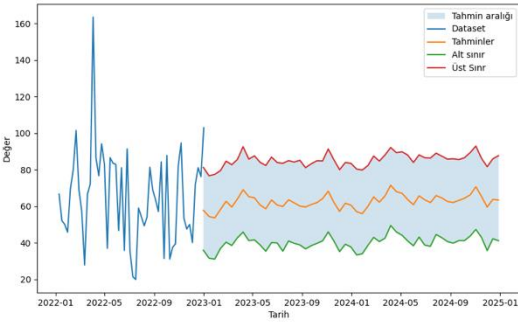
Şekil 33. Tuzla istasyonu 2023-2025 tahmini (Tuzla station 2023-2025 forecast)



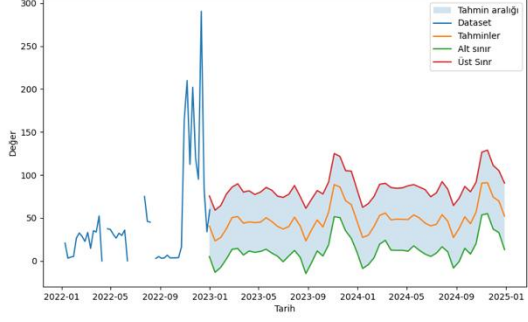
Şekil 30. Bağcılar istasyonu 2023-2025 tahmini (Bağcılar station 2023-2025 forecast)



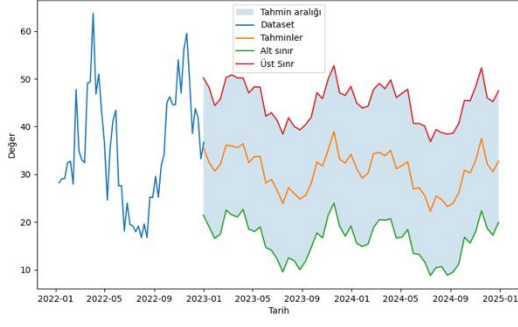
Şekil 34. Ümraniye1 istasyonu 2023-2025 tahmini (Ümraniye1 station 2023-2025 forecast)



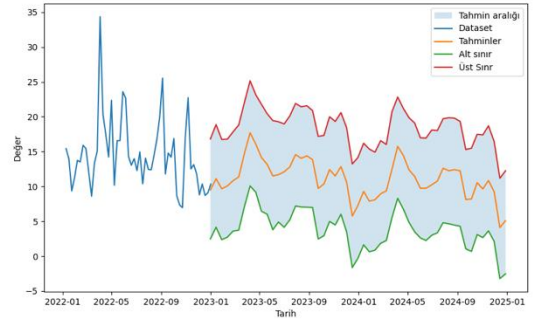
Şekil 31. Göztepe istasyonu 2023-2025 tahmini (Göztepe station 2023-2025 forecast)



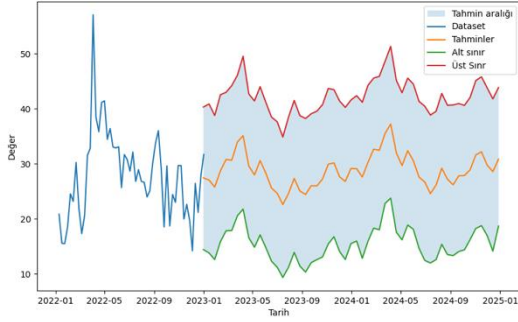
Şekil 35. Mobil istasyonu 2023-2025 tahmini (Mobil station 2023-2025 forecast)



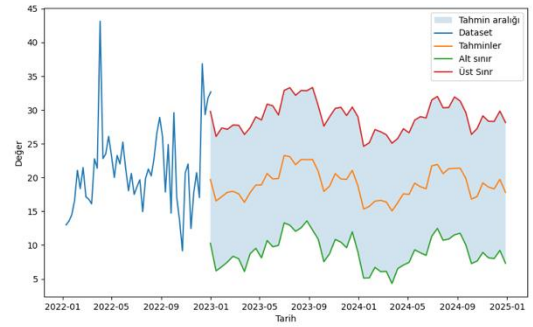
Şekil 36. Yenibosna istasyonu 2023-2025 tahmini (Yenibosna station 2023-2025 forecast)



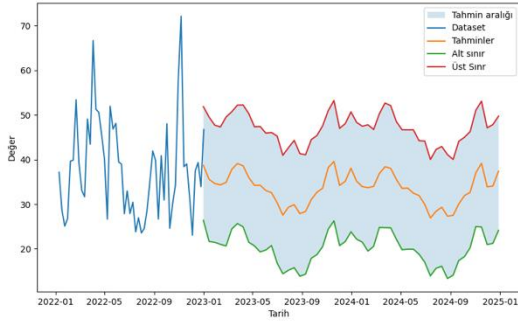
Şekil 40. Büyükada istasyonu 2023-2025 tahmini (Büyükada station 2023-2025 forecast)



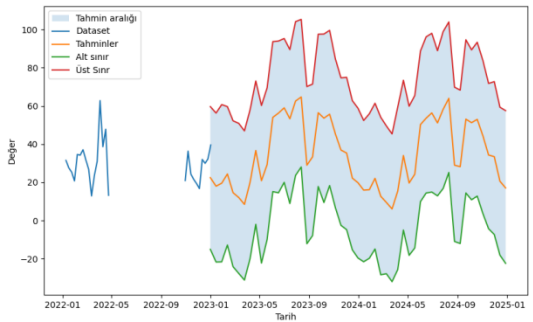
Şekil 37. Sarıyer istasyonu 2023-2025 tahmini (Sarıyer station 2023-2025 forecast)



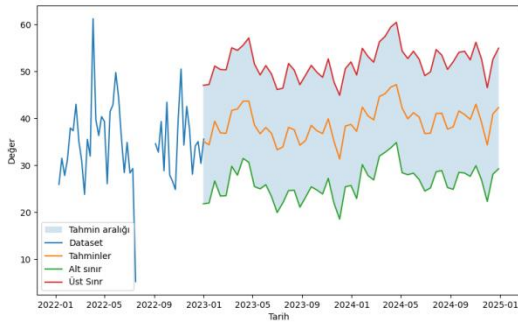
Şekil 41. Kumköy istasyonu 2023-2025 tahmini (Kumköy station 2023-2025 forecast)



Şekil 38. Esenler istasyonu 2023-2025 tahmini (Esenler station 2023-2025 forecast)



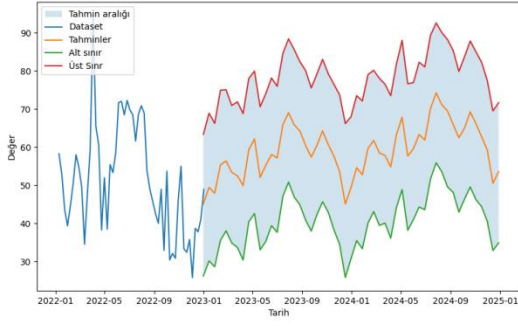
Şekil 42. Sultangazi istasyonu 2023-2025 tahmini (Sultangazi station 2023-2025 forecast)



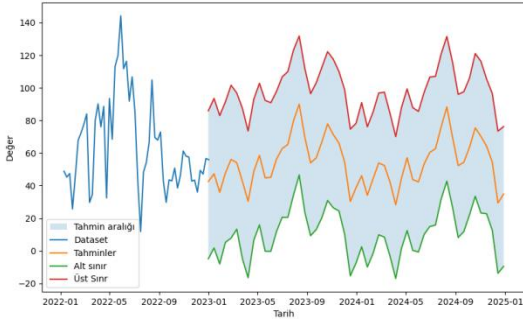
Şekil 39. Çatladıkapı istasyonu 2023-2025 tahmini (Çatladıkapı station 2023-2025 forecast)



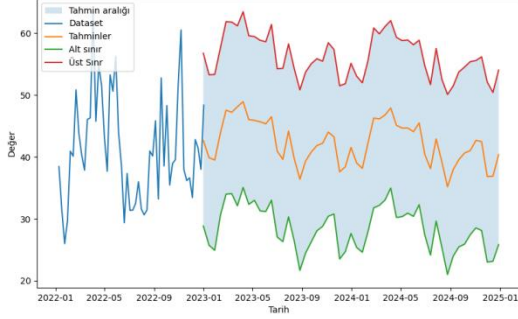
Şekil 43. Maslak istasyonu 2023-2025 tahmini (Maslak station 2023-2025 forecast)



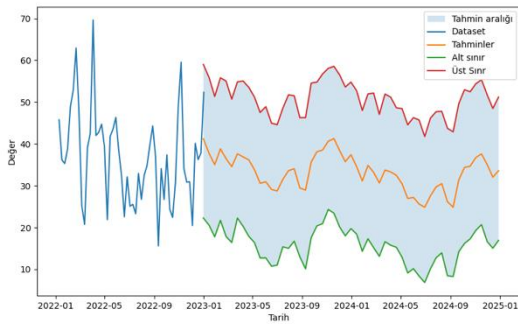
Şekil 44. Sultangazi2 istasyonu 2023-2025 tahmini (Sultangazi2 station 2023-2025 forecast)



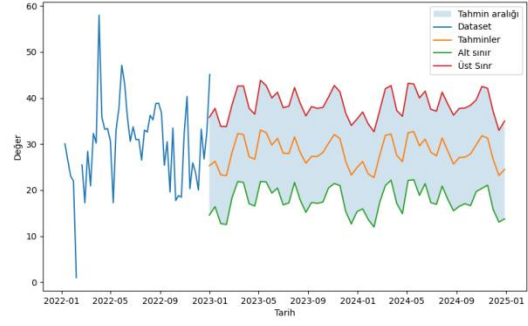
Şekil 45. Sultangazi3 istasyonu 2023-2025 tahmini (Sultangazi3 station 2023-2025 forecast)



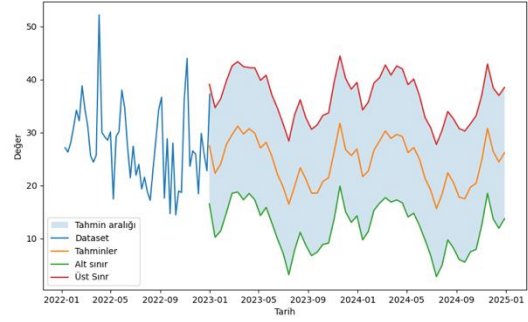
Şekil 46. Aksaray istasyonu 2023-2025 tahmini (Aksaray station 2023-2025 forecast)



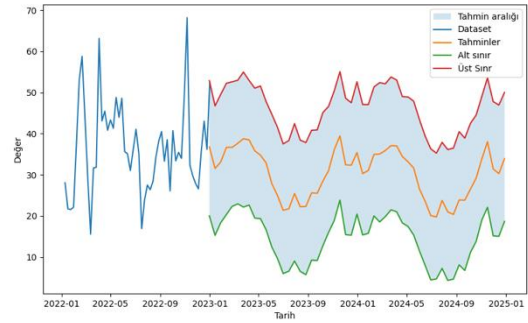
Şekil 47. Alibeyköy istasyonu 2023-2025 tahmini (Alibeyköy station 2023-2025 forecast)



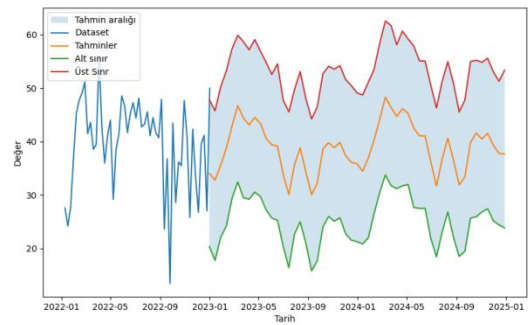
Şekil 48. Avcılar istasyonu 2023-2025 tahmini (Avcılar station 2023-2025 forecast)



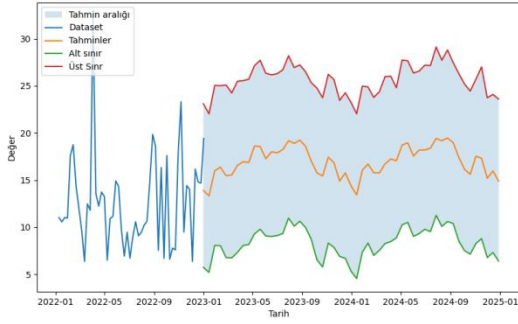
Şekil 49. Üsküdar1 istasyonu 2023-2025 tahmini (Üsküdar1 station 2023-2025 forecast)



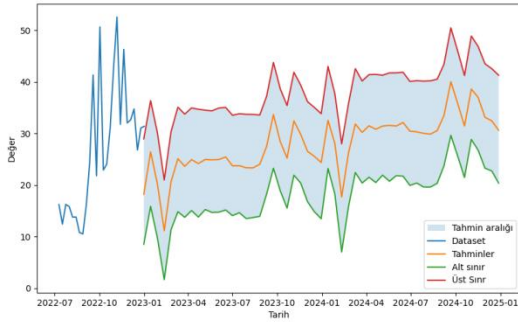
Şekil 50. Kadıköy istasyonu 2023-2025 tahmini (Kadıköy station 2023-2025 forecast)



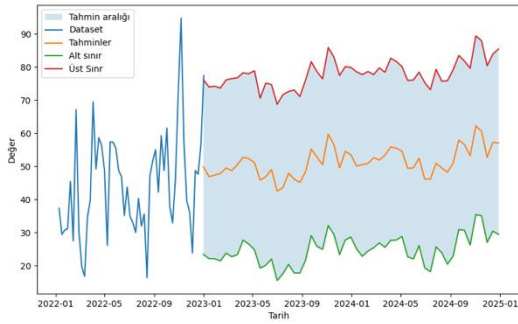
Şekil 51. Selimiye istasyonu 2023-2025 tahmini (Selimiye station 2023-2025 forecast)



Şekil 52. Kandilli istasyonu 2023-2025 tahmini (Kandilli station 2023-2025 forecast)



Şekil 53. Beylikdüzü istasyonu 2023-2025 tahmini (Beylikdüzü station 2023-2025 forecast)



Şekil 54. Kağıthane1 istasyonu 2023-2025 tahmini (Kağıthane1 station 2023-2025 forecast)

BÖLÜM 3: SONUÇ VE ÖNERİLER (CONCLUSION AND RECOMMENDATIONS)

Bu bölüm, çalışmanın özgün sonuçlarının paylaşıldığı, tartışıldığı bölümdür.

3.1. Model Sonuçlarının Yorumu (Interpretation of Model Results)

Çalışmada her ilçe için yapıla her modele eklenen MAPE değerlerinin ortalaması 0.316495 olarak bulunmuştur ki bu da alanda yapılan başka çalışmaların MAPE değerleri ile karşılaştırıldığında kabul edilebilir bir modelleme olduğunu göstermektedir [53]. Ancak literatürde uzun

sürekli verilerle çalışılmış ve uzun süreli tahmin yapmış kapsamlı bir örnek bulunmadığı ve tahminlemede daha çok diğer değişkenlerin ayrı ayrı değişimine odaklanıldığı için tam bir karşılaştırma mümkün olamamıştır.

Saatlik, günlük, haftalık zaman dilimlerinde kurulan modellerin hata oranı oldukça yüksek olabilmektedir. Çalışma kapsamında veriler bir yıllık süreyle aylık olarak tahmin edilmiştir. Böylelikle kısa zaman dilimlerinde hava kirliliğini etkileyebilecek fazla sayıda unsur ile model performansında oluşabilecek olası olumsuz etkilerden korunmuştur. Her ilçe için yapılmış olan modelleme tek tek yukarıda şekillerle gösterilmiştir. Çalışma gerektiği taktirde daha fazla veri eklenerek daha iyi tahminler elde edilebilir. Böylece karar vericilerin alacağı önlemler ile gelecek zaman dilimlerinde model tahmini ile ölçülen değerler kıyaslanabilecektir.

Sultangazi, Mobil, Çatladıkapı istasyonlarında eksik verilerin varlığı göze çarpmaktadır. Prophet modeli az miktarda eksik veriler barından veri setlerinde de başarılı tahmin yapabilecek modellerden biridir. Yine de bu çevre problem ele alınırken geçmiş verilerin önemi göz ardı edilmemeli, ölçüm istasyonlarının sayısı ve kalitesi sürekli artırılmalıdır.

Mobil istasyonu sürekli veriye sahip olmadığı ve sabit konumda olmadığı için bölgesel fikirler verememektedir. Ancak, nedenselliği sorgulamak, olası özel durumlarda hava kalitesini ölçebilmek için bunun gibi istasyonlar şarttır ve sayısı artırılmalıdır. Ölçüm yaptığı lokasyonların tam konumu da ilgili verilerde paylaşılmalıdır.

Tahmin sonuçlarına göre AQI değeri artış trendine giren bir istasyon bulunmamaktadır. Bu durum en azından kirliliğin agresif bir şekilde artmamasını öngörse de, bazı bölgeler (Sultangazi, Sarıyer, Üsküdar) için yılın belirli zamanlarında yükseliş beklentisi göze çarpmaktadır. Bu artışın sebepleri o zaman diliminde detaylıca incelenmelidir. Özellikle bu zamanlarda yerel yönetimlerin yapacağı farkındalık çalışmaları ve alınacak önlemlerle bu artışın önüne geçilmeye çalışılmalıdır.

Modelin çıktı olarak ürettiği; alt sınır, üst sınır değerleri gözetilmeli, özellikle üst sınırın üzerine çıkılan zamanlarda bir sonraki bölümde verilen bölgesel öneriler dikkate alınabilir.

Bu çalışmada verilerin ve değerlendirmenin il geneli için kapsayıcı olması nedeniyle başka modellerin denenip karşılaştırılması ve model doğruluğu konusunda başka metriklere yer verilmesi mümkün olmamıştır.

3.2. Öneriler (Suggestions)

Birçok çalışma büyük şehirlerde hava kirliliğine karşı alınabilecek önlemlerden bahsetmiştir [54][55][56][57]. Buna karşın bu çalışmadan çıkan sonuçlara göre başka iller için benzer çalışmaların aynı ya da başka kategorileri dikkate alarak, aynı ya da başka modeller de kullanılarak

yapılması bu şehirlerde yaşayan insanların hava kirliliğinin etkilerini daha iyi anlamasına yardımcı olacaktır. Çalışmaya rüzgar, nem, yağış gibi meteoroloji ve trafik yoğunluğu gibi veriler de eklenerek tahmin doğruluğu desteklenebilir ve eş zamanlı tahminler yapılabilir ve paydaşlarla eş zamanlı paylaşımına açılabilir.

KAYNAKLAR (REFERENCES)

- [1] Internet: J. Mackenzie, Air Pollution: Everything You Need to Know, NRDC, <https://www.nrdc.org/stories/air-pollution-everything-you-need-know#whatis>, 19.07.2023.
- [2] G. Kara, B. Çiğdem, Y. Çay, "Konya'da Hava Kirliliği Konusunda Farkındalık Düzeyinin Belirlenmesi", *Ulusal Çevre Bilimleri Araştırma Dergisi*, 2(2), pp. 91-94, 2019.
- [3] W. Nazar ve M. Niedozytko, "Air Pollution in Poland: A 2022 Narrative Review with Focus on Respiratory Diseases", *International Journal of Environmental Research and Public Health*, 19(2), p. 895, 2022.
- [4] H. Zamri, Z. A. Shaffiei, N. A. Daud, D. N. Ahmad, "AirAwareMalaysia: Data Visualization and Air Quality Awareness on Air Pollution in Selangor Using Big Data Analytics", *International Conference on Soft Computing and Data Mining*, 457, pp. 223-233, 2022.
- [5] Internet: D. Kim, Media Portrayal and Public Awareness of Air Pollution in Los Angeles and Seoul, Digital Commons Connecticut College, 2022, <https://digitalcommons.conncoll.edu/sip/39/>, 19.07.2023.
- [6] S. Illari, R. T. Lado, L. Marrodan, "Traffic and Pollution Modelling for Air Quality Awareness: An Experience in the City of Zaragoza", *SN Computer Science*, 3(3), p. 281, 2022.
- [7] G. Kanchelashvili, L. Gulbani, A. Dekanosidze, L. Kvachantiradze, G. Kamkamidze, L. Sturua, "Knowledge Of Georgian Population Towards Air Pollution And Health Effects Of Lead Contamination", *Georgian Med News*, 322, pp. 58-62, 2022.
- [8] A. Piracha, M. T. Chaudhary, "Urban Air Pollution, Urban Heat Island and Human Health: A Review of the Literature", *Sustainability*, 14(15), p. 9234, 2022.
- [9] M. Banach, R. Dlugosz, T. Talaska, W. Pedrycz, "Air Pollution Monitoring System with Prediction Abilities Based on Smart Autonomous Sensors Equipped with ANNs with Novel Training Scheme", *Remote Sensing*, 14(2), p. 413, 2022.
- [10] C. Giorio, "Energy efficiency vs indoor air quality conundrum and possible solutions", *ECG Bulletin*, p. 20, 2022.
- [11] E. Di Pirro, L. Sollustio, J. A. C. Castellar, G. Sgrigna, M. Marchetti, B. Lasserre, "Facing Multiple Environmental Challenges through Maximizing the Co-Benefits of Nature-Based Solutions at a National Scale in Italy", *Forests*, 13(4), p. 548, 2022.
- [12] A. Bhatnagar, "Cardiovascular Effects of Particulate Air Pollution", *Annual Reviews*, 73, pp. 393-406, 2022.
- [13] T. Jin, Q. Di, W. J. Requia, M. D. Yazdi, E. Castro, T. Ma, Y. Wang, "Associations between long-term air pollution exposure and the incidence of cardiovascular diseases among American older adults", *Environment international*, 170, p. 107594, 2022.
- [14] R. D. Brook, J. R. Brook ve S. Rajagopalan, "Air pollution: The "heart" of the problem", *Current Hypertension Reports*, 5, pp. 32-39, 2003.
- [15] X. Xu, H. Yang, C. Li, "Theoretical Model and Actual Characteristics of Air Pollution Affecting Health Cost: A Review", *International Journal of Environmental Research and Public Health*, 19(6), p. 3532, 2022.
- [16] R. Fuller, P. J. Landrigan, K. Balakrishnan, G. Bathan, M. Brauer, J. Caravanos ve T. Chiles, "Pollution and health: a progress update", *The Lancet Planetary Health*, 6(6), pp. 535-547, 2022.
- [17] B. Helene, S. Gabet, G. Kerneis, A. Provent, C. Rieux, N. B. Salem, H. Dupont, et al., "Designing local air pollution policies focusing on mobility and heating to avoid a targeted number of pollution-related deaths: Forward and backward approaches combining air pollution modeling, health impact assessment and cost-benefit analysis", *Environment International*, 159, p. 107030, 2022.
- [18] Ş. Güzel, P. Özer, "Türkiye'de Hava Kirliliği ve Sağlık Harcamaları", *Sağlık ve Sosyal Refah Araştırmaları Dergisi*, 4(2), pp. 186-202, 2022.
- [19] S. Ünalı, N. Yalçın, "Hava Kirliliğinin Makine Öğrenmesi Tabanlı Tahmini: Başakşehir Örneği", *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), pp. 35-44, 2022.
- [20] D. Sanjeev, "Implementation of Machine Learning Algorithms for Analysis and Prediction of Air Quality", *International Journal of Engineering Research & Technology (IJERT)*, 10(3), pp. 533-538, 2021.
- [21] M. Öklü ve P. Canbay, "Makine Öğrenmesi Yöntemleri ile Şehirlerin Hava Kalitesi Tahmini". *International Journal of Advances in Engineering and Pure Sciences*, 35(1), pp.39-53, 2023.
- [22] E. Çetin Yağmur, "Atmosferik Partikül Maddelerin Makine Öğrenmesi İle Tahmini: Beşiktaş, İstanbul Örneği", *KONJES*, 10(4), pp. 807-826, 2022. doi: 10.36306/konjes.1082866.
- [23] Oğuz, K., & Pekin, M. A. (2022). Makine Öğrenme Algoritmaları ile PM10 Konsantrasyon Tahmini. *Journal of Advanced Research in Natural and Applied Sciences*, 8(2), 201-213. <https://doi.org/10.28979/jarnas.981202>
- [24] B. Duygu Çelik ve N. Arıcı, "Covid-19 Salgın Sürecinde Hava Kalitesi Tahmini: Zonguldak Örneği", *GJES*, vol. 7, no. 3, pp. 222-232, 2021.
- [25] Akgün, M., & Barlık, N. (2023). Makine Öğrenmesi Algoritmaları Kullanılarak Hava Kalitesi İndeksinin Tahmini. *Avrupa Bilim Ve Teknoloji Dergisi* (51), 97-107. <https://doi.org/10.31590/ejosat.1241170>
- [26] Y. Gültepe, "Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme", *European Journal of Science and Technology*, 16, 8-15, 2019.
- [27] L.K. Wang, N.C. Pereira and T.S. Hung, **Air pollution control engineering**, 1, New Jersey, Humana Press, 100-105, 2004.
- [28] Y.T. Tsai, Y. R. Zeng and Y. S. Chang, "Air Pollution Forecasting Using RNN with LSTM" 2018 The 16th IEEE International Conference on Pervasive Intelligence and Computing, 1074-107, 2018.

- [29] Y. Rybarczyk and R. Zalakeviciute, "Machine learning approach to forecasting urban pollution" 2016 IEEE Ecuador Technical Chapters Meeting, ETCM, 2016.
- [30] <https://api.ibb.gov.tr/havakalitesi/OpenDataPortalHandler/GetAQIByStationId>
- [31] H. Gong, "Health Effects Of Air Pollution: A Review of Clinical Studies", *Clinics in Chest Medicine*, 13(2), 201-214, 1992. ISSN 0272-5231, [https://doi.org/10.1016/S0272-5231\(21\)00851-0](https://doi.org/10.1016/S0272-5231(21)00851-0).
- [32] R. Chen, H. Kan, B. Chen, W. Huang, Z. Bai, G. Song, G. Pan, "Association of Particulate Air Pollution With Daily Mortality: The China Air Pollution and Health Effects Study", *American Journal of Epidemiology*, 175(11), 1173-1181, 2012. <https://doi.org/10.1093/aje/kwr425>
- [33] Impactsey and V. Ramanathan, "Recent climate and air pollution impacts on Indian agriculture." *Proceedings of the National Academy of Sciences* 111(46), 16319-16324, 2014.
- [34] J. L. Domingo and R. Joaquim, "Effects of air pollutants on the transmission and severity of respiratory viral infections", *Environmental research*, 187 (2020), 109650, 2020.
- [35] J. Q. Koenig, "Air pollution and asthma", *Journal of allergy and clinical immunology*, 104.4, 717-722, 1999.
- [36] K. Donaldson, M. Gilmour, W. MacNee, "Asthma and PM10", *Respiratory Research*, 1(1), pp. 12-15, 2000.
- [37] C. Mallik, "Anthropogenic sources of air pollution", *Air pollut. sources, impacts control*, 6-25, 2019.
- [38] S. A. Horn and P. K. Dasgupta, "The Air Quality Index (AQI) in historical and analytical perspective a tutorial review", *Talanta*, 125260, 2023.
- [39] W. Krewitt, A. Trukenmüller, T.M. Bachmann, et al., "Country-specific damage factors for air pollutants", *Int J LCA*, 6, pp. 199-210 2001.
- [40] E. Bales, N. Nikzad, N. Quick, et al., "Personal pollution monitoring: mobile real-time air quality in daily life", *Pers Ubiquit Comput*, 23, pp. 309-328, 2019.
- [41] C. Fan, Y. Li, J. Guang, Z. Li, A. Elnashar, M. Allam, G. de Leeuw, "The Impact of the Control Measures during the COVID-19 Outbreak on Air Pollution in China", *Remote Sens.*, 12, 1613, 2020.
- [42] Internet: "Adalar Turizm Raporu Ocak 2020", Şehir Planlama Müdürlüğü, <https://sehirplanlama.ibb.istanbul/wp-content/uploads/2020/07/adalar-turizm-raporu-IBB-icin.pdf>
- [43] Internet: Üsküdar "Şehir Kimliği" Çalıştayı Sonuç Raporu, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2014, <https://cevresehiriklimkutuphanesi.csb.gov.tr/SourcePDF/52dcf1a1-9c69-4149-9cdc-067dfe15de1f>
- [44] Internet: Türkiye News, "Air pollution high in areas with traffic congestion in Istanbul", *Hürriyet Daily News*, 2023. <https://www.hurriyetdailynews.com/air-pollution-high-in-areas-with-traffic-congestion-in-istanbul-187565>
- [45] D. Jandacka, D. Durcanska, M. Bujdos, "The contribution of road traffic to particulate matter and metals in air pollution in the vicinity of an urban road", *Transportation Research Part D: Transport and Environment*, 50, pp.397-408, 2017.
- [46] Internet: H. F. Başer, "İstanbul'da hava kirliliğinin en büyük nedeni "trafik"", Anadolu Ajansı, <https://www.aa.com.tr/tr/cevre/istanbulda-hava-kirliliginin-en-buyuk-nedeni-trafik/2425106>, 19.07.2023.
- [47] Internet: Taşıt-kilometre İstatistikleri, 2019, Türkiye İstatistik Kurumu, <https://data.tuik.gov.tr/Bulten/Index?p=Tasit-kilometre-Istatistikleri-2019-37409>, 19.07.2023.
- [48] S. J. Taylor, B. Letham, "Forecasting at scale", *The American Statistician*, 72(1), pp. 37-45, 2018.
- [49] A. H. Setianingrum, N. Anggraini, and M. F. D. Ikram, "Prophet model performance analysis for Jakarta air quality forecasting." In 2022 10th international conference on cyber and it service management (citsm), 1-7, IEEE, 2022.
- [50] F. Cai, "The Prediction of the Air Quality based on the Prophet Algorithm" *Highlights in Science, Engineering and Technology*, 39, 1056-1060, 2023.
- [51] W.Y. Zhang, W.C. Hong, Y. Dong, G. Tsai, J.T. Sung and G.F. Fan, "Application of SVR with chaotic GASA algorithm in cyclic electric load forecasting" *Energy*, 45(1), 850-858, 2012.
- [52] V. Gupta, S. Kapadia and C. Bhadane, "Time Series Analysis and Forecasting of Air Quality in India", In 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1-5, IEEE, 2023.
- [53] S. Ünalı ve N. Yalçın, "Hava kirliliğinin makine öğrenmesi tabanlı tahmini: Başakşehir örneği", *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), 35-44, 2022.
- [54] S. Ertürk Atabey, E. Eroğlu, "Kentsel Negatif Dışsallık Olarak Trafığe Bağlı Hava Kirliliği: Kentsel Kamu Maliyesi Bağlamında Çözüm Önerisi", *Çağdaş Yerel Yönetimler Dergisi*, 30(1), 71-98, 2021.
- [55] R. Akdur, "Hava Kirliliği ve Çözümler", *Halk Sağlığı ile İlgili Güncel Sorunlar ve Yaklaşımlar*, Ankara Tabipler Odası, 151-155, 2009.
- [56] M. E. Akıner, "Akıllı Şehirlerde Hava Kirliliğine Yönelik Mobil Algılama Sistemleri ve Sürdürülebilir Binalarda İç Hava Kalitesine Genel Bakış", *Teoriden Pratiğe Akıllı Şehirler*, 2024.
- [57] K. B. Schnelle, M. E. Ternes, R. F. Dunn, **Air pollution control technology handbook**, ABD: Crc Press, 2016.

Derin Öğrenme İle Beyin Tümör Segmentasyonu *

Araştırma Makalesi/Research Article

 Bilal TAŞDEMİR¹,  Necaattin BARIŞCI^{2*}

¹Fen Bilimleri Enstitüsü, Teknoloji Fakültesi, Bilgisayar Mühendisliği, Gazi Üniversitesi, Ankara, Türkiye

²Fen Bilimleri Enstitüsü, Teknoloji Fakültesi, Bilgisayar Mühendisliği, Gazi Üniversitesi, Ankara, Türkiye

tsdmr.bilal@gmail.com, nbarisci@gazi.edu.tr

(Geliş/Received:27.11.2023; Kabul/Accepted:15.05.2024)

DOI: 10.17671/gazibtd.1396872

Özet— Artan nüfus ile birlikte her geçen gün daha fazla insan beyin tümöründen etkilenmektedir. Diğer hastalıklar ile kıyaslandığında beyin tümörünün ölüm oranı çok daha yüksektir. Ayrıca beyin tümörü hastalığına yakalanan bireyler günlük yaşamlarında önemli zorluklarla karşılaşmaktadır. Beyin tümörünün tanısı doktorlar için zorlu bir süreçtir. Teşhis aşamasında doktorların karşılaştığı zorluklar ve bu sorunların üstesinden gelmek adına bir çözüm önerisi olarak, beyin MR görüntülerinin otomatik segmentasyonunu sağlayacak bir modelin tasarlanması hedeflenmiştir. Bu çalışmada beyin MR görüntülerinin segmentasyonundaki zorluğu aşmak için topluluk öğrenimi yöntemi kullanılmıştır. Topluluk öğrenimi yönteminde derin öğrenme tabanlı dikkat mekanizmalı u-net ve u-net modelleri kullanılmıştır. Bu yöntem ile iki farklı modelden gelen tahmin değerlerinin ortalamasının alınması ve daha kararlı bir modelin geliştirilmesi amaçlanmıştır. Model eğitimi için BRATS veri setinin 2018, 2019 ve 2020 versiyonları kullanılırken, model testleri için 2017 versiyonu tercih edilmiştir. Veri setindeki dengesiz sınıf dağılımı problemine çözüm olarak farklı veri ön işleme adımları kullanılmıştır ve topluluk öğrenimi modeli ile beyin MR görüntülerinin segmentasyon problemi ele alınmıştır. Elde edilen topluluk öğrenimi yöntemi ile BRATS2017 veri seti üzerinde %87,33 ortalama zar skoru, %81,74 nekrotik sınıfı zar skoru, %91,57 ödem sınıfı zar skoru, %76,03 artırılmış tümör sınıfı zar skoru, %99,96 arka plan sınıfı zar skoru ve Tüm Tümör (TT), Çekirdek Tümör (ÇT) ve Artırılmış Tümör (AT) için sırasıyla %83,11, %78,88 ve %76,03 zar skoru elde edilmiştir.

Anahtar Kelimeler— segmentasyon, tümör, derin öğrenme, u-net, görüntü işleme, normalizasyon, topluluk öğrenimi

Brain Tumor Segmantation With Deep Learning *

Abstract— With the increasing population, more and more people are affected by brain tumors every day. Compared to other diseases, the death rate of brain tumors is much higher. In addition, people suffering from brain tumor disease have important difficulties in their daily lives. The diagnosis of brain tumors poses a challenging process for medical professionals. To address the difficulties faced by doctors during the diagnostic phase and propose a solution, the objective is to design a model that enables the automatic segmentation of brain MR images. In this study, Ensemble learning method was used to overcome the difficulty in segmentation of brain MRI images. Deep learning base attention u-net and u-net models were used in ensemble learning method. The aim is to develop a more stable model by averaging the prediction values from two different models. While the 2018, 2019 and 2020 versions of BRATS dataset were used for model training, the 2017 version was preferred for model testing. With different data preprocessing steps and ensemble learning model, the difficulty of segmentation of brain MR images has been overcome on imbalanced dataset. With the ensemble learning method created, 87.33% average zar score, 81.74% necrotic class zar score, 91.57% edema zar score, 76.03% enhancing zar score, 99.96% background class zar score and zar score of 83.11%, 78.88% and 76.03% for Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET) were obtained on BRATS2017 dataset.

Keywords— segmantation, tumor, deep learning, u-net, image processing, normalization, ensemble learning

1. GİRİŞ (INTRODUCTION)

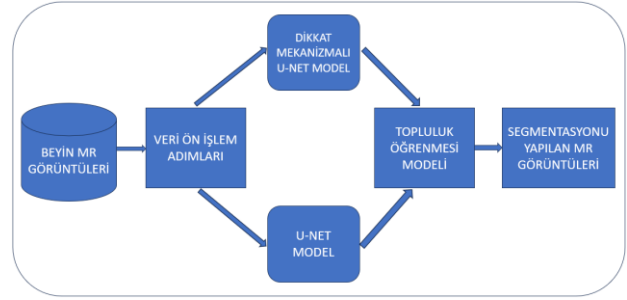
Beyin, sinir sisteminin merkezi olarak hizmet eden bir organdır. Her insanın beyni komplike bir yapıya sahiptir, bilgiler beyin tarafından çağrılır toplanır ve değerlendirilir. Beyin veya omurilikte anormal bir hücre büyümesi sonucunda oluşan kitlelere beyin tümörü denilmektedir. Beyin tümörü günümüzde yaygın olarak görülen bir hastalık türüdür. Doğrudan beyin üzerinde meydana gelen zararlı tümör hücreleri insanların yaşamlarını da olumsuz etkilemektedir. Beyin tümörü ile mücadele eden hastalarda nörolojik semptomlar sebebiyle hareket bozuklukları, denge kaybı, kas güçsüzlüğü, duyu kaybı, konuşma güçlüğü gibi belirtiler ortaya çıkabilmektedir. Beyin tümörünün teşhisi ve tedavisi için doktorlar beyin manyetik rezonans (MR) görüntülerini incelemektedirler. MR görüntüleri değerli tamamlayıcı bilgiler sağlamaktadır ve beyin tümörü segmentasyonunun performansını önemli ölçüde artırmaktadır [1]. Beyin tümörü hastalığında ölüm oranı çok yüksektir, Çin’de beyin tümöründen ölen insan sayısı yıllık 50,000 ile 100,000 arasındadır [2]. Amerikada her yıl 80,000 kişide beyin tümörü teşhis edilmektedir [3]. Beyin tümörü insanlar açısından çok önemli bir hastalık olduğundan üzerinde birçok çalışma yapılmaktadır.

Beyin tümörlerinin tanısı genellikle MR veya bilgisayarlı tomografi (BT) görüntüleri ile konulmaktadır. Beyin tümörü tanı sürecinde uzman bir nöroloji doktoruna ihtiyaç duyulmaktadır. Tedavi yöntemleri tümörün tipine, boyutuna ve konumuna göre değişmektedir.

Beyin tümörünün tanısı oldukça zorlu bir süreçtir, tanı esnasında tümörün boyutu ve konumu manuel olarak dikkatlice tespit edilmelidir. Bazı tümörlerin boyutu küçüktür ve belli bir görüntüleme tekniği ile net bir şekilde görüntülenemeyebilir. Ayrıca bazı tümörlerin iyi veya kötü huylu olması da görüntüleme sonuçlarını etkileyebilmektedir.

Beyin MR görüntülerinin doğru ve güvenilir bir şekilde segmentasyonun yapılması kanser teşhisinde kritik rol oynamaktadır [4]. Beyin tümörü hastalığının teşhis ve tedavi süreçlerindeki olası hatalar insanların ölümü ile sonuçlanabilir.

Beyin MR görüntülerinin yapay zeka desteğiyle otomatik olarak segmente edilmesi hastalığın teşhisi ve tedavisi açısından son derece kritiktir. Bu nedenle sağlık alanında hataları minimuma indirmek amacıyla yapay zeka kullanımı giderek artmaktadır. Bu yaklaşım, uzman doktor işgücüne olan ihtiyacı azaltmanın yanı sıra beyin MR görüntülerinin doğru bir şekilde segmentasyonunu sağlamaktadır. Bu araştırma çalışmasının amacı Şekil 1’de görüldüğü üzere beyin MR görüntüsünü otomatik olarak tarayıp, beyin farklı bölümlerinde hangi tümör sınıfının bulunduğunu otomatik olarak segmente edecek ve Şekil 2’de görüldüğü üzere doktorlar tarafından verilecek beyin MR görüntülerinin segmentasyonunu yapıp doktorlara sonuçları iletecek model elde etmektir.



Şekil 1. Beyin MR görüntülerinin segmentasyon süreci
(Segmentation process of brain MR images)

Bu çalışmanın diğer çalışmalardan ilk farkı, Brain Tumor Segmentation (BRATS) veri setinin 2018, 2019 ve 2020 versiyonları ile derin öğrenme modelinin eğitilip, 2017 versiyonu ile testler yapılmasıdır. Bu sayede modelin hiç görmediği bir versiyon üzerindeki başarısı test edilmektedir. İkinci farkı ise veri ön işleme adımında dengesiz sınıf dağılımı problemini ele almasıdır.



Şekil 2. Segmentasyon modelinin sonuç üretmesi
(Segmentation model outcome producing)

Bu çalışmada 4 adet amacımız bulunmaktadır. Çalışmanın hedefleri şu şekilde özetlenebilir:

- Farklı derin öğrenme yöntemlerini kullanarak beyin MR görüntülerinin segmentasyon problemini çözmek
- Beyin tümörü hastalığının tanı ve teşhis aşamasında uzman doktorlara doğru bilgi sağlamak
- Dengesiz veri dağılımına sahip beyin MR görüntüsü veri seti için özel veri ön işleme adımları kullanarak derin öğrenme yöntemlerinin başarı oranını artırmak
- Beyin MR görüntülerini segmente ederek hastaların tedavi süreçlerine destek olmak

Bu çalışmanın ana katkıları aşağıda özetlenmiştir.

- Beyin MR görüntülerinin segmentasyonu otomatikleştirilmektedir.
- Beyin MR görüntülerinin segmentasyon işlemi otomatikleştirilerek hastalığın tanısı ve teşhisinde harcanan süre kısaltılmaktadır.
- Doktorlar tarafından hataya açık şekilde manuel olarak yapılan segmentasyon işleminin daha doğru ve güvenilir olması sağlanmaktadır.
- Doktorların teşhis ve tanı aşamasında daha az zaman harcayarak hastalığın tedavisine yoğunlaşmasını sağlamaktadır.
- Hastaların tedavi sürecine katkıda bulunarak iyileşme oranını artırmaktadır.

2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Yapay zeka alanında yapılan çalışmalarda derin öğrenme yöntemleri yaygın olarak kullanılmaktadır. Çok büyük miktarda veri üreten kaynakların artması ile birlikte derin öğrenme yöntemleri giderek önem kazanmaktadır. Sağlık alanında da günümüzde çok büyük veriler üretilerek çeşitli derin öğrenme yöntemleri ile bazı hastalıklara çözüm aranmaktadır, beyin tümörü de bu hastalıklardan birisidir. Günümüz dünyasında beyin tümörü birçok insanda görülmektedir, beyin tümöründen etkilenen insan sayısı yaş ve cinsiyet farketmeksizin gün geçtikçe artmaktadır. Beyin MR görüntülerinin doğru bir şekilde segmentasyonunun yapılması beyin tümörünün tanı ve tedavisi için önemlidir [5]. Derin öğrenme yöntemleri kullanılarak beyin tümörü üzerinde çeşitli çalışmalar gerçekleştirilmiştir. Zhang ve ark. [5] modalite çeşitlerinden farklı ve tamamlayıcı özneliklerin çıkarılmasını ve birleştirilmesini içeren derin öğrenme yöntemi önermiştir, çalışma kapsamında BRATS2021 veri seti kullanmıştır. Huang ve ark. [6] literatürdeki beyin MR görüntüleri üzerine yapılan yapay zeka çalışmalarını incelemiştir, inceleme sonucunda bu alanda yapılan çalışmaların son yıllarda arttığını, fakat yapılan çalışmaların klinik ortamlardaki pratiğe adopte olması gerektiği sonucunu elde etmiştir. Wang ve Chung [7] beyin MR görüntülerinin segmentasyonu için hibrit bir Evrişimli Sinir Ağı (ESA) modeli önermiştir, çalışmayı BRATS2019 veri seti ile gerçekleştirmiştir, veri setindeki dengesiz sınıf dağılımı problemini çözmek için ise dinamik odak kayıp fonksiyonu önermiştir. Jiang ve ark. [2] Çift Akış Kod Çözmeli U-net ismini verdikleri ESA mimarisi ile U-net mimarisini kombineledikleri bir model önermiştir, model içerisinde dengesiz sınıf problemini çözmek için ağırlıklı çapraz entropi kayıp fonksiyonu kullanmıştır, çalışmayı BRATS2017 ve BRATS2018 veri setleri ile gerçekleştirmiştir. Goceri [8] Kapsül Ağı mimarisi ile ikili sınıflandırma yöntemi geliştirmiştir. Kullandığı beyin MR görüntülerini normal ve anormal olarak sınıflandırarak %92,65 doğruluk oranı elde etmiştir. Alzahrani [9] beyin MR görüntülerinin sınıflandırılması için Evrişimsel Dikkat Mekanizmalı Karıştırıcı yöntemini önermiştir. Önerilen model %97,94 doğruluk oranı elde etmiştir. Zhang ve ark. [10] iki aşamalı bir model önermiştir. Çapraz Modalite Özellik Geçiş aşamasını bilgiyi farklı türler arasında aktararak zengin özellik temsilleri elde etmek için, Çapraz Modalite Özellik Füzyonu aşamasını ise farklı türden gelen bilgileri birleştirmek için kullanmıştır. Naser ve Deen [11] U-net modeli ile Görsel Geometri Grup 16 modelini kombineleyerek bir model 6 önermiştir. Modeli normal ve anormal olacak şekilde ikili sınıflandırma yöntemi ile oluşturmuştur. Dengesiz sınıf dağılımı problemi için kayıp fonksiyonunda sınıflara ağırlık vermiştir. Diao ve ark. [1] beyin MR görüntülerindeki eksik bilgileri tamamlamak için yeniden yapılandırma yöntemi ve eksik modaliteli MR görüntü segmentasyonu için ortak öğrenmeye dayalı özellik yeniden yapılandırma ve geliştirme yöntemi önermiştir. Önerilen yöntem ile BRATS2018 veri seti üzerinde testler yapılmıştır. Yapılan testlerde Tüm Tümör (TT), Çekirdek Tümör (ÇT), Artırılmış Tümör (AT) sınıf kümeleri için sırasıyla %86,28, %77,02, %59,64 zar skoru

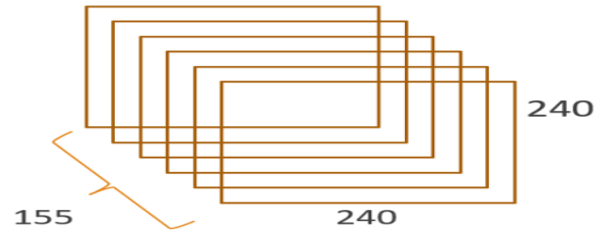
elde etmiştir. Rammurthy ve Mahesh [12] beyin MR görüntülerinden tümörü tespit etmek için Balina Harris Hawks Optimizasyonu yöntemini önermiştir. Balina Optimizasyon Algoritması ve Harris Hawks Optimizasyon algoritmalarının kombinlenmesiyle oluşturulan yöntem Derin Evrişimsel Sinir Ağı modeli ile birlikte ikili sınıflandırmada %81,6 doğruluk oranı elde etmiştir. Hashemzahi ve ark. [13] beyin MR görüntüleriyle ikili sınıflandırıcı hibrit model önermişlerdir. Önerilen hibrit modeli Sinirsel Otoregresif Dağılım Tahmini ve ESA modelini kombineleyerek oluşturmuşlardır. Önerilen model %95 doğruluk oranı elde etmiştir. Nalepa ve ark. [14] geliştirdikleri derin öğrenme modeline ek olarak hata oranını düşüren kübik model geliştirmiştir. Naceur ve ark. [15] beyin MR görüntülerinin segmentasyonu için ESA modeli kullanmıştır, dengesiz sınıf dağılımı problemini çözmek için ağırlıklı çapraz entropi kayıp fonksiyonu, görüntüler arası uzaysal ilişkiyi yakalamak için ise Çakışan Yamalar yöntemini önermiştir. Çalışmada sırasıyla her bir sınıf için %90, %83 ve %83 zar skoru elde edilmiştir. Deepak ve Ameer [16] beyin MR görüntülerinin segmentasyonunda ilk olarak GoogLeNet modeli ile öznelik çıkarımı gerçekleştirmiştir. Sonrasında ise Destek Vektör Makinesi (DVM) sınıflandırıcı algoritması kullanmıştır. Maharjan ve ark. [17] beyin tümör tespiti için ESA modeli geliştirmiştir, kayıp fonksiyonu olarak özelleştirilmiş softmax kayıp fonksiyonu kullanmıştır. Saba ve ark. [18] beyin MR görüntülerinin segmentasyonunda Yakalama Kesme metodunu kullanmıştır, öznelik çıkarımında ise Görsel Geometri Grup 19 hazır modelini kullanmıştır. Veri seti olarak BRATS2015, BRATS2016 ve BRATS2017 veri setlerini kullanmıştır. Mittal ve ark. [19] öznelik çıkarımı için Sabit Dalgacık Dönüşümü kullanmıştır, Büyüyen Evrişimli Sinir Ağı kullanarak segmentasyon modelini oluşturmuştur, önerilen hibrit yöntem %98 kesinlik oranına sahiptir. Zhao ve ark. [4] Tam Evrişimli Sinir Ağı ve Koşullu Rastgele Alanlar yöntemini kullanarak beyin MR görüntülerinin segmentasyonu için bir framework geliştirmiştir. Geliştirilen framework ile 3 farklı beyin MR görüntüsünü kullanarak 3 farklı model eğitmiştir ve oylama tabanlı yöntem ile segmentasyon yapmıştır. Model eğitimini ve testlerini BRATS2013, 7 BRATS2015 ve BRATS2016 veri setleri üzerinde gerçekleştirmiştir. Yang ve ark. [20] Küçük Çekirdekli İki Yollu Evrişimsel Sinir Ağı ve Rastgele Orman algoritması ile hibrit model önermiştir. Küçük Çekirdekli İki Yollu Evrişimsel Sinir Ağı modelinin öznelik çıkarımı yapılarak Rastgele Orman algoritması ile sınıflandırma işlemi yapılmıştır. Sonuçlar BRATS2015 veri seti ile değerlendirilmiştir. Önerilen method sırasıyla TT zarskoru bazında %89, ÇT zar skoru bazında %80 ve AT zar skoru bazında ise %87 başarı oranı elde etmiştir. Mehrotra ve ark. [21] derin öğrenme algoritmalarını kullanarak normal ve anormal olarak ikili sınıflandırma şeklinde bir sınıflandırıcı önermiştir. Çeşitli öğrenme aktarımı modelleri ile testler yapmıştır. Veri seti olarak The Cancer Imaging Archive (TCIA) veri setini kullanarak %99 oranında doğruluk oranı elde etmiştir. Xu ve ark. [22] beyin tümör segmentasyonu için hibrit bir model önermiştir. Köşe Dikkat Mekanizmalı Modül ile dilimler arası bilgiyi yakalamayı amaçlamıştır.

Algısal benzerliği keşfeden ve lokal tutarlılığı muhafaza eden Yüksek Boyutlu Algısal Kayıp kayıp fonksiyonu kullanmıştır, segmentasyon modeli için U-net kullanmıştır. Model üzerindeki testleri BRATS2018, BRATS2019 ve BRATS2020 veri setleri üzerinde gerçekleştirmiştir. Liu ve ark. [23] beyin tümör segmentasyonu için Ölçeğe Uyarlanabilir Süper Özellik Metrik U-net modelini önermiştir. Ayrıca önerdiği Ölçeğe Uyarlanabilir Metrik kayıp fonksiyonu ile modelin daha fazla bilgi öğrenmesini sağlamıştır. Model testlerini BRATS2019 veri seti üzerinde gerçekleştirmiştir. Drai ve ark. [24] beyin MR görüntüleri üzerinden öznelik çıkarımı için Otomatik MR Beyin Çıkarma Algoritması ve otomatik tümör segmentasyonu için Otomatik Tümör Segmentasyonu modelini önermiştir. Model başarısını ölçümlemek için zar skoru kullanmıştır. Bidkar ve ark. [25] beyin tümörünü sınıflandırmak için Salp Su Optimizasyonu Temelli Derin İnanç Ağı modelini önermiştir. Veri seti olarak BRATS2018 kullanmıştır. Önerilen model ile %93,3 doğruluk oranı elde etmiştir. Zhou ve ark. [26] bir veya daha fazla beyin MR görüntüsünü dışarıda tutarak yeni bir U-net tabanlı beyin tümör segmentasyon ağı önermiştir. Model testlerini BRATS2018 veri seti ile gerçekleştirmiştir. Model performansı zar skoru ile ölçülmüştür. Sırasıyla zar skoru bazında her bir sınıf için %82, %74,9 ve %59,1 başarı oranı elde etmiştir. Huang ve ark. [27] ilk olarak arkaplan kaldırma ve görüntü normalizasyonunu iteratif olarak uygulamıştır. Ardından önerilen Kanallar Arası Dikkat Mekanizmalı Kalıntı U-net segmentasyon modelini uygulamıştır. Önemli öznelik gruplarını vurgulamak için ise Grup Kanallar Arası Dikkat Mekanizması modelini önermiştir. Model eğitimi ve testleri için BRATS2017 ve BRATS2018 veri setleri kullanılmıştır. Cinar ve ark. [28] U-net ile DenseNet121 modellerini hibrit şekilde kullanarak öğrenme aktarımı tabanlı hibrit bir sistem önermiştir. Dengesiz sınıf dağılımı 8 problemini çözmek için MR görüntüsündeki tümörlü alanı otomatik tespit edip kesen bir algoritma önermiştir. Model eğitimi ve testleri için BRATS2019 veri seti kullanmıştır. Mazumdar ve Mukherjee [29] mevcutta önerilen ESA modellerinden daha hızlı ve daha az parametreye sahip, U-net tabanlı Verimli Mekansal Dikkat Ağı modelini önermiştir. Dengesiz sınıf dağılımı problemini çözmek için zar, odak ve Hausdorff Uzaklık kayıp fonksiyonlarının bileşimi olan yeni bir kayıp fonksiyonu önermiştir. Veri seti olarak BRATS2021 kullanmıştır.

3. VERİ SETİ (DATASET)

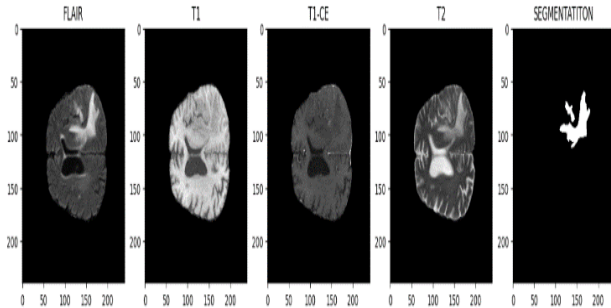
BRATS veri setleri nörolojik radyoloji alanında kullanılan, hastalara ait beyin tümörü MR görüntülerini içeren veri setleridir. BRATS2018 veri seti 285 hastadan, BRATS2019 veri seti 335 hastadan ve son olarak BRATS2020 veri seti ise farklı görüntüleme merkezlerine başvuran 369 farklı hastadan elde edilmiş MR görüntülerini içermektedir. Özellikle yapay zeka ve derin öğrenme gibi çalışmalarda sıklıkla kullanılmaktadır. BRATS veri setleri ilk olarak 2012 yılında düzenlenmeye başlamıştır.

BRATS veri setleri T1, T1-CE, T2 ve FLAIR gibi farklı MR görüntü modalitelerinden elde edilen beyin tümörü görüntülerini içerisinde bulundurmaktadır. Her modalite çeşidi farklı görüntüleme teknolojisi ile elde edildiğinden dolayı beyne ait farklı özellikleri içerisinde barındırır [30]. Ayrıca Şekil 3'te görüldüğü üzere her bir modalite görüntüsü 155 dilimden oluşmaktadır, her bir dilim ise 240 genişlik ve 240 yükseklik boyutlarına sahiptir.



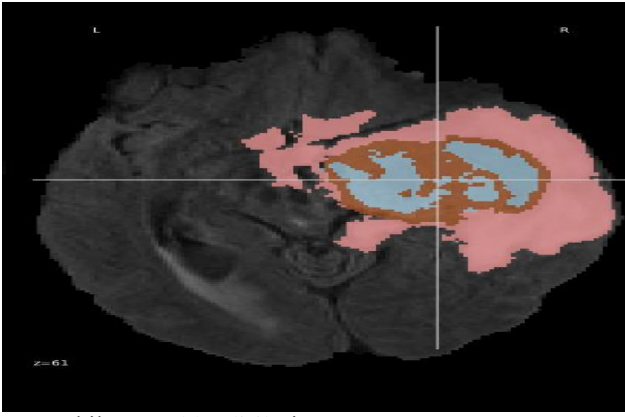
Şekil 3. Beyin MR görüntüsü boyutları
(Shape of brain MR images)

Farklı modalite çeşitleri farklı görüntüleme teknolojileri ile oluşturuldukları için her bir modalite çeşidinde beyin farklı bir bölgesi ön plana çıkmaktadır. Şekil 4'te farklı modalite çeşitlerinin aynı diliminden elde edilmiş görüntüler ve görüntülere karşılık uzman kişilerin yapmış olduğu segmentasyon gösterilmiştir.



Şekil 4. Farklı modalite çeşitlerinin aynı dilimindeki görüntüler
(Images of different modalities in same slice)

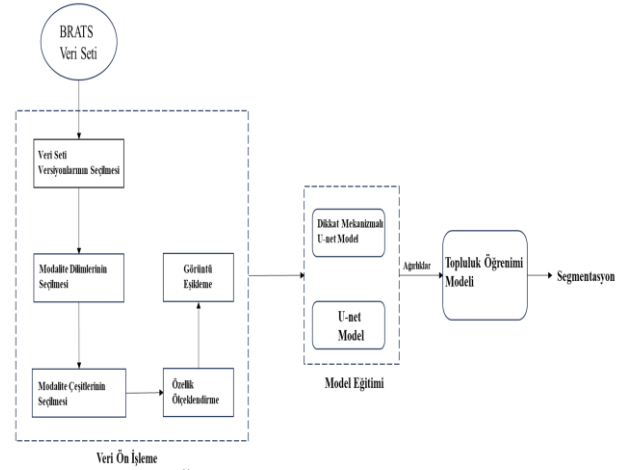
Uzmanlar tarafından etiketlenen beyin segmentasyon görüntüleri içerisinde toplamda 4 farklı sınıf bulundurmaktadır. Bu sınıflar tümör sınıflarını ifade etmektedir. Sınıf 0 arka plan, sınıf 1 nekrotik, sınıf 2 ödem ve sınıf 3 ise artırılmış tümör olarak isimlendirilmektedir. Ek olarak literatürde yapılan çalışmalarda farklı sınıf kümelerine farklı isimlendirmeler yapılarak model sonuçları bu isimlendirmeler ile değerlendirilmektedir. Tüm Tümör (TT) ile isimlendirilen küme içerisinde artırılmış, ödem ve nekrotik sınıfları bulunmaktadır. Çekirdek Tümör (ÇT) ile isimlendirilen küme içerisinde artırılmış ve nekrotik tümör sınıfı bulunmaktadır. Artırılmış Tümör (AT) ile isimlendirilen kümede ise yalnızca artırılmış tümör sınıfı bulunmaktadır [37]. Şekil 5'te uzmanlar tarafından etiketlenen segmentasyon görüntüsünün her bir sınıfının farklı renkte renklendirilmesiyle elde edilen görüntü gösterilmiştir.



Şekil 5. Renklendirilmiş segmentasyon görüntüsü
(Colored segmentation image)

4. YÖNTEM (METHOD)

Önerilen modelin temel amacı, BRATS veri setinin farklı versiyonlarını kullanarak beyin tümör segmentasyonu yapan bir sistem oluşturmaktır. Bu sistem için sırası ile uygulanan işlemler şu şekildedir; BRATS veri seti versiyonlarından hangi kombinasyonun eğitim veri seti olarak seçileceğine karar verilmesi, veri seti içerisinde bulunan modalitelerin hangi dilim aralıklarının kullanılacağına karar verilmesi, hangi modalitelerin model eğitim aşamasında kullanılması gerektiğinin tespit edilmesi, görüntü eşikleme ve özellik ölçeklendirme yöntemlerinin uygulanması ve son olarak farklı derin öğrenme modelleri ile segmentasyon işleminin gerçekleştirilmesidir. Çalışmada önerilen modelin mimarisi Şekil 6'da gösterilmiştir. Çalışma kapsamında hangi veri seti versiyonlarının model eğitim aşamasına dahil olması gerektiği ile ilgili yapılan testler sonucunda BRATS veri setinin 2018, 2019 ve 2020 versiyonları model eğitim aşamasında kullanılırken, 2017 versiyonu modelin testleri için kullanılmıştır. (240 x 240 x 155) boyutlarına sahip olan modaliteler ile yapılan testlerde 155 adet dilim içerisinden 15 – 143 arası dilimler aşağı örnekleme yöntemi ile seçilmiştir. BRATS veri seti içerisinde 4 farklı modalite çeşidi bulunmaktadır, bunlar; Flair, T1, T1-CE, T2. Yapılan testler sonucunda 4 farklı modalite çeşidinden Flair, T1 ve T1-CE modalite çeşitleri seçilmiştir. Minimum-maksimum özellik ölçeklendirme yöntemi ile veri seti üzerinde bulunan farklı örneklerin birbirleri üzerinde üstünlük kurması engellenmiştir. Sıfıra eşikle görüntü eşikleme yöntemi ile de beyin MR görüntüsü üzerindeki tümör içeren bölgenin ayrıştırılarak daha belirgin olması sağlanmıştır. Veri ön işleme adımları tamamlandıktan sonra segmentasyon problemlerinde sıklıkla tercih edilen derin öğrenme tabanlı 2 farklı model oluşturulmuştur, bunlar; dikkat mekanizmalı u-net ve u-net modelleridir. Eğitilen 2 farklı model ile birlikte daha kararlı ve güçlü bir model ortaya çıkarabilmek için topluluk öğrenimi modeli oluşturulmuştur. Oluşturulan topluluk öğrenimi modeli 2 farklı u-net modelinden gelen sınıf tahminlerinin ortalamasını hesaplayarak her bir sınıf için yeni ortalama bir tahmin üretmektedir. Bu şekilde 2 farklı modelin tahmin gücü ile tahminler gerçekleştirilmiştir.

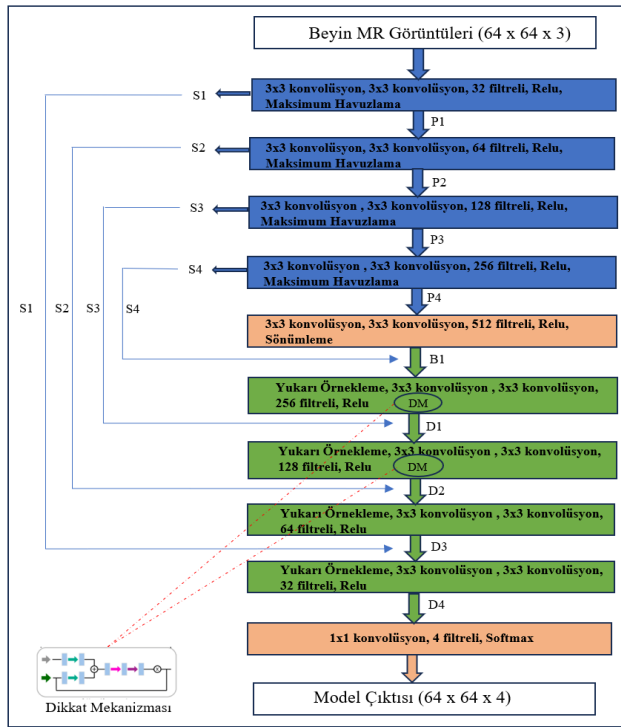


Şekil 6. Önerilen modelin detaylı mimarisi
(Detailed architecture of proposed model)

4.1. Derin Öğrenme Tabanlı Modeller (Deep Learning Based Models)

4.1.1. Dikkat mekanizmalı u-net modeli (Attention u-net model)

Çalışma kapsamında geliştirilen dikkat mekanizmalı u-net modeli Şekil 7'de gösterilmiştir. İlk olarak modele 3 farklı modalite çeşidinden oluşan veri girdi olarak verilmektedir. Daha sonra 4 parçadan oluşan kodlayıcı bölümünün ilk parçası gerçekleşerek iki farklı çıktı üretmektedir. S1 ile isimlendirilen çıktı konvolüsyon işleminden elde edilen çıktı, P1 ise maksimum havuzlama işleminden sonra elde edilen çıktıdır. Kodlayıcı bölümünün ikinci, üçüncü ve dördüncü parçalarında da sırasıyla S2, P2, S3, P3 ve S4, P4 çıktıları elde edilmektedir. S2, S3 ve S4 çıktıları konvolüsyon işlemlerinden elde edilen çıktılar olurken, P2, P3 ve P4 çıktıları maksimum havuzlama işlemlerinden elde edilen çıktılar olmaktadır. Kodlayıcı bölümünün her parçasında maksimum havuzlama işleminden elde edilen çıktılar bir sonraki parçaya iletilmektedir. Kodlayıcı bölümünden sonra orta bölümde konvolüsyon işlemi uygulanır ve burada elde edilen çıktı ile kodlayıcı katmanının dördüncü parçasında konvolüsyon işlemi ile elde edilen çıktı çözücü katmanının ilk parçasına girdi olarak verilir. Daha sonra girdilere sırasıyla yukarı örnekleme, dikkat mekanizması ve konvolüsyon işlemi uygulanır. Dikkat mekanizması çözücü katmanının yalnızca ilk iki bölümünde kullanılmaktadır. Bu işlemler sırasıyla çözücü bölümündeki tüm parçalarda gerçekleştirilir. Sonuç olarak (64 x 64 x 4) boyutlarında çıktı elde edilmektedir. Çalışma kapsamında aktivasyon fonksiyonu olarak ara katmanlarda Relu çıkış katmanında ise softmax tercih edilmiştir. Relu ve softmax tercih edilme sebebi literatürdeki çalışmalarda yoğun olarak kullanılmasıdır.

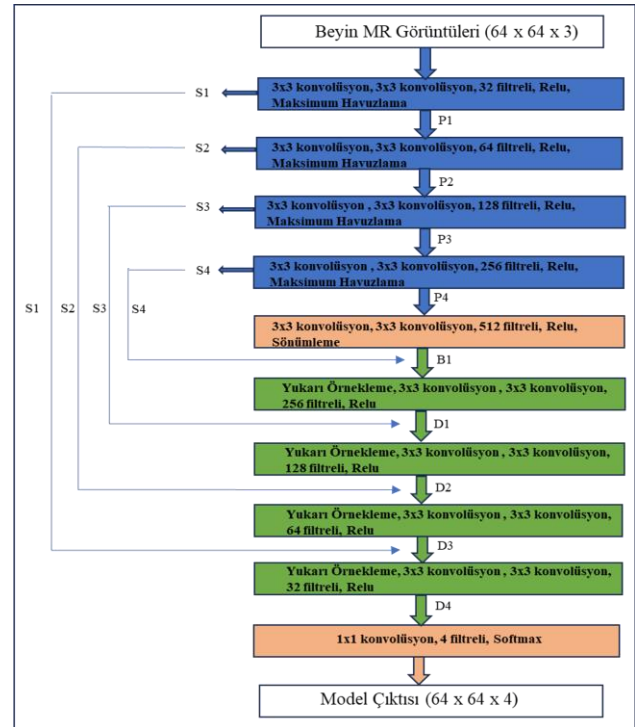


Şekil 7. Geliştirilen dikkat mekanizmalı u-net modeli
(Developed u-net model with attention mechanism)

Model üzerinde eğitim gerçekleştirilirken BRATS veri setinin 2018, 2019 ve 2020 versiyonları kullanılmaktadır. Test aşamasında ise BRATS veri setinin 2017 versiyonu kullanılmaktadır.

4.1.2. U-net modeli (U-net model)

Çalışma kapsamında geliştirilen u-net modeli Şekil 8'de gösterilmiştir. İlk olarak modele 3 farklı modalite çeşidinden oluşan veri girdi olarak verilmektedir. Daha sonra 4 parçadan oluşan kodlayıcı bölümünün her bir parçası gerçekleşerek iki farklı çıktı üretmektedir. S1, S2, S3 ve S4 ile isimlendirilen çıktılar konvolüsyon işleminden elde edilen çıktılar, P1, P2, P3 ve P4 ile isimlendirilen çıktılar ise maksimum havuzlama işleminden sonra elde edilen çıktılar ifade etmek için kullanılmaktadır. Kodlayıcı bölümünün her parçasında maksimum havuzlama işleminden elde edilen çıktılar bir sonraki parçaya iletilmektedir. Kodlayıcı bölümünden sonra orta bölümde konvolüsyon işlemi uygulanır, bu bölümdeki çıktı B1 ile isimlendirilmektedir. Orta bölümde elde edilen çıktı ile kodlayıcı katmanının dördüncü parçasında konvolüsyon işlemi ile elde edilen çıktı çözücü katmanının ilk parçasına girdi olarak verilir. Daha sonra girdilere yukarı örnekleme ve konvolüsyon işlemi uygulanır. Bu işlemler sırasıyla çözücü bölümündeki tüm parçalarda gerçekleştirilir. Sonuç olarak (64 x 64 x 4) boyutlarında çıktı elde edilmektedir. Çalışma kapsamında aktivasyon fonksiyonu olarak ara katmanlarda Relu çıkış katmanında ise softmax tercih edilmiştir. Relu ve softmax tercih edilme sebebi literatürdeki çalışmalarda yoğun olarak kullanılmasıdır.

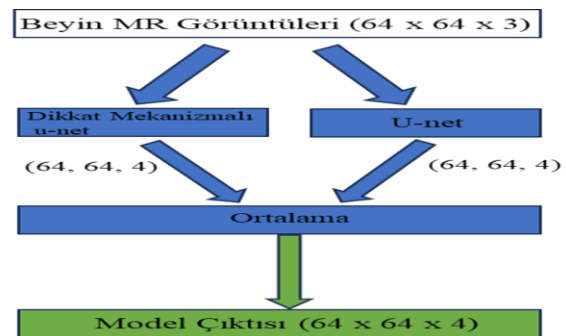


Şekil 8. Geliştirilen u-net modeli
(Developed u-net model)

Model üzerinde eğitim gerçekleştirilirken BRATS veri setinin 2018, 2019 ve 2020 versiyonları kullanılmaktadır. Test aşamasında ise BRATS veri setinin 2017 versiyonu kullanılmaktadır.

4.1.3. Topluluk öğrenimi modeli (Ensemble learning model)

Çalışma kapsamında geliştirilen dikkat mekanizmalı u-net ve u-net modellerinden elde edilen çıktılar ile topluluk öğrenimi modeli oluşturulmaktadır. Şekil 9'da topluluk öğrenimi modeli gösterilmektedir. Dikkat mekanizmalı u-net ve u-net modellerinin eğitim işlemi tamamlandıktan sonra ağırlık dosyaları ile model ağırlıkları yeniden yüklenerek oluşturulan ortalama katmanına çıktılar iletilmektedir. Ortalama katmanında her bir modelin her bir sınıf için yaptığı olasılık tahminlerinin ortalaması alınarak sonuç tahmin çıktısı üretilmektedir. Bu şekilde her iki modelin de sınıf tahminine katkı yaptığı topluluk öğrenimi modelinde başarı oranı modellerin bireysel başarılarından daha iyi olmaktadır.



Şekil 9. Topluluk öğrenimi modeli mimarisini
(Architecture of ensemble learning model)

5. SONUÇLAR VE TARTIŞMA (RESULTS AND DISCUSSION)

Literatürde BRATS veri setleri kullanılarak yapılan birçok çalışma bulunmaktadır. Bu çalışma kapsamında beyin tümör segmentasyonu ile ilgili yapılan güncel çalışmaların incelenmesi sonucunda BRATS veri setlerinin kullanılmasına karar verilmiştir. Derin öğrenme tabanlı mimariler beyin tümör segmentasyonu konusunda en çok kullanılan yöntemlerden biridir, sıklıkla kullanılan mimarilerden birisi de u-net mimarisidir [38]. Beyin tümör segmentasyonu doktorlara tümörlerin yerini ve boyutunu doğru bir şekilde tanımlama olanağı tanıdığı için kritik bir işlemdir [39]. Medikal alanda yapılan segmentasyon çalışmalarının değerlendirilmesi sonucunda u-net tabanlı model mimarilerinin yoğun olarak kullanıldığı görülmüştür. Bu kapsamda u-net tabanlı modeller ile deneysel çalışmalar yapılmaktadır.

5.1. Veri Ön İşlemleri (Data Preprocessing)

Veri ön işleme adımı yapay zeka projelerinin en temel ve en önemli adımıdır. Bu adımda makine öğrenmesi modellerine girdi olarak verilecek olan veri seti üzerinde çeşitli işlemler yapılarak verinin model için hazır hale getirilmesi sağlanır. Aynı zamanda bu adımda yapılacak işlemler makine öğrenmesi modellerinin başarı oranlarını doğrudan etkilemektedir.

5.1.1. Veri seti versiyonlarının seçilmesi (Dataset versions selection)

Çalışma kapsamında BRATS veri setinin 4 farklı versiyonu kullanılmaktadır. Bu versiyonlardan 2017 versiyonu test için ayrılmıştır. 2018, 2019 ve 2020 versiyonlarından hangi kombinasyonun kullanılacağına karar vermek için ise 40 adımlık temel bir u-net modeli ile testler yapılmıştır. 2018, 2019 ve 2020 versiyonlarının olası tüm kombinasyonları ile aynı model 40 adım olacak şekilde eğitilmiştir. Eğitim sonuçları tablo 1' de gösterilmektedir. Yapılan testler sonucunda en başarılı performansı 2018, 2019 ve 2020 versiyonlarının bulunduğu kombinasyon sağladığı için 3 versiyon da eğitim verisi olarak seçilmiştir.

Tablo 1. Farklı veri seti kombinasyonları ile elde edilen sonuç tablosu

(Result table obtained with different data set combinations)

Veri Seti Kombinasyonu	Adım Sayısı	Eğitim Ortalama Zar Skoru	Doğrulama Ortalama Zar Skoru
2018	40	0,5160	0,5022
2019	40	0,4280	0,4177
2020	40	0,5986	0,4336
2018 - 2020	40	0,7763	0,7150
2018 - 2019	40	0,7814	0,7184

2019 - 2020	40	0,7824	0,7382
-------------	----	--------	--------

5.1.2. Modalite dilimlerinin seçilmesi (Selection of modalities slices)

Çalışma kapsamında kullanılan BRATS veri setine ait tüm versiyonlarda her bir modalite çeşidi (240x240x155) boyutlarından oluşmaktadır. 155 adet dilimden oluşan beyin MR görüntülerine karşılık gelen segmentasyon görüntüsü incelendiğinde etiketli piksellerin büyük bir çoğunluğunu arka plan sınıfının oluşturduğu görülmüştür. Tablo 2'de segmentasyon görüntülerindeki belirli dilim aralıklarında sınıfların bulunma oranları paylaşılmıştır.

Tablo 2. Segmentasyon görüntüleri dilim aralıklarının sınıfları bulundurma oranları

(Classification rates of slice intervals in segmentation images)

Dilim	Arka Plan	Artırılmış Tümör	Tümör Çekirdeği	Tüm Tümör
0-15	1,0	0	0	0
15-30	0,99	0	0	0
30-45	0,99	0,0008	0,0021	0,0008
45-60	0,98	0,0037	0,0085	0,0036
60-75	0,97	0,0059	0,0145	0,0054
75-90	0,9705	0,0067	0,0171	0,0055
90-105	0,9758	0,0056	0,0143	0,0041
105-120	0,9876	0,0028	0,0076	0,0018
120-135	0,9968	0,0007	0,0021	0,0003
135-143	0,9998	0	0,0001	0
143-155	1,0	0	0	0

Makine öğrenmesi modellerinde dengesiz sınıf problemi modelin başarısını olumsuz yönde etkilemektedir. Dengesiz sınıf dağılımı probleminin modeller üzerindeki negatif etkisini azaltmak için yukarı örnekleme veya aşağı örnekleme yöntemi kullanılmaktadır. Segmentasyon görüntülerinde arka plan sınıfı diğer sınıflara oranla daha fazla bulunmaktadır.

Tablo 3. Orjinal veri seti ile aşağı örnekleme uygulanan veri setinin sınıf oranlarının karşılaştırılması
(Comparison of class ratios of the original data set and the downsampled data set)

	Arka Plan Sınıfı Bulunma Oranı	Artırılmış Tümör Sınıfı Bulunma Oranı	Tümör Çekirdeği Sınıfı Bulunma Oranı	Tüm Tümör Sınıfı Bulunma Oranı
Orjinal Veri Seti	0,9888	0,0025	0,0064	0,0021
Aşağı Örnekleme Yöntemi Uygulanan Veri Seti	0,9865	0,0031	0,0078	0,0025

Tablo 3'te orjinal veri setinde arka plan sınıfının çoğunluğu oluşturduğu görülmektedir. Bu sebeple dengesiz sınıf dağılımı problemi oluşmaktadır. Tablo 2'de 0-15 arasındaki dilimlerin ve 143-155 arasındaki dilimlerin yalnızca arka plan sınıfını içerdiği görülmektedir. Arka plan sınıfının çoğunlukta olduğu dengesiz sınıf dağılımı problemini çözmek için 0-15 arasındaki tüm dilimler ve 143-155 arasındaki tüm dilimler veri setinden çıkarılarak aşağı örnekleme yöntemi uygulanmıştır. Tablo 4'te aşağı örnekleme yönteminin 10 adımlık temel u-net modeline yaptığı katkı gösterilmiştir.

Tablo 4. Orjinal veri seti ve aşağı örnekleme yöntemi uygulanmış veri setleri ile elde edilen model başarı oranları

(Model success rates obtained with the original data set and data sets with downsampling method applied)

Dilim Aralığı	Adım Sayısı	Ortalama Eğitim Zar Skoru	Ortalama Doğrulama Zar Skoru
0 - 155	10	0.6213	0.6031
15 - 143	10	0.6345	0.6341

5.1.3. Modalite çeşitlerinin seçilmesi (Selection of modality types)

Çalışma kapsamında kullanılan BRATS veri seti versiyonlarında Flair, T1, T1-CE ve T2 olmak üzere 4 farklı modalite çeşidi bulunmaktadır. Eğitim aşamasında hangi modalite kombinasyonunun daha iyi sonuç verdiğini tespit edebilmek için temel bir u-net modeli ile 10 adımlık testler yapılmıştır.

Tablo 5. Farklı modalite kombinasyonlarının 10 adımlık temel u-net modelinde göstermiş oldukları performans değerleri

(Performance values of different modality combinations in the 10 step basic u-net model)

Modalite Kombinasyonu	Kayıp Skoru	Ortalama Eğitim Zar Skoru
Flair	0,0039	0,3346
T1	0,0066	0,2280
T1-CE	0,0057	0,2326
T2	0,0054	0,1211
Flair, T1-CE	0,0034	0,4659
Flair, T1	0,0036	0,3750
Flair, T2	0,0039	0,3463
T1, T1-CE	0,0052	0,3302
T1, T2	0,0051	0,1861
T2, T1-CE	0,0048	0,3430
Flair, T1, T1-CE	0,0033	0,5111
Flair, T1, T2	0,0037	0,3715
Flair, T1-CE, T2	0,0033	0,4828
T1, T1-CE, T2	0,0047	0,3650
Flair, T1, T1-CE, T2	0,0031	0,5093

Tablo 5'te yapılan testler gösterilmektedir. Elde edilen sonuçlarda Flair, T1 ve T1-CE modalite kombinasyonunun hem kayıp skorunda hem de ortalama eğitim zar skorunda diğer kombinasyonlara oranla daha iyi sonuç elde ettiği görülmektedir.

5.1.4. Özellik ölçeklendirme (Feature scaling)

Özellik ölçeklendirme yöntemleri makine öğrenmesi çalışmalarında sıklıkla başvurulan veri ön işlemlerinden birisidir. Çalışma kapsamında farklı özellik ölçeklendirme yöntemleri ile 10 adımlık temel u-net modeli eğitilerek özellik ölçeklendirme yöntemlerinin modellere yaptığı katkılar araştırılmıştır. Tablo 6'da farklı özellik ölçeklendirme yöntemleri ile yapılan testler gösterilmektedir.

Tablo 6. Farklı ölçeklendirme yöntemleri ile yapılan test sonuçları

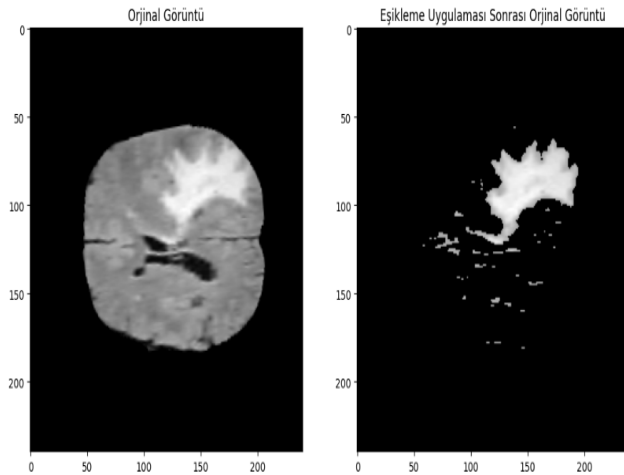
(Test results made with different scaling methods)

	Adım Sayısı	Ortalama Eğitim Zar Skoru	Ortalama Doğrulama Zar Skoru
Ölçeklendirme Kullanılmadan	10	0,3138	0,3120
Minimum Maksimum Ölçeklendirme	10	0,6692	0,6684
Standardizasyon	10	0,6389	0,6336

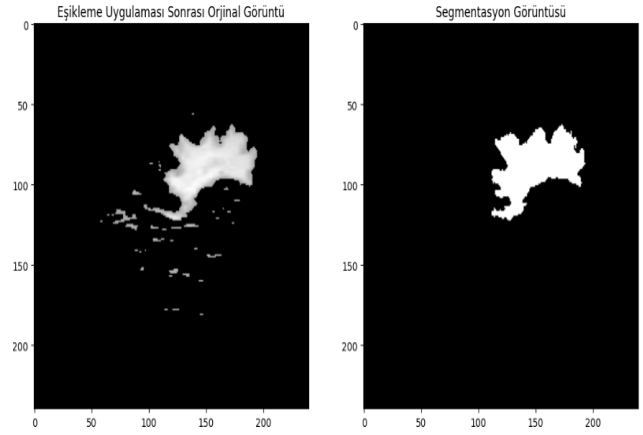
Tablo 6’da elde edilen test sonuçları neticesinde çalışma kapsamında minimum maksimum ölçeklendirme yöntemi kullanılmıştır.

5.1.5. Görüntü eşikleme (Image thresholding)

Çalışma kapsamında modelin başarı oranını artırmak için Sıfıra Eşikle görüntü eşikleme yöntemi uygulanmıştır. Uygulanan bu görüntü eşikleme uygulamasında bir eşik değeri belirlenmektedir. Belirlenen bu eşik değerden büyük olan piksel değerleri sabit kalırken, eşik değerden küçük olan piksel değerleri belirlenen bir değer ile yer değiştirilmektedir. Beyin MR görüntüsü segmentasyon probleminde tümörlü alana daha çok yoğunlaşabilmek için görüntü eşikleme yöntemi ile tümörlü bölgenin ön plana çıkması sağlanmıştır. Aynı zamanda farklı tümör sınıflarına ait özellik değerleri sabit bırakılarak modelin sınıf bazlı özellik değerlerini kaybetmemesi sağlanmıştır. Şekil 10’da görüntü eşikleme uygulanmadan önce ve uygulandıktan sonraki MR görüntüleri gösterilmektedir. Şekil 11’de ise görüntü eşikleme uygulandıktan sonra elde edilen MR görüntüsü ile aynı görüntüye karşılık gelen uzmanlar tarafından etiketlenerek segmentasyonu yapılmış MR görüntüsü arasındaki benzerlik gösterilmektedir.



Şekil 10. Görüntü eşikleme uygulaması sonrasında elde edilen görüntü ile orjinal görüntü arasındaki farklar (Differences between the image obtained after image thresholding application and the original image)

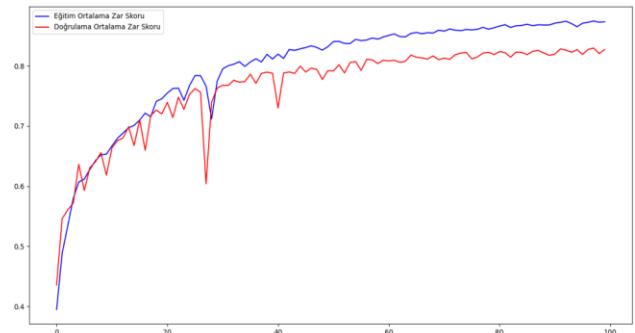


Şekil 11. Görüntü eşikleme uygulaması sonrası elde edilen görüntü ile segmentasyon görüntüsü arasındaki benzerlik

(Similarity between the image obtained after image thresholding application and the segmentation image)

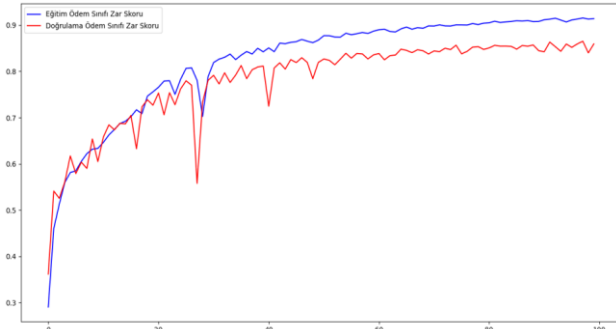
5.2. Dikkat Mekanizmalı U-net Modeli (Attention U-net Model)

Çalışma kapsamında kullanılan veri seti toplamda 4 sınıftan oluşmaktadır. Sınıflardan biri arka planı ifade ederken diğer 3 sınıf tümör çeşitlerini ifade etmektedir. Eğitilen modelin başarısını test etmek için 4 sınıfın zar skorlarının ortalamasına ek olarak 3 sınıf (nekrotik, ödem, artırılmış) ayrıca incelenmiştir. Ek olarak modelin başarısını test etmek amacıyla literatürde farklı sınıf kümelerini temsil etmek için kullanılan TT, ÇT ve AT kümeleri de kullanılmıştır. TT kümesi artırılmış, ödem ve nekrotik sınıflarını, ÇT kümesi artırılmış ve nekrotik sınıflarını, AT kümesi ise artırılmış sınıfını içermektedir. Model eğitim aşamasında kullanılan kategorik çapraz entropi kayıp skoru da model performansını ölçmek için kullanılmıştır. 100 devir sonunda %87,33 eğitim ortalama zar skoruna ulaşırlarken %82,73 doğrulama ortalama zar skoruna ulaşılmıştır. Test veri seti ile yapılan testlerde nekrotik, ödem ve artırılmış sınıfları için sırasıyla %69,56, %86,58 ve %72,83 zar skoru elde edilirken, arka plan sınıfı için ise % 99,93 zar skoru elde edilmiştir. Ek olarak TT, ÇT ve AT sınıf kümeleri için sırasıyla, %76,32, %71,19, %72,83 zar skoru elde edilmiştir. Şekil 12’de 100 devir sonunda eğitim ortalama zar skoru ve doğrulama ortalama zar skoru için elde edilen grafik gösterilmektedir.



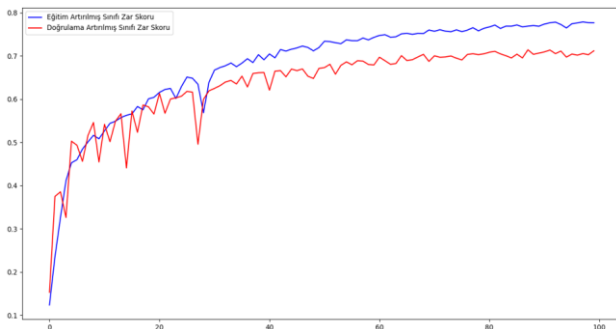
Şekil 12. Dikkat mekanizmalı u-net modeli eğitim ve doğrulama ortalama zar skoru (Training and validation of u-net model with attention mechanism average dice score)

Şekil 13'te ödem sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %91,40 eğitim ödem sınıfı zar skoru ve %85,94 doğrulama ödem sınıfı zar skoru elde edilmiştir.



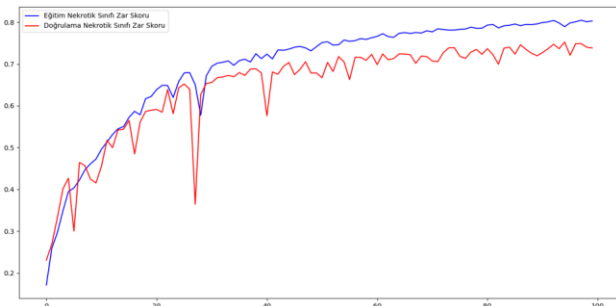
Şekil 13. Dikkat mekanizmalı u-net modeli eğitim ve doğrulama ödem sınıfı zar skoru
(Training and validation of u-net model with attention mechanism edema class dice score)

Şekil 14'te artırılmış sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %77,65 eğitim artırılmış sınıfı zar skoru ve %71,20 doğrulama artırılmış sınıfı zar skoru elde edilmiştir.



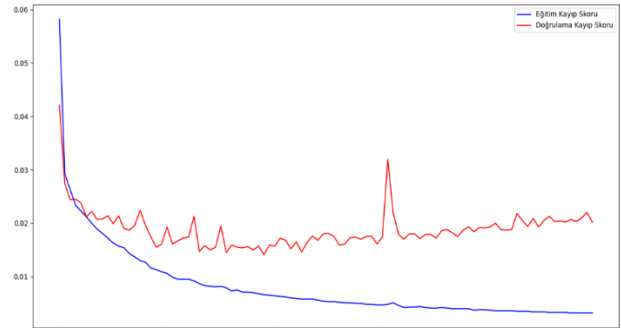
Şekil 14. Dikkat mekanizmalı u-net modeli eğitim ve doğrulama artırılmış sınıfı zar skoru
(Training and validation of u-net model with attention mechanism enhancing class dice score)

Şekil 15'te nekrotik sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %80,29 eğitim nekrotik sınıfı zar skoru ve %73,83 doğrulama nekrotik sınıfı zar skoru elde edilmiştir.



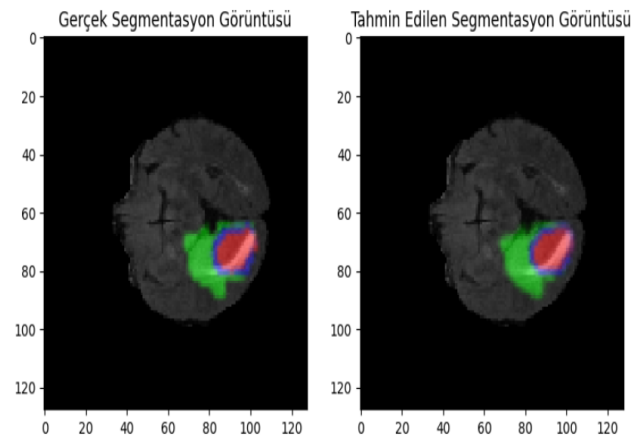
Şekil 15. Dikkat mekanizmalı u-net modeli eğitim ve doğrulama nekrotik sınıfı zar skoru
(Training and validation of u-net model with attention mechanism necrotic class dice score)

Şekil 16'da dikkat mekanizmalı u-net modelinin 100 devir sonundaki kategorik çapraz entropi kayıp skoru performansı gösterilmektedir. 100 devir sonunda 0,0061 eğitim kayıp skoru ve 0,0137 doğrulama kayıp skoru elde edilmiştir.



Şekil 16. Dikkat mekanizmalı u-net modeli eğitim ve doğrulama kayıp skoru
(Training and validation of u-net model with attention mechanism loss score)

Şekil 17'de örnek bir test verisinin rastgele bir dilimi seçilerek bu dilim üzerindeki gerçek segmentasyon görüntüsü ile tahmin edilen segmentasyon görüntüsü paylaşılmaktadır.

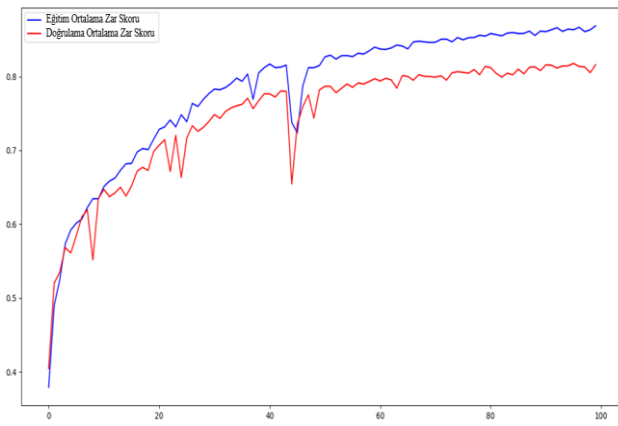


Şekil 17. Dikkat mekanizmalı u-net modeli tahmin edilen segmentasyon görüntüsü ile gerçek segmentasyon görüntüsünün karşılaştırılması
(Comparison of the u-net model predicted segmentation image with attention mechanism and the actual segmentation image)

Dikkat mekanizmalı u-net modelinin eğitim ve doğrulama sonuçlarına ait grafikler incelendiğinde doğrulama sonuçlarının eğitim sonuçlarından bir miktar geride kaldığı görülmektedir. Dengesiz sınıf dağılımı problemi mevcut veri setinde tam anlamıyla çözülemediği için model arka plan sınıfına daha fazla ağırlık vermektedir. Bu sebeple de modelin doğrulama sonuçları eğitim sonuçlarına oranla bir miktar geride kalmaktadır fakat Şekil 12 ve Şekil 17 incelendiğinde modelin test verisi ile yaptığı tahminin gerçek değerler ile benzer olduğu görülmektedir. Dikkat mekanizmalı u-net modeli test verisi üzerinde başarılı bir performans göstermektedir, ezberleme problemi oluşmamaktadır.

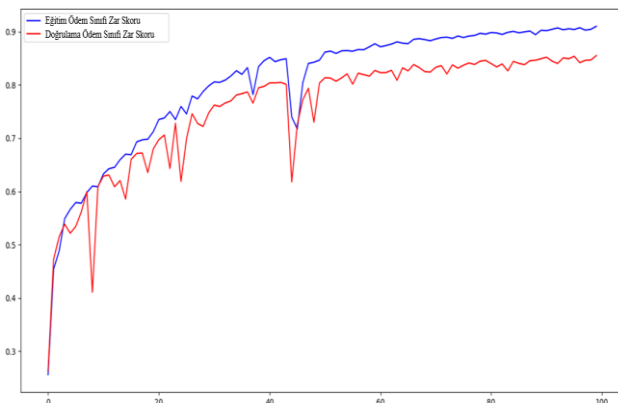
5.3. U-net Modeli (U-net Model)

Modelin başarısını test etmek amacıyla sırasıyla ortalama zar skoru, artırılmış sınıfı zar skoru, ödem sınıfı zar skoru ve nekrotik sınıfı zar skoru kullanılmıştır. Ek olarak model eğitim aşamasında kategorik çapraz entropi kayıp skoru kullanılmıştır. 100 devir sonunda %86,32 eğitim ortalama zar skoruna ulaşılırken %81,61 doğrulama ortalama zar skoruna ulaşılmıştır. Test veri seti ile yapılan testlerde nekrotik, ödem ve artırılmış sınıfları için sırasıyla %70,24, %84,26 ve %66,03 zar skoru elde edilirken arka plan sınıfı için %99,92 zar skoru elde edilmiştir. TT, ÇT ve AT sınıf kümeleri için sırasıyla, %73,51, %68,13, %66,03 zar skoru elde edilmiştir. Şekil 18’de 100 devir sonunda eğitim ortalama zar skoru ve doğrulama ortalama zar skoru için elde edilen grafik gösterilmektedir.



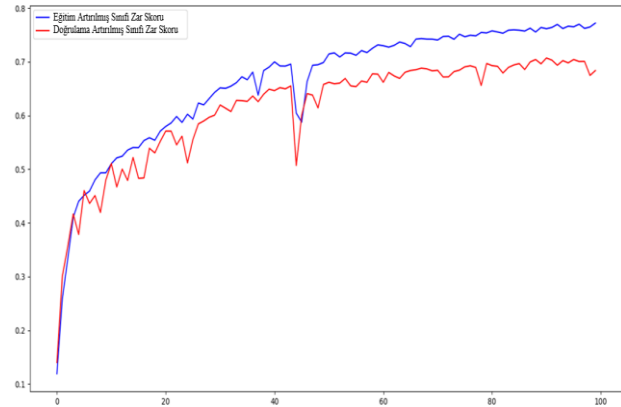
Şekil 18. U-net modeli eğitim ve doğrulama ortalama zar skoru
(Training and validation of u-net model average dice score)

Şekil 19’da ödem sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %91,01 eğitim ödem sınıfı zar skoru ve %85,52 doğrulama ödem sınıfı zar skoru elde edilmiştir.



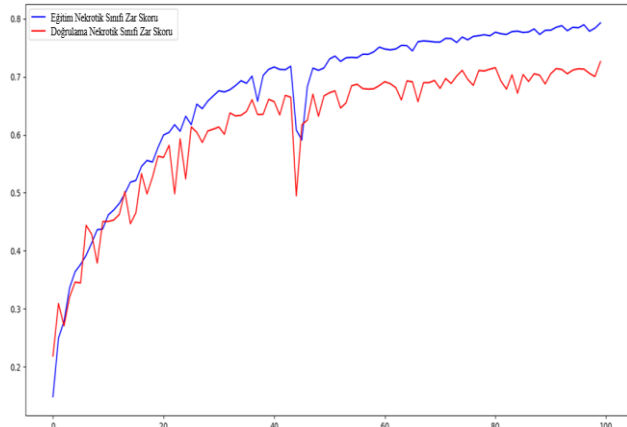
Şekil 19. U-net modeli eğitim ve doğrulama ödem sınıfı zar skoru
(Training and validation of u-net model edema class dice score)

Şekil 20’de artırılmış sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %77,22 eğitim artırılmış sınıfı zar skoru ve %68,38 doğrulama artırılmış sınıfı zar skoru elde edilmiştir.



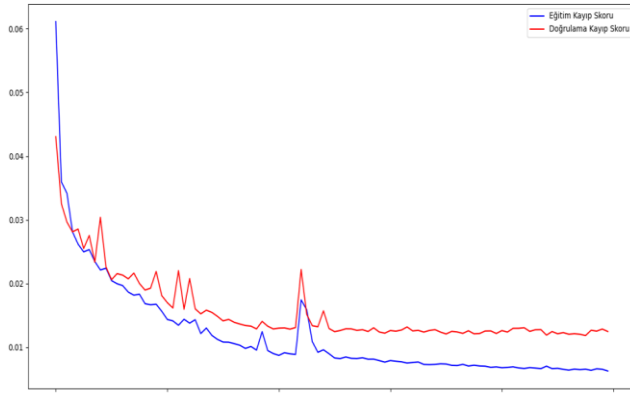
Şekil 20. U-net modeli eğitim ve doğrulama artırılmış sınıfı zar skoru
(Training and validation of u-net model enhancing class dice score)

Şekil 21’de nekrotik sınıfının 100 devirlik zar skoru performansı gösterilmektedir. 100 devir sonunda %79,28 eğitim nekrotik sınıfı zar skoru ve %72,60 doğrulama nekrotik sınıfı zar skoru elde edilmiştir.



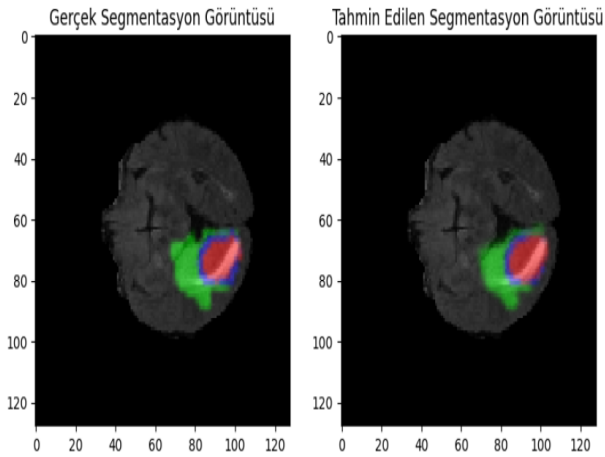
Şekil 21. U-net modeli eğitim ve doğrulama nekrotik sınıfı zar skoru
(Training and validation of u-net model necrotic class dice score)

Şekil 22’de u-net modelinin 100 devir sonundaki zar kayıp skoru performansı gösterilmektedir. 100 devir sonunda 0,0063 eğitim kayıp skoru ve 0,0125 doğrulama kayıp skoru elde edilmiştir.



Şekil 22. U-net modeli eğitim ve doğrulama kayıp skoru
(Training and validation of u-net model loss score)

Şekil 23'te örnek bir test verisinin rastgele bir dilimi seçilerek bu dilim üzerindeki gerçek segmentasyon görüntüsü ile tahmin edilen segmentasyon görüntüsü paylaşılmaktadır.

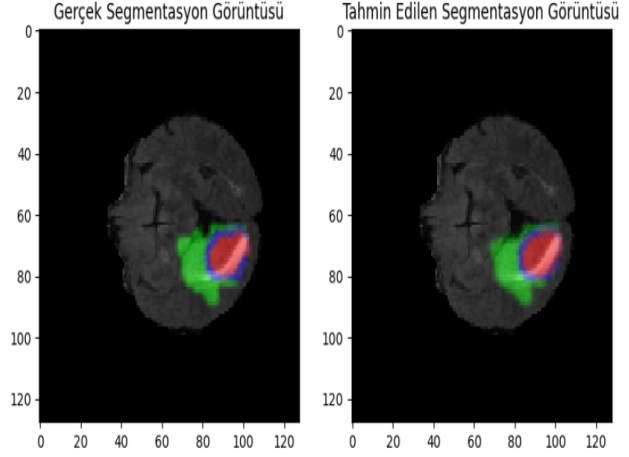


Şekil 23. U-net modeli tahmin edilen segmentasyon görüntüsü ile gerçek segmentasyon görüntüsünün karşılaştırılması

(Comparison of the u-net model predicted segmentation image and the actual segmentation image)

5.4. Topluluk Öğrenimi Modeli (Ensemble Learning Model)

Şekil 24'te topluluk öğrenimi modeli ile örnek bir test verisinin bir dilimi üzerinde yapılan test gösterilmektedir. Gerçek segmentasyon görüntüsü ile topluluk öğrenimi modelinin yaptığı tahminleri içeren segmentasyon görüntüsü karşılaştırılmıştır.



Şekil 24. Topluluk öğrenimi modeli tahmin edilen segmentasyon görüntüsü ile gerçek segmentasyon görüntüsünün karşılaştırılması

(Comparison of the ensemble learning model predicted segmentation image and the actual segmentation image)

Topluluk öğrenimi modeli oluşturulurken dikkat mekanizmalı u-net ve u-net modellerinin çıktularından faydalanılmıştır. Dikkat mekanizmalı u-net ve u-net modelleri eğitilip ağırlık dosyaları kaydedildikten sonra topluluk öğrenimi modeli oluşturulurken kullanılmıştır ve test aşamasında BRATS2017 veri seti ile testler yapılmıştır. Yapılan testlerde topluluk öğrenimi modeli ile ortalama zar skoru, arka plan sınıfı zar skoru, nekrotik sınıfı zar skoru, ödem sınıfı zar skoru ve artırılmış sınıfı zar skoru için sırasıyla %87,33, %99,96, %81,74, %91,57, %76,03 zar skoru elde edilmiştir. Ek olarak TT, ÇT ve AT sınıf kümeleri için sırasıyla %83,11, %78,88, %76,03 zar skoru elde edilmiştir. Kayıp skoru olarak ise 0,0063 elde edilmiştir.

Tablo 7, dikkat mekanizmalı u-net ve u-net modellerine ait eğitim skorlarını içermektedir. Aynı veri seti ile eğitilen iki modelin eğitim aşamasındaki elde ettiği skorlar kıyaslandığında dikkat mekanizmalı u-net modelinin u-net modelinden daha iyi sonuçlar verdiği görülmektedir. İki model nekrotik, artırılmış ve ödem tümör sınıfları bazında incelendiğinde en büyük fark 0,1496 ile nekrotik sınıfı zar skorunda, en küçük fark ise 0,0665 ile ödem sınıfı zar skorunda elde edilmiştir.

Tablo 7. Dikkat mekanizmalı u-net ve u-net modellerine ait eğitim skorları

(Training scores of u-net and u-net models with attention mechanism)

	Dikkat mekanizmalı u-net	U-net
Eğitim Ortalama Zar Skoru	0,8733	0,8687
Eğitim Arka Plan Sınıfı Zar Skoru	0,9996	0,9996

Eğitim Nekrotik Sınıfı Zar Skoru	0,8029	0,7928
Eğitim Ödem Sınıfı Zar Skoru	0,9140	0,9101
Eğitim Artırılmış Sınıfı Zar Skoru	0,7765	0,7722
Eğitim TT Skoru	0,8311	0,8250
Eğitim ÇT Skoru	0,7897	0,7825
Eğitim AT Skoru	0,7765	0,7722
Eğitim Kayıp Skoru	0,0061	0,0063

Tablo 8, dikkat mekanizmalı u-net ve u-net modellerine ait doğrulama skorlarını içermektedir. İki modele ait doğrulama sonuçları incelendiğinde tüm skorlarda dikkat mekanizmalı u-net modeli doğrulama sonuçlarının u-net modeli doğrulama sonuçlarına oranla daha yüksek olduğu görülmektedir. İki model nekrotik, artırılmış ve ödem tümör sınıfları bazında incelendiğinde en büyük fark 0,0282 ile artırılmış sınıfı zar skorunda, en küçük fark ise 0,0042 ile ödem sınıfı zar skorunda elde edilmiştir. Dikkat mekanizmalı u-net modeli hem eğitim skorlarında hem de doğrulama skorlarında u-net modeline göre daha iyi performans göstermiştir.

Tablo 8. Dikkat mekanizmalı u-net ve u-net modellerine ait doğrulama skorları

(Validation scores of u-net and u-net models with attention mechanism)

	Dikkat mekanizmalı u-net	U-net
Doğrulama Ortalama Zar Skoru	0,8273	0,8161
Doğrulama Arka Plan Sınıfı Zar Skoru	0,9993	0,9993
Doğrulama Nekrotik Sınıfı Zar Skoru	0,7383	0,7260
Doğrulama Ödem Sınıfı Zar Skoru	0,8594	0,8552
Doğrulama Artırılmış Sınıfı Zar Skoru	0,7120	0,6838
Doğrulama TT Skoru	0,7699	0,7550
Doğrulama ÇT Skoru	0,7251	0,7049
Doğrulama AT Skoru	0,7120	0,6838
Doğrulama Kayıp Skoru	0,0137	0,0125

Tablo 9’da dikkat mekanizmalı u-net, u-net ve topluluk öğrenimi modellerine ait test skorları paylaşılmaktadır. Test skorları incelendiğinde dikkat mekanizmalı u-net modeli nekrotik sınıfı hariç diğer tüm sınıflarda u-net modelinden daha iyi performans göstermiştir. Birbirlerine farklı sınıflar bazında üstünlük kuran iki farklı model ile oluşturulan topluluk öğrenimi modeli tüm skorlarda dikkat mekanizmalı u-net ve u-net modellerinden üstün bir performans göstermiştir. 3 model TT, ÇT ve AT sınıf kümelerine ait zar skorları bazında incelendiğinde, dikkat mekanizmalı u-net modeli, u-net modelinden belirgin bir şekilde daha iyi performans gösterirken, topluluk öğrenimi modeli yüksek bir farkla iki modelden de daha iyi bir performans göstermiştir.

Tablo 9. Dikkat mekanizmalı u-net, u-net ve topluluk öğrenimi modellerine ait test skorları

(Test scores of u-net and u-net models with attention mechanism)

	Dikkat mekanizmalı u-net	U-net	Topluluk Öğrenimi Modeli
Test Ortalama Zar Skoru	0,8222	0,8011	0,8733
Test Arka Plan Sınıfı Zar Skoru	0,9993	0,9992	0,9996
Test Nekrotik Sınıfı Zar Skoru	0,6956	0,7024	0,8174
Test Ödem Sınıfı Zar Skoru	0,8658	0,8426	0,9157
Test Artırılmış Sınıfı Zar Skoru	0,7283	0,6603	0,7603
Test TT Skoru	0,7632	0,7351	0,8311
Test ÇT Skoru	0,7119	0,6813	0,7888
Test AT Skoru	0,7283	0,6603	0,7603
Test Kayıp Skoru	0,0144	0,0147	0,0063

Tablo 10’da literatürde gerçekleştirilen çalışmaların sonuçları ile önerilen modelin sonuçları paylaşılmaktadır. Önerilen topluluk öğrenimi modeli ile literatürde gerçekleştirilen çalışmalar karşılaştırıldığında önerilen modelin sonuçlarının literatürde gerçekleştirilen bazı çalışmaların sonuçlarını yakaladığı görülürken bazı çalışmalar ise daha üstün performans göstermektedir. Literatürde gerçekleştirilen çalışmalar TT sınıfında önerilen modelden genel olarak daha iyi performans gösterirken ÇT ve AT sınıflarında önerilen model

literatürdeki bazı çalışmalardan daha üstün performans göstermektedir. Bu durum gerçekleştirilen çalışmanın literatürdeki sonuçları destekler nitelikte olduğunu göstermektedir.

Tablo 10. Literatürde gerçekleştirilen çalışmaların sonuçları ve önerilen model ile elde edilen sonuç (The results of the studies carried out in the literature and the result obtained with the proposed model)

	BRATS Veri Seti Versiyonu	TT	ÇT	AT
Tam Evrişimli Ağ [34]	2017	0,8760	0,7630	0,6420
Kademeli Anizotropik Evrişimsel Sinir Ağı [35]	2017	0,9050	0,8378	0,7859
Çift Akış Kod Çözmeli U-net [2]	2018	0,8986	0,8388	0,7837
Çapraz Modaliteli Derin Özellik Öğrenme Çerçevesi [10]	2017	0,8900	0,8230	0,7620
Ortak Öğrenmeye Dayalı Özellik Yeniden Yapılandırma ve Geliştirme Yöntemi [1]	2018	0,8628	0,7702	0,5964
Derin Evrişimsel Sinir Ağları [15]	2018	0,9000	0,8300	0,8300
Ardışık 3D U-net [36]	2017	0,8820	0,7320	0,7300
Süper Özellik Metrik U-net [23]	2019	0,8304	0,7625	0,8048
Geliştirilmiş Özellikli Nesil ve Çok Modlu Füzyon Tabanlı Derin Sinir Ağı [26]	2018	0,8290	0,7490	0,5910
Kanallar Arası Dikkat Mekanizmalı Kalıntı U-net [27]	2017	0,8720	0,7960	0,7810
Önerilen Model	2017	0,8311	0,7888	0,7603

6. SONUÇ

Bu çalışmada, derin öğrenme yöntemlerinin performansını değerlendirmek amacıyla çeşitli testler yapılmıştır. Gerçekleştirilen çalışmada BRATS veri setinin 2018, 2019 ve 2020 versiyonları eğitim aşamasında, 2017 versiyonu ise test aşamasında kullanılmıştır. Çalışma kapsamında aşağı örnekleme, özellik ölçeklendirme ve görüntü işleme gibi veri ön işleme yöntemleri uygulanmıştır. Eğitim aşamasında kullanılacak veri setine karar vermek için

temel bir u-net modeli üzerinde veri seti versiyonlarının tüm kombinasyonları ile testler yapılmıştır. Modalite dilimleri seçilirken dengesiz sınıf dağılımı gözetilerek örnekler içerisinde fazla sayıda bulunan sınıfı azaltacak şekilde aşağı örnekleme yöntemi uygulanmıştır. Modalite grubu seçilirken modalite türlerinin tüm kombinasyonları ile temel u-net modeli kullanılarak testler yapılmıştır. Veri seti içerisindeki farklı örneklerin birbirine üstünlük kurmasını engellemek için minimum-maksimum özellik ölçeklendirme yöntemi kullanılmıştır. Sıfıra Eşikle görüntü eşikleme yöntemi ile beyin MR görüntülerindeki tümör içeren bölgelerin daha fazla görünür olması sağlanmıştır. Veri seti üzerindeki ön işlemler tamamlandıktan sonra dikkat mekanizmalı u-net, u-net ve topluluk öğrenimi modelleri oluşturulmuştur. Dikkat Mekanizmalı u-net ve u-net modelleri eğitilirken BRATS veri setinin 2018, 2019 ve 2020 versiyonları kullanılmıştır. Eğitilen dikkat mekanizmalı u-net ve u-net modellerinin ağırlıkları ile topluluk öğrenimi modeli oluşturulmuştur. Önerilen topluluk öğrenimi modeli ile eğitilen iki modelin her bir sınıf için yaptığı tahmin olasılık değerlerinin ortalaması alınmıştır. Eğitilen modeller ve topluluk öğrenimi modellerinin başarıları veri kümesi üzerinde test edilmiştir. Dikkat mekanizmalı u-net, u-net ve topluluk öğrenimi modelleri ile test veri kümesi üzerinde sırasıyla %82,22, %80,11 ve %86,87 ortalama zar skoru oranı elde edilmiştir. Ayrıca dikkat mekanizmalı u-net modeli ile test veri seti üzerinde TT, ÇT ve AT sınıf kümelerinde sırasıyla %76,32, %71,19, %72,83 zar skoru, u-net modeli ile test veri seti üzerinde TT, ÇT ve AT sınıf kümelerinde sırasıyla %73,51, %68,13, %66,03 zar skoru ve son olarak topluluk öğrenimi modeli ile test veri seti üzerinde TT, ÇT ve AT sınıf kümelerinde sırasıyla %83,11, %78,88, %76,03 zar skoru elde edilmiştir. Literatür çalışmalarından farklı olarak modeller eğitilirken veri setinin 2018, 2019 ve 2020 versiyonları kullanılırken test aşamasında modellerin daha önce hiç görmediği 2017 versiyonu ile testler yapılmıştır ve farklı görüntü işleme teknikleri ile modelin daha kararlı ve doğru çalışması sağlanmıştır. Önerilen yöntemin dezavantajı arka plan sınıfının oluşturduğu dengesiz sınıf dağılımı probleminin üstesinden yeteri kadar gelememesi ve arka plan sınıfına daha fazla yoğunlaşmasıdır. Gelecekteki araştırmalarda dengesiz veri seti dağılımı problemini çözmek için yeni yöntemlerin nasıl kullanılması gerektiği konusu araştırılacaktır, ayrıca önerilen topluluk öğrenimi modelinin daha fazla eğitilmiş model ile beslenmesi amaçlanmaktadır.

KAYNAKLAR

- [1] Diao, Y., Li, F., Li, Z. (2023). Joint learning-based feature reconstruction and enhanced network for incomplete multi-modal brain tumor segmentation. *Computers in Biology and Medicine*, 163, 107234.
- [2] Jiang, M., Zhai, F., Kong, J. (2021). A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images, *Artificial Intelligence in Medicine*, 121, 102180.

- [3] Zhou, T. (2023). Feature fusion and latent feature learning guided brain tumor segmentation and missing modality recovery network. *Pattern Recognition*, 141, 109665.
- [4] Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis*, 43, 98-111.
- [5] Zhang, G., Zhou, J., He, G. and Zhu, H. (2023). Deep fusion of multi-modal features for brain tumor image segmentation. *Heliyon*, 9 (8).
- [6] Huang, J., Shlobin, N. A., Lam, S. K. and DeCuypere, M. (2022). Artificial Intelligence Applications in Pediatric Brain Tumor Imaging: A Systematic Review. *World Neurosurgery*, 157, 99-105.
- [7] Wang, P. and Chung, A. C. S. (2022). Relax and focus on brain tumor segmentation. *Medical Image Analysis*, 75 (102259).
- [8] Gocer, E. (2020). CapsNet topology to classify tumours from brain images and comparative evaluation. *IET Image Processing*, 14, 882-889.
- [9] Alzahrani, S. M. (2023). ConvAttenMixer: Brain Tumor Detection and Type Classification using Convolutional Mixer with External and Self-Attention Mechanisms. *Journal of King Saud University - Computer and Information Sciences*, 35 (10).
- [10] Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J. and Yu, Y. (2021). Cross-modality deep feature learning for brain tumor segmentation. *Pattern Recognition*, 110 (107562).
- [11] Naser, M. A. and Deen, M. J. (2020). Brain tumor segmentation and grading of lowergrade glioma using deep learning in MRI images. *Computers in Biology and Medicine*, 121 (103758).
- [12] Rammurthy, D. and Mahesh, P. K. (2022). Whale Harris hawks optimization based deep learning classifier for brain tumor detection using MRI images. *Journal of King Saud University - Computer and Information Sciences*, 34, 3259-3272.
- [13] Hashemzahi, R., Mahdavi, S. J. S., Kheirabadi, M. and Kamel, S. R. (2020). Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. *Biocybernetics and Biomedical Engineering*, 40, 1225-1232. 58
- [14] Nalepa, J., Lorenzo, P. R., Marcinkiewicz, M., Billewicz, B. B., Wawrzyniak, P., Walczak, M., Kawulok, M., Dudzik, W., Kotowski, K., Burda, I., Machura, B., Mrukwa, G., Ulrych, P. and Hayball, M. P. (2020). Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. *Artificial Intelligence in Medicine*, 102 (101769).
- [15] Naceur, M. B., Akil, M., Saouli, R. and Kachouri, R. (2020). Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Medical Image Analysis*, 63 (101692).
- [16] Deepak, S. and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine*, 111 (103345).
- [17] Maharjan, S., Alsadoon, A., Prasad, P. W. C., Al-Dalain, T. and Alsadoon, O. H. (2020). A novel enhanced softmax loss function for brain tumour detection using deep learning. *Journal of Neuroscience Methods*, 330 (108520).
- [18] Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J. and Sharif, M. (2020). Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research*, 59, 221-230.
- [19] Mittal, M., Goyal, L. M., Kaur, S., Kaur, I., Verma, A. and Hemanth, D. J. (2019). Deep learning based enhanced tumor segmentation approach for MR brain images. *Applied Soft Computing*, 78, 346-354.
- [20] Yang, T., Song, J. and Li, L. (2019). A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI. *Biocybernetics and Biomedical Engineering*, 39, 613-623.
- [21] Mehrotra, R., Ansari, M. A., Agrawal, R. and Anand, R. S. (2020). A Transfer Learning approach for AI-based classification of brain tumors. *Machine Learning with Applications*, 2 (100003).
- [22] Xu, W., Yang, H., Zhang, M., Cao, Z., Pan, X. and Liu, W. (2022). Brain tumor segmentation with corner attention and high-dimensional perceptual loss. *Biomedical Signal Processing and Control*, 73 (103438).
- [23] Liu, Y., Du, J., Vong, C., Yue, G., Yu, J., Wang, Y., Lei, B. and Wang, T. (2022). Scaleadaptive super-feature based MetricUNet for brain tumor segmentation. *Biomedical Signal Processing and Control*, 73 (103442).
- [24] Draï, M., Testud, B., Brun, G., Hak, J. F., Scavarda, D., Girard, N. and Stellman, J. P. (2022). Borrowing strength from adults: Transferability of AI algorithms for paediatric brain and tumour segmentation. *European Journal of Radiology*, 151 (110291).
- [25] Bidkar, P. S., Kumar, R. and Ghosh, A. (2022). SegNet and Salp Water Optimizationdriven Deep Belief Network for Segmentation and Classification of Brain Tumor. *Gene Expression Patterns*, 45 (119248).
- [26] Zhou, T., Canu, S., Vera, P. and Ruan, S. (2021). Feature-enhanced generation and multimodality fusion based deep neural network for brain tumor segmentation with missing MR modalities. *Neurocomputing*, 466, 102-112. 59
- [27] Huang, Z., Zhao, Y., Liu, Y. and Song, G. (2021). GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation. *Biomedical Signal Processing and Control*, 70 (102958).
- [28] Cinar, N., Ozcan, A. and Kaya, M. (2022). A hybrid DenseNet121-UNet model for brain tumor segmentation from MR Images. *Biomedical Signal Processing and Control*, 76 (103647).
- [29] Mazumdar, I. and Mukherjee, J. (2022). Fully automatic MRI brain tumor segmentation using efficient spatial attention convolutional networks with composite loss. *Neurocomputing*, 500, 243-254.
- [30] Zhou, T., Zhu, S. (2023). Uncertainty quantification and attention-aware fusion guided multi-modal MR brain tumor segmentation. *Computers in Biology and Medicine*, 163, 107142.
- [31] Hu, J., Gu, X., Wang, Z., Gu, X. (2023). Active consistency network for multi-source domain generalization in brain tumor segmentation. *Biomedical Signal Processing and Control*, 86, 105132.

- [32] Akbar, A. S., Fatichah, C., Suciati, N. (2022). Single level UNet3D with multipath residual attention block for brain tumor segmentation. *Journal of King Saud University - Computer and Information Sciences*, 34, 3247-3258.
- [33] Taşdemir, B., Barışçı, N. (2023). Dynamic Image Scaling for Imbalanced Brain MRI Dataset. International Conference on Global Practice of Multidisciplinary Scientific Studies-IV Turkish Republic of Northern Cyprus, 2710-2719.
- [34] Li, H., Li, A., Wang, M. (2019). A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Computers in Biology and Medicine*, 108, 150-160.
- [35] Wang, G., Li, W., Ourselin, S. (2018). Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks. International MICCAI Brainlesion Workshop, 178-190.
- [36] Beers, A., Chang, K., Brown, J. (2017). Sequential 3D U-Nets for Biologically-Informed Brain Tumor Segmentation. arXiv:1709.02967.
- [37] Hu, J., Gu, X., Wang, Z. and Gu, X. (2023). Active consistency network for multi-source domain generalization in brain tumor segmentation. *Biomedical Signal Processing and Control*, 86 (105132).
- [38] Akbar, A. S., Fatichah, C. and Suciati, N. (2022). Single level UNet3D with multipath residual attention block for brain tumor segmentation. *Journal of King Saud University - Computer and Information Sciences*, 34, 3247-3258.
- [39] Taşdemir, B. and Barışçı, N. (2023). Dynamic Image Scaling for Imbalanced Brain MRI Dataset. International Conference on Global Practice of Multidisciplinary Scientific Studies-IV Turkish Republic of Northern Cyprus, 2710-2719.

A Systematic Literature Review of Machine Learning Applications for Team Formation Problems

Literatür Makalesi/Literature Article

 Soner KARATAŞ¹,  Hüseyin ÇAKIR²

¹Gazi University, Graduate Schools of Informatics, Ankara, Turkey

²Gazi University, Gazi Faculty of Education, Ankara, Turkey

soner.karatas@gazi.edu.tr, hcakir@gazi.edu.tr

(Geliş/Received:04.01.2024; Kabul/Accepted:16.05.2024)

DOI: 10.17671/gazibtd. 1414527

Abstract — With the development of technology, the variety and number of data held for any process has increased exponentially. By processing and analyzing this data, it is possible to solve many problems. Selection of the most appropriate team member and correct team formation in the activities carried out by the team are the factors that affect the success and result of teamwork. For this reason, the problem of team member selection and team formation has become one of the increasing research topics in recent years. Researchers from different disciplines are trying to develop tools, techniques and methodologies to ensure a successful team building process. Machine Learning (ML) methods have become one of the methods that have started to be used in team formation and team member selection problems in recent years. The successful outcome of this problem depends on the correct collection and processing of data and the selection of appropriate machine learning methods. The aim of this article is to present a systematic literature review of machine learning methods applied in team formation and team member selection problems, and to show which machine learning methods are applied in this field and their performance. Articles on the subject were searched in six scientific databases. In addition to providing fundamental information about ML methods, this review also supports new research efforts on team formation problems.

Keywords— team formation, player selection, machine learning, systematic literature review

Ekib Oluşturma Sorunlarına Yönelik Makine Öğrenimi Uygulamalarına İlişkin Sistemik Bir Literatür Taraması

Özet— Teknolojinin gelişmesiyle birlikte herhangi bir sürece ait tutulan veri çeşitliliği, veri sayısı katlanarak arttı. Bu verilerin işlenmesi ve analiz edilmesiyle bir çok problemin çözümü mümkün olabilmektedir. Ekib tarafından gerçekleştirilen faaliyetlerde en uygun ekib üyesinin seçimi ve doğru ekib oluşumu ekib çalışması başarısını ve sonucunu etkileyen unsurdur. Bu nedenle ekib üyesi seçimi, takım oluşturma problemi son yıllarda artan araştırma konularında biri olmuştur. Farklı disiplinlerden araştırmacılar, başarılı bir ekib oluşturma sürecini sağlayabilmek için araçlar, teknikler ve metodolojiler geliştirmeye çalışmaktadırlar. Makine Öğrenmesi (ML) yöntemleri takım oluşumu, ekib üyesi seçimi problemlerinde son yıllarda kullanılmaya başlayan yöntemlerden biri olmuştur. Bu problemin başarılı sonucu verilerin doğru bir şekilde toplanması, işlenmesi ve uygun makine öğrenme yöntemlerinin seçimine bağlıdır. Bu makalenin amacı takım oluşumu, ekib üyesi seçimi problemlerinde uygulanan makine öğrenme yöntemlerinin sistemik bir literatür taramasını sunmak, bu alanda hangi makine öğrenme metodlarının uygulandığını ve bunların performansını göstermektir. Altı bilimsel veri tabanında konuyla ilgili makaleler araştırılmıştır. Bu inceleme ML yöntemleri hakkında temel bilgiler sağlamanın yanı sıra takım oluşumu problemlerinde yeni araştırma çalışmalarını da desteklemektedir.

Anahtar Kelimeler — takım oluşumu, oyuncu seçimi, makine öğrenimi, sistemik literatür taraması

1. INTRODUCTION

In addition to the activities that can be carried out individually, there are activities planned to be carried out by the team. When it comes to choosing the team members who will carry out this activity among the alternatives, the success and performance of the team will vary depending on the choice. Team formation is crucial because team success depends largely on the appropriate assignment of team members to the teams [1]. Some teams could be more effective than others, only because of the composition of the characteristics of its members [2]. Attributes such as communication skills, teamwork experience, and personality traits, are criteria that affect team effectiveness and have an impact on team collaboration, efficiency and productivity [1]. The assignment will need to be made according to the characteristics of the candidate to be selected for the task. Therefore, it is important to implement and apply an effective technique to ensure (to some extent) the optimal team composition [1]. Many academic personnel have studied on to research novel techniques and methodologies to solve this problem. Team formation, which is not based on any basis and is based solely on the selection of team members based on human instinct, is an issue that is criticized for many reasons such as loss of time, accuracy and effort. Therefore, demand and orientation for methodology, techniques and tools that enable team formation by identifying the best team member is increasing recently.

Data production is increasing exponentially every day. Using machine learning to extract information from this data in different fields has been widely used in recent years. Machine Learning algorithms and associated Artificial Intelligence technologies are helpful in various fields such as prediction and decision making [3]. ML approaches have the ability to handle high dimensional and multivariate data, and to extract hidden relationships within data in complex and dynamic environments (such as, industrial environments) [4]. Using the machine learning approach in selecting the most suitable team member and team building problems has become one of the popular areas of study in recent years. The results show that machine learning algorithms can be used for player selection and team formation process [5]. However, the success of these applications depends on which ML technique is applied to which data and under which conditions.

The purpose of this article is to present a Systematic Literature Review (SLR) of articles whose topic is team formation and team member/player selection using machine learning techniques. This article serves as an important resource on machine learning techniques, the characteristics of the data involved in their use, problems in implementation, and recent advances, inspiring new research efforts on team formation and team member selection.

The contents of other chapters of the article are given below. Chapter 2 explains how SLR is implemented.

Chapter 3 provides an overview of machine learning. Chapter 4 includes considerations regarding the research questions and brief information of the literature reviewed. Finally, in chapter 5, the conclusions and contributions of this article are presented.

2. LITERATURE REVIEW PLANNING PROTOCOL

2.1. Research Methodology

SLR is a process that allows the collection of relevant evidence that meets predetermined eligibility criteria on a particular topic and answers to formulated research questions [6]. Although they are similar in general terms, methodologies of systematic literature reviews may differ. For this purpose, different methodologies have been included and applied in many sources. In this article, the five-stage steps for the research methodology suggested by Petersen et al. [7] were applied. This systematic mapping method proposed by Peterson et al. aims to provide an overview of the area of interest, reduce systematic errors, and increase the legitimacy of the analyzed data for more reliable results [1], [7].

2.2. Search Strings

Keywords for search strings were selected based on words commonly taking part in the literature. For the systematic literature study, searches were made in 6 different online databases containing publications in this field. These are Scopus, IEEE, Web of Science, Ebsco (Academic Search Ultimate), Science Direct, Springer Link databases. Since database search functions work differently, search strings have determined as follows.

- Scopus: Article title-abstract-keywords (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)
- IEEE: Abstract (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)
- Web of Science: Abstract (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)
- EBSCO (Academic Search Ultimate): Abstract or author-supplied abstract (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)
- Science Direct: Title, abstract or author-specified keywords (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)

- Springer Link: (“team formation” AND “machine learning”) OR (“player selection” and “machine learning”) OR (“player ranking” and “machine learning”)

6 databases were searched. As a result of this search, 376 articles that fell within the search criteria were found. The exclusion criteria specified in item 2.4 were applied to these articles. It was determined that 28 articles were appropriate for the subject of this article.

2.3. Research Questions

Research questions prepared for machine learning applications in team formation problems are given below. While determining the research questions, similar literature studies on machine learning and the needs of new research studies in this field were taken into considerations.

Q1) What are the current studies and research on team formation problems using ML techniques?

Q2) What are the machine learning methods that are being used on team formation/player selection problems?

Q3) What are the data/datasets used on machine learning and what are their properties?

2.4. Screening of Papers for Inclusion & Exclusion

After examining publications obtained through the search string search, only 28 articles were included in the systematic literature review. Different criteria have been created to determine articles suitable for the research topic and objective. Criteria for inclusion and exclusion are set out below.

Inclusion Criteria:

- Include papers related to team formation studies using machine learning methods
- Include papers from 2000 (year)
- Include papers only in English language

Exclusion Criteria:

- Papers containing team formation studies using different methods other than machine learning
- Not relevant to the research questions
- Papers in languages other than English
- Only one of the same publications in different databases is taken into account and the others are excluded.

2.4. Data Extraction & Mapping

All articles were reviewed and the information listed below was extracted; title, author, year, keywords, country, publication type, publisher, research type, used machine learning methods, data source and type, study summary,

contribution and future work proposal. These extracted data were analyzed for the research questions given in the previous sections.

3. MACHINE LEARNING FUNDAMENTALS

In daily life, people make many decisions. In which areas they invest in different periods, which study subjects they pursue, which school they prefer, etc. When making these decisions, it may be difficult for them to make logical decisions due to reasons such as not having access to the correct information, not evaluating the correct information in an appropriate and methodological way, being influenced by their emotions, and missing details due to the mass of data. At this stage, machine learning, which models human thought structure and decision-making ability, evaluates all situations, examines millions of data very quickly and enables rational decisions to be made in a very short time, is the solution for such problems. For machine learning, it can be said that it is a developing branch of computational algorithms designed to imitate human intelligence [8], and it is the modeling of a problem with an algorithm suitable for data [9], [10].

Machine learning involves many steps, starting from data to applying the model. Below is the figure for this.

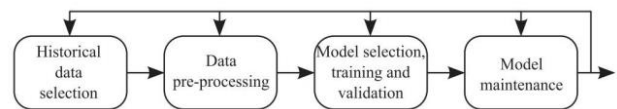


Figure 1. Machine learning steps [4]

One of the main factors affecting the results of machine learning studies is working with the most appropriate and accurate data set. Obtaining this data is the first step in machine learning. The second step is the data preprocessing step, where the data is converted and processed into a suitable form for the ML model. Below are some of these processes with examples.

- Data transformation,
- Data cleaning
- Data reduction

The third step consists of three main components. In model selection, the appropriate model that can solve the problem is selected. The next component is training the model based on the data. In the last component, the model's potential to represent the system is evaluated. The final step of machine learning is model maintenance and includes checks and

improvements to application performance that may change over time.

4. RESULTS OF SLR

4.1. What is the latest studies and research on team formation problems using ML techniques? (Q1)

Totally 28 articles were found in the search made in 6 large databases according to the criteria specified in Chapter 2. The table including general information about these articles is given below.

In the investigation of the field in which the studies were carried out, 10 articles on the selection-ranking of football players, 7 articles on the selection of cricket players, 4 articles on the selection of software team members, 3 articles on the selection of expert team members, 3 article on the selection-ranking of basketball players and 1 article on the selection-ranking of ice hockey players were identified. From these results, it has been seen that studies mainly focus on creating sports and software teams. It has been concluded that machine learning methods have not been studied for other topics in team formation problems and that there are areas where they can be conducted.

In the investigation, it was seen that although studies on team formation problems took place before 2015, studies using machine learning techniques were not available. Figure 2 shows the number of articles published after 2000 (using this article's extraction criteria). It appears that the first article was published in 2016. This research shows that articles on ML techniques in team formation and team member selection have become more intense, especially after 2020. Therefore, it is possible to say that the

application of machine learning for team formation problems is a new application area. It can be said that interest in this field of research has increased after 2020. On the other hand, when the studies on publication type are examined, the number of articles is 18 and the number of conference papers is 10.

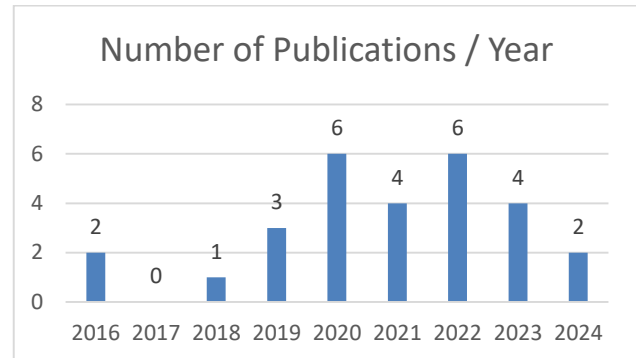


Figure 2. Number of publications / Year

It was examined in which countries the articles were prepared. India has been the country where the most intensive studies in this field have been carried out, with 9 articles. Thailand, where similar articles were published on the same subject, came in second place with 3 articles. Turkey, Bangladesh and Iran are the countries that have published 2 publications in this field. Taiwan, Ireland, Palestine, Canada, Australia, Austria, Greece, Pakistan, Saudi Arabia, USA, Sweden, Italy are among the countries that publish in this field with 1 article each. Considering that the subject of machine learning is a developing field in recent years, it is seen that researchers in India tend to use machine learning techniques in solving team formation problems in the 2020s.

Table 1. Information about studies

Ref. No	In which field?	Publication Type	Database	Country	Year
3	Cricket player selection	Article	Scopus	India	2022
5	Football player selection	Article	Scopus	Turkey	2021
11	Football player selection	Conference Paper	Scopus	India	2022
12	Football player selection	Conference Paper	Scopus	Bangladesh	2019
13	Cricket player selection	Conference Paper	Scopus	India	2023
14	Cricket player selection	Conference Paper	IEEE	India	2020
15	Cricket player selection	Conference Paper	IEEE	India	2020

Ref. No	In which field?	Publication Type	Database	Country	Year
16	Software team member selection	Article	Scopus	Thailand	2022
17	Expert team member selection	Article	Scopus	Taiwan	2022
18	Software team member selection	Conference Paper	Scopus	Palestine	2022
19	Software team member selection	Article	Scopus	Thailand-Pakistan	2021
20	Football player selection	Conference Paper	Scopus	India	2021
21	Expert team member selection	Article	Scopus	Ireland	2020
22	Software team member selection	Conference Paper	Scopus	Thailand	2020
24	Expert team member selection	Article	Scopus	India	2016
25	Football player selection	Article	Academic Search Ultimate (EBSCO)	Canada	2018
26	Football player selection	Article	Academic Search Ultimate (EBSCO)	Turkey	2023
27	Football player selection	Article	Academic Search Ultimate (EBSCO)	Iran	2019
28	Cricket player selection	Article	Science Direct	India	2023
29	Basketball player selection	Article	Science Direct	Australia	2024
30	Cricket player selection	Article	Science Direct	Bangladesh	2022
31	Football player selection	Article	Springer link	Iran-Austria	2023
32	Basketball player ranking	Article	Springer link	Greece	2024
33	Basketball player ranking	Article	Science Direct	Pakistan-Saudi Arabia	2021
34	Football player ranking	Conference Paper	Web of Science	USA	2016
35	Ice hockey player ranking	Conference Paper	Scopus	Sweden	2020
36	Football player ranking	Article	Scopus	Italy	2019
37	Cricket player ranking	Article	Scopus	India	2020

4.2. What are the Machine learning methods that are being used on team formation/player selection problems? (Q-2)

One of the search questions of the article was to determine which ML methods are used in team formation problems. In the review conducted for this purpose, it was seen that more than one machine learning algorithm was used in many articles. It has been stated that the purpose of this is to achieve the most efficient result by using different machine learning algorithms and to compare these algorithms. The figure showing the number of machine learning algorithms used is given below in Fig.3. Accordingly, it was seen that the most used machine learning algorithms were Random forest (16), Decision Trees (13), Support Vector Machines (13). Since these results were obtained in 28 article reviews, it is seen that these algorithms were used in %57,1 - %46,4 of the articles. On the other hand, the least used machine learning algorithms were Polynomial Regression, Q Learning with 1 article each (Polynomial Regression was stated in the others groups). It has been observed that the use of boosting algorithms such as XGBoost, Catboost, etc. has increased in recent years. When the machine learning classes were examined, it was seen that classification class algorithms were used in 21 articles, regression class algorithms were used in 16 articles, and the clustering class algorithm was used in 3 articles (Figure 4). From these results, it has been determined that ML methods are used extensively for classification and regression purposes in team formation problems. Another important data was obtained from machine learning types (Fig 5). Accordingly, supervised learning methods were used in 25 articles (%89,2). Unsupervised learning method was used in only 3 articles (%10,7). It was determined that the reinforcement method was used in 1 article (%3,5). As a result, it has been observed that supervised learning methods are used extensively in team formation problems.

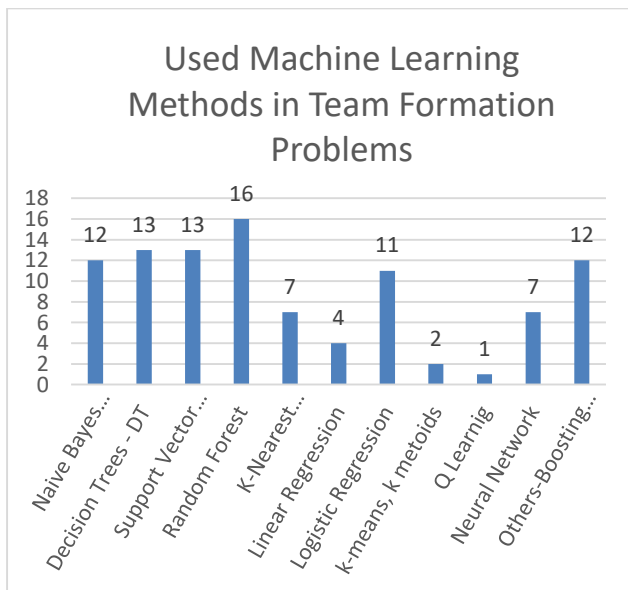


Figure 3 Used machine learning methods in Team formation problems

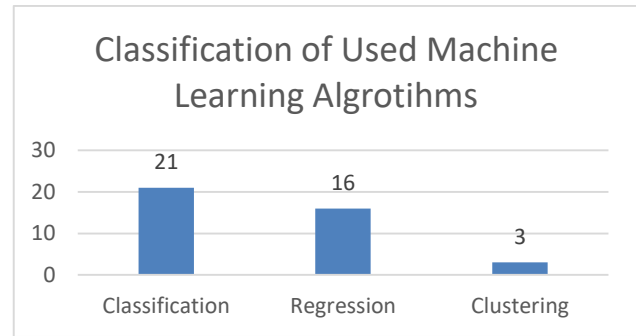


Figure 4. Classification of used machine learning algorithms

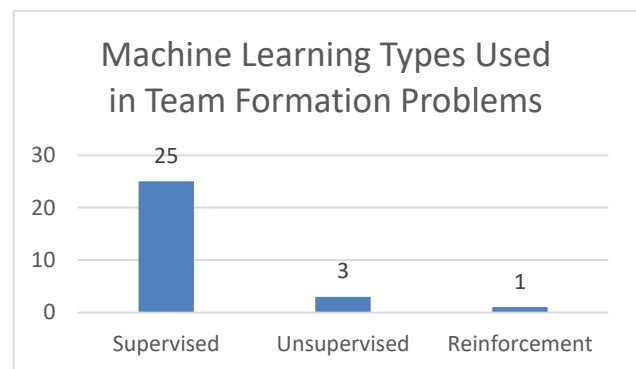


Figure 5. Machine Learning Types

In the study of Datta and Rudra [11], the researchers evaluated the performance of players in football clubs. For performance evaluation of players, they used five machine learning models and compared their efficiency values. These machine learning methods are Decision Tree (DT), Support Vector Machine (SVM), Linear Regression, Random Forest (RF) and XGBoost. The XGBoost method is superior to other methods with low error values. When the article is examined, it is seen that the technique with the lowest performance was obtained in support vector machine.

Shahriar and his colleagues [12] studied to develop a player classification and selection method for a football team. The players were classified based on their performances using multiple ML techniques. Five algorithms were used for the classification. These are: Support Vector Machine (SVM), Decision Tree, Naïve Bayes Classifier (NB), Random Forest Classifier and KNN Classifier. A small comparison was made between them after the classification. Five factors were included in the comparison. As a result of the comparison, the best performance was obtained from the Decision Tree algorithm, and the weakest performance was obtained from the Support Vector Machine algorithm.

Sumathi, Prabu and Rajkamal [13] developed a system to evaluate the performance of cricket players. Linear Regression technique was used to predict the players' performance. K-means algorithm was used to find players with similar performance. Random Forest technique was applied to generate the ranking list.

Shetty and her colleagues [14] developed the selection of the best 11 players in the Indian cricket team with their proposed model. Different techniques such as Random Forest, SVM, Decision Tree and Logistic Regression were used. Random Forest were the most successful techniques obtained results.

Santra and colleagues [15] developed a technique to predict the batsman's (cricket player) order by evaluating the batsman's past performance. The novelty of the technique lies in designing the algorithm and taking into account multiple cricket parameters instead of a single parameter in predicting the best batsman. Polynomial Regression and Linear Regression models were created for prediction. Polynomial Regression was found to outperform linear regression in every aspect.

Random Forest, Decision Tree, Second Order Discriminant Analysis (QDA), Neural Network, Naive Bayes, k-Nearest Neighbors (kNN), Logistic Regression machine learning algorithms were used and compared for the team effectiveness scoring function for software development team in Assavakamhaenghan et al. [16] study. In this study, the most successful results were obtained with the Random Forest model.

Chang et al. [17] proposed an approach using the Reinforcement Learning (RL) method to create a solution to the team formation problem.

In Ishi and his colleagues' paper [3], the hybrid approach of CS-PSO with machine learning models was used to find the right set of team combinations from the group of players. 9 machine learning methods, including K Nearest Neighbors, Random Forest, Decision Tree, Gradient Boosting Algorithm, Logistic Regression, Naive Bayes, XGBoost, CatBoost and Support Vector Machine were applied and the most successful results were obtained from SVM.

Tanbour et al. [18] developed a ML technique that can better match software development experts with software project tasks. Three different models were tested in their studies: Random Forest Classifier, Decision Tree Classifier and Logistic Regression. Random Forest Classifier has been the most successful model.

In the study of Tuarob's and his colleagues [19], they proposed a machine learning model that can recommend suitable software team members for software development

tasks. Seven classification algorithms were tested and Random Forest received the best score for all datasets.

Ghar et al. [20] designed a ML model that predicts the future performance (VAEP value) by examining the past performances of football players.

Abidin [5] applied ML algorithms for team formation and player selection for the players of a football team. 7 machine learning algorithms were used and their performances were compared. These algorithms were Logistic regression, Random Forest (RF), Classification and Regression Tree (CART), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM) and Bayes Theorem.

In their proposed framework, Keane et al. [21] incorporated both the individual's attributes and the topological features of the individual's network into a ML link prediction task. They developed models such as LR, SVM and CART Using Python's XGBoost module.

Assavakamhaenghan et al. [22] studied on possibility to adopt the team recommendation algorithm proposed by Liu et al. [23] to develop a software team recommendation. They proposed the logistic regression approach to recommend suitable software team members. The approach take both individual strength and collaborative efficiency among team members into account to give a recommendation.

Three machine learning techniques were applied to select the best employees for effective team formation and their results were compared in Krishankumar and Ravichandran's [24] article. The three methods were Ensemble Decision Tree (EDT), Artificial Neural Network (ANN) and Decision Tree. The results reveal that the EDT approach performs better.

In Tosato and Wu's article [25], Projective Adaptive Resonance Theory (PART) was used to provide data-based sports decisions. PART is the machine learning projective clustering algorithm and based on neural network. It was seen that PART provides a purely data-driven analysis to identify attributes for a group of players in an unsupervised way.

Buyrukoglu and Savas [26] studied on a technique to classify footballer positions using a stacked ensemble ML model. Firstly they used 4 filter based feature selection methods to choose optimal feature subsets. Then 2 level stacked ensemble machine learning algorithms were used to determine the position of the football player. In Level-0, Gradient Boosting, Random Forest and Deep Neural Network algorithms were used as based models and then in

Level-1, the Logistic Regression algorithm was used as a meta-model.

Machine learning approaches were employed to establish a ranking for players in Maanijou and Mirroshande's article [27]. Different algorithms (SVM, Logistic Regression, PART, Naive Bayes, etc.) were used to classification and ranking. Experiments were done in the Persian soccer league. The study showed promising results for ranking.

Manju and Philip [28] studied on a novel framework that ranks batsmen (cricket players) on their performances. New performance index was created using different machine learning algorithms (Logistic Regression, Support Vector Machine, RF, XGB Classifier, CART and Naive Bayes). Then players were clustered with k-means clustering algorithm to identify the best players.

Ke, Bian and Chandra [29] developed a unified framework to categorize the players and built the optimal team model. Principal Components Analysis (PCA) was used to reduce the number of features describing player performance in the unsupervised phase. Then simple neural network was used to build a team model in the supervised learning phase.

Tirtho, Rahman and Mahbub [30] studied on to forecast player performance for future cricket competitions. Different machine learning methods were used for this purpose (Random Forest, Decision Trees, K-nearest Neighbors, Support Vector Machine and Naive Bayes). It was determined that Random Forest produced the best accurate prediction models for cricket players.

Nourai, Eslahchi and Baca [31] developed deep learning models to obtain right scores for players' positions. They designed a procedure that can identify the best players for each position.

To forecast NBA player performance several ML models (Decision tree, Linear regression, Random forest, etc.) were employed in the study of Papageorgiou et al. [32]. It was seen that the approach of blending data from the last ten seasons and last three seasons increased the prediction accuracy and model stability. They emphasized that the use of standard statistics, advanced statistics datasets, long-term and short-term data, is crucial for obtaining right predictions.

Mahmood et al. [33] presented prediction mechanism to explore rising stars in basketball using machine learning methods. Different machine learning methods like Maximum Entropy Markov Model (MEMM), Classification and Regression Trees (CART), Support Vector Machines (SVM) and Naïve Bayes (NB) were applied to find a function which can assign class label to a

feature set. It was seen that the Maximum Entropy Markov Model was dominant in all data sets in terms of F-measure score.

Brooks et al. [34] proposed the Pass Shot Value (PSV) which is a metric to predict the importance of a pass resulting with a shot. They used support vector machine model to estimate whether or not a given pass generates a shot. When they ranked players in La Liga with more than 200 passes with stated metric, they saw some of the outstanding players at the top.

Persson et al. [35] compared which features best predict the success of ice hockey players in different positions. 6 different machine learning methods were used (Logistic regression, Bayesian network, k-Nearest neighbor, Naive Bayes, Decision tree and Random forest) to predict players' ranking tier for 3 player positions. They compared the result of the models and they concluded that two Bayesian classifiers had best performance and sensitivity.

PlayeRank which is a data-driven framework that offers evaluation of the performance of football players has been presented by Papalodios et al. [36]. They used Linear Support Vector Machine (SVM) to solve classification problem. They compared PlayeRank with known algorithms for performance evaluation in soccer and they saw that PlayeRank outperforms the competitors.

Kaviya, Mishra and Valarmathi [37] ranked the players, based on the Player Ranking Index using machine learning techniques. They used various algorithms for ranking which include Decision Tree, JRIP reduced error pruning algorithm, Support Vector Machine, Random Forest, and Naïve Bayes classifier, etc. JRIP was seemed the most promising amongst all the algorithms.

In order to compare the results of this study, other literature studies on machine learning were also examined. Carvalho et al. [38] explored a systematic literature review, covering the main papers of Predictive Maintenance (PdM) using ML techniques. In this study, the most used ML methods to perform PdM were found to be Random Forest (RF), Artificial Neural Network (ANN) and Support Vector Machine (SVM). Another literature study is the work of Xames et al. [39] on recent trends of machine learning applications in additive manufacturing. The most used ML methods in this literature study were ANN, SVM, Ensemble (including RF) based ML methods. Similarly, in our literature review in the field of team formation, RF, SVM and Decision Trees were among the most frequently used ML techniques. It is considered that the reason for this choice is that RF and SVM can be used in classification and regression problems and provide versatile flexible solutions. However, in the literature studies examined, it is seen that the use of the Boosting algorithm family

(XGBoost, Catboost etc.) has increased, especially in recent years.

To answer question Q3, all articles were examined on this axis and the results were shared below.

4.3. What are the data/datasets used on machine learning and what are their properties?

Table 2. Properties of Data

Ref. No.	Data Type	Source (from)	Volume of data	Data cleaning and Pre-processing	Train/Test Data
3	Real Data	ESPN Cricinfo website.	101*4 players. The number of columns could not be determined.	Pre-processing was carried out.	%70 training- %30 testing
5	Real and Synthetic data	From Hit/it Assistant and the coach evaluations. Synthetically generated data also.	Football team has only 21 players. Synthetic data generation is done as 10 synthetic instances for one player. So, the size of the dataset becomes 231 instances in total. Feature : 28	Preprocessing and standardization of two data sources	Not mentioned
11	Real Data	SoFIFA.com	89 columns (attributes) and 18.207 rows (players).	Filter and clean data	%80 training - %20 testing
12	Real Data	www.footystats.org	The number of rows could not be determined. Eight main attributes.	Not mentioned	Multiple training datasets.
13	Real Data	https://www.kaggle.com	304 records, 13 attributes.	Remove noisy and missing data.	Not mentioned
14	Real Data	espnricinfo.com	The number of rows could not be determined.	It is stated that preprocessing is applied.	%80 of the dataset was used for training, %20 of the dataset was used to test.
15	Real Data	IPL website.	The number of rows and columns could not be determined.	Not mentioned	Train regression models to predict batsman positions. Test data from ball by ball data of 2019
16	Real Data	Jira (Apache and Atlassian datasets)	507.319 and 238.322 Rows	Cleaning of data	%80 training and %20 testing
17	Real Data	SNAP and SIoT	The number of rows and columns could not be determined.	Not mentioned clearly	Not mentioned

Ref. No.	Data Type	Source (from)	Volume of data	Data cleaning and Pre-processing	Train/Test Data
18	Real Data	https://data.world website.	1.000 computer development-related experts (Rows). Columns: could not be determined.	Not mentioned	%60 training and %40 testing
19	Real Data	Jira(Atlassian, Apache, and Moodle)	The number of rows could not be determined. There are about 20 features.	The collected data are then prepossessed (e.g. removing incomplete issues).	Each dataset are separated into %80 training and %20 testing sets.
20	Real Data	EA Sports, FIFA, FBref, Wyscout	511 players. 104 attributes	Clean not mentioned, preprocess yes	%80 training and %20 testing.
21	Real Data	USPTO data set.	Not mentioned	Not mentioned	New datasets were created for the purpose of training machine learning models. The model was tested on unseen dataset.
22	Real Data	Jira (Moodle).	Total number of projects about 26.800. Feature number: 6	Not mentioned clearly	%80 training, %20 testing
24	Real Data	Several websites like Odesk, Elance etc.	474 recipients (row), eight predictors (Column)	Normalizing the data using max-min normalization.	The training and testing data are split equally for training and validating methods.
25	Real Data	Football Manager 2018	Forty seven attributes for twenty four soccer player	Not mentioned	Not mentioned
26	Real Data	https://www.kaggle.com/karangadiya/fifa19	89 columns and 18.207 rows	Missing/null values are removed from the dataset.	%80 training and %20 testing
27	Real Data	Online freely sources	495 players, 20 features	Missing and noisy data are removed from the dataset	Not mentioned
28	Real Data	Websites ESPNcrinfo and IPLT20	283 players	Data extraction and data cleaning are performed	%70 training and %30 testing
29	Real Data	5 data sets from the NBA official website	8.511 entries and 71 features	Dimensional reduction and visualization	Not mentioned

Ref. No.	Data Type	Source (from)	Volume of data	Data cleaning and Pre-processing	Train/Test Data
30	Real Data	www.howstat.com, www. cricmetric.com	152 players	Dimensional reduction	%80 training and %20 testing
31	Real Data	Sofifa dataset	18.034 players and 48 attributes	Not mentioned	%80 training data, %20 testing
32	Real Data	National Basketball Association: www.nba.com	79.036 instances, 67 features, 203 players	Cleaning and pre-processing are performed	%70 training data, %30 testing data
33	Real Data	A sports website www.espn.com/nba/	100 players	Not mentioned clearly	Not mentioned clearly
34	Real Data	2012-13 La Liga season	Not mentioned clearly	Normalization is performed	%80 training data, %20 testing data
35	Real Data	https://www.hockey-reference.com/	Not mentioned clearly	Some data was removed. Numerical data was normalized.	%80 training data, %20 testing data
36	Real Data	https://wyscout.com	31.496.332 events, capturing 19.619 matches, 296 clubs and 21.361 players.	Normalization is performed	%80 training data, %20 testing data
37	Real Data	Dataset obtained from a cricsheet website	Approximately 550 rows, 22 features	It states that the data sets are processed, but does not clearly state the processes.	Not mentioned clearly

When the data sources are examined, it is seen that the majority of them are collected from public websites. Many websites that keep sports player statistics were used in the studies. Open source software systems hosted on the Jira platform, such as Apache, Atlassian and Moddle, were preferred in software team member selection. When the data sets were examined, it was evaluated that the most distinctive source was in Abidin's work [5]. In this study, datasets were obtained from 2 main sources. The first source is the Hit/it device, an electronic sports system that can measure a player's skills. The coach's evaluations are the second source.

When the data type issue was examined, it was seen that real data was used in all articles. Only in Abidin's article [5] synthetic data were used as well as real data. Since the number of players was 21 and it was thought that this number was not enough for classification, 10 synthetic data were produced for each player. According to the results of these two examinations, we can state that it is preferred to use statistical real data obtained from open sources in team building problems.

When the articles were evaluated according to their data size, it was determined that some of them did not include

this issue clearly. In article no. [16], a total of around 700.000 lines of data were studied. However, the number of features cannot be clearly seen in this article. In article no [11], a large data set consisting of 18.207 rows (18.207 players) and 89 columns (attributes and skills of the players) was used. In Abidin's article [5], where the lowest data set was used and therefore synthetic data addition was needed, 231 players and 28 features were included.

When the majority of the articles were examined, there were statements that data clean and data preprocessing processes were used. For example, in the study of Sumathi et al. [13] before analysis the dataset is preprocessed to remove noisy and missing data. In this article, each numeric column is normalized because it has a different representation.

The last issue examined regarding the data was the ratio of training and testing data. While this rate information was not clearly shared in 11 articles, the amount of this rate was given in 17 articles. Accordingly, the number of articles using the training test ratio %80-%20 is 12, the number of articles using %70%30 is 3, the number of articles using %60-%40 is 1 and the number of articles using %50-%50

was 1. According to these values, we can say that the %80-%20 scale is commonly used for the training test data usage rate in team formation problems.

When other literature studies on machine learning were examined, it was determined that in the studies of Carvalho et al. [38], %89 real data was used and %11 synthetic data was used in the literature in the field of PdM. In this study, it was observed that real data was used in all studies on team formation and synthetic data was additionally used in one study. It was evaluated that this similarity resulted from the availability of historical real data in both study areas.

5. CONCLUSION

In this article, a systematic literature review on team formation problems was conducted using machine learning techniques. It was aimed to answer the research questions specified in the planning protocol. As a result, it has been seen in many articles that different ML techniques are used to evaluate performance among them. The majority of articles stated that the best performance result was obtained from Random Forest (RF) technique. It has been seen that Supervised Machine Learning methods are dominant in practice. However, it has been determined that ML algorithms are used extensively for classification and regression purposes in team formation problems.

It has been revealed in the reviewed articles that the application of ML techniques to team formation problems is mostly in sports and software fields. Looking at the article publication momentum, articles subject on team formation using ML techniques have an increasing momentum after 2020. For this reason, it is anticipated that the volume of articles in this scope will continue to increase.

Within the scope of this study, the data/datasets used in machine learning and their characteristics were also investigated. Data in machine learning was mostly obtained from public websites and was real. In terms of data size, there were examples where a large amount of data was used, as well as articles where a small amount of data was used. Similar to the results of this literature study, it was observed that real data was used instead of synthetic data in other literature studies on machine learning. Since the success of machine learning depends on obtaining accurate data, the data acquisition processes were also examined and it was determined that the data preprocessing stages was explained clearly and in detail in many articles. It was considered that not explaining this issue clearly in some articles was a deficiency, and it was evaluated that these preprocessing stages should be given in detail in

articles related to machine learning. In the articles reviewed, examples of data cleaning, data reduction and data normalization were mostly included in the preprocessing stages. When the separation of training and test data sets was examined, it was seen that the %80-%20 scale is commonly used in team formation problems.

It can be stated that machine learning techniques such as RF, SVM and Decision Tree have been implemented extensively to team formation problems. When other literature review studies on machine learning were examined, it was seen that RF, SVM, ANN methods were most commonly used in solving other problems. It has been evaluated that the reason for this similarity is that these methods provide versatile and flexible solutions. Similar to the findings of this literature study, it has been observed in other literature studies that the use of ensemble algorithms has increased.

However, it is considered that there are still some aspects that can be further researched in solving team formation problems with machine learning. In this regard, suggestions for future research are as follows;

- develop works that apply machine learning methods other than sports and software teams, such as quality assessment teams, and comparing the results
- comparison of performance rates by applying different data sets for the same problem.

In summary, this comprehensive review highlights the applicability of machine learning techniques to team formation problems and sets the stage for future research efforts.

REFERENCES

- [1] G. Stavrou, P. Adamidis, J. Papathanasiou, K. Tarabanis "Team Formation: A Systematic Literature Review", *Int. Journal of Business Science and Applied Management*, 18(2), 2023.
- [2] J. Juárez, C. Santos, F. A. A. M. N. Soares, R. Vita, R. P. Francisco, J. P. Basto, S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance", *ACM Computing Surveys*, 54(7), 2021.
- [3] M. Ishi, J. Patil, J. Jhang, V. Patil, "An efficient team prediction for one day international matches using a hybrid approach of CS-PSO and machine learning algorithms", *Array* 14, 2022.
- [4] T. P. Carvalho, C. Santos, C. A. Brizuela, "A Comprehensive Review and a Taxonomy Proposal of Team Formation Problems", *Computers & Industrial Engineer*, 2019.
- [5] D. Abidin, "A case study on player selection and team formation in football with machine learning", *Turkish Journal of Electrical Engineering & Computer Sciences*, 29, 1672 – 1691, 2021.

- [6] W. Mengist, T. Soromessa, G. Legese, "Method for conducting systematic literature review and meta-analysis for environmental science research", *MethodsX* 7, 2020.
- [7] K. Petersen, S. Vakkalanka, L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update" *Information and Software Technology*, 64, 1–18, 2015
- [8] I. El Naqa, M. J. Murphy, "What is machine learning?" In *Machine Learning in Radiation Oncology*, 3-11, 2015
- [9] M. Atalay, E. Çelik, "Artificial Intelligence and Machine Learning Applications in Big Data Analysis", *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172, 2017
- [10] E. Tosunoğlu, R. Yılmaz, E. Özeren, Z. Sağlam, "Eğitimde Makine Öğrenmesi: Araştırmalardaki Güncel Eğilimler Üzerine İnceleme", *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 3(2), 178-199. 2021
- [11] M. Datta, B. Rudra, N. Mead, C. Rolland, "An Intelligent Decision Support System for Bid Prediction of Undervalued Football Players", 2nd International Conference on Intelligent Technologies (CONIT), 2022.
- [12] T. Shahriar, Y. Islam, N. Amin, "Player Classification Technique Based on Performance for a Soccer Team Using Machine Learning Algorithms", 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019.
- [13] M. Sumathi, S. Prabu, M. Rajkamal, "Cricket Players Performance Prediction and Evaluation Using Machine Learning Algorithms", 2023 International Conference on Networking and Communications (ICNWC), 2023.
- [14] M. Shetty, S. Rane, C. Pandita, S. Salvi, "Machine learning-based Selection of Optimal sports Team based on the Players Performance", Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020), 2020.
- [15] A. Santra, A. Sinha, P. Saha, A. K. Das, "A Novel Regression based Technique for Batsman Evaluation in the Indian Premier League", 2020 IEEE International Conference for Convergence in Engineering, 2020.
- [16] N. Assavakamhaenghan, W. Tanaphantaruk, P. Suwanworaboon, M. Choetkiertikul, S. Tuarob, "Quantifying effectiveness of team recommendation for collaborative software development", *Automated Software Engineering*, 2022.
- [17] C. Chang, M. Chang, J. Jhang, L. Yeh, C. Shen "Learning to Extract Expert Teams in Social Networks", *IEEE Transactions On Computational Social Systems*, 9(5), 2022.
- [18] Z. Tanbour, D. Khudarieh, H. Abuodeh, A. Hawash, "Forming Software Development Team: Machine-Learning Approach", 2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS), 2022.
- [19] S. Tuarob, N. Assavakamhaenghan, W. Tanaphantaruk, P. Suwanworaboon, S. Hassan, M. Choetkiertikul, "Automatic team recommendation for collaborative software development", *Empirical Software Engineering*, 2021.
- [20] S. Ghar, S. Patil, W. Tanaphantaruk, V. Arunachalam, "Data Driven football scouting assistance with simulated player performance extrapolation", 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021.
- [21] P. Keane, F. Ghaffar, D. Malone, "Using machine learning to predict links and improve Steiner tree solutions to team formation problems - a cross company study", *Applied Network Science*, 2020.
- [22] N. Assavakamhaenghan, P. Suwanworaboon, W. Tanaphantaruk, S. Tuarob, M. Choetkiertikul, "Towards Team Formation in Software Development: A Case Study of Moodle", 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2020.
- [23] H. Liu, M. Qiao, D. Greenia, R. Akkiraju, S. Dill, T. Nakamura, Y. Song, and H. M. Nezhad, "A machine learning approach to combining individual strength and team features for team recommendation," *Proceedings of The 13th International Conference on Machine Learning and Applications*, 12, pp. 213–218, 2014.
- [24] R. Krishankumar, K. S. Ravichandran, "A Novel Trio Combo Strategy For Efficient Team Formation Using Hybrid Triangulation Mechanism", *ARPN Journal of Engineering and Applied Sciences*, 11(5), 2016.
- [25] M. Tosato, J.Wu, "An Application Of Part To The Football Manager Data For Players Clusters Analyses To Inform Club Team Formation", *Big Data & Information Analytics*. 3 (1), 43-54, 2018.
- [26] S. Buyrukoglu, S. Savas, "Stacked-Based Ensemble Machine Learning Model for Positioning Footballer", *Arabian Journal for Science and Engineering*. 48, 1371-1383, 2022.
- [27] R. Maanjou, S. A. Mirroshandel, "Introducing an expert system for prediction of soccer player ranking using ensemble learning", *Neural Computing and Applications*. 31, 9157-9174, 2019.
- [28] M. K. Manju, A. O. Philip, "Novel method for ranking batsmen in Indian Premier League", *Data Science and Management*. 6, 158-173, 2023.
- [29] Y. Ke, R. Bian, R. Chandra, "A unified machine learning framework for basketball team roster construction: NBA and WNBA", *Applied Soft Computing*. 153, 2024.
- [30] D. Tirtho, S. Rahman, S. Mahbub, "Cricketer's tournament-wise performance prediction and squad selection using machine learning and multi-objective optimization", *Applied Soft Computing*. 129, 2022.
- [31] M. Nouraie, C. Eslahchi, A. Baca, "Intelligent team formation and player selection: a data-driven approach for football coaches", *Applied Intelligence*. 53, 30250-30265, 2023.
- [32] G. Papageorgiou, V. Sarlis, C. Tjortjis, "An innovative method for accurate NBA player performance forecasting and line-up optimization in daily fantasy sports", *International Journal of Data Science and Analytics*, 2024.
- [33] Z. Mahmood, A. Daud, R. A. Abbasi, "Using machine learning techniques for rising star prediction in basketball", *Knowledge-Based Systems*, 211, 2021.
- [34] J. Brooks, M. Kerr, J. Guttag, "Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights", *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49-55, 2016.

- [35] T. L. Persson, H. Kozlica, N. Carlsson, P. Lambrix, "Prediction of Tiers in the Ranking of Ice Hockey Players", 7th International Workshop on Machine Learning and Data Mining for Sports Analytics, MLSA 2020, 89-100, 2020
- [36] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, F. Giannotti "PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach", ACM Transactions on Intelligent Systems and Technology, 10(5), 1-27, 2019
- [37] A. Kaviya, A. S. Mishra, B. Valarmathi, Comprehensive Data Analysis and Prediction on IPL using Machine Learning Algorithms", *International Journal on Emerging Technologies*, 2020.
- [38] T. P. Carvalho , A. A. Fabrizzio, M. N. Soares, , V. Roberto, P. F. Roberto, J. P. Bastoc , S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance", *Computers & Industrial Engineering*. 137, 2019.
- [39] D. Xames, F. K. Torsha, F. Sarwar, "A systematic literature review on recent trends of machine learning applications in additive manufacturing", *Journal of Intelligent Manufacturing*. 34, 2529–2555, 2023F

Graf Teorisi ve Malatya Merkezilik Algoritmasına Dayalı Haber Metinlerinin Özetlemesi

Araştırma Makalesi/Research Article

 Cevher Tayyib BAKAN^{1*},  Selman YAKUT^{1*}

¹Yazılım Mühendisliği, İnönü Üniversitesi, Malatya, Türkiye
cevherbakan@gmail.com, selman.yakut@inonu.edu.tr
(Geliş/Received:01.04.2024; Kabul/Accepted:16.05.2024)
DOI: 10.17671/gazibtd.1463107

Özet— Günümüzde internetin yaygın kullanımıyla, bilgi kaynaklarındaki doğru bilgiye erişimi önemli kılmaktadır. Bilgi kaynaklarının artmasıyla birlikte özgün içeriğe sahip bilgiye erişim güçleşmektedir. Bu nedenle metin özetleme yöntemlerinin önemi giderek artmaktadır. Haber metinleri gibi önemli temel bilgi kaynaklarının etkili bir şekilde özetlenmesi günümüzde bir gereklilik haline gelmiştir. Bu çalışmada haber metinlerinin etkili bir şekilde özetlenmesi için Malatya merkezilik algoritmasını temel alan bir özetleme yaklaşımı önerildi. Önerilen yaklaşımda orijinal metin tanımlayıcıların çıkarılması, kelime köklerinin elde edilmesi gibi çeşitli ön işlemlerden geçirilerek graf yapısına dönüştürülür. Graf'a dönüştürülen metin için Malatya merkezilik algoritması kullanılarak graftaki düğümlerin Malatya merkezilik değerleri hesaplanır. Bu değerler dikkate alınarak metin özetini oluşturan özetler seçilir. Seçilen özetler graftan çıkarılır. Oluşan yeni graf yapısı için merkezilik değeri hesaplanarak seçim işlemleri devam ettirilir. Graf Teorisi ve Malatya merkezilik algoritmasının birlikte kullanımı, haber metinlerinin özetlenmesinde verimliliği artırdığı gösterildi. Bununla birlikte haber içeriklerinin anlamlı bir şekilde özetlenmesi sağlandı. Bu yaklaşımın başarısını değerlendirmek amacıyla BBC veri seti üzerinde toplamda 2224 İngilizce haber metniyle kapsamlı bir şekilde test edildi. Çalışmada haber metinleri etkili bir şekilde özetlendiği yapılan testlerle ve alınan etkili rouge değerleriyle gösterildi. Graf teorisi ve Malatya merkezilik algoritması, bilgiye erişimi kolaylaştırmak ve anlam düzeyinde etkileşimi artırmak adına önemli bir potansiyele sahip olduğu gösterildi. Elde edilen uygulama sonuçları, haber metinlerini daha anlamlı bir şekilde sunabileceğini ve etkili özetler üretilabileceğini göstermektedir.

Anahtar Kelimeler— haber metinleri özetleme, graf tabanlı özetleme, malatya merkezilik algoritması, çıkarım tabanlı özetleme yöntemleri, metin işleme

Summarization of News Texts Based on Graph Theory and Malatya Centrality Algorithm

Abstract— The widespread use of the internet today highlights the importance of accessing accurate information from reliable sources. With the proliferation of information sources, accessing original content has become increasingly challenging. Therefore, the importance of text summarization methods is steadily increasing. Effectively summarizing essential information sources such as news articles has become a necessity. In this study, an approach based on the Malatya centrality algorithm is proposed for effectively summarizing news articles. In the proposed approach, various preprocessing steps such as extracting descriptive terms and obtaining word roots are applied to transform the original text into a graph structure. The Malatya centrality algorithm is then utilized to calculate the centrality values of nodes in the graph representing the text. Based on these values, summaries constituting the text summary are selected. The selected summaries are removed from the graph. Centrality values are then calculated for the resulting new graph structure, and the selection process continues. The combined use of Graph Theory and the Malatya centrality algorithm is shown to enhance efficiency in summarizing news articles. Additionally, meaningful summarization of news content is achieved. To evaluate the success of this approach, it was comprehensively tested on a total of 2224 English news articles from the BBC dataset. The study demonstrates effective summarization of news articles through conducted tests and achieved effective ROUGE scores. The utilization of Graph Theory and the Malatya centrality algorithm is shown to have significant potential in facilitating information access and increasing interaction at the semantic level. The obtained application results indicate the ability to present news articles in a more meaningful manner and produce effective summaries.

Keywords— news texts summarization, graph based summarization, malatya centrality algorithm, inference based summarization methods, text processing

1. GİRİŞ (INTRODUCTION)

Günümüzde, bilgi kaynaklarının ve internetin gelişimiyle birlikte birçok bilgi kaynağına erişilmekte ancak bu kaynaklar içinde doğru ve anlamlı içeriğe ulaşmak giderek zorlaşmaktadır. Metin özetleme, bu bilgi yoğunluğu altında önemli bir araç haline geldi. Özellikle haber metinleri gibi uzun ve detaylı içeriklerin özünü kavramak, anlamlı bir şekilde özetlemek, günümüzün bilgi çağında vazgeçilmez bir gerekliliktir. Bu bağlamda, geleneksel metin özetleme yöntemleri zaman zaman sınırlamalarla ve hatalarla karşılaşır. İnternet teknolojisinin gelişmesiyle birlikte, kullanıcılar geniş bir bilgi yelpazesine erişebilmekte ancak doğru, özgün ve öne çıkan içeriği belirlemek oldukça zor hale gelmektedir. Metinlerde tekrar eden ifadeler, gereksiz detaylar, yoğunluk asıl mesajı bulmayı ve önemli bilgileri belirlemeyi zorlaştırır.

Metin özetlemede çıkarıma dayalı ve yoruma dayalı olmak üzere temel olarak iki yöntem kullanılır. Çıkarıma dayalı özetleme, temelde metindeki önemli ifadelerin tespitini amaçlar. Bu süreçte, kelimelerin yinelenmesi ve benzerleri göz önünde bulundurularak ana fikirler belirlenir. Öte yandan, yoruma dayalı özetleme ise derin öğrenme ve yapay sinir ağları gibi yöntemlerle, verilen metinden yeni ve anlamlı cümlelerle özet çıkarır. Her iki yönteminde olumlu ve olumsuz yanları bulunur. Yoruma dayalı özetlemede, yeni cümleler oluştururken anlam bozuklukları ortaya çıkabilir. Yoruma dayalı özetleme, çıkarıma dayalı özetlemeye göre çok daha uzun sürebilmektedir. Çıkarıma dayalı özetlemelerde, yeni cümle oluşturamazken metindeki en anlamlı cümleleri bir araya getirerek özet oluşturur. Bu yöntem çok daha hızlı ve anlam bozukluğu olmadan özetleri oluşturabilir.

Minimum tepe örtme problemi, bir graftaki bütün kenarların kapsandığı düğümlerinin minimum alt kümesini seçmeyi amaçlar. Bu seçilen düğümler, grafa ait tüm kenarları en az bir kez kapsamalıdır. Yani, seçilen düğümler sayesinde grafin tüm kenarlarıyla temas edilir. Bu problem, çeşitli uygulamalarda, özellikle ağ tasarımı ve optimizasyonunda kullanılır. Örneğin, bir iletişim ağındaki düğümleri temsil eden bir graf düşünülürse bu düğümleri bir alt küme olarak seçilip herhangi iki düğüm arasında iletişim kuracak bir bağlantı oluşturulabilir. Ancak bu bağlantıların oluşturulması maliyetlidir. En az maliyetle tüm düğümler arasında iletişim kurulmalıdır. Minimum tepe örtme problemi, bu durumda bize en az sayıda düğüm seçerek bu bağlantıların kurulmasını sağlayacak bir çözüm sunar. Bu problem, polinom zamanında çözülemeyen bir problem olarak bilinir. Ancak pratikte, çeşitli yaklaşımlar ve algoritmalarla çözümüne yaklaşmak mümkündür. Bu nedenle, Minimum tepe örtme problemi, bilgisayar biliminde ve birçok alanda önemli bir optimizasyon sorunudur. Yakut ve diğerleri, Minimum tepe örtme problemi için, çalışmasında Malatya vertex-cover algoritmasını sundular. [1] Bu algoritmanın metin özetleme üzerinde de başarılı olabileceği düşünüldü. Haber metinlerini özetlerken bu algoritma kullanılarak algoritmanın metin özetlemedeki başarısı da ortaya çıkarıldı.

Bu çalışma, BBC veri seti üzerinde gerçekleştirilen 2224 haber metni ile Malatya merkezilik algoritması ve graf teorisi gibi tekniklerin performansını değerlendirerek haber metinlerinin özetlenmesinde yeni bir yaklaşım sunar. Önerilen yaklaşımda haber metinleri tanımlayıcıların

atılması, kelime köklerinin belirlenmesi gibi belirli önışlemlerden geçirilerek graf yapısına dönüştürülür. Daha sonra oluşturulan bu graf yapısına Malatya merkezilik algoritması uygulanarak graftaki düğümlerin Malatya merkezilik değerleri hesaplanır. Bu algoritma, metin içindeki temel yapıları vurgulayarak özünü koruyan, anlaşılır bir özetleme süreci sunabilmektedir. Ayrıca çalışma, daha önceki araştırmalarda belirtilen graf temelli çıkarımsal ve yoruma dayalı yöntemleri de göz önüne alır. Bilgi çağının gerekliliklerine uygun olarak daha etkili ve anlaşılır metin özetleme yöntemlerinin geliştirilmesine katkı sağlamayı hedefler. BBC veri seti üzerinde gerçekleştirilen 2224 haber metni üzerinde yapılan deneyler ve üretilen Rouge değerleri önerilen yöntemin haber kaynakları gibi farklı metinleri içerebilen metinlerde etkili metin özetleme yapıldığını gösterir.

Bu makale çalışmasının devamında şu bölümler yer almaktadır: bölüm 2’de önerilen çalışma ile ilgili literatüre yer verildi. Bölüm 3’te önerilen algoritma yer almaktadır. Bölüm 4’te çalışmanın deneysel sonuçlarına yer verildi. Bölüm 5’te ise sonuçlar yer almaktadır.

2. LİTERATÜRDE İLGİLİ ÇALIŞMALAR (LITERATURE RELATED STUDIES)

Metin özetleme, akademik literatürde farklı şekillerde ele alınan geniş bir konu başlığıdır. Bu alandaki çalışmalar, graf tabanlı çıkarma yöntemlerini kullananlar olduğu gibi, makine öğrenimi yöntemleri ile yorumlama odaklı yaklaşımları benimseyenleri de barındırır. Literatürdeki bu yaklaşımları ve uygulama alanlarını farklı şekillerde sınıflandırmak ve ele almak mümkündür. Bununla beraber bu makalede daha çok literatürdeki çıkarsama odaklı yöntemlere ilişkin örnek çalışmalar incelendi.

Bakan ve Yakut, metin özetleme işlemi için graf teorisi ve Malatya merkezilik algoritması kullanmıştır. [2] Bu çalışmada örnek bir metin graf yapısına dönüştürülerek bu metinde özet çıkarma işlemi gerçekleştirilmiştir. Önerilen algoritmanın, metin özetleme işleminin hızlı ve tam olarak gerçekleştirdiği örnek uygulamayla gösterilmiştir. Önerilen algoritma Malatya merkezilik algoritması kullanılarak metinler için etkili özetler üretildiğini göstermektedir.

Yakut ve diğerleri, minimum tepe örtme problemi için Malatya vertex-cover algoritması önermişlerdir. Bu algoritma, grafin düğümleri için Malatya merkezilik değerleri kullanılarak oluşturulur. İlk adımda grafin düğümlerinin Malatya merkezilik değerleri, Malatya merkezilik algoritması kullanılarak hesaplanır. İkinci adımda minimum tepe örtme çözümü için düğümler, graf içindeki maksimum Malatya merkezilik değerine sahip düğüm seçilerek ve çözüm kümesine eklenerek seçilir. Daha sonra bu düğüm ve bu düğüme bağlı olan kenarlar graftan çıkarılır. Kalan düğümlerden oluşan graf için Malatya merkezilik değerleri tekrar hesaplanır ve seçim işlemi devam eder. İşlem, grafin tüm kenarları kaplandığında sona erer. [3]

Tülek, Türkçe metinlerin özetlerini oluşturmak için kök çıkarma algoritmaları ve kök-ek analizini kullanılmış ve ardından özetlenecek cümleler belirlenmiş. [4] Khushboo ve diğerleri ise graf sıralama algoritmalarını en kısa yol algoritmalarıyla birleştirerek otomatik cümle çıkarma tekniklerini geliştirdi. [5] Erkan ve Dragomi ise LexRank yaklaşımını benimseyerek cümlelerin önemini, cümleler

arası kosinüs benzerliğine dayalı olarak belirlendi.[6] Moawad ve Aref, orijinal belgeyi özetlemek için zengin bir anlamsal graf kullandı ve bu grafik üzerinden daha kısa bir belge oluşturdu. [7] Ferreira ve diğerleri, metin işleme uygulamaları için dört ayrı boyutta temellendirilmiş bir graf modeli sundu. [8]

Mallick ve diğerleri, her bir web sayfasına özgü olarak düzenlenmiş bir TextRank metodu kullanarak belgenin özünü yakalayan bir graf-tabanlı metin özetleme yöntemi geliştirdi.[9] Sankarasubramaniam vd. Vikipedi'yi grafik sıralama ile birleştirerek orijinal metne başlık cümleleri ekleyerek yeni bir yaklaşım geliştirdi.[10] Alguliev ve Aliguliyev ise belgesiz bir özetleme yöntemi üzerinde çalıştı, orijinal belgeden cümlelerin kümelmesi ve seçilerek özet oluşturulması konusunda yeni kriter fonksiyonları sundular.[11] Nagwani ve Verma, Java programlama dilinde sık kullanılan terimlere dayalı olarak bir metin özetleme algoritması tasarladı ve uyguladı.[12] Mihalcea'nın çalışmasında, otomatik cümle çıkarma için graf-tabanlı sıralama algoritmaları kullanılan bir yöntem sunulmaktadır.[13] Akter ve diğerleri, tek veya çoklu Bengalce belgelerinden önemli cümleleri çıkaran bir metin özetleme yöntemi önermişlerdir. Bu yaklaşımda, giriş belgeleri tokenizasyon ve kök çıkarma gibi ön işlemlerden geçmekte; daha sonra Terim Frekansı-Ters Belge Frekansı (TF/IDF) kelime puanlarını hesaplamakta ve bu kelimelerin puanları cümle puanlarını belirlemek için toplanmaktadır. Ayrıca, çalışma K-ortalama kümeleme algoritmasını kullanmaktadır.[14]

Literatürdeki diğer çalışmalar, duygusal temelli yaklaşımların önemli ilgi gördüğünü göstermektedir. Babar ve Patil'in çalışması, metin özetleme için Fuzzy Mantık Çıkarımı ve Latent Semantic Analysis gibi anlamsal yaklaşımlara odaklanmaktadır.[15] Ayrıca Ozsoy ve diğerleri, Latent Semantic Analysis (LSA) tabanlı özetleme algoritmalarını tanıtmış ve bu öneriler de makalenin yazarlarına aittir. [16]

Makine öğrenimi ve derin öğrenme gibi teknikleri kullanan bazı çalışmalar da incelendi. Erhandı, derin öğrenme yaklaşımını kullanarak Türkçe ve İngilizce metinler üzerinde bir özetleme çalışması yürütmüş.[17] Neto ve diğerleri, orijinal metinden doğrudan çıkarılan özellikler kullanarak eğitilebilir makine öğrenimi algoritmaları kullanarak metin özetleme sürecini sunmuşlar.[18] Silla vd. metin özetleme sorununu bir sınıflandırma problemi olarak ele alarak otomatik özetleme yöntemi geliştirmişlerdir. Çalışmanın amacı, sınıflandırma algoritmalarının performansını artırmak için Genetik Algoritma tabanlı özellik seçiminin etkinliğini araştırmaktır.[19]

Genetik algoritmaları kullanan birçok çalışmada incelenmiştir. Kaynar'ın çalışmasında, metin tabanlı özetleme için genetik algoritma kullanılmıştır. Bu genetik algoritma, veri setleri üzerinde sistemi eğitmek için kullanılmıştır.[20] Al-Abdallah ve Al-Taani, Arapça belge özetleme için Parçacık Sürü Optimizasyonu algoritmasını önermişler. Önerilen yaklaşım, Evrimsel Algoritmalar ve Harmony Search yöntemleriyle karşılaştırılmıştır.[21]

Karicioğlu ve Yaşa, otomatik metin özetleme için genetik algoritmaların başarı oranlarının yüksek olduğu ve bu alanda daha iyi bireyleri seçerek problemi çözebileceği gösterilmiştir. Gelecekte, makine öğrenimi algoritmalarıyla genetik algoritmanın uyum fonksiyonunu geliştirmeyi ve

farklı yapay zeka yöntemlerini bir araya getirerek sistem performansını artırmayı hedeflemektedirler.[22]

Karakoç ve Yılmaz, derin öğrenme yöntemleriyle Türkçe haberlerin özetlenmesi üzerine odaklanılmıştır. Makalede, haber başlıklarının tahmin edilmesinde kodlayıcı-kod çözücü modeli kullanılarak yoruma dayalı metin özetleme yöntemi uygulanmıştır. Yapılan deneysel çalışmalar, haber metninin tamamının eğitimiyle elde edilen modelin, özetleme performansı açısından diğer modellere göre daha başarılı olduğunu göstermiştir.[23]

Torun ve İner, internet haberciliğinin geleneksel medya üzerindeki etkilerini inceleyerek, TextRank algoritmasıyla Türkçe haber metinlerini özetleyerek haber vektörleri oluşturmuş ve benzerlik tespiti için TF-IDF ve Doc2Vec kullanmıştır. Uzmanlarca değerlendirilen sistem, detaylı bir başarı analizi sunmuştur.[24]

Kartal ve Kutlu, haber metinlerini en iyi temsil eden cümleleri seçerek otomatik bir özetleme sistemi öneriyor. Bu model, cümle konumu, anlatım ifadeleri, varlık isimleri, kelime sıklığı ve başlık benzerliği gibi öznitelikleri kullanarak geliştirilmiş ve LSA tabanlı metotlardan daha başarılı sonuçlar elde etmiştir.[25]

Hark ve diğerleri, entropi tabanlı metin özetleme yöntemi üzerine yapılan çalışma inceleniyor. Bu metodun, metinlerin yapısını ve içerdiği bilgi miktarını değerlendirerek ROUGE metrikleriyle başarılı bir performans gösterebildiği belirtiliyor. Entropi tabanlı bu yöntem, diğer metotlara göre umut verici sonuçlar sunarak araştırmacıların dikkatini çekiyor.[26]

Wang vd. biyomedikal literatür ve elektronik sağlık kayıtları üzerindeki özetleme tekniklerini ve değerlendirme yöntemlerini inceledi. 2013 ile 2021 arasındaki 7235 makaleden 58'i incelenerek, mevcut sistemlerin çoğunlukla literatür özetlemeye odaklandığı, hibrit yöntemlerin başarılı olduğu ve elektronik sağlık kayıtlarının özetlenmesine dair çalışmaların arttığı belirlendi.[27]

Kumar vd. farklı dillerdeki metin özetleme yöntemlerini inceleyerek Hindistan ve yabancı dillerde gerçekleştirilen çalışmaları karşılaştırır. Genellikle makine ve derin öğrenme temelli sınıflandırma yöntemlerinin başarılı sonuçlar verdiğini belirtirken, gelecekteki araştırma alanları olarak özellik çıkarma, sınıflandırma yöntemlerinin geliştirilmesi ve farklı dil veri tabanlarının bulunabilirliği gibi konulara işaret eder.[28]

Uçkan vd. çoklu belgelerden özet çıkarmak için CatSumm adında yeni bir yöntem tanıtıyor. Bu yöntem, metinlerin temsili, spektral grafik bölütleme ve cümlelerin puanlanması üzere üç aşamadan oluşuyor. Yapılan deneyler, önerilen CatSumm metodunun DUC-2002 ve DUC-2004 veri setlerinde %44.073 ile %56.513 arasında başarılı özetleme sonuçları verdiğini gösteriyor 2021.[29]

Aydın ve Uçkan, metin özetleme için farklı benzerlik ölçütleri ve çizge temsilleri kullanarak çıkarımsal metin özetleme yöntemlerini incelemiştir. Kosinüs Benzerliği, Jaccard Benzerliği, Levenshtein Benzerliği ve Pearson Korelasyon Katsayısı gibi ölçütlerle çizgeler oluşturulmuş, ardından Arasındalık Merkeziliği, Yakınlık Merkeziliği, Derece Merkeziliği ve Özvektör Merkeziliği ölçütleriyle en değerli cümleler tespit edilerek 200 ve 400 kelimelik özetler üretilmiştir. DUC-2002 veri seti üzerinde yapılan deneysel

çalışmalar sonucunda, Jaccard Benzerliği ve Yakınlık Merkeziliği ile 200 kelimelik, Kosinüs Benzerliği ve Özvektör Merkeziliği ile 400 kelimelik özetlerin en yüksek performansı sergilediği belirlenmiştir.[30]

Kuş ve Acı, Türkçe tıbbi metinlerin özetlenmesi üzerine odaklanıyor. COVID-19'a ilişkin 84 makale üzerinde çalışarak, Term Frekansı ve LexRank gibi algoritmalarla genişletilmiş özetler elde etti. Yapılan değerlendirmeler, özetlerin orijinal metinlere benzediğini gösteriyor. Gelecekte daha derin öğrenme metodlarıyla çalışmayı hedefliyor.[31]

Literatürde metin özetleme için farklı şekillerde sınıflandırılabilir birçok yaklaşım önerilmekle beraber bu yaklaşımları genel olarak yorumlayıcı ve çıkarıcı metin özetleme yaklaşımları olarak ele almak mümkündür. Yorumlayıcı metin özetleme yaklaşımları metinde anlamı dikkate alarak özetleme yaparken çıkarımsal yaklaşımlar metindeki ifadeleri özet belirlemede kullanılmaktadır. Bu yaklaşımların kullanım amacı, kullanılan dilin türü, başarı ölçütü gibi parametreler dikkate alınarak değerlendirilebilir. Benzer algoritmaların farklı dil türlerinde veya farklı içeriğe ve amaca sahip veri setlerinde metin özetleme gerçekleştirilebilmektedir. Özellikle haber kaynakları gibi farklı alanlara hitap edebilen yaklaşımların kullanılması etkili metin özetleme açısından önemlidir.

3. ÖNERİLEN YÖNTEM (RECOMMENDED METHOD)

Bu çalışmada haber metinlerinin özetlenmesinde etkili bir algoritma önerildi. Malatya merkezilik algoritması, metinlerin anlamını temsil eden etkili özetlerin elde edilmesi için önemli kelimeleri belirlenmesi ve metin özetlerinin oluşturulması amacıyla kullanıldı. Bu süreç, birkaç adımdan oluşur. İlk olarak, gelen haber metinleri ön işleme adımları ile işlenir. Bu adımlar arasında metinlerin ayrıştırılması, tokenlara bölünmesi, gereksiz kelimelerin (stopwords) çıkarılması ve kelimelerin köklerinin belirlenmesi yer alır. Ardından, Malatya merkezilik algoritması, bu ön işlemde geçmiş metinleri kelime düğümlerinden oluşan bir graf olarak temsil eder.

Malatya merkezilik algoritması, kelime düğümleri arasındaki bağlantıları ve kelimelerin önemini belirlemek için kullanılır. Graf yapısındaki her kelimenin bağlantıları, metindeki diğer kelimelerle olan ilişkileri gösterir. Bu ilişkilerin incelenmesi ve analiziyle Malatya merkezilik skorları hesaplanır.

Skorlar, bir kelimenin metindeki konumu, sıklığı ve diğer kelimelerle olan ilişkileri temel alınarak belirlenir. Malatya merkezilik skorlarına göre en yüksek skorlara sahip kelimeler seçilir. Bu kelimeler, metnin ana fikrini ve en önemli noktalarını yansıtan kelimelerdir. Bu seçilen kelimelerin içerdiği cümleler birleştirilerek özet metin oluşturulur. Özet metin, özgün metnin temel noktalarını içerir ve orijinal metnin anlamını korurken daha kısa ve öz bir yapı sunar. Bu yöntem, Malatya merkezilik algoritmasının kullanımıyla haber metinlerinin özetlenmesini sağlar. Graf tabanlı bu yaklaşım, metnin içeriğini temsil eden önemli kelimeleri belirler. Bu önemli kelimeler üzerinden metnin özeti oluşturulur.

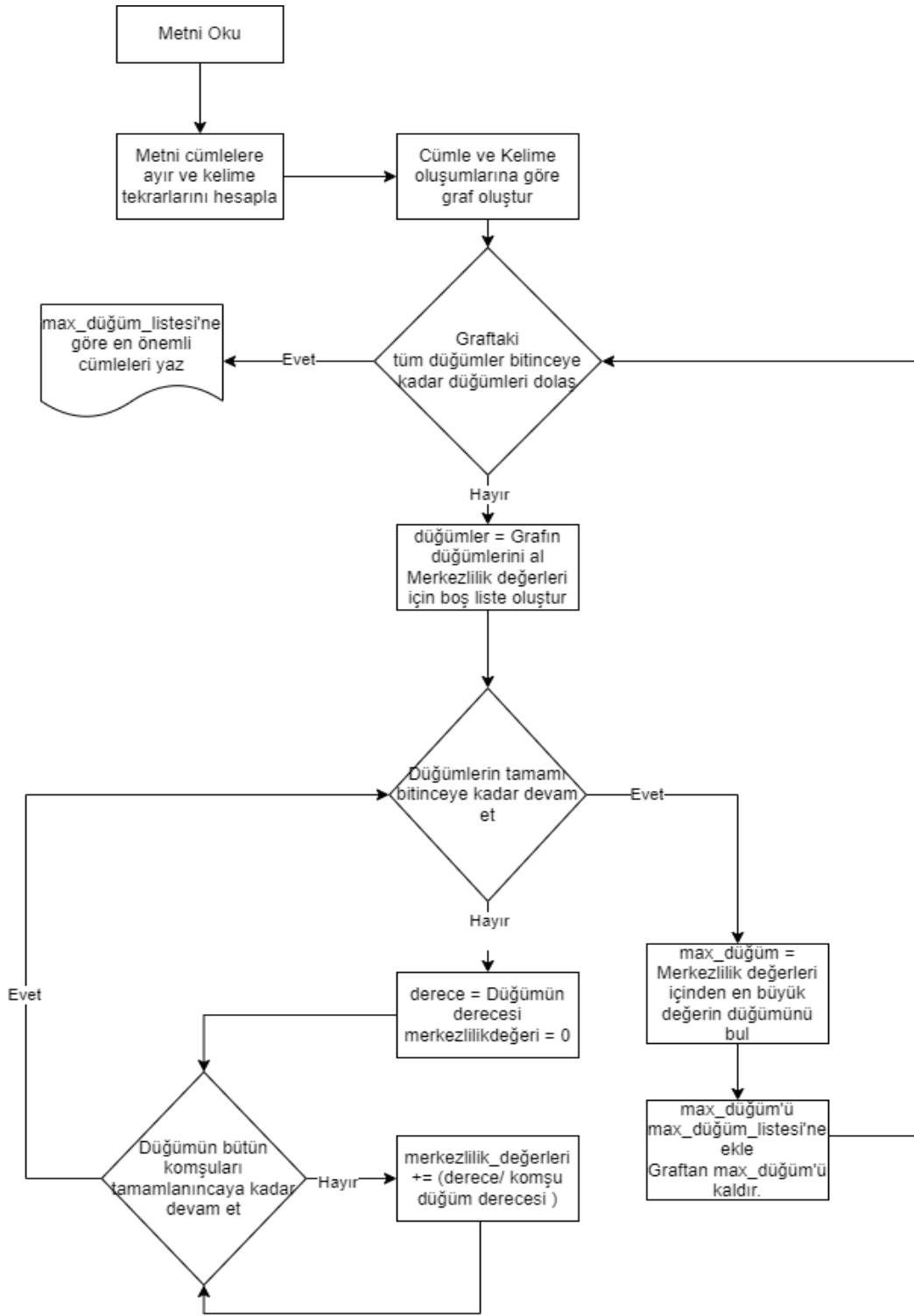
Şekil-1'de, metinde yer alacak kelimelerin belirlenmesi sürecini gösteren algoritmanın akış şeması bulunmaktadır.

Bu şemada, öncelikle durak kelimelerinin belirlenmesi ve bunların graftan çıkarılması yer alır. Graf oluşturulduktan sonra her cümle graf üzerine yerleştirilir. Malatya merkezilik algoritması uygulanarak en etkili düğüm seçilir. Her aşamada, en etkili düğüm graftan çıkarılır ve yeni graftaki en etkili düğüm bulunur. Bu işlem düğümler bitinceye kadar tekrarlanır. Ardından, bu belirlenen en etkili düğümler birleştirilerek özetlenmiş metin oluşturulur. Şekil-1 de gösterilen şema, yeni yöntemin ayrıntılı adımlarını gösterir. Ayrıca metin özetlemede Malatya merkezilik algoritmasının kullanımını açıklar.

Önerilen algoritmaya ait sözde kodlar verildi. Bu kodda, başlangıçta orijinal metinden tanımlayıcı unsurların çıkarılması ve kelime köklerinin belirlenmesi gibi çeşitli ön işlemler uygulanır. Bu ön işlemler sonucunda metin, bir graf yapısına dönüştürülür. Daha sonra, bu graf yapısındaki düğümlerin Malatya merkezilik değerleri, Malatya merkezilik algoritması kullanılarak hesaplanır. Bu değerler doğrultusunda, metin özetini oluşturacak özetler belirlenir. Seçilen özetler graf yapısından çıkarılır. Ardından, oluşan yeni graf yapısı için merkezilik değeri hesaplanarak seçim işlemleri sürdürülür. Metin özeti elde edildiğinde merkezilik değerlerinin hesaplanması ve seçim işlemleri tamamlanır.

Önerilen algoritmanın sözde kodu

```
Fonksiyon: MerkezilikHesapla(graf)
düğümler = GrafDüğümünüAl(graf)
merkezilik_değerleri = []
Her bir düğüm için:
    derece = GrafDerecesiniAl(graf, düğüm)
    merkezilik_değeri = 0
    Her bir komşu için:
        merkezilik_değeri = merkezilik_değeri +
(derece/GrafDerecesiniAl(graf, komşu))
    merkezilik_değeri merkezilik_değerleri'ne ekle
merkezilik_değerlerini döndür
Fonksiyon: MaksMerkezilikDeğeriBul(graf)
merkezilik_değerleri = MerkezilikHesapla(graf)
maks_indeks = MaksIndexiBul(merkezilik_değerleri)
maks_düğüm = GrafDüğümünüIndexleAl(graf, maks_indeks)
GrafDüğümüKaldır(graf, maks_düğüm)
maks_düğümü döndür
Fonksiyon: MaxDüğümListesiHesapla(graf)
Graftaki tüm düğümler bitinceye kadar:
    maks_düğüm = MaksMerkezilikDeğeriBul(graf)
    max_düğüm'ü listeye ekle
    maks_düğüm_listesini sırala ve döndür
Fonksiyon: GrafOluştur(cümleler, kelime_oluşumları)
durak_kelimeler = DurakKelimeleriniAl()
graf = BoşGrafOluştur()
Her bir cümle için:
    GrafDüğümüEkle(graf, cümle)
Her bir kelime, oluşum_sayısı içinde:
    Eğer KelimeCümledeVar(cümle, kelime) ve
KelimeDurakKelimesiDeğil(kelime):
    GrafKenarıEkle(graf, cümle, kelime, ağırlık=oluşum_sayısı)
graf'ı döndür
Fonksiyon: RougeHesapla(referans, özet)
RougeMetrikleriniHesapla ve döndür
Fonksiyon: Main()
metin = MetniOku()
cümleler = CümleleriAyr(metin)
kelime_oluşumları = KelimeOluşumlarınıSay(metin)
graf = GrafOluştur(cümleler, kelime_oluşumları)
maks_düğüm_listesi = MaxDüğümListesiHesapla(graf)
etkili_cümleler = CümleleriIndexle(cümleler,
min_max_düğüm_listesi)
referans_metin = "Örnek Referans Metin"
skorlar = RougeHesapla(referans_metin, etkili_cümleler)
```



Şekil 1. Önerilen Yöntemin Akış Şeması

4. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

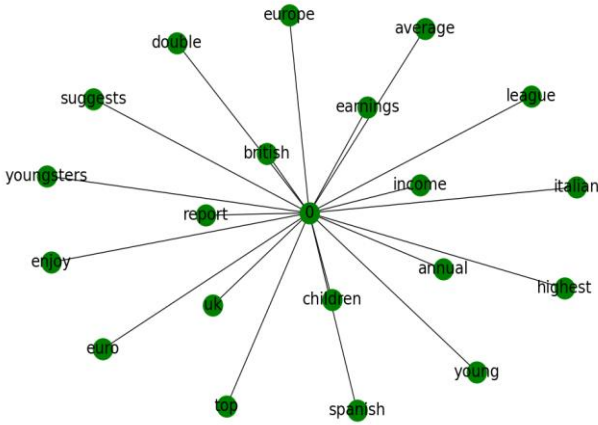
Bu çalışmada, metin özetleme sürecinde Malatya merkezilik algoritmasının etkinliği değerlendirildi. Bu amaçla, toplamda 2224 haber metni, Textrank, Lexrank, Pagerank ve önerilen yöntem aracılığıyla özetlendi. Elde edilen sonuçlar, rouge değerleri üzerinden objektif bir değerlendirmeye tabi tutuldu.

Malatya merkezilik algoritması, bu çalışmanın odak noktasını oluşturan temel algoritma olarak önerilen yöntem kullanıldı. Malatya merkezilik algoritmasının özetleme sürecindeki başarı oranları, metinlerin en etkili noktalarını belirleme konusundaki yeteneğiyle ön plana çıkmaktadır.

Rouge değerleri incelendiğinde, Malatya merkezilik algoritmasının Textrank, Lexrank ve Pagerank gibi diğer yaygın kullanılan algoritmalarla kıyaslandığında genel olarak daha yüksek performans sergilediği belirlendi. Bu durum, Malatya merkezilik algoritmasının haber metinlerini özetleme sürecinde sağladığı avantajları vurgular. Textrank, Lexrank ve Pagerank algoritmaları da önerilen yöntemdeki gibi graf tabanlı algoritmalar olduğu için bu algoritmalar ile karşılaştırıldı. Bu algoritmalar yaygın olarak birçok çalışmada kullanılmış, diğer yöntemler ile karşılaştırılmıştır.

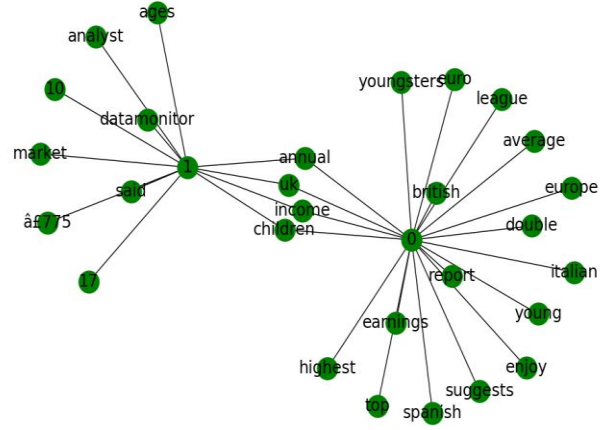
Önerilen yöntem, Malatya merkezilik algoritmasını içerdiği için diğer yöntemlere göre daha başarılı bir özetleme sağlandı. Araştırmanın bulguları, metin özetleme alanında Malatya merkezilik algoritmasının etkili bir çözüm sunabileceği yönünde önemli bir perspektif sunmaktadır.

Bu çalışma, Malatya merkezilik algoritmasının metin özetleme sürecindeki etkinliğini vurgular ve bu algoritmanın gelecekteki çalışmalarda daha fazla kullanılmasını teşvik eder.



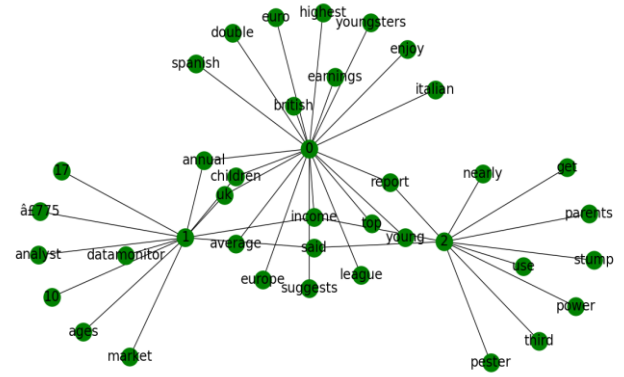
Şekil 2. İlk Cümledeki Graf Yapısı

Şekil-2 İlk cümledeki her kelimenin 1.düğüme eklendiğini gösteren bir örnek.

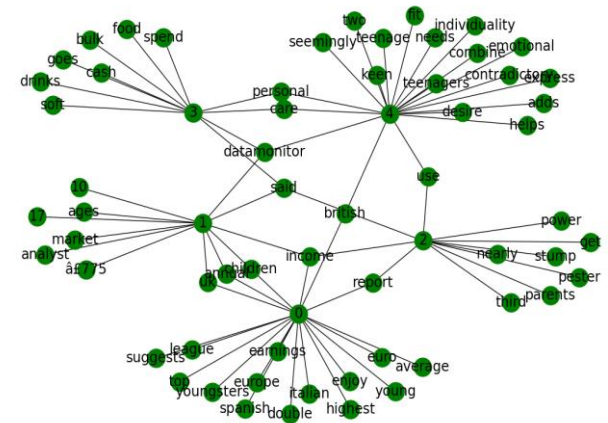


Şekil 3. İkinci cümle eklendikten sonraki graf yapısını gösterir.

Şekil 3'de ikinci cümledeki paylaşımın graf'a eklenmesiyle ilk cümledeki paylaşılan düğümler belirlenerek dahil edilmektedir. Bu süreç cümleler arasındaki ilişkiyi ortaya çıkarır.



Şekil 4. Üçüncü Cümle Eklendikten Sonra Graf Yapısı.



Şekil 5. Beşinci Cümle Eklendikten Sonra Graf Yapısı.

Şekil-5 'te grafikteki 5. cümleye ulaştıkça cümleler arasındaki ilişkilerin daha da arttığı gözlemlenebilir. İlerledikçe daha sık kullanılan kelimeler ortaya çıkacak ve bu da metnin içinden en etkili cümlelerin seçilmesini sağlayacaktır.

Aynı örnek metin, metin özetlemede kullanılan diğer algoritmalarla test edildi. Önerilen yöntemle karşılaştırılan algoritmalar TextRank, LexRank ve PageRank algoritmalarıdır. Bu algoritmalarından elde edilen sonuçların rouge puanları, önerilen yöntemden elde edilen sonuçlarla karşılaştırıldı. Karşılaştırmalar Tablo-1'de Rouge-1 metriği, Tablo-2'de Rouge-2 metriği ve Tablo-3'te Rouge-1 metriği temel alınarak yapıldı.

Tablo 1. ROUGE-1 Metriğine Dayalı Metin Özetleme için Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

Özetleme Yöntemleri	Rouge-1		
	Recall	Precision	F-Score
TextRank	0,3286	0,7946	0,4526
LexRank	0,276	0,6919	0,3846
PageRank	0,3514	0,7606	0,4694
Önerilen Yöntem	0,6088	0,4265	0,4869

Tablo-1'de ROUGE-1 metriklerine dayalı karşılaştırmaya göre TextRank, LexRank ve PageRank algoritmaları tarafından uygulanan özetleme yöntemleri, bir önerilen yöntemle karşılaştırıldı. Rouge-1, özet metin ile referans metin arasındaki tek kelime uzunluğundaki parçacıkların benzerliğini değerlendirir. Bu puanlar kelime düzeyindeki benzerliği ölçer.

Elde edilen sonuçlara göre, her bir özetleme yönteminin Recall (hatırlama) değeri, özetin gerçek metindeki önemli bilgilerin yüzdesini ifade eder. TextRank, LexRank ve PageRank sırasıyla %32,86, %27,6 ve %35,14 Recall değerine sahipken, önerilen yöntem bu değeri %60,88 olarak göstermektedir. Bu, önerilen yöntemin daha fazla önemli bilgiyi özetle koruduğunu ortaya çıkarır.

Precision, özetin içinde bulunan bilgilerin ne kadarının gerçekten önemli olduğunu ölçer. Özetin ne kadarının gerçekten gerekli ve doğru bilgilerle doldurulduğunu gösterir. Yüksek bir Precision değeri, özetin genellikle

gereksiz bilgiler içermediğini ve sadece gerekli olanların bulunduğunu gösterir. Yapılan çalışmada Rouge-1 metriğine göre Precision değerleri; TextRank, LexRank ve PageRank sırasıyla %79,46, %69,19 ve %76,06 Precision değerine sahipken, önerilen yöntem bu değeri %42,65 olarak göstermektedir. Rouge-1 metriğine göre Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özette gereksiz kelimelerin de var olduğu sonucunu çıkarmaktadır.

F-Score değeri, Recall ve Precision'ın harmonik ortalamasıdır. Bu değer, özetlemenin genel başarısını yansıtır. TextRank, LexRank ve PageRank sırasıyla %45,26, %38,46 ve %46,94 F-Score'a sahipken, önerilen yöntem bu değeri %48,69 olarak göstermektedir. Bulgular, önerilen yöntemin diğerlerine kıyasla daha yüksek bir Recall ancak daha düşük bir Precision ile birlikte daha dengeli bir F-Score sağladığını ortaya koymaktadır.

Tablo 2. ROUGE-2 Metriğine Göre Metin Özetlemede Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

Özetleme Yöntemleri	Rouge-2		
	Recall	Precision	F-Score
TextRank	0,2514	0,6914	0,3566
LexRank	0,1835	0,551	0,2667
PageRank	0,2626	0,6561	0,3639
Önerilen Yöntem	0,469	0,296	0,3513

Tablo-2'de ROUGE-2 metriklerine dayalı bir karşılaştırma gerçekleştirilmiştir. TextRank, LexRank ve PageRank algoritmaları ile uygulanan özetleme yöntemleri, önerilen yöntemle kıyaslandı.

Rouge-2, iki kelime uzunluğundaki parçacıkların benzerliğini ölçer. Rouge-2 metriğine göre Recall (hatırlama) değeri, metnin orijinalinde bulunan önemli iki kelimelik öbeklerin yüzdesini ifade eder. TextRank, LexRank ve PageRank sırasıyla %25,14, %18,35 ve %26,26 Recall değerine sahipken, önerilen yöntem bu değeri %46,9 olarak göstermektedir. Bu sonuç Rouge-2 metriğine göre önemli bilgilerin özet içinde daha fazla bulunduğunu gösterir.

Precision (kesinlik) değeri ise özetin doğruluğunu ölçer. TextRank, LexRank ve PageRank sırasıyla %69,14, %55,1 ve %65,61 Precision değerine sahipken, önerilen yöntem bu değeri %29,6 olarak göstermektedir. Bu sonuç özette gerekli olmayan bilgilerin olabileceğini gösterir.

F-Score değeri, Recall ve Precision'ın harmonik ortalamasıdır ve bu değer, özetlemenin genel başarısını yansıtır. TextRank, LexRank ve PageRank sırasıyla %35,66, %26,67 ve %36,39 F-Score'a sahipken, önerilen yöntem bu değeri %35,13 olarak göstermektedir. Bu sonuca göre de önerilen yöntemin Rouge-2 metriğinin ortalama olarak iyi olduğu sonucu ortaya çıkar.

Tablo 3. ROUGE-L Metriğine Göre Metin Özetlemede Önerilen Yöntemin TextRank, LexRank ve PageRank Algoritmaları ile Karşılaştırılması

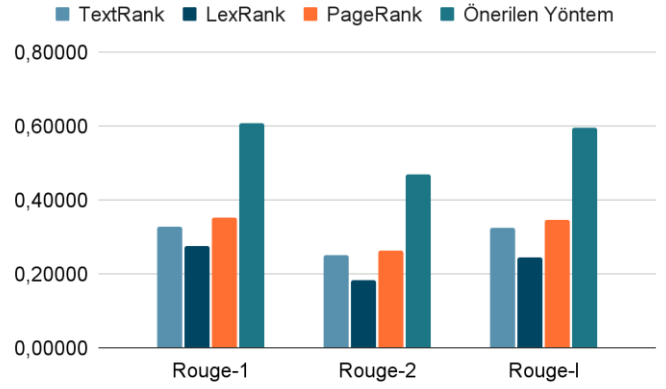
Özetleme Yöntemleri	Rouge-l		
	Recall	Precision	F-Score
TextRank	0,323	0,7811	0,4448
LexRank	0,2445	0,6638	0,3687
PageRank	0,3446	0,7462	0,4605
Önerilen Yöntem	0,5953	0,4141	0,4744

Tablo-3'de ROUGE-L metriklerine dayalı bir karşılaştırma gerçekleştirildi. TextRank, LexRank ve PageRank algoritmaları tarafından uygulanan özetleme yöntemleri, bir önerilen yöntemle karşılaştırılmıştır. Bulgulara göre, her bir özetleme yönteminin ROUGE-L metrikleri altında değerlendirilmiştir. Rouge-1, tüm özet ve referans metinleri arasındaki en uzun paylaşılan alt dizinin benzerliğini ölçen En Uzun Ortak Alt Dizi (LCS) tabanlı benzerlik ölçüsünü kullanır.

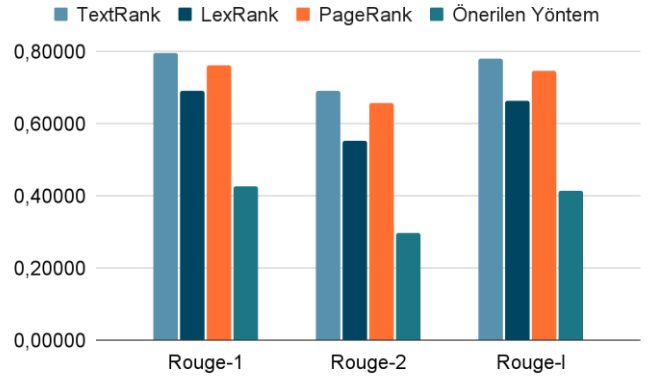
Recall (hatırlama) değeri, TextRank, LexRank ve PageRank sırasıyla %32,3, %24,45 ve %34,46 Recall değerine sahipken, önerilen yöntem bu değeri %59,53 olarak göstermektedir. Bu, önerilen yöntemin daha fazla önemli bilgiyi özetle koruduğunu gösterir.

Precision (kesinlik) değeri, TextRank, LexRank ve PageRank sırasıyla %78,11, %66,38 ve %74,62 Precision değerine sahiptirken, önerilen yöntem bu değeri %41,41 olarak göstermektedir. Rouge-1 metriğine göre Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özetle gereksiz kelimelerin bulunabileceği sonucu ortaya çıkar.

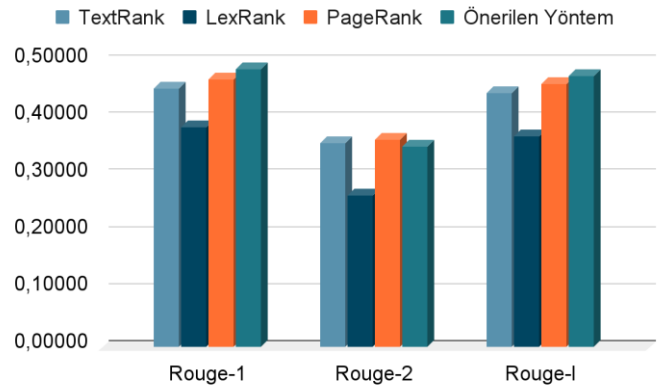
F-Score değeri, TextRank, LexRank ve PageRank sırasıyla %44,48, %36,87 ve %46,05 F-Score'a sahipken, önerilen yöntem bu değeri %47,44 olarak göstermektedir. F-Score değeri hesaplanırken hem Recall hem de Precision kullanıldığı için bu sonuca göre Rouge-1 metriğinin diğer yöntemlere kıyasla daha iyi olduğu sonucuna varılır.



Şekil 6. ROUGE Metriklerine Göre Recall Değerlerinin Karşılaştırılması



Şekil 7. ROUGE Metriklerine Göre Precision Değerlerinin Karşılaştırılması



Şekil 8. ROUGE Metriklerine Göre F-Skor Değerlerinin Karşılaştırılması

5. SONUÇLAR (CONCLUSIONS)

Bu çalışmada, Malatya merkezilik algoritmasının graf tabanlı bir yaklaşım kullanarak İngilizce haber metinleri ile etkili bir şekilde özetleme işlemi yapılabildiği gösterildi. Algoritma, metnin içeriğini temsil eden önemli kelimeleri belirler. Bu kelimeler üzerinden metnin özetini oluşturur. Bu graf tabanlı yaklaşım, metin özetleme alanında önemli bir ilerleme sağlamaktadır.

Malatya merkezilik algoritmasının etkinliğini değerlendirmek amacıyla bu çalışma, geniş bir veri kümesi olan bbcnews veri seti üzerinde gerçekleştirilmiştir. Toplamda 2224 haber metni kullanılarak yapılan testlerle algoritmanın performansı incelendi. Bu veri seti, farklı konulardaki zengin ve çeşitli haber içeriklerini içermektedir. Bu da Malatya merkezilik algoritmasının genel uygulanabilirliğini değerlendirmek için uygun bir ortam sağlar. Algoritma, kelime düğümleri arasındaki bağlantıları ve kelimelerin önemini belirlemek üzere Malatya merkezilik skorlarını kullanır. Bu skorlar, metin içindeki kelimenin belirginliğini gösterir. Yüksek skorlara sahip kelimeler, metnin ana fikrini ve kritik noktalarını yansıtan önemli kelimeler olarak seçilir. Seçilen kelimelerin içerdiği cümleler birleştirilerek özet metin oluşturulur.

Sonuçlar değerlendirildiğinde, Rouge-1 metriğinde Recall ve f-skore değerlerinde iyi olduğu fakat Precision değeri karşılaştırılan diğer algoritmalara göre biraz daha düşük çıktığı için özet de gereksiz kelimelerin de var olduğunu gösteriyor. Algoritma Rouge-1 de genellikle başarılı olsa da bu konuda biraz daha geliştirilmesi gerekir. Rouge-2 metriğine göre Precision ve f-skore değerlerinde diğer algoritmalara kıyasla yeteri kadar başarılı olamadığı görüldü. Bu metrik, özetlenen metinle referans metin arasındaki aynı iki kelimenin yan yana gelme oranının diğer algoritmalara göre biraz daha düşük olduğunu gösterir. Rouge-1 metriğine göre algoritmanın genel olarak başarılı olduğu görüldü. Rouge-1 de en uzun ortak alt-dizi benzerliği diğer algoritmalara göre daha başarılı sonuç verdiği anlaşıldı.

Bu çalışma ile geliştirilmesi öngörülen kısımlar geliştirilerek önerilen yöntemin daha iyi sonuçlar verebileceği ortaya koyuldu. Malatya merkezilik algoritmasının bu çalışmadaki başarıları, metin özetleme tekniklerinin geliştirilmesine önemli bir katkı sunacaktır. Bu yöntem, gelecekteki araştırmalar için bir temel oluşturabilir ve haber metinlerinin daha etkili bir şekilde özetlenmesi için yeni stratejilere yol açabilir.

KAYNAKLAR (REFERENCES)

- [1] Yakut, S., Öztemiz, F., & Karci, A. "A New Approach Based on Centrality Value in Solving the Minimum Vertex Cover Problem: Malatya Centrality Algorithm." *Computer Science*. Volume Vol:7, Issue Issue:2, 81 - 88. (07.12.2022)
- [2] Bakan, C. & Yakut, S. "Development of Text Summarization Method based on Graph Theory and Malatya Centrality Algorithm." *Computer Science*. Volume: IDAP-2023. (18.10.2023)
- [3] Yakut, S., Öztemiz, F., & Karci, A. "A New Approach Based on Centrality Value in Solving the Maximum Independent Set Problem: Malatya Centrality Algorithm." *Computer Science*, Vol:8(Issue:1), 16-23. (2023)
- [4] Tülek, M. "Türkçe için Metin Özetleme." (2007)
- [5] Khushboo S. Thakkar, R.V. Dharaskar, & M.B. Chandak. "Graph-Based Algorithms for Text Summarization." *IEEE*. 10.1109/ICETET.2010.104. (2010)
- [6] Güneş Erkan, & Dragomir R. Radev. "LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization." (2004)
- [7] Ibrahim F. Moawad, & Mostafa Aref. "Semantic Graph Reduction Approach for Abstractive Text Summarization." *IEEE*. 10.1109/ICCES.2012.6408498. (2013)
- [8] Rafael Ferreira, Frederico Freitas, Luciano de Souza Cabral, Rafael Dueire Lins, Rinaldo Lima, Gabriel França, Steven J. Simske, & Luciano Favaro. "A Four Dimension Graph Model for Automatic Text Summarization." *IEEE*. 10.1109/WI-IAT.2013.55. (2013)
- [9] Chirantana Mallick, Ajit Kumar Das, Madhurima Dutta, Asit Kumar Das, & Apurba Sarkar. "Graph-Based Text Summarization Using Modified TextRank." *Springer*. (2018)
- [10] Yogesh Sankarasubramaniam, Krishnan Ramanathan, & Subhankar Ghosh. "Text Summarization using Wikipedia." *ScienceDirect*. (2014)
- [11] Rasim Alguliev, & Ramiz Aliguliyev. "Evolutionary Algorithm for Extractive Text Summarization." *Scientific Research*. (2009)
- [12] Naresh Kumar Nagwani, & Dr. Shrish Verma. "A Frequent Term and Semantic Similarity based Single Document Text Summarization Algorithm." *ResearchGate*. (2011)
- [13] Rada Mihalcea. "Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization." (2004)
- [14] Sumya Akter, Aysa Siddika Asa, Md. Palash Uddin, Md. Delowar Hossain, Shikhor Kumer Roy, & Masud Ibn Afjal. "An Extractive Text Summarization Technique for Bengali Document(s) Using K-means Clustering Algorithm." *IEEE*. 10.1109/ICIVPR.2017.7890883. (2017)
- [15] S.A. Babar, & Pallavi D. Patil. "Improving Performance of Text Summarization." *ScienceDirect*. (2015)
- [16] Makbule Gulcin Ozsoy, & Ferda Nur Alpaslan. "Text Summarization using Latent Semantic Analysis." 10.1177/0165551511408848. (2011)
- [17] Erhandi, B. "Derin Öğrenme ile Metin Özetleme." (2020)
- [18] Joel Larocca Neto, Alex A. Freitas, & Celso A. A. Kaestner. "Automatic Text Summarization Using a Machine Learning Approach." *Springer*. (2003)
- [19] Carlos N. Silla Jr., Gisele L. Pappa, Alex A. Freitas, & Celso A. A. Kaestner. "Automatic Text Summarization with Genetic Algorithm-Based Attribute Selection." *Springer*. (2004)
- [20] O. Kaynar, Y. E. Işık, Y. Görmez, ve F. Demirkoparan, "Otomatik Metin Özetleme için Genetik Algoritma Tabanlı Cümle Çıkarımı", *Yönetim Bilişim Sistemleri Dergisi*, c. 3, sy. 2, ss. 62–75, 2017.
- [21] Raed Z. Al-Abdallah, & Ahmad T. Al-Taani. "Arabic Single-Document Text Summarization Using Particle Swarm Optimization Algorithm." *ScienceDirect*. (2017)

- [22] Abdullah Ammar Karcioğlu & Ahmet Cahit Yaşa. "Automatic Summary Extraction in Texts Using Genetic Algorithms." IEEE. (2021)
- [23] Enise Karakoç, Burcu Yılmaz. "Deep Learning Based Abstractive Turkish News Summarization." IEEE. (2019)
- [24] Hakan Torun, A. Burak İner. "A Method for Similarity Detection in Vector Space by Summarizing News Articles." IEEE. (2022)
- [25] Yavuz Selim Kartal, Mucahid Kutlu. "Machine Learning Based Text Summarization for Turkish News." IEEE. (2021)
- [26] Hark, C., Taner Uçkan, T., Seyyarer, E., & Karcı, A. "Metin Özetlemesi için Düğüm Merkezliklerine Dayalı Denetimsiz Bir Yaklaşım." dergipark, 8(3). (30.09.2019)
- [27] Mengqian Wang, Manhua Wang, Fei Yu, Yue Yang, Jennifer Walker, Javed Mostafa. "A Systematic Review of Automatic Text Summarization for Biomedical Literature and EHRs." *Journal of the American Medical Informatics Association*, Volume 28, Issue 10, October. (2021)
- [28] Kumar, Y., Kaur, K. & Kaur, S. "Study of Automatic Text Summarization Approaches in Different Languages." *Artif Intell Rev* 54, 5897–5929. (2021)
- [29] Taner Uçkan, Cengiz Hark & Ali Karcı. "SSC: Clustering of Turkish Texts by Spectral Graph Partitioning." *Politeknik Dergisi*. (2021)
- [30] Aydın, A., & Uçkan, T. "Çizgeler Üzerinde Farklı Ağırlıklandırma Yöntemleri Ve Merkezilik Ölçütleri İle Çıkarımsal Metin Özetleme." *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 35(1), 71-79. (2023)
- [31] Kuş, A., & Acı, Ç. İ. "An Extractive Text Summarization Model for Generating Extended Abstracts of Medical Papers in Turkish." *Bilgisayar Bilimleri Ve Teknolojileri Dergisi*, 4(1), 19-26. (2023)

Hitit Çivi Yazılı Tabletlerin Yapay Zeka Algoritmaları ile Okunması*

Araştırma Makalesi/Research Article

 Baran BİNGÖL¹,  Gülgüney MASALCI ŞAHİN²,  Özlem SİR GAVAZ²,  Savaş TAKAN¹

¹ Ankara Üniversitesi, Mühendislik Fakültesi, Yapay Zeka ve Veri Mühendisliği, Ankara, Türkiye

² Ankara Üniversitesi, Dil ve Tarih-Coğrafya Fakültesi, Eskiçağ Dilleri ve Kültürleri Bölümü, Ankara, Türkiye

22290007@ogrenci.ankara.edu.tr, gmasalci@ankara.edu.tr, gavaz@ankara.edu.tr, stakan@ankara.edu.tr

(Geliş/Received:09.02.2024; Kabul/Accepted:08.06.2024)

DOI: 10.17671/gazibtd.1434452

Özet— Çivi yazısı gibi özel yazı biçimlerinin yapay zeka ile okunması, görüntü işleme teknolojisinin hızla gelişmesiyle son yıllarda önemli bir çalışma alanı haline gelmiştir. Çeşitli dillerde yazılmış tabletler üzerinde gerçekleştirilmiş pek çok önemli literatür bulunmaktadır. Buna karşın, her ne kadar dünyada farklı merkezlerde 3D modelleme ve dijitalleştirme çalışmaları başlamış olsa da yapay zeka kullanılarak okunan çivi yazılı diller arasında Hititçe dili bulunmamaktadır. Literatürdeki bu boşluğu doldurmak, Hititologların tablet okuma hızlarını artırarak bilimsel çalışmalarını daha derinlemesine analiz etmelerine olanak tanımak ve manuel yöntemlerle okunan tabletlerin elle temasını azaltarak binlerce yıl öncesinden günümüze ulaşan tabletlerin tahrip olma durumunu en aza indirmek amacıyla çalışmamızda, Hitit çivi yazısı üzerinde transfer öğrenme yöntemi ile geliştirilmiş derin öğrenme modelleri denenmiştir. Bunun sonucunda model topluluğu yöntemiyle %89 doğruluk elde edilmiştir. Çalışmamızın, Hitit çivi yazısı uzmanlarının tabletler üzerindeki çalışma sürelerini kısaltması ve daha verimli işleyecek bir bilimsel sürece katkı sunması beklenmektedir.

Anahtar Kelimeler— hitit tabletleri, yapay zeka, derin öğrenme, model topluluğu, görüntü işleme

Hittite Cuneiform Tablets with Artificial Intelligence Algorithms * (İngilizce)

Abstract—Artificial intelligence reading of special scripts such as cuneiform has become an important field of study in recent years with the rapid development of image processing technology. There are many important literatures on tablets written in various languages. However, although 3D modelling and digitisation studies have started in different centres around the world, there is no Hittite language among the cuneiform languages read using artificial intelligence. In order to fill this gap in the literature, to enable Hittitologists to analyse their scientific studies in more depth by increasing the tablet reading speed, and to minimise the destruction of the tablets that have survived thousands of years ago by reducing the manual contact of the tablets read by manual methods, deep learning models developed with transfer learning method were tested on Hittite cuneiform in our study. As a result, 89% accuracy was obtained with Ensembling method. Our study is expected to shorten the working time of Hittite cuneiform experts on tablets and contribute to a more efficient scientific process.

Keywords—hittite tablets, artificial intelligence, deep learning, model ensembling, image processing

1. GİRİŞ (INTRODUCTION)

Görüntü işleme teknolojisinin hızla gelişmesi, birçok uygulama alanına yeni olanaklar sunmaktadır. Bu alanlardan biri de karakter tanıma ve nesne tespiti çalışmalarıdır. Bu çalışmalar çivi yazısı gibi özel yazı biçimlerinin yapay zeka ile okunması konusunda büyük potansiyel taşımaktadır [1], [2]. Söz konusu potansiyelden yola çıkılan çalışmamızda, Hitit çivi yazısının, görüntü işleme teknikleriyle yapay zeka ile okunmasına odaklanılmıştır. Hititçe tabletler üzerinde dünyada farklı merkezlerde 3D modelleme ve dijitalleştirme çalışmaları başlamış olmasına karşın yapay zeka kullanılarak okunan çivi yazılı diller (Sümerce, Akadça, Eblaca, Babilce) arasında Hititçe'nin bulunmaması, bu çalışmanın özgün yönünü oluşturmaktadır.

Hitit çivi yazılı tabletler, Hititlerin başkenti Boğazkale/Hattuşa başta olmak üzere Çorum-Ortaköy / Şapinuwa, Samsun Vezirköprü- Oymağaç / Nerik, Sivas-Kuşaklı / Şarişsa, Sivas-Kayalıpınar / Şamuha, Kırıkkale-Karakeçili Büklükale, Çorum-Alacahöyük, Eskişehir, Yozgat-Sorgun Uşaklı Höyük, Kilis Oylum Höyük gibi Anadolu'nun farklı bölgelerinde ve Türkiye dışındaki arkeolojik kazılarda elde edilmiştir [3, 5]. Hitit çivi yazılı tabletlerinde, Hitit dili dışında Palaca, Luvice, yerli halkın dili olan Hattice, Hurrice, Sümerce, Akadça ve Kikkullu yönetmeliğinde geçen Sanskritçe öğelerle birlikte toplam sekiz farklı dil yer almaktadır. Bunlara ilaveten Hititlerin başkenti Hattuşa'da 2023 kazı sezonunda açığa çıkarılan çift dilli bir tablette (KBo 71.145: CTH 470) Hititçe ile birlikte kaydedilmiş yeni bir Anadolu dili olan Kalaşmaca tespit edilmiştir [4]. Bu keşif, kadim Anadolu dillerine bir yenisinin eklenmesine ve dünyanın gözünü Hititçe tabletlere çevirmesine neden olmuştur. Hitit çivi yazılı metinler incelendiğinde; hem hece işaretlerinin fazla olması, hem farklı kâtipler tarafından yazılmış olmaları hem de aynı tablet üzerinde yukarıda bahsi geçen farklı dillerde kelimelerin olması tabletlerin çözülmesini zorlaştırmaktadır. Ayrıca kil üzerine yazılan yaklaşık üç bin beş yüz yıllık tabletler, günümüze maalesef kırık, yanmış ya da fazla tahrip olmuş şekilde ulaşmıştır. Bu da hece işaretlerinin güçlüğüle okunması ya da okunamamasına sebep olmakta, tabletlerin transliterasyon ve tercümelerini önemli ölçüde zorlaştırmaktadır.

Bu metinler bir ölü dile ait olması sebebiyle, günümüzde kullanılmayan hece ve karakterler içerir ve bu nedenle Hitit çivi yazısıyla yazılmış bir tableti okumak Hititoloji konusunda eğitilmiş bir insan için çokça zaman ve emek gerektirmektedir. Öte yandan bu yazı sistemini çözümlmek, dilbilimciler ve eskiçağ tarihi uzmanları için hayati önem arz eder. Zira, günümüzden yaklaşık üç bin beş yüz yıl önce Anadolu coğrafyasına hükmetmiş ve büyük bir medeniyet kurmuş olan Hititlerin siyasi tarihteki rolleri, sosyal ve kültürel yaşamları, dini yaşamları, dönemin devletleri ile olan diplomatik ilişkileri ve günümüze aktardıkları kültürel öğeler ile ilgili en önemli belgeler Hitit arşivinde bulunmaktadır. Öyle ki okunacak ve yorumlanacak her tablet Hitit tarihinin, dolayısıyla Eski Anadolu tarihinin aydınlatılmasında son derece önemlidir. Ayrıca Ankara Anadolu Medeniyetleri Müzesi başta olmak

üzere, İstanbul ve Çorum Müzelerinde okunmayı ve yayınlanmayı bekleyen binlerce tablet bulunmaktadır [5]. Tüm bu bilgilerden yola çıkarak, çalışmanın amacı, yapay zeka tabanlı bir yöntem ile Hitit çivi yazısı tabletlerinin otomatik olarak transliterasyonunu yapabilmek ve bu sayede Hititologların tabletleri çözümü ve yayınlama hızlarının artırılması ile daha fazla yayınlanmamış tabletin bilim camiasına kazandırılmasını sağlamaktır. Ayrıca manuel yöntemlerle okunan tabletlerin elle temasını azaltarak binlerce yıl öncesinden günümüze ulaşan tabletlerin tahrip olma durumunu da en aza indirmek amaçlanmaktadır.

Çalışmada bu doğrultuda, Hitit çivi yazısı üzerinde derin öğrenme tabanlı görüntü işleme tekniklerinin kullanılması incelenmiş, Hititçe veya başka ölü dillerdeki çivi yazısı sistemlerini kullanmak isteyen araştırmacılar için çözüm önerileri sunulmuştur. Çalışmamızın, çivi yazısı araştırmalarını desteklemek için önemli bir adım olacağı düşünülmektedir. Öte yandan bu çalışmanın, Hitit çivi yazısı uzmanlarının tabletler üzerindeki çalışma sürelerini kısaltması ve daha verimli işleyecek bir bilimsel sürece katkı sunması da öncül ve beklenen hedefler arasında bulunmaktadır.

2. İLGİLİ ÇALIŞMALAR (RELATED WORK)

Derin öğrenme kullanan bildiğimiz ilk harf tanıma çalışmasında LeCun vd. [6], el yazısıyla yazılmış posta kodlarını tanımak için geri yayılım (backpropagation) algoritmasını kullanan bir sinir ağı mimarisini incelemiştir. Ağ, ABD Posta Servisi'nin verileri üzerinde test edilmiş ve el yazısı rakamlarını yüksek doğrulukla sınıflandırabilmiştir. Ağ tasarımında yenilikçi özellikler, özellikle ağırlık paylaşımı ve özel mimari katkıları vurgulanmıştır. Bulgular, geri yayılım (backpropagation) yönteminin el yazısı tanıma görevlerinde pratik ve etkili olduğunu göstermiştir [6].

Williams vd. [7], çivi yazısı işaretlerini tanımlamak ve sınıflandırmak için DeepScribe adında bir yapay zeka sistemi geliştirmiştir. Sistem, Ahameniş İmparatorluğu dönemine ait kil tabletler üzerindeki yazıtları analiz etmek için kullanılmış ve bu işaretleri doğru bir şekilde yerleştirebilme ve tanımlayabilme kabiliyetine sahip olduğunu göstermiştir. Sistem, bu tabletlerin fotoğrafları üzerinde yüksek doğrulukta işaret önerileri sunarak araştırmacılara transkripsiyon konusunda yardımcı olmayı amaçlamıştır [7].

Stötzner, Homburg ve Mara'nın [8] çalışmasında, çivi yazısını dijital olarak işleyen araçlar geliştirilmiştir. Çivi yazısı, günümüzden yak. 5. Bin yıl önce icat edilmiş, çok geniş bir coğrafyada ve en az sekiz farklı dilde kullanılmış 3D bir yazı tipidir. Araştırma, 500 etiketlenmiş tablet içeren HeiCuBeDa ve MaiCuBeDa veri setlerini kullanmış ve fotoğraflarla 3D renderlar arasında etiketlemeyi aktarabilen bir araç sunmuştur. Çivi yazısı karakterlerinin yerlerini tahmin etmek için RepPoints dedektörü kullanılmıştır. 3D render edilmiş görüntülerin, fotoğraflar

üzerindeki karakter tespitinde daha iyi sonuç verdiği bulunmuştur [8].

Diğer taraftan, Homburg vd. [9], çivi yazılı kil tabletlerin 3D taramalarını ve bu taramalar üzerindeki 3D notasyonları içeren bir veri seti tanımlamıştır. Çalışmada, iki farklı tarihi döneme ait tabletlerin, özellikle arkeoloji ve dil bilimi çalışmaları için kullanılabilecek biçimde işlenmesi ve yorumlanması üzerine odaklanılmıştır. 3D notasyonların, karakter tanıma ve dil çalışmalarına katkı sağlamak için makine öğrenimi görevlerinde kullanılabileceği öne sürülmüştür. Veri seti, farklı yazılımlarla kullanım için çeşitli formatlarda yayımlanmış ve çivi yazılı araştırmalarda 3D verinin modern işlenişine dair öneriler sunulmuştur [9].

Tel Aviv Üniversitesi'nde Gai Gutherz isimli bir araştırmacının yüksek lisans tezi için bir tez projesi olarak başlayan proje çalışmasının son adımı, 2023 Mayıs ayında Oxford University Press'in PNAS Nexus dergisinde, Akadca'dan İngilizce'ye nöral makine çevirisini anlatan bir (Translating Akkadian to English with Neural Machine translation) araştırma makalesi olarak yayımlanmıştır. Google Translate, Baidu translate ve diğer çeviri motorları tarafından da kullanılan nöral makine çevirisi, kelimeleri bir sayı dizisine dönüştürerek çalışan ve başka bir dildeki bir cümleyi kelimesi kelimesine çevirmekten daha doğru ve dilin doğal yapısına uygun bir sinir ağı adı verilen karmaşık bir matematiksel formül kullanılmıştır [10].

Horst Kremers'in editörlüğünde 2020 yılında yayımlanan Digital Cultural Heritage adlı kitapta, Charles University (Prague) ekibi tarafından (Petr Zemánek, Jana Mynářová, Petra Štefcová & Jaroslav Valach) yayımlanan 'Virtual Collection of Cuneiform Tablets as a Complex Multilevel System with Interdisciplinary Content' adlı çalışmada, sanal bir çivi yazısı tablet koleksiyonunun mimarisi ele alınmıştır. Kültepe kazılarında elde edilen yaklaşık 400 Eski Asurca tableti içeren koleksiyon, çivi yazısıyla yazılmış ek bir metin bilgisi katmanına sahip özel bir koleksiyonun dijitalleştirilmesi, eserlerin dijital modellerini (3D modeller, stereometrik ve standart fotoğraflar, Structure-from-Motion) ve aynı zamanda tek tek nesnelere ilişkin ek verileri (boyutlar, renk, şekiller, kompozisyon gibi fiziksel özellikler) içeren çalışmada, çivi yazılı metin kısmı da bir corpus olarak dilbilimcilerin kullanımına sunulmuş ve filolojik veriler bazı önemli nitelikler (kişiler, mallar, bağlantılar) için bir araya getirilmiştir. Koleksiyondan elde edilen veriler, başka verilerle ilişkilendirilerek veritabanı kültürel ve tarihsel gelişim bağlamına yerleştirilmiştir [11].

Dijital dokümantasyona dair bir diğer önemli çalışma ise 'Cuneiform Tablets Micro-Surveying in an Optimized Photogrammetric Configuration' yani Çivi Yazılı Tabletlerin Optimize Edilmiş Fotogrametrik Konfigürasyonda Mikro-Ölçülmesi adlı çalışmadır. Böylelikle çivi yazılı tabletlerin üzerindeki her bir işaret mikroskobik ölçümleri ile alınmakta, okunmakta ve arşivlenebilmektedir. Università degli Studi di Salerno, Politecnico di Milano ve Department of Languages and Cultures, Ghent University bölümlerinden araştırmacılara

göre, mikrometrik ölçme sistemi, çivi yazılı tabletlerin sayısallaştırılması ve belgelenmesi için verimli ve güvenilir çözümler sunmaktadır [12].

Tyndall'ın [13] çalışması, Hititçe dilinde yazılmış çivi yazılı tablet parçalarını daha büyük metinlere otomatik olarak toplamayı amaçlayan bir yaklaşımı sunmaktadır. Çalışma, parçalı (fragman tabletler) ve hasar görmüş metinlerin sınıflandırılmasında bilinen metin sınıflandırma metriklerinin ve Hititçe çivi yazılı metinlerin yapısına dair bazı bilgilerin kullanılmasını önermektedir. Bu sayede, çivi yazılı tablet parçalarını, var olan tam metinlerin altında sınıflandırmak mümkün olabilir. Çalışmada, Naif Bayes (Naive Bayes) ve Maksimum Entropi (Maximum Entropy) sınıflandırıcıları kullanılarak bu parçalar analiz edilmiş ve çivi yazısındaki Sumer ve Akad ideogramları kullanılarak sınıflandırma performansı artırılmıştır [13].

He vd. [14] derin sinir ağlarını eğitmenin zorluğunu gidermek için artık(residual) öğrenme çerçevesini tanıtmıştır. Yazarlar, katmanları, girdilere göre artık(residual) fonksiyonlar olarak yeniden formüle ederek ağların daha derin olmasına rağmen daha kolay optimize edilmesini sağlamıştır. 152 katmana kadar derinlikteki artık(residual) ağlar, düşük karmaşıklıkla ImageNet'te yüksek doğruluk elde etmiş ve çeşitli yarışmalarda birincilik kazanmıştır [14].

Tan ve Le'nin [15] çalışmasında, ConvNets'in performansını artırmak için derinlik, genişlik ve çözünürlüğün eşit oranda ölçeklendirilmesine dayanan yeni bir yöntem önerilmiştir. Tasarlanan EfficientNets, ImageNet ve diğer veri setlerinde daha az parametreye ile mevcut modellere göre daha yüksek doğruluk ve verimlilik sağlanmıştır [15].

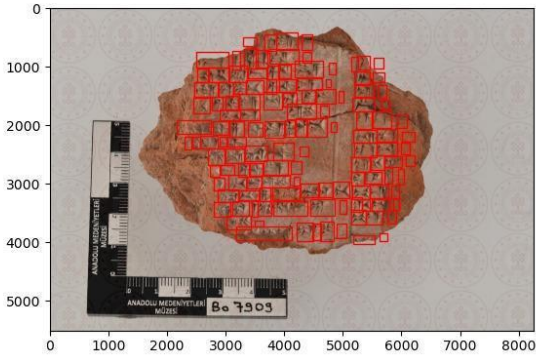
Yapay zeka kullanılarak tablet okuma konusunda önemli çalışmalar yapan bir diğer merkez ise Chicago Oriental Institute ile yine Chicago Üniversitesi Bilgisayar Bilimleri Bölümü'dür (The Department of Computer Science). 1990'lardan bu yana 3D çizim konusunda ve okuma konusunda bilişim teknolojilerinden faydalanan araştırmacılar, tablet okumada %80 başarıya ulaştıklarını ancak tekrar eden fiiller ve kalıplar dışında, yorum gerektiren kelime ve ifadelerde başarı oranının düştüğünü dile getirmişlerdir [16].

Mevcut literatürde, çeşitli dillerde yazılmış çivi yazılı tabletlerin yapay zeka ile transliterasyonuna dair çalışmalar bulunmaktadır. Bu çalışmalardan farklı olarak bizim çalışmamız, Hitit çivi yazılı tabletler üzerinde gerçekleştirilmiştir. Öte yandan Başkent Üniversitesi Bilgisayar Mühendisliği Bölümü'nde Hitit Çiviyazısının Bilgisayar Desteği ile okunması ile ilgili yürütülen çalışmalar sonucunda bir teknik rapor ve yüksek lisans tezleri tamamlanmıştır. Çalışmalarda Hitit tabletlerinden bilgisayarlı görü uygulamalarının çalışma ilkeleri ayrıntılı olarak açıklanarak, algoritmalar dilin işaretlerine uygulanmıştır ve Hitit işaretlerine karşılık gelen sözcüklerin dilbilgisi yapısı üzerinde durulmuştur. [17, 18, 19] Ayrıca Tyndall [13], Hititçe tabletler üzerinde çalışmış

ancak kırık tabletleri birleştirmek üzerine bir çalışma yürütmüştür. Bunu yaparken de sınıflandırma algoritmalarından yararlanmışır. Bundan farklı olarak bizim çalışmamızda, Hitit tabletlerinin transliterasyonu gerçekleştirilmiştir.

3. METOT (METHOD)

Yapay zeka tabanlı bir yöntem ile Hitit çivi yazısı tabletlerinin otomatik olarak transliterasyonunu yapabilmek amacıyla çalışmamızda, öncelikle Hititçe tabletlerin yüksek çözünürlüklü fotoğrafları çekilmiş ve dijital ortama aktarılmıştır. Daha sonra, geliştirdiğimiz web sayfası aracılığıyla tabletler, Hititologlar tarafından etiketlenmiştir. Her bir heceyi içine alan bir dörtgen çizilerek etiketleme gerçekleştirilmiştir. Şekil 1’de buna örnek gösterilmiştir.



Şekil 1. Hititçe işaretlerin dörtgen etiketleme yöntemi (Quadrilateral labelling of Hittite signs)

Etiketleme sonucunda, fotoğraflar parçalara ayrılmış, aynı hecelerin farklı fotoğrafları gruplandırılmıştır. Ancak Hitit işaret listesinde bazı hecelerin görünümü aynı olduğu için, bu işaretler projede bir işaret olarak kabul edilmiştir. Bunlara örnek bazı işaretler Tablo 1’de verilmiştir.

Tablo 1 Gösterimi aynı olan işaretler

DINGİR	<i>an</i>
<i>at</i>	AD
<i>ag</i>	ak
<i>bu</i>	pu
<i>pé</i>	pi
<i>har</i>	HUR
GIŞ	ez
GIŞ	iz
<i>ke</i>	ki
<i>ši</i>	LIM
<i>lu</i>	UDU

Bu işaretlere örnek görseller Şekil 2’de verilmiştir. Burada Hitit çivi yazılı metinlerde geçen Sümerce DINGİR ve Hititçe *an* işaretlerinin aynı olduğunu görülmektedir.



“DINGİR” işareti



“an” işareti

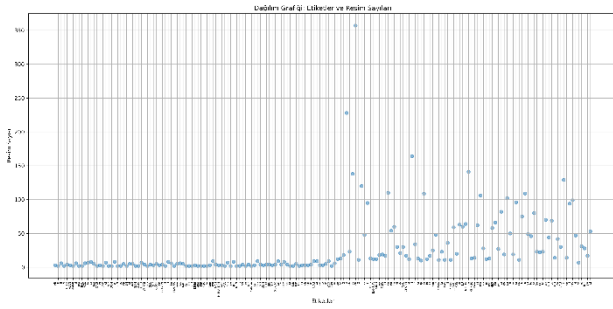
Şekil 2. DINGİR ve an işaretlerinin görünümü (View of DINGİR and *an* signs)

Etiketleme sonuçlarına göre Hititçe işaretlerin fotoğrafları incelendiğinde, veri setimizdeki işaretlerin fotoğraflarının sayıca dağılımının dengesiz olduğu gözlenmiştir. Bu durum Şekil 3’te gösterilmektedir.



Şekil 3. Hititçe işaretlerin elde edilen veri setindeki dağılımı (Distribution of Hittite signs in the obtained data set)

Ayrıca veri setinin daha detaylı bir grafiği Şekil 4’te verilmiştir. Bu grafikte x ekseninde etiketler ve isimleri görülmektedir. Y ekseninde ise bu etiketlerin sahip olduğu toplam resim sayısı verilmiştir. Bu grafik bir saçılım grafiğidir ve sahip olunan veri setinin dağılımı hakkında detaylı bir bakış sunmaktadır.



Şekil 4. Veri setinin saçılım grafiği
(Scatter plot of Hittite signs in the obtained data set)

Veri setindeki dengesizliği ortadan kaldırmak için veri artırma(data augmentation) yöntemi kullanılmıştır. Veri artırma yöntemleri seçilirken uygulanacak olan hecenin anlamının değişmemesi dikkate alınmıştır.

Veri setini iyileştirmek için çeşitli yöntemler denenmiştir. Öncelikle eğitim verileri her bir grup 400x400 piksel'e yeniden ölçeklendirilmiştir. Sonrasında her bir gruptan rastgele 360x360 piksellik bir bölge alınmıştır. Böylece modelin daha iyi genelleşebilmesi amaçlanmıştır. Alınan her 360x360 piksellik resimlere %30 olasılıkla 5 derece döndürme uygulanmıştır. %30 olasılıkla rastgele afin

dönüşümü(random affine transformation) yöntemi uygulanmıştır. En son işlem olarak resimler ImageNet istatistikleriyle normalize edilmiştir çünkü kullandığımız her model imagenet üzerinde ön eğitilmiştir. Test verileri için, fotoğraflar öncelikle 400x400 piksel boyutlarına indirilmiştir. Sonrasında, 360x360 piksel kalacak şekilde resmin orta yerinden kırılmış ve imagenet istatistikleriyle normalize edilmiştir.

Veri setinin ResNet101 modelinde veri artırma yapılmamış ve yapılmış değerleri Tablo 2'de gösterilmiştir.

ResNet101 modeli karşılaştırırken kullanılan hiperparametreler aynıdır, dolayısıyla metrikler arasındaki fark sadece veri artırma etkisiyle oluşmuştur. Buradan da anlaşılacağı üzere, veri artırma yöntemi derin öğrenme modelinin performansını önemli ölçüde iyileştirmiştir.

İşaretlerin, 3930 tanesi eğitim ve 860 tanesi test olmak üzere ikiye ayrılmıştır. Hititçe işaret sayısı 1 olan fotoğraflar, modeli eğitirken eğitim ve test diye ayırlamayacağı için, bu harfler veri setinden çıkarılmıştır. Buna ek olarak, Tablo 1'de verilen aynı gösterime sahip işaretler de aynı sayıldıktan sonra, toplamda 181 farklı etikete sahip bir veri seti elde edilmiştir.

Tablo 2 Veri setinin ResNet101 modelinde veri artırma yapılmamış ve yapılmış değerleri

MODELLER	Eğitim Yitimi (Train Loss)	Geçerleme Yitimi (Validation Loss)	Doğruluk (Accuracy)	İlk-3-Doğruluk (Top-3-Accuracy)	İlk-5-Doğruluk (Top-5-Accuracy)	Kesinlik (Precision)	Duyarlılık (Recall)	F1-Skoru (F1-Score)
ResNet101 181 Etiket (veri artırma yokken)	0.01	0.72	%84.3	%93.7	%95.6	0.811	0.843	0.826
ResNet101 181 Etiket (veri artırma varken)	0.05	0.64	%85.7	%93.5	%96.1	0.836	0.857	0.846

3.1 Kullanılan modellerin açıklamaları (Descriptions of the models used):

ResNet: Her bir katman kendinden önceki katmanın hem çıkışını hem de girişini alır. Bu yapı sayesinde katmanlar arası bilgi akışı artar ve patlayan gradyan sorunu çözünür.

DenseNet: Her bir katman kendisinden önceki tüm katmanların çıkışını giriş olarak alır. Bu yoğun bağlantılar modelin daha iyi öğrenmesini sağlar.

EfficientNet: Modelin derinliği ve genişliği ölçek katsayılarıyla dengelendiğinden, daha hızlı ve daha iyi performansla çalışan evrişimli sinir ağlarından oluşur.

ConvNext: Geleneksel evrişimli sinir ağlarının yanı sıra özel yöntemlerle tasarlanmış yeni katmanlar içerir. Bu katmanlar daha etkili özellik çıkarır ve modele daha iyi öğrenme imkanı sağlar.

Model Topluluğu (Model Ensembling) Yöntemi: Farklı modellerin çıktılarını birleştirerek daha güçlü ve genelleşmiş bir model elde edilmesini sağlar. Bu yöntem

genellikle başarıyı artırır fakat daha fazla hesaplama istediğinden dolayı daha fazla zaman harcar.

Oluşturduğumuz modellerde yaptığımız deneyler sonucunda, derin öğrenme modellerinin önceden eğitilmesi (ImageNet üzerinde), veri setimiz üzerinde daha iyi sonuç vermiş ve daha hızlı yakınsamasını sağlamıştır. Bu yüzden oluşturduğumuz modellerinin hepsi imagenet ağırlıklarıyla başlatılarak eğitilmiştir.

Oluşturduğumuz model tablosunda bazı modellerin diğer modellere göre daha büyük olmasından dolayı daha iyi genelleşmeleri için küçük modellere uyguladığımızdan daha fazla düzenleme (regularization) gerektirmiştir. Daha fazla düzenleme gerektirmediği yitim fonksiyonu(loss function) ve iyileştirici(optimizer) için etiket düzleştirme (label smoothing) ve L2 cezası(L2 penalty) kullanılmıştır. Model, Tablo 3'te * işaretiyle belirtilmiştir.

Tablo 3 Modellerin Performans Değerleri

Model İsimleri	Eğitim Yitimi (Train Loss)	Geçerleme Yitimi (Validation Loss)	Doğruluk (Accuracy)	İlk-3-Doğruluk (Top-3-Accuracy)	İlk-5-Doğruluk (Top-5-Accuracy)	Keskinlik (Precision)	Duyarlılık (Recall)	F1-Skoru (F1-Score)	Model Parametre Sayısı
ResNet50	0.03	0.70	% 84.4	%93.2	%96.0	0.816	0.844	0.830	23,878,901
ResNet101	0.05	0.64	% 85.7	%93.5	%96.1	0.836	0.857	0.846	42,871,029
ResNet152	0.01	0.64	% 86.9	%94.4	%96.1	0.836	0.868	0.852	58,514,677
ResNet50-101-152 topluluğu (ensemble)	-	-	% 87.1	%95.5	%96.5	0.843	0.871	0.856	-
DenseNet201	0.01	0.59	% 87.6	%94.3	%96.4	0.848	0.875	0.861	18,440,629
EfficientNetB5	0.01	0.77	% 86.1	%93.3	%94.7	0.833	0.860	0.846	28,711,653
ConvNext-base	0.01	0.59	% 87.7	%95.5	%97.4	0.847	0.877	0.862	87,751,989
ConvNext-large	0.02	0.61	% 85.6	%95.1	%96.9	0.827	0.856	0.841	196,508,533
ConvNext-large*	0.48	0.94	% 87.9	%95.6	%97.7	0.853	0.879	0.865	196,508,533
En iyi 4 model topluluğu (ensemble)	-	-	% 89.0	%96.0	%97.3	0.861	0.889	0.875	-

4. SONUÇLAR (CONCLUSION)

Yapay zeka tabanlı bir yöntem ile Hitit çivi yazısı tabletlerinin otomatik olarak transliterasyonunu yapabilmek ve bu sayede Hititologların tablet okuma hızlarını artırarak, bilimsel çalışmalarını daha derinlemesine analiz etmelerine olanak tanımak amacıyla gerçekleştirilen çalışmada, derin öğrenme tabanlı görüntü işleme tekniklerinin kullanılması incelenmiş, Hititçe veya başka ölü dillerdeki çivi yazısı sistemlerini kullanmak isteyen araştırmacılar için çözüm önerileri sunulmuştur.

Çalışma kapsamında öncelikle Hititçe'de aynı gösterime sahip işaretler tek işaret olarak kabul edilmiştir. Veri setimizde sadece 1 adet bulunan heceler, test ve eğitim olarak ayrılmayacağı için veri setinden çıkarılmıştır. Veri setinin dengesizlik problemini çözmek için veri artırma uygulanmıştır. Bu uygulamalardan sonra ortaya çıkan veri seti ile ResNet, DenseNet, EfficientNet, ConvNext, model topluluğu(model ensemble) yöntemleri ile modeller oluşturulmuştur. Çıkan sonuçlara göre, model topluluğu yöntemi, %89 ile en yüksek başarıyı göstermiştir. Modellerin başarımlarının farklarının arkasındaki nedeninin model dizayn yapısı ve içerdikleri parametre sayısı olduğu düşünülmektedir. Ortaya çıkan ilginç bir sonuç ise DenseNet201 yapısının karşılaştırılan modeller arasında en az parametre sayısına sahip olmasına rağmen en iyi sonuç veren algoritmalarından birisi olmasıdır. DenseNet yapısında her katmanın ileri yönde tüm diğer seviyelere yoğun bir şekilde bağlandığı bilinmektedir bu sayede sonraki tüm katmanların resimlerin daha düşük seviye özelliklerden yararlanmasını sağlar. DenseNet bu yapısı gereği Hititçe işaretleri sınıflandırma görevinde iyi sonuç verdiği düşünülmektedir. Ayrıca tabloda görülmektedir ki EfficientNet modeli ResNet101 modelinden %34 daha az parametre sayısına sahip olmasına rağmen doğruluk oranı olarak daha iyi sonuç vermiştir ayrıca F1-skorları aynıdır. Bunun sebebi ise EfficientNet modelinin derinliğinin ve genişliğinin ölçek

katsayılarıyla dengelenmesidir dolayısıyla bu model daha az parametreyle daha iyi sonuç vermiştir.

Elimizdeki veri, 4790 adet işaret ile sınırlıdır. Sonraki çalışmalarda, daha fazla işaret ile çalışıldığında başarımlarının artması beklenmektedir. Ayrıca etiketleme yöntemlerinde de geliştirilme yapılması planlanmaktadır. Mevcut literatürde, çeşitli dillerde yazılmış çivi yazılı tabletlerin yapay zeka ile transliterasyonuna dair çalışmalar bulunmaktadır. Bu çalışmalardan farklı olarak bizim çalışmamız, Hitit çivi yazılı tabletler üzerinde gerçekleştirilmiştir. Çalışmamızın, çivi yazısı araştırmalarını desteklemek için önemli bir adım olacağı düşünülmektedir. Öte yandan bu çalışmanın, Hitit çivi yazısı uzmanlarının tabletler üzerindeki çalışma sürelerini kısaltması ve daha verimli işleyecek bir bilimsel sürece katkı sunması da beklenen etkiler arasındadır.

TEŞEKKÜR (Acknowledgement)

Bu makale, Ankara Üniversitesi bünyesinde yürütülen 3219 Proje ID'li, SBA-2024-3219 Proje kodlu, ADEP projesi kapsamında yayınlanmıştır.



KAYNAKLAR (REFERENCES)

- [1] Collins, D. J., Ma, Z., Han, J., & Ai, Y. (2017). Continuous micro-vortex-based nanoparticle manipulation via focused surface acoustic waves. *Lab on a Chip*, 17(1), 91-103.
- [2] Homburg, T., Zwick, R., Mara, H., & Bruhn, K. C. (2022). Annotated 3D-models of cuneiform tablets. *J. Open Archaeol. Data*, 10, 1-8.
- [3] Erkut, S., (1997). Çiviyazılı Hitit Tabletleri, *Bellekten*, C. LXI, Aralık, S.232, Ankara 1997, 495-498.

- [4] İnternet: Schwemer D., (2023-10-29) Thesaurus Linguarum Hethaeorum digitalis, hethiter.net/: TLHdig KBo 71.145 (2021-12-31).
- [5] Özkan, S. (2017). Türkiye Müzelerinde Korunan Çivi Yazılı Tabletler, Prof.Dr. Recep Yıldırım'a Armağan, Ed. Pınar Pınarcık, Bilcan Gökce, Mehmet Salih Erkek, Sena Coşğun Kandal, Bilgin Kültür Sanat Yayınları, Ankara 2017, 173-193.
- [6] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541-551.
- [7] Williams, E. C., Su, G., Schloen, S. R., Prosser, M. C., Paulus, S., & Krishnan, S. (2023). DeepScribe: Localization and Classification of Elamite Cuneiform Signs Via Deep Learning. arXiv preprint arXiv:2306.01268.
- [8] Stötzner, E., Homburg, T., & Mara, H. (2023). CNN based Cuneiform Sign Detection Learned from Annotated 3D Renderings and Mapped Photographs with Illumination Augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1680-1688).
- [9] Homburg, T., Zwick, R., Bruhn, K. C., & Mara, H. (2022). 3D data derivatives of the Haft Tappah processing pipeline. *Cuneiform Digital Library Journal*, 2022(1).
- [10] Gutherz – Gordin – Sáenz – Levy – Berant, (2023), “Translating Akkadian to English with neural machine Translation”, *PNAS Nexus* 2, 1–10.
- [11] Zemánek – Mynářová – Štefcová, – Valach, “Virtual Collection of Cuneiform Tablets as a Complex Multilevel System with Interdisciplinary Content” in. *Digital Cultural Heritage* (ed. Horst) p. 183– 194, 2020.
- [12] Antinozzi – Fiorillo – Surdi, (2022), “Cuneiform Tablets Micro-Surveying in an Optimized Photogrammetric Configuration”, *HERITAGE*, 5 (4), p.3133-3164.
- [13] Tyndall, S. (2012, July). Toward automatically assembling Hittite-language cuneiform tablet fragments into larger texts. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 243-247).
- [14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [15] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [16] İnternet: Ancient Language Processing: Teaching Computers to Read Cuneiform Tablets ,
<https://datascience.uchicago.edu/news/ancient-language-processing-teaching-computers-to-read-cuneiform-tablets/>, 18.05.2024
- [17] Aktaş – Aşuroğlu (2016) Hitit Çiviyazılarının Bilgisayar Desteği ile Okunması Ve Çevirisi İle İlgili El Kitabı, Başkent Üniversitesi Bilgisayar Mühendisliği Bölümü Teknik Yayın, Temmuz 2016.
- [18] Aşuroğlu, T. (2015), **Hitit Çiviyazısı İşaretlerinin Bilgisayar Desteği ile Okunması ve Veri Madenciliği Uygulama Örnekleri**, Yüksek Lisans Tezi, Başkent Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı, Ankara 2015.
- [19] Yeşiltepe, M.B. (2015) Hitit Çiviyazılı İşaretlerin Okunmasında Uzman Sistem Uygulama Örnekleri, Başkent Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı, Ankara 2015.

Üretken Yapay Zekaya Dayalı Bireysel Emeklilik Bilgilendirme ve Öneri Sistemi

Araştırma Makalesi/Research Article

 Ezgi AVCI^{1*},  Mehmet Furkan ATİK²,  Nur Muazzez MEMİŞ³

¹Department of Applied Data Science, TED University, Ankara, TURKEY

²Department of Digital Product Management, Strategy and System Development, Pension Monitoring Center, İstanbul, Türkiye

³Department of Advanced Analytics and Research Services, Pension Monitoring Center, İstanbul, Türkiye

ezgi.avci@tedu.edu.tr, furkan.atik@egm.com.tr, muazzez.memis@egm.org.tr

(Geliş/Received:29.05.2024; Kabul/Accepted:25.06.2024)

DOI: 10.17671/gazibtd.1475239

Özet— Bu makale, üretken yapay zeka (GenAI) ile güçlendirilmiş yenilikçi bir bireysel emeklilik bilgi ve tavsiye sisteminin tasarımını sunmaktadır. Sistem, kullanıcı verilerini analiz etmek ve kişiselleştirilmiş emeklilik planlama tavsiyeleri üretmek için gelişmiş AI algoritmalarını kullanacak şekilde özelleştirilmiştir. GenAI entegrasyonu ile sistem, kullanıcılar arasında finansal okuryazarlığı önemli ölçüde artırmayı, emeklilik planlaması ve finansal ürünler hakkında daha derin bir anlayış sağlamayı hedeflemektedir. GenAI destekli içgörüler, kullanıcıların uzun vadeli emeklilik hedefleri ve risk tercihleriyle uyumlu bilinçli kararlar alabilmelerini sağlayacak şekilde özelleştirilmiş yatırım stratejilerini kolaylaştıracaktır. Bu yaklaşım, sadece bireysel finansal sonuçları iyileştirmeyi amaçlamakla kalmayıp, geleneksel olarak yalnızca finansal danışmanlar aracılığıyla erişilebilen finansal tavsiyeye erişimi demokratikleştirmeyi de hedeflemektedir. Sistem geliştikçe, değişen ekonomik koşullara ve kişisel durumlara uyum sağlaması, kullanıcıların yaşam değişiklikleriyle uyumlu dinamik tavsiyeler sunması beklenmektedir. Bu sistemin amacı, emekliliğe yaklaşırken ve emekliliğe girerken kullanıcılarının finansal refahını ve güvenliğini artıracak şekilde proaktif bir emeklilik planlaması yaklaşımını teşvik etmektir.

Anahtar Kelimeler— bireysel emeklilik, üretken yapay zeka, yapay zeka, karar destek sistemi

GenAI-Based Private Pension Information and Recommendation System

Abstract— This paper presents the design of an innovative private pension information and recommendation system powered by generative artificial intelligence (GenAI). The system is tailored to leverage GenAI algorithms to analyze user data and generate personalized retirement planning advice. By integrating GenAI, the system seeks to significantly enhance financial literacy among users, providing them with a deeper understanding of retirement planning and financial products. The GenAI-driven insights will facilitate tailored investment strategies, enabling users to make informed decisions that align with their long-term retirement goals and risk preferences. This approach aims not only to improve individual financial outcomes but also to democratize access to financial advice that is traditionally available only through financial advisors. As the system evolves, it is expected to adapt to changing economic conditions and personal circumstances, offering dynamic advice that keeps pace with users' life changes. The goal of this system is to foster a proactive approach to retirement planning, thereby enhancing the financial well-being and security of its users as they approach and enter retirement.

Keywords— private pension, generative artificial intelligence, artificial intelligence, decision support system

1. INTRODUCTION

The importance of private pension planning cannot be overstated in today's financial landscape. In an era marked by rapid economic changes and increasing life expectancy, effective retirement planning has become crucial for ensuring long-term financial security. Retirement planning involves understanding and managing various financial elements, including savings, investments, pensions, and other income streams, to achieve a financially secure retirement. However, financial literacy—the ability to understand and effectively use various financial skills, including personal financial management, budgeting, and investing—is a critical component that influences the effectiveness of individual retirement planning. Financial literacy is not merely about understanding basic financial concepts; it's about applying that knowledge to make informed decisions about saving, investing, and managing money, which ultimately affects one's quality of life in retirement. Unfortunately, many individuals lack the financial literacy required to navigate the complex landscape of retirement options effectively. Studies have shown that a significant portion of the population lacks the knowledge necessary to make decisions that optimize their retirement outcomes. This lack of understanding can lead to poor financial choices, such as inadequate savings, inappropriate investments, and a general lack of preparation for retirement.

In response to these challenges, the emergence of generative artificial intelligence (GenAI) in financial services presents a transformative opportunity to enhance decision-making and improve individual outcomes. GenAI, particularly models like Generative Pre-trained Transformer (GPT), has revolutionized various sectors by providing advanced data processing capabilities and generating human-like text based on the input they receive. In the context of financial services, GenAI can analyze vast amounts of financial data, recognize patterns, and provide personalized advice based on an individual's financial status, goals, and risk tolerance. The application of GenAI in financial decision-making extends beyond mere automation. It involves understanding complex financial scenarios, adapting to new financial information, and evolving with the user's changing financial conditions and life stages. For retirement planning, this means that GenAI can assist users in crafting a comprehensive, dynamic financial plan that adjusts over time, providing recommendations on savings rates, investment choices, and other financial decisions that align with the user's retirement goals and financial situation.

The primary objective of the proposed GenAI-driven system is to develop a tool that harnesses the power of GenAI to deliver personalized retirement planning advice directly to end-users, empowering them with tailored financial insights and strategies. While the system primarily targets individual users seeking to enhance their retirement planning, it also offers valuable support for financial advisors and intermediaries by providing them with advanced tools to better serve their clients. This

system aims to bridge the gap between complex financial knowledge and the average individual's understanding. By integrating GenAI, the system can process detailed personal financial information provided by users, including their current financial status, retirement goals, risk tolerance, and other relevant data. Utilizing advanced algorithms, the system can generate customized advice that guides users through the myriad options available, suggesting strategies to increase their retirement savings, optimize their investment portfolio, and manage risks associated with their financial decisions. Moreover, the GenAI system is designed to be interactive, allowing users to receive explanations and clarifications on financial concepts and the reasoning behind specific advice. This educational component is crucial as it helps to enhance the user's financial literacy over time, empowering them to make informed decisions independently. The system's ability to learn from user interactions and feedback also enables it to improve its advice, ensuring that it remains relevant as economic conditions change and as the user's personal financial situation evolves.

2. BACKGROUND and LITERATURE REVIEW

2.1. Rationale to enter the Individual Pension System (IPS)

2.1.1. A Comfortable Retirement

The IPS has been created as a supplement to the social security system to increase one's wealth by providing them with an additional income during retirement. One can participate in the IPS by acquiring a pension product from their choice of pension company. Eligibility for retirement requires to have been stayed in the system for at least 10 years and attained the age of 56. There are three different options to be chosen from after becoming eligible for retirement. (1) One can simply request a lump-sum payment of some or all the accumulations in the individual pension and state contribution account (2) One choose to stay in the IPS and receive accumulations in part from the pension company according to the reimbursement plan they decide. (3) One can earn regular income by purchasing an annuity with some or all the accumulations.

2.1.2. Transparency

One can access their daily individual pension account information via two different platforms: (1) from the BES Mobile application, provided by the Pension Monitoring Center (EGM) (2) from the customer portal on one's pension company's website (3) from the mobile app provided by the pension company. Since the individual pension accounts are kept by the Istanbul Settlement and Custody Bank (TakasBank), participants can also keep track of their daily fund shares from TakasBank's accounts.

2.1.3. Right to Choose the Fund

The pension company offers participants a pension plan suitable for their pension expectations, income level and

age. It is up to the participants to decide on the funds in which they wish to invest their contributions. If one does not choose any funds, their contributions shall be invested in the funds stipulated by the legislation. Also, Pension Fund Trading Platform (BEFAS) enables participants to purchase funds from different companies without requiring them to change their pension companies. Fund distribution can be changed 12 times per year.

2.1.4. *State Contribution*

A state contribution is paid into the participant's account in the amount of up to 30 percent of the contributions they made to the IPS. A participant can benefit from the state contribution for the contribution payments up to the total amount of the gross minimum wage determined for the relevant year within a calendar year. Contribution payments exceeding the limit will be considered in the calculation of the state contribution as carryover contribution shares transferred to the company accounts on the first day of each year. Separate state contribution limits have been set for the contracts within the scope of Auto Enrollment System (AES) and IPS, respectively.

2.1.5. *Professional Fund Management*

Participant's accumulations are managed by the portfolio management companies overseen by the Capital Markets Board of Turkey (SPK).

2.1.6. *Purchasing Service from Licensed Intermediaries*

Only licensed intermediaries can engage in the IPS promotion, marketing and sales activities. The professional competency of these intermediaries is tested with an exam conducted by the Pension Monitoring Center.

2.1.7. *Flexible Contribution Payments*

In addition to the regular contributions, participants can also make initial or additional contribution payments in interim periods to the IPS; they can also suspend contributing anytime.

2.1.8. *Transfer to Another Pension Company*

If one wishes, savings can be transferred to another pension company at least two years after the effective date of the contract. The waiting period is one year for subsequent transfers.

2.1.9. *Tax Advantage for Employers*

Employers can enter employees into the IPS by drawing up a group pension contract. In this case, the contributions paid on behalf of employees are recorded as expenses when calculating business income without associating them with the salary. Employers may determine the vesting period for the accumulations arising from contributions paid for the

employees and from their earnings on the condition of not exceeding seven years. This is a good way to achieve employee motivation and loyalty to the company.

2.1.10. *Reliability*

All elements of the system are under supervision. The operations of all the pension companies are supervised by the Insurance and Private Pension Regulation and Supervision Agency (SEDDK) while the operations of the funds, portfolio managers and custodians are supervised by the SPK. The company's accounts and transactions are subject to independent external audits at least once a year. EGM monitors all pension companies' operations electronically daily. Assets that belong to pension companies and to pension mutual funds are separate from each other. Fund assets cannot be pledged, used as collateral except for portfolio-related transactions, seized, or included in bankruptcy assets. SEDDK may impose sanctions against a pension company if the financial structure of the pension company is deemed to have been weakened to a level that may put the participants' rights and interests in danger and if it is impossible to remedy the situation. In the event of the death of the participant during the contract period, their accumulations and all of the state contributions in their account shall be paid to the beneficiaries stated in the pension contract or lawful heirs stated.

The procedures and principles regarding instant monitoring of the return performance of pension mutual funds, rewarding the managers of the funds with relatively high returns during the evaluation period, and imposing sanctions on the managers of relatively unsuccessful funds have been regulated. The task of operating the system has been entrusted to EGM stakeholders.

2.2. *Challenges of entering the IPS*

2.2.1. *Complexity and Accessibility*

One of the primary challenges is the inherent complexity of many retirement systems. The options available can be bewildering, with a range of investment choices, contribution levels, and rules regarding withdrawal. This complexity can act as a barrier to effective participation, particularly for those with lower levels of financial literacy.

2.2.2. *Lack/Level of Financial Literacy*

Financial literacy plays a critical role in effective retirement planning. It influences an individual's ability to make informed financial decisions, which in turn affects their financial well-being in retirement. However, financial literacy levels vary significantly across different demographics, impacting the effectiveness of retirement planning efforts. Studies have shown that financial literacy is not uniformly distributed across populations. Younger people, women, and those with lower income levels often exhibit lower levels of financial literacy. This demographic

variation presents a challenge to universal retirement planning efforts, as these groups are often the most vulnerable to poverty in retirement. Lower levels of financial literacy are strongly correlated with poor retirement planning. Individuals with limited financial knowledge are less likely to calculate their retirement needs accurately, less likely to invest in diversified portfolios, and more likely to opt out of beneficial retirement plans.

3. THEORETICAL FRAMEWORK AND METHODOLOGY

3.1. Generative AI

GenAI signifies a major leap in AI, especially for its capacity to create new content. This technology leverages *large language models* (LLMs) that process extensive datasets of text and documents to produce coherent outputs in various formats, including text, images, and languages. Its capabilities have substantial implications across multiple sectors, particularly in finance. A pivotal advancement in LLMs was the introduction of the Transformer architecture by Vaswani et al. in their seminal paper "Attention is All You Need" in 2017 [1]. This architecture, using the *self-attention mechanism* [2], supports parallel processing and effectively manages long-range dependencies. This framework underpins models like OpenAI's GPT series [3] and Google's BERT [4], which have achieved remarkable results across various language tasks. These models are designed to interpret entire contexts, providing genuine contextual understanding [5]. Notable variations of the GPT architecture include models like ChatGPT, Llama, and Mistral [6] which share a consistent training approach: During *pre-training*, the models learn grammar, information, reasoning abilities, and some level of common-sense knowledge from a wide range of internet text [7][8]. This broad linguistic understanding equips them well for various tasks. *Fine-tuning* then focuses these models on specific applications, like conversational contexts for ChatGPT, making them suitable for applications such as chatbots and virtual assistants [9][10][11]. While models like Llama and Mistral are less well-known, they may represent advancements or specialized variants designed for specific research or practical uses. These models are at the forefront of NLP developments, facilitating more natural human-like interactions through advanced AI-driven language models [12][13][14].

3.2. Generative AI in Financial Advisory

Robo-advisors are digital platforms that provide financial investment services through algorithmic processes, modernizing and automating traditional financial advising and wealth management using advanced information technologies [15] [16]. These platforms enhance passive indexing strategies to reduce costs, risks, and the time required for portfolio rebalancing. They also mitigate behavioral biases and improve financial transaction

efficiency [17]. The Covid-19 pandemic has boosted the popularity of robo-advisors, which currently manage over \$1 trillion in assets and are expected to grow annually by 16.72% from 2022 to 2025 [18][19] noted that the pandemic highlighted the value of online investment services, prompting even those who preferred traditional in-person banking or human advisors to shift towards digital solutions. This shift has led to a significant increase in robo-advisor-related research, particularly over the past six years.

Research on robo-advisors covers several domains, including their impact on financial literacy, adoption and acceptance, investment performance, risk management, regulation and compliance, and human-robot interactions. [16] examined design principles, factors influencing adoption and acceptance, and portfolio recommendations. Additionally, the ongoing digitalization of the economy and the rise of fintech have accelerated the integration of digital experiences in financial services, transforming how services and channels converge [20]. The high-quality and effective services offered by robo-advisors ensure client satisfaction by providing tailored and personalized solutions [21][22]. *Personalization*, a key competitive edge, enhances sales by improving conversion rates, strengthening customer relationships, fostering loyalty, and boosting revenues. It involves tailoring offers to meet individual customer needs, but understanding these unique needs can be challenging. Personalization anticipates needs based on previously gathered data, whereas customization directly responds to specific client requests. AI and digital technology enable these personalized experiences by analyzing patterns from contextual and personal data [23] [24]. As technology and the industry continue to develop, the research focus within the robo-advisory services sector is expected to adapt and expand. Despite increased academic attention, there is still a notable research gap in understanding how robo-advisory services can be more effectively cater to the distinct preferences and requirements of individual investors.

3.3. Impact on Financial Literacy and Retirement Planning

GenAI and robo-advisors significantly influence financial literacy and retirement planning. Individuals need to possess sufficient financial knowledge about pension products and information provided by pension companies and be capable of making informed investment decisions. [25] developed a private pension literacy scale to determine the pension literacy levels of academic and administrative staff and found that 36.7% of respondents had medium and 8% had high levels of individual retirement literacy. The findings suggest that participants have a low level of private pension literacy. A significant difference was observed between individual retirement literacy levels and variables such as gender, marital status, and age. [26] studied the relationship between financial literacy levels of students and their awareness of the individual pension system. They found that students who attend financial courses more frequently have higher financial literacy scores and greater awareness of the private pension system.

A positive but weak correlation was found between students' financial literacy levels and their awareness of the private pension system. The influence of GenAI and robo-advisors extends into the realm of financial literacy and retirement planning. Literature suggests that AI tools can enhance users' understanding and engagement with their retirement plans, simplifying complex financial concepts and promoting proactive management of retirement savings.

3.4. Trust and Satisfaction in Retirement Systems

A critical factor in adopting new technologies in personal finance is user trust and satisfaction. [27] study on trust and satisfaction perceptions within individual retirement systems indicates that AI's personalization capabilities can lead to higher satisfaction and trust among participants, fostering more robust engagement with the retirement planning process. The Private Pension System has been promoted through recent regulations and legislative changes. Examining these applications reveals that the system has a solid legislative structure, which creates trust. This trust and the profitability provided by the system are significant factors in customer satisfaction. The study found that participants' perceptions of trust and satisfaction were slightly above average and that customer trust positively impacted customer satisfaction.

To establish trust in the Individual Retirement System, accurate and complete information must be provided to potential participants in a clear and understandable manner. Individuals are reluctant to commit to a system they do not fully understand. Therefore, it is recommended that investment advisors assigned by banks provide personalized and frequent updates rather than just general investment bulletins. This approach enhances the perception of helpfulness, which significantly impacts trust. Additionally, investment advisory services are a key advantage of individual pension funds. Through fund management, individuals can generate additional income beyond their contributions. In investment and savings systems, the perception of competence stands out more than accuracy/honesty and helpfulness, as the primary goal of these systems is to enable individuals to save in addition to their social security retirement income. Thus, increasing individuals' perception of competence can be achieved by ensuring they believe they are part of a profitable investment system [27].

4. DECISION SUPPORT SYSTEM DEVELOPMENT

The proposed GenAI-driven decision support system (DSS) is designed to address the complexities of retirement planning by integrating advanced GenAI technologies. This system aims to provide dynamic, personalized retirement planning advice tailored to individual user profiles, enhancing both the precision of financial guidance and the user's financial literacy over time.

4.1. System Architecture

The architecture of the proposed system (Figure 1) is structured around a core GenAI engine that leverages a GenAI model, similar in functionality to models like ChatGPT, which is fine-tuned for financial advising and retirement planning.

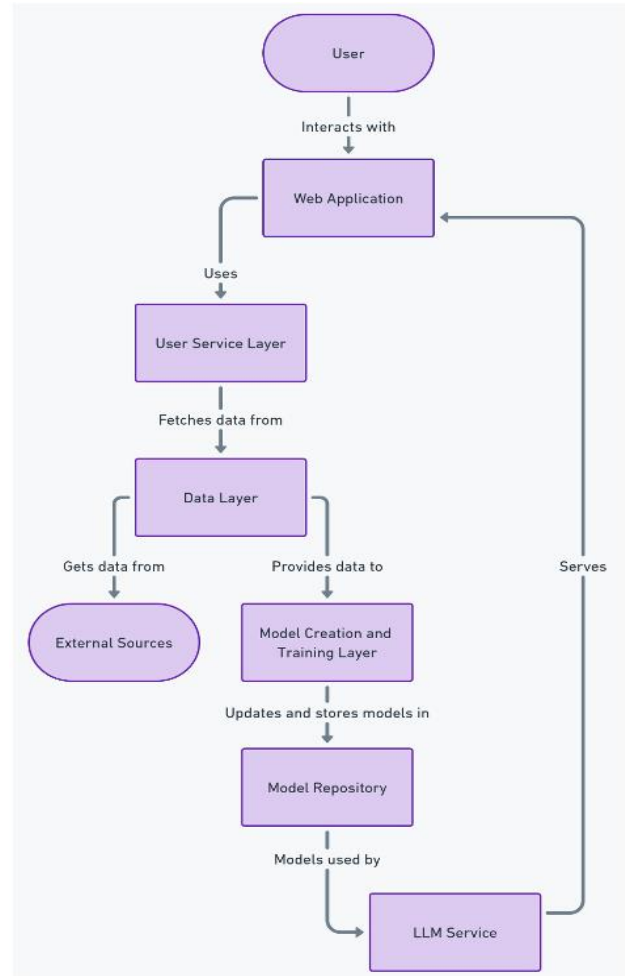


Figure 1. System architecture

The system architecture is composed of several key components:

- **User:** This component represents the end-user who interacts with the system. Users are the primary actors who initiate queries and receive responses, driving the operational flow of the application.
- **Web Application:** Serves as the interface through which users interact with the system. It is designed to be user-friendly and responsive, facilitating easy navigation and efficient access to services provided by the underlying architecture.
- **User Service Layer:** This layer acts as a mediator between the user and the data layer. It processes user requests, retrieves the necessary data from the data

layer, and ensures that responses are tailored to the user's needs. This layer improves user experience by optimizing data retrieval and response generation.

- **Data Layer:**

Central to the system's operation, this layer manages the acquisition, storage, and provision of data. It fetches data from both internal and external sources and serves it to other parts of the system for further processing. The data layer is essential for ensuring that the GenAI system and its agents have access to high-quality, relevant data necessary for generating accurate and personalized retirement planning advice.

The Extract, Transform, Load (ETL) process is a crucial component of the data layer, converting raw data into a structured format suitable for machine learning models. The ETL process consists of three main stages:

Extract:

Data is extracted from various internal and external sources, including user input, financial databases, market data feeds, and other relevant financial information. The data layer connects to external databases or APIs to fetch data that enriches the application's responsiveness and functionality. This includes integrating real-time data feeds or accessing large datasets relevant to the user's queries. The data can come in various formats such as CSV, JSON, XML, and SQL databases. Extraction tools and custom-built scripts are used to gather data efficiently from the aforementioned multiple sources.

Transform: Raw data is cleaned to remove inconsistencies, duplicates, and errors. This step ensures the quality and reliability of the data. Data is standardized to a consistent format, making it easier to integrate and analyze. This includes converting currencies, normalizing date formats, and ensuring consistent units of measurement. Additional relevant information is added to the data to enhance its value. This includes deriving new features, aggregating data, or integrating supplementary data from external sources. Transformation tools (Apache Spark, AWS Glue) and Python libraries (Pandas, NumPy) are used to process and transform the data.

Load: Transformed data is loaded into a data warehouse or a data lake, where it can be accessed by the GenAI system and its agents for further analysis and processing. For our system, we chose a hybrid approach using both a data warehouse and a data lake.

We use data lake for storing raw data, log files, market data feeds, and other large datasets. This allows for flexible data exploration and transformation before loading into the data warehouse for detailed analysis. We chose Amazon S3 which provides a highly scalable and flexible storage solution for unstructured and semi-structured data. It allows for the storage of large volumes of data in its raw form.

We use data warehouse for storing structured data, such as user profiles, financial transactions, and historical performance metrics. We chose Amazon Redshift which offers high-performance, scalable, and cost-effective data warehousing solutions. It supports complex queries and integrates seamlessly with other AWS services.

Model Creation and Training Layer: This critical layer receives data from the data layer and utilizes it for training and refining language models. The layer incorporates machine learning algorithms and data processing techniques to develop models that understand and generate language effectively.

- **Model Repository:** Acts as a storage unit for all the language models created and refined by the model creation and training layer. This repository ensures that models are readily accessible for use, updates, and further refinements.
- **LLM Service:** This service utilizes the models stored in the model repository to process user queries and generate appropriate responses. It is an essential component that applies natural language understanding and generation capabilities to meet user needs effectively.

The architecture illustrated in Figure 1 underscores a structured approach to managing user interactions, data processing, and language model utilization. Each component is strategically placed to optimize data flow and service delivery, ensuring the system is efficient, scalable, and responsive to the dynamic needs of users. This system architecture supports continuous improvement and adaptation of language models, enhancing the overall quality and effectiveness of the web application.

4.2. User Interface Design

The UI is designed to be intuitive and accessible, catering to users with varying levels of financial literacy. It provides clear navigation, straightforward data input methods, and the presentation of advice in a digestible format.

4.2.1. Front end

Figure 2 details the process flow from a user query to the response delivery in a web application environment. Each step in the sequence diagram is explained below:

- **User:** Represents the individual or system initiating the interaction. The user inputs a query through the user interface (UI), initiating the sequence of operations. The user's role is crucial as the originator of requests that drive the application's functional responses.
- **UI:** The front-end component where the user inputs their query. This interface is designed to be intuitive

and user-friendly, providing a seamless experience that efficiently captures user inputs and communicates them to the back-end systems.

User Service Layer

- **Forward Query:** Once the query is received from the UI, it is forwarded to the User Service Layer. This layer acts as a conduit, passing the query deeper into the system for processing.
- **Authentication & Authorization checks:** Critical for security, the User Service Layer performs checks to verify the user's identity and ensure they have the necessary permissions to perform the query. This step prevents unauthorized access and ensures compliance with security policies.

LLM Service

- **Forward Query:** After passing authentication and authorization, the query is forwarded to the LLM Service. This transition is crucial as it moves the query to the processing phase.
- **Process Query:** The LLM Service processes the query using pre-trained language models. This involves interpreting the query's intent and generating an appropriate response based on the information available to the model.
- **Send Response:** Once the response is generated, it is sent back through the User Service Layer to the UI. This marks the completion of the query processing phase.

UI (User Interface)

- **Display Response:** The final step in the sequence where the processed response is displayed to the user. This phase is critical as it provides the user with the information or feedback they sought, completing the interaction cycle.

This sequence diagram illustrates a streamlined interaction flow within a web application, emphasizing the critical steps of query processing and response generation. Each component—from the UI to the LLM Service—plays a pivotal role in ensuring the system's responsiveness and security. The process effectively demonstrates how user queries are handled, authenticated, processed, and responded to within a modern web application architecture, providing a clear and efficient pathway from user input to system output.

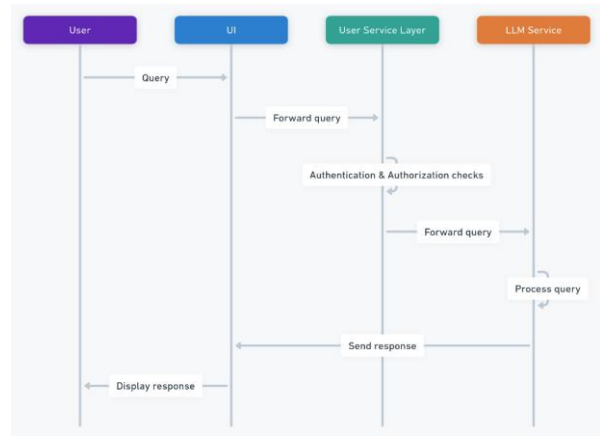


Figure 2. Front-end Sequence Diagram

4.2.2. *Back-end*

Figure 3 represents a back-end sequence diagram detailing the workflow from data collection to model utilization in a system that employs LLMs.

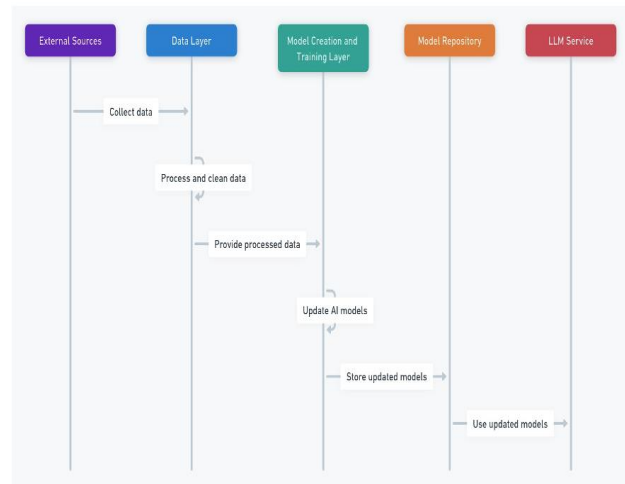


Figure 3. Back-end sequence diagram

External Sources

- **Collect Data:** This initial step involves gathering data from various external sources. These sources may include databases, web services, APIs, and other data repositories that provide the raw data necessary for training and refining language models. The efficiency and breadth of data collection are crucial for building robust and versatile models.

Data Layer

- **Process and Clean Data:** Once data is collected, it moves to the data layer, where it undergoes processing and cleaning. This step is essential to ensure the quality and usability of the data. Processing involves converting raw data into a more structured format, while cleaning involves removing inaccuracies,

duplicates, and irrelevant information, which helps in enhancing the data's reliability for training purposes.

- **Provide Processed Data:** After processing and cleaning, the data is made available to the next layer. This provision ensures that the data is in a usable state, formatted correctly, and free from elements that could degrade the model's performance.

Model Creation and Training Layer

- **Update AI Models:** This layer receives the processed data and uses it to update or train new AI models. Updating models involves adjusting existing models based on new data, which can include retraining or fine-tuning. This step is crucial for keeping the models current and effective, allowing them to adapt to new information or changes in data trends.

Model Repository

- **Store Updated Models:** Once the models are updated, they are stored in the Model Repository. This storage is vital for version control, accessibility, and deployment. The repository ensures that all model versions are tracked and can be retrieved or rolled back if needed. This feature supports continuous improvement and testing of models before they are deployed into production.

LLM Service

- **Use Updated Models:** The final step in the sequence involves the LLM Service utilizing the updated models to perform tasks such as data analysis, prediction, or any other language-related processing required by the application. This step highlights the practical application of the models, where they serve the end-user or automated systems by providing intelligent responses based on learned patterns and data.

Each step in the back-end sequence is geared towards enhancing the data quality, model accuracy, and overall system efficiency, ensuring that the end-service remains relevant and effective in real-world applications. This workflow not only underscores the systematic approach to data-driven model training but also highlights the continuous cycle of improvement and adaptation that characterizes advanced AI-driven systems.

4.3. Integration of LLM for Dynamic, Personalized Pension Planning

The integration of GenAI into the system allows for dynamic and personalized pension planning in several ways. Firstly, the system can simulate various financial scenarios based on current trends and possible future events, such as economic downturns, inflation changes, or shifts in the stock market. This helps users understand

potential risks and the impact of different retirement planning decisions. Secondly, as financial markets fluctuate and personal financial situations change, the AI updates its advice in real-time. This ensures that users always have access to the most current and relevant financial guidance.

4.3.1. User Query Life Cycle

Figure 4 represents the User Query Life Cycle, detailing the process and infrastructure required for managing user queries within the system. This comprehensive breakdown of the User Query Life Cycle illustrates how each component is crucial for ensuring the system not only responds accurately to user queries but also manages these interactions securely and efficiently. This cycle is pivotal for maintaining high user satisfaction and system reliability.

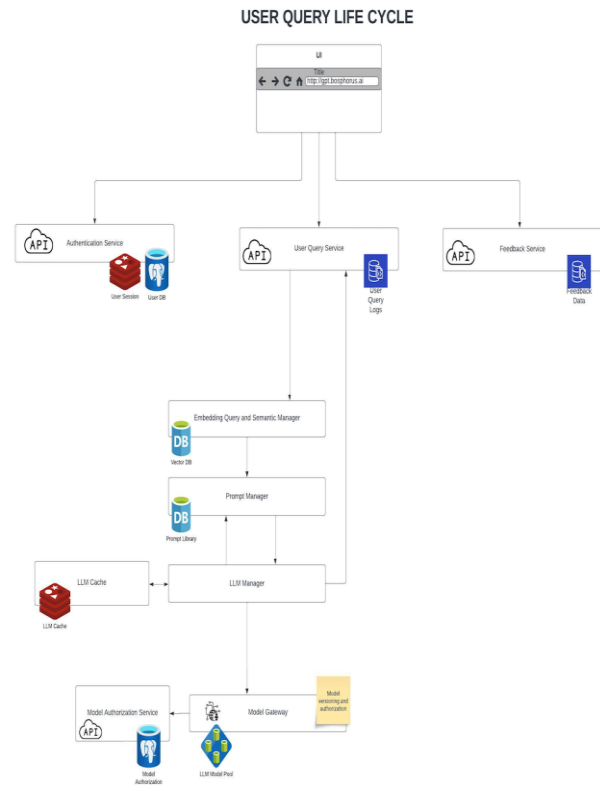


Figure 4. User Query Life Cycle

1. User Interface (UI): The interface serves as the primary interaction point between the user and the system. It's designed to be intuitive, enabling effortless navigation and query input. Innovation in UI design incorporates adaptive interfaces that adjust to user preferences and behaviors, enhancing engagement and satisfaction.

2. Authentication Service

- **API for Authentication Service:** This component is critical for security and personalization. It ensures that user interactions are secure and that data is only

accessible to authenticated users. The innovative use of session management and advanced encryption enhances user trust and system integrity.

3. User Query Service

- **User Query Service API:** Central to the query processing framework, this service efficiently manages the distribution and logging of user queries. Its innovative aspect lies in its ability to handle massive volumes of queries simultaneously while maintaining high performance and reliability.

4. Feedback Service

- **Feedback Service API:** This service collects user feedback to continuously refine and improve the response quality. The innovative use of machine learning algorithms to analyze feedback helps in dynamically adjusting query responses based on real-time user input, greatly enhancing the learning capabilities of the system.

5. Data Management and Processing

- **Embedding Query and Semantic Manager:** By converting queries into semantic embeddings, this manager plays a crucial role in understanding user intent more deeply. The use of cutting-edge natural language processing technologies to generate and manage these embeddings represents a significant innovation, leading to more precise and context-aware responses.
- **Prompt Manager:** This component manages a library of query prompts to streamline response generation. The innovation here includes the use of advanced retrieval algorithms that can quickly match user queries with the most effective prompts, greatly speeding up response times and improving accuracy.

6. LLM Cache: The caching mechanism is vital for enhancing system efficiency. It reduces latency by storing frequently or recently accessed data. The innovative application of smart caching algorithms predicts which data is likely to be reused, thus optimizing resource allocation and speeding up query processing.

7. LLM Manager: It intelligently coordinates between cached and new data processing, ensuring optimal system performance. This manager's innovative feature includes the use of decision algorithms that automatically determine the most efficient way to process a query, considering current system load and query complexity.

8. Model Authorization and Access

- **Model Authorization Service API:** This ensures that all operations within the system comply with legal and ethical standards. The innovative use of policy-driven

access controls and real-time monitoring tools safeguard sensitive data and ensure compliance with global data protection regulations.

- **Model Gateway:** Acts as a secure conduit between the query services and the LLMs, managing access permissions and resource allocation. Its innovative aspect is the dynamic scaling capability, allowing it to handle varying loads seamlessly and ensuring that resources are efficiently utilized.

9. LLM Model Pool: A sophisticated collection of language models tailored to various aspects of query processing. The innovation here lies in the models' ability to learn from interactions continuously and adapt to new information, improving their accuracy and relevance over time.

4.3.2. Integration of Agents within the GenAI System

Agents are employed to perform specific, well-defined tasks such as mathematical calculations, data preprocessing, and basic financial modeling. For instance, calculating the future value of retirement savings based on current contributions and expected interest rates can be handled by a simple machine learning model or even a deterministic algorithm. By delegating these tasks to smaller models, the system can operate more efficiently, reserving the computational power of the GenAI model for more complex, context-aware tasks such as generating personalized advice and interpreting user queries.

LangChain Agents facilitate the integration of multiple machine learning models and APIs into a cohesive system. This approach allows for modularity, where each agent performs a specific function, and their outputs are fed into the GenAI system for further processing. Using LangChain Agents, the system optimizes workflows by chaining together different models and tools, ensuring that each task is handled by the most appropriate component.

In our system, a user's query about their retirement savings initiate the following sequence: (1) The user inputs their current savings, expected contributions, and retirement goals. (2) An agent model calculates the projected future value of the savings using a basic financial formula. (3) The result is passed to the GenAI system, which integrates this information with broader financial data and generates a comprehensive retirement plan. The GenAI model provides personalized advice, considering the user's risk tolerance and investment preferences. (4) The user interacts with the GenAI system to refine their plan, ask questions, and explore different scenarios.

4.3.3. LLM Training Life Cycle

This life cycle (Figure 5) demonstrates the sequential and structured processes involved in preparing, training, and deploying LLMs:

- **Data Sources:** This is the initial phase where diverse data sets are collected to train the LLMs. The data encompasses:
 - **Documents:** Written content sourced from various formats and domains.
 - **User Query Logs:** Data collected from user interactions, providing insight into user intent and language usage.
 - **Feedback Data:** Information derived from user responses to the system, crucial for iterative model improvements.
 - **Oracle EGM Data Repo:** A specialized repository that might include structured data, relevant for domain-specific training.
- **Data Crawler Layer:**

This layer is designed to handle the systematic retrieval and processing of data. This layer encompasses both web scraping (for real-time data extraction from the web) and the transfer of "cold data" or file-based data into a structured data catalogue.

1. **Web Scraping for Real-Time Data Retrieval:** Aims to extract real-time financial data, market trends, and news relevant to retirement planning from various online sources. There are three types of patterns: (1) Factory pattern is used to create different types of web crawlers based on the source and data structure. For example, one factory might create a crawler for financial news websites, while another creates a crawler for market data APIs (2) Iterator pattern allows the crawler to systematically iterate through the pages or API responses to retrieve data. (3) Visitor pattern applies specific operations to each element of the retrieved data, such as parsing HTML, extracting relevant information, and transforming it into a usable format.

The process initializes by crawler' identifying the target websites or APIs and the data structures used. The crawler navigates through web pages or API endpoints, extracting the necessary data. This can include parsing HTML content, handling pagination, and managing API rate limits. The extracted data is processed to remove unnecessary elements, normalize formats, and ensure consistency. The cleaned and structured data is then stored in the data warehouse or data lake for further analysis and use by the GenAI system.

Use Case: The system initializes a web scraper for a financial news website. The scraper navigates through the website, extracting headlines, summaries, and publication dates. The HTML content is parsed, and the extracted data is cleaned and structured into a JSON format. The structured data is stored in the data lake for real-time analysis and integration into user recommendations.

2. **Crawling for Cold Data Transfer:** Aims to transfer large volumes of historical or "cold" data from various sources such as databases, file systems, and data repositories into a structured data catalogue. There are three types of patterns: (1) Factory Pattern is used to create different types of crawlers for various data sources, such as SQL databases, CSV files, or cloud storage. (2) Iterator pattern enables the crawler to systematically process files or database records, ensuring no data is missed. (3) Visitor pattern applies specific operations to each data item, such as validating, cleaning, and transforming it before storage.

The process initializes by crawler identifying the source locations (e.g., database connections, file directories). The crawler reads data from the source, handling different formats such as CSV, JSON, XML, or SQL tables. The retrieved data undergoes transformation processes, including cleaning, normalization, and enrichment. The processed data is organized into a data catalogue, stored in the data.

Use Case: The system initializes a crawler for an on-premises SQL database containing historical financial transaction records. The crawler reads data from the database, processing one table at a time. The retrieved data is cleaned to remove duplicates, normalized to ensure consistent formats, and enriched with additional metadata. The processed data is uploaded to a data warehouse, indexed in a data catalogue, and made available for query and analysis.

The data crawler layer plays a vital role in the GenAI-based private pension information and recommendation system by ensuring comprehensive and flexible data retrieval. Whether through web scraping for real-time data or transferring cold data into a structured data catalogue, this layer ensures that the system has access to high-quality, relevant data necessary for providing accurate and personalized retirement planning advice. By leveraging design patterns like Factory, Iterator, and Visitor, the data crawler layer is both robust and adaptable, capable of handling diverse data sources and structures effectively.

- **Data Storage:** The crawled data is stored in a robust system like Amazon S3 bucket, ensuring security, scalability, and accessibility. This setup supports various data types and is critical for large-scale data operations.

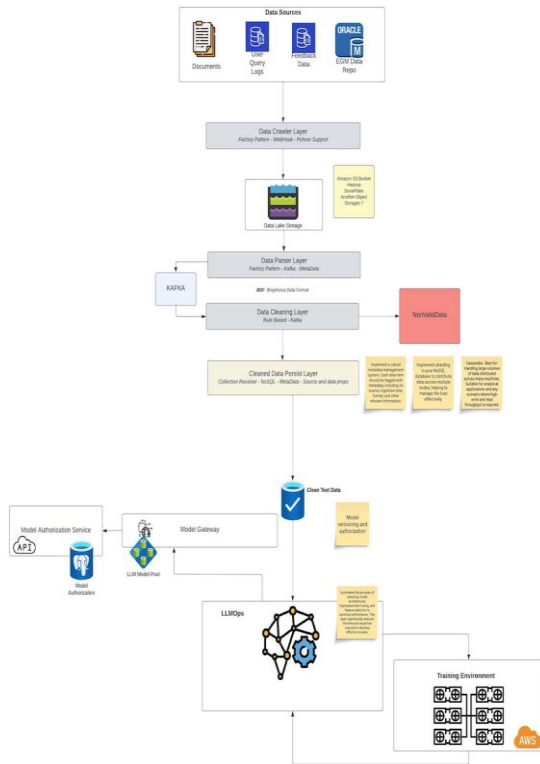


Figure 5. LLM Training Life-Cycle

- Data Parser Layer:** Following storage, data is parsed using Factory and Lambda/Mediator patterns into structured formats like AVRO or Parquet. This transformation is crucial for standardizing data before it undergoes cleaning and further processing.
- Data Cleaning Layer:** This layer focuses on refining the data using defined rules and real-time processing via Kafka to ensure quality and relevance. Non-valid data is identified, logged, and handled appropriately to maintain the integrity of the training set.
- Cleaned Data Persist Layer:** Post-cleaning, the data is persisted in systems that support efficient retrieval and management, employing NoSQL databases with ACID compliance to ensure reliability and performance during the data retrieval phase.
- Model Gateway:** Acts as a checkpoint that integrates with model authorization services to ensure all data used for training complies with regulatory and ethical standards, safeguarding against the misuse of sensitive information.
- LLM Model Pool:** This is the core repository where various pre-trained and to-be-trained models are stored. The models available here are ready to be fine-tuned or trained with the newly cleaned and structured data.

- LLM Ops:** This operational unit manages the continuous integration and deployment of language models, ensuring they are up-to-date and perform optimally. It incorporates advanced automation tools for monitoring, scaling, and managing the models within a production environment.
- Training Environment:** The actual training of models occurs in a controlled environment, often cloud-based platforms like AWS, which provide the necessary computational power and scalability to handle extensive model training sessions efficiently.

4.3.4. Selection of LLMs for the Proposed System

The choice of LLMs for the GenAI-based private pension information and recommendation system is critical for its effectiveness and reliability. The selection should be based on several factors including the model’s ability to understand and generate human-like text, its performance on financial data, and its adaptability to the specific needs of the system. Most used LLMs and their capabilities are:

GPT-4 (Generative Pre-trained Transformer 4) by OpenAI

GPT-4 has demonstrated superior capabilities in understanding and generating human-like text across various contexts. It can maintain context over long conversations, making it suitable for interactive financial advising. Fine-tuning capabilities allow the model to be adapted specifically for financial data and personalized advice. GPT-4 has been extensively tested and validated across numerous applications, ensuring reliability. Its ability to handle a wide range of queries and provide detailed explanations makes it ideal for a complex domain like financial planning. GPT-4 can handle large-scale deployments, making it suitable for use in a system intended for a broad user base.

BERT (Bidirectional Encoder Representations from Transformers) by Google

BERT processes text in both directions, allowing for a deeper understanding of context. It is trained on extensive text corpora, making it robust in various language tasks. It can be fine-tuned for specific tasks, such as financial data analysis and personalized recommendations. BERT’s bidirectional nature enhances its ability to understand nuanced financial queries. Fine-tuning BERT on financial datasets improve its performance in providing accurate and relevant advice. For institutions already using Google’s cloud services, BERT offers integration and optimization.

LLaMA (Large Language Model Meta AI) by Meta

LLaMA is designed to be efficient and effective in various NLP tasks, making it suitable for experimental applications. It delivers high accuracy and performance in generating human-like text. Can be fine-tuned for specific applications, including financial advising. LLaMA’s design ensures it can provide accurate responses quickly, enhancing user experience. Its optimization for research allows continuous improvement and adaptation based on

feedback. Suitable for integration into systems requiring high performance and adaptability.

For our system, GPT-4 by OpenAI is the most suitable LLM. Several factors contribute to this selection. Firstly, GPT-4's advanced natural language understanding capabilities enable it to comprehend and generate human-like text with high accuracy, essential for interpreting complex financial queries and providing detailed, personalized advice. Secondly, its ability to maintain contextual awareness over long conversations ensures that users receive consistent and coherent advice throughout their interaction with the system. Thirdly, GPT-4's customization potential allows it to be fine-tuned specifically for financial data and personalized recommendations, enhancing its relevance and accuracy for users. Additionally, GPT-4's proven performance across numerous applications demonstrates its reliability and effectiveness in delivering high-quality responses. Finally, its scalability makes it suitable for large-scale deployments, ensuring that the system can accommodate a broad user base without compromising performance. These attributes make GPT-4 the optimal choice for enhancing the user experience and effectiveness of the GenAI-based private pension information and recommendation system.

4.3.5. Language Model Operational Framework

Figure 6 represents a streamlined structure within a language model operational framework, emphasizing the integration of model management and prompt categorization to enhance response generation.

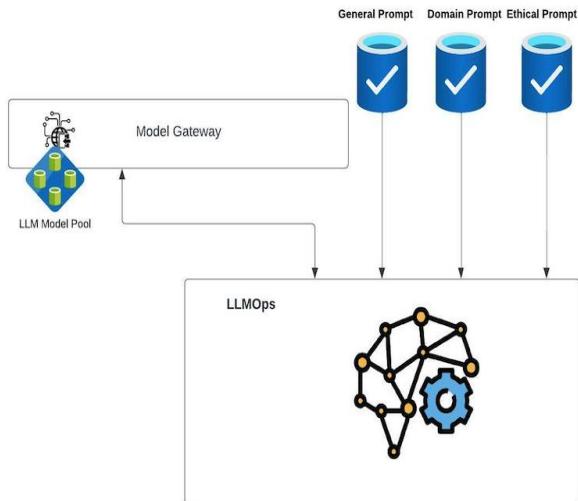


Figure 6. LLM Operational Framework

- **Model Gateway:** This is the primary interface between the stored language models and the operational environment. It functions as a regulatory node that ensures all requests for model utilization pass through a uniform gateway, providing a centralized point of control. This setup allows for

efficient management of model resources, security protocols, and compliance checks. The gateway's innovation lies in its ability to dynamically route requests based on model availability, query complexity, and priority, ensuring optimal allocation of computational resources.

- **LLM Model Pool:** A robust repository of pre-trained language models that cater to a wide range of tasks and domains. This pool is fundamental in providing diversity in response capabilities and specialization in content generation. The innovative aspect of the model pool is its extensive coverage of different linguistic styles and domain-specific knowledge, enabling highly accurate and contextually appropriate outputs.
- **Prompt Categorization**
 - **General Prompt:** These prompts cater to a broad range of queries without specific domain constraints. They are designed to handle general knowledge questions and everyday informational needs, making them versatile and widely applicable.
 - **Domain Prompt:** Specifically tailored to address queries within a particular field or industry, these prompts enhance the accuracy and relevancy of responses by utilizing domain-specific language models. This allows for specialized knowledge processing, which is crucial in fields like medicine, law, or finance.
 - **Ethical Prompt:** Designed to ensure responses adhere to ethical guidelines and social norms. These prompts are crucial for maintaining the integrity and trustworthiness of the system, especially when dealing with sensitive topics. They incorporate filters and checks that prevent the generation of harmful or biased content.

- **LMM Operations (LLMOps):**

The LLMOps component is crucial for the ongoing maintenance, monitoring, and optimization of language models within the GenAI-based private pension information and recommendation system. This component encompasses a comprehensive suite of tools and processes designed to ensure that the models perform at their peak efficiency, adapting to changing usage patterns and requirements. Key processes in LLMOps include: (1) **Automated Model Tuning:** Continuously improve the language model's performance through regular fine-tuning based on new data and user feedback. The process begins with gathering new data from user interactions, financial markets, and other relevant sources. Then we periodically retrain the model with this updated data to enhance its accuracy and relevance. We utilize machine learning frameworks (TensorFlow, PyTorch, or Hugging Face's Transformers library) for efficient model tuning. (2) **Real-Time Performance Analytics:** Monitor the model's performance metrics in real-time, ensuring it meets the desired accuracy and efficiency standards. We use BLEU

score to evaluate the quality of generated text against reference texts. We monitor the response time of the model (latency) to ensure it meets performance expectations. Then we implement our custom dashboards to visualize and track these metrics. (3) Adaptive Deployment Strategies: Deploy the language model in a manner that optimally balances performance and resource usage. The process initializes with distributing requests across multiple instances of the model to prevent overloading any single instance (Load Balancing). Then the system automatically scales the number of active model instances based on current demand. We use serverless computing solutions (Kubernetes and AWS Lambda) for scalable and adaptive deployment. (4) Monitoring Post-Training Performance: To ensure that the language model maintains its performance after retraining or fine-tuning. The process initializes with comparing the model's performance metrics (accuracy, BLEU score) against previous versions to detect any regressions (Baseline Comparison). Then we implement continuous integration/continuous deployment (CI/CD) pipelines to automatically test and deploy updates. We use tools MLflow for tracking and versioning model performance. (5) Debugging and Validation: To identify and resolve issues that may arise from model updates or changes in data patterns. The process initializes with analyzing incorrect predictions to understand the root causes and make necessary adjustments. Then we conduct thorough validation tests using a separate validation dataset to ensure the model's robustness. We employ debugging frameworks (TensorBoard, integrated with Jupyter notebooks) for in-depth analysis. As an example, suppose the GenAI model is retrained after a year to incorporate the latest financial data and user feedback. Post-training, the model's performance is tracked using metrics such as accuracy and BLEU scores. These metrics are compared to the baseline established before retraining. If any performance drop is detected, the system triggers an error analysis process to identify the issues. Continuous monitoring tools ensure that any significant deviations are promptly addressed, and adaptive deployment strategies help in balancing the load and scaling the resources as needed.

By integrating advanced prompt management with robust model control and operations, the system ensures that outputs not only meet the diverse needs of users but also adhere to high standards of accuracy, relevance, and ethical consideration. This framework is pivotal in deploying GenAI-driven applications that require nuanced understanding and generation of human-like text.

5. DECISION SUPPORT SYSTEM IMPLEMENTATION and USE CASE

5.1. Implementation Challenges and Strategies

5.1.1. Integration with Existing Financial Platforms

Integrating the GenAI-based private pension information and recommendation system with existing financial

platforms presents several technical and operational challenges. One major challenge is ensuring compatibility with diverse systems that use different technologies and data formats. Financial platforms often have established infrastructures, proprietary software, and specific data standards that can complicate seamless integration. To deal with these challenges we will start with a phased (incremental) integration approach, beginning with less critical components and gradually expanding to more complex systems. This allows for the identification and resolution of integration issues without disrupting the entire system. We will then develop robust APIs that allow seamless communication between the GenAI system and existing financial platforms. These APIs are designed to handle different data formats and protocols, ensuring compatibility and ease of integration. We will also implement data standardization practices to ensure that the data exchanged between the GenAI system and financial platforms is consistent and accurate. This involves using industry-standard data formats and protocols. Moreover we ensure that robust security measures are in place to protect sensitive financial data during integration. This includes encryption, secure data transmission protocols, and regular security audits.

5.1.2. Ethical Implications of Using GenAI in Financial Advice

The first ethical risk is that: the system collects and processes sensitive financial data, raising concerns about data privacy and security. To mitigate this risk we use advanced encryption techniques to protect data both at rest and in transit. Also we collect only the data necessary for providing accurate recommendations, reducing the risk of data breaches. Moreover we clearly communicate data usage policies to users, ensuring they are informed about how their data is handled.

The second risk is; AI algorithms can inadvertently perpetuate biases present in training data, leading to unfair or discriminatory recommendations. To mitigate this risk we implement regular audits to identify and correct biases in the algorithms. We use diverse and representative datasets to train the AI models, reducing the risk of biased outcomes. Moreover we develop and monitor fairness metrics to ensure that recommendations do not disproportionately favor or disadvantage any user group.

Transparency and explainability is the third risk. Users need to understand how and why specific recommendations are made to trust the system. To mitigate this risk we develop models that can provide clear, understandable explanations for their recommendations. We implement feedback mechanisms that allow users to ask questions and receive explanations about the advice given.

Accountability is another risk that we should consider. Determining accountability for financial advice provided by an AI system can be complex. To mitigate this risk we establish clear policies outlining the roles and

responsibilities of the AI system, financial advisors, and the institution. We ensure that human advisors are available to review and validate the recommendations provided by the AI system.

5.1.3. Ensuring User Adoption

User adoption is another significant challenge, as users may be hesitant to trust and rely on an AI-based system for their retirement planning. Factors such as lack of familiarity with AI, concerns about data privacy, and resistance to change can hinder adoption. To deal with these challenges we start with designing a user-friendly interface that is intuitive and easy to navigate. The system provides clear instructions, feedback, and support to help users feel comfortable and confident in using it. Then we will conduct educational campaigns to inform users about the benefits and functionalities of the GenAI system. This includes webinars, tutorials, and detailed user manuals that explain how the system works and how it can enhance their retirement planning. We will also implement and communicate strong data privacy measures to assure users that their personal and financial data is secure. This includes transparent data handling policies and compliance with relevant data protection regulations.

5.1.4. Real-World Testing

We will perform a pilot program involving a small group of users to test the system in a real-world setting. This pilot will target a diverse user base to gather comprehensive feedback. We will choose participants based on demographics, financial literacy levels, and retirement planning needs to ensure a representative sample. We will establish a robust feedback mechanism (survey and where necessary, interview) to collect user experiences, identify issues, and gather suggestions for improvement. We will define clear performance metrics to evaluate the system's effectiveness. These metrics include user satisfaction, accuracy of recommendations, ease of use, and impact on financial literacy. We will continuously collect data on these metrics during the pilot phase to assess performance and make necessary adjustments. We will use the feedback and performance data to iteratively improve the system. This involves refining algorithms, enhancing user interface design, and addressing any technical issues identified during the pilot phase.

5.1.5. Scaling

Collaborating with a range of partners facilitates broader adoption and integration thus we will establish partnerships with multiple financial institutions to expand the system's reach. Moreover we will invest in scalable infrastructure to support increased user loads and data processing demands. This includes cloud-based solutions that offer flexibility and scalability. Also we will ensure compliance with financial regulations and standards. This involves continuous monitoring of regulatory changes and adapting the system accordingly. Then we will launch a

comprehensive marketing campaign to promote the system to a wider audience. This includes digital marketing, participation in financial industry conferences, and leveraging social media platforms. Finally we will provide ongoing support and updates to users to maintain engagement and satisfaction. This includes customer service, regular system updates, and continuous improvements based on user feedback.

5.1.6. Use Case

Objective of our use case is to enhance retirement planning services using GenAI technology to provide personalized advice and improve user engagement. There are three actors: (1) User: The end-user seeking personalized retirement advice. (2) System: GenAI-based private pension information and recommendation system. (3) Financial Advisor: A SecurePension financial advisor providing support and additional consultation. Our primary goals are: (1) To provide personalized retirement planning advice. (2) To improve financial literacy and user engagement. (3) To ensure adaptive recommendations. Steps and LLM utilization are explained below:

Step 1: User Profile Creation

Activity: User logs into the platform and creates a profile by providing personal and financial information (e.g., age, income, current savings, retirement goals).

LLM Utilization: The LLM analyzes user input to ensure completeness and clarity, providing prompts for missing or unclear information. Example output: "It looks like you haven't specified your current savings. Can you please provide this information to tailor your retirement plan better?"

Step 2: Initial Analysis and Recommendation

Activity: The system uses the provided data to generate an initial retirement plan recommendation.

LLM Utilization: The LLM processes the user data, applying financial models and algorithms to generate personalized advice. The LLM provides a summary and detailed explanation of the recommended retirement plan.

Example output: "Based on your current savings and retirement goals, we recommend increasing your monthly savings by \$200 to meet your target. Here's a detailed breakdown of how this will impact your retirement fund."

Step 3: Scenario Simulation

Activity: The user can simulate various financial scenarios (e.g., changes in income, market fluctuations) to see how they impact the retirement plan.

LLM Utilization: The LLM generates and explains multiple financial scenarios, highlighting potential risks and benefits. Example output: "If you experience a 10% increase in income next year, your retirement fund could grow by an additional \$50,000 by the time you retire. Conversely, a 5% market downturn could reduce your fund by \$20,000."

Step 4: Interactive Q&A

Activity: The user asks specific questions about the retirement plan and financial concepts they don't understand.

LLM Utilization: The LLM provides real-time, context-aware responses to user queries, enhancing financial literacy. Example prompt: "What are the benefits of diversifying my investment portfolio?" The LLM responds, example output: "Diversifying your portfolio reduces risk by spreading investments across various asset classes, which can help protect your retirement fund from market volatility."

Step 5: Feedback and Continuous Improvement

Activity: The user provides feedback on the system's recommendations and overall experience.

LLM Utilization: The LLM collects and analyzes user feedback to improve its algorithms and user interactions continuously. Example: "How satisfied are you with the clarity of the advice provided? Your feedback will help us enhance our recommendations."

This use case illustrates the practical application of a GenAI-based private pension information and recommendation system within a financial institution. By leveraging LLMs, the system can provide personalized, real-time, and adaptive retirement planning advice while enhancing user engagement and financial literacy. The LLM's capabilities in processing and analyzing data, generating detailed explanations, simulating scenarios, and interacting with users in a context-aware manner are crucial for delivering a comprehensive and user-friendly experience. This approach ensures that users receive the most relevant and effective financial advice tailored to their unique needs and circumstances.

6. CONCLUSION

The development and implementation of the GenAI-based private pension information and recommendation system outlined in this paper represents a significant step forward in individual retirement planning. The system's use of GenAI algorithms to process and analyze user data has demonstrated substantial potential in enhancing users' financial literacy and ability to make well-informed retirement planning decisions. Through personalized advice, the system facilitates a deeper understanding of various financial products and investment strategies tailored to individual risk preferences and retirement goals. One of the notable outcomes of this system is its ability to democratize financial advice, historically the domain of professional financial advisors. By making personalized financial guidance accessible and affordable, the system breaks down barriers to financial knowledge and planning tools, potentially leveling the playing field for a broader range of individuals. Furthermore, the system's dynamic adaptability to both macroeconomic conditions and personal life changes ensures that the advice remains relevant and timely, providing users with confidence in managing their retirement planning proactively. The continuous learning capability of the GenAI algorithms

ensures that the system evolves with the latest financial trends and data, enhancing its reliability and accuracy over time.

Potential limitations of this study are: (1) Different users have diverse financial goals, risk tolerances, and investment preferences. A one-size-fits-all approach may not adequately address individual needs. To mitigate this limitation, we will utilize the system's personalization capabilities to tailor advice to individual profiles and preferences. (2) Economic conditions and cultural attitudes towards savings and investments vary across regions and demographics. Recommendations that are effective in one context may not be suitable in another. To mitigate this limitation, we will incorporate localized economic data and culturally relevant advice to ensure applicability across diverse contexts.

Future developments should focus on: (1) Exploring advanced personalization techniques using deeper machine learning models to cater to highly specific user needs. Since there is a potential for improved user satisfaction and engagement through more accurate and relevant recommendations. (2) Investigating methods for seamless integration with a wider range of financial platforms and services. Since providing a unified financial management experience will increase system adoption and user convenience. (3) Developing algorithms capable of processing and acting on real-time financial data more efficiently. This will enhance responsiveness and relevance of financial advice in dynamic market conditions.

TEŞEKKÜR (ACKNOWLEDGMENTS)

This study was supported with the project number 3235018 under the program of "TÜBİTAK 1707".

REFERENCES

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [2] Shaw, P., Uszkoreit, J., Vaswani, A. (2018). Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.
- [3] Ghojogh B., Ghodsi, A. (2020). Attention mechanism, transformers, bert, and gpt: tutorial and survey
- [4] 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.
- [5] Liu, Q., Kusner, M. J., Blunsom, P. (2020). A survey on contextual Embeddings. arXiv preprint arXiv:2003.07278.
- [6] Roumeliotis, K. I., Tselikas, N. D., Nasiopoulos, D. K. (2023). Llama 2: Early adopters' utilization of meta's new open-source pretrained model.
- [7] Baladn, A., Sastre, I., Chiruzzo, L., Ros, A. (2023). Retuyt-inco at bea 2023 shared task: Tuning open-source llms for generating teacher responses. In: Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications

- (BEA 2023), 756–765.
- [8] Nay, J. J. (2023). Large language models as fiduciaries: A case study toward robustly communicating with artificial intelligence through legal standards.
- [9] Sennrich, R., Haddow, B., Birch, A. (2015). Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.
- [10] Zhuo, T. Y., Li, Z., Huang, Y., Li, Y.-F., Wang, W., Haffari, G., Shiri, F. (2023). On robustness of prompt-based semantic parsing with large pretrained language model: An empirical study on codex. arXiv preprint arXiv:2301.12868.
- [11] Chiang, W.-L., Li, Z., Lin, Z., Sheng, Y., Wu, Z., Zhang, H., Zheng, L., Zhuang, S., Zhuang, Y., Gonzalez, J.E., Stoica, I., Xing, E.P. (2023). Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. March.
- [12] Yao, Z., Yazdani Aminabadi, R., Zhang, M., Wu, X., Li, C., He, Y. (2022). Zeroquant: Efficient and affordable post-training quantization for large-scale transformers. *Advances in Neural Information Processing Systems*, 35, 27168–27183.
- [13] Zou, A., Wang, Z., Kolter, J. Z., Fredrikson, M. (2023). Universal and transferable adversarial attacks on aligned language models,” arXiv preprint arXiv:2307.15043.
- [14] Katz, D. M., Bommarito, M. J., Gao, S., Arredondo, P. (2023). GPT-4 Passes the Bar Exam. March.
- [15] Jung, D., Dorner, V., Glaser, F. and Morana, S. (2018a). Robo-advisory: digitalization and automation of financial advisory. *Business and Information Systems Engineering*, 60(1),81-86. <https://doi.org/10.1007/s12599-018-0521-9>.
- [16] Jung, D., Dorner, V., Weinhardt, C. and Pusmaz, H. (2018b). Designing a robo-advisor for risk-averse, low-budget consumers. *Electronic Markets*, 28(3), 367-380. <https://doi.org/10.1007/s12525-017-0279-9>.
- [17] D’Acunto, F., Prabhala, N. and Rossi, A.G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5), 1983-2020. <https://doi.org/10.1093/rfs/hhz014>.
- [18] Isaia, E. and Oggero, N. (2022). The potential use of RAs among the young generation: evidence from Italy. *Finance Research Letters*, 48, 103046. <https://doi.org/10.1016/j.frl.2022.103046>.
- [19] Au, C.D., Klingenberg, L., Svoboda, M. and Frere, E. (2021). Business model of sustainable robo-advisors: empirical insights for practical implementation. *Sustainability*,13(23),13009. <https://doi.org/10.3390/su132313009>.
- [20] Oh, S., Park, M.J., Kim, T.Y. and Shin, J. (2022). Marketing strategies for fintech companies: text data analysis of social media posts. *Management Decision*, 61(1), 243-268. <https://doi.org/10.1108/md-09-2021-1183>.
- [21] Rodrigues, L.F., Oliveira, A. and Rodrigues, H. (2023). Technology management has a significant impact on digital transformation in the banking sector. *International Review of Economics and Finance*, 88, 1375-1388. <https://doi.org/10.1016/j.iref.2023.07.040>.
- [22] McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- [23] Huang, E.Y. and Lin, C.Y. (2005). Customer-oriented financial service personalization”, *Industrial Management and Data Systems*, 105(1), 26-44. <https://doi.org/10.1108/02635570510575171>.
- [24] Gao, Y. and Liu, H. (2022). Artificial intelligence-enabled personalization in interactive marketing: a customer journey perspective. *Journal of Research in Interactive Marketing*, 17(5), 1-18, <https://doi.org/10.1108/jrim-01-2022-0023>.
- [25] Güneş, H. (2022). akademik ve idari personelin bireysel emeklilik okuryazarlığının tespiti: makü örneği. *Muhasebe Ve Finans İncelemeleri Dergisi*, 5(1), 39-51. <https://doi.org/10.32951/mufider.1000689>.
- [26] Onat, O. K., & Yöntem, H. (2022). Finansal okuryazarlığın bireysel emeklilik sistemi tercihlerine etkisi: burdur mehmet akif ersoy üniversitesinde bir araştırma. *Finansal Araştırmalar Ve Çalışmalar Dergisi*, 14(26), 165-192. <https://doi.org/10.14784/marufacd.1055196>.
- [27] Akgün, M. K., & Bozkurt, Ö. (2023). Bireysel emeklilik sistemindeki katılımcıların memnuniyet ve güven algılarının incelenmesi. *Çalışma İlişkileri Dergisi*, 14(1), 1-19.

Siber Güvenlik Çalışmaları Üzerine Bibliyometrik Analiz

Araştırma Makalesi/Research Article

 Yusuf Bera SANLI¹,  Fırat BALTACI¹,  Emel GÜVEN¹,  Tamer EREN^{1*}

¹Endüstri Mühendisliği, Kırıkkale Üniversitesi, Kırıkkale, Türkiye

beraasanli@gmail.com, firatbaltaci5@gmail.com, emel-gvn@hotmail.com, tamereren@gmail.com

(Geliş/Received:24.04.2024; Kabul/Accepted:28.06.2024)

DOI: 10.17671/gazibtd.1473206

Özet—İşletmeler açısından rekabetin hızla yükseldiği günümüzde dijital dönüşüm ile ilgili yapılan altyapı çalışmaları oldukça fazladır ve işletmelerin verileri koruma süreçlerini sağlıklı bir şekilde yönetmesinde de siber güvenliğin etkili olduğu bilinmektedir. Bu çalışmada, siber güvenlik anahtar kelimesi adı altında yapılan bibliyometrik analiz sunulmuştur. Yükseköğretim Kurumu Ulusal Tez Merkezi (YÖKTEZ) arşivinde yer alan 408 lisansüstü tez ile ilgili karşılaştırmalı veriler analiz edilerek yorumlanmıştır. “Siber Güvenlik” alanında yapılan analiz sonucunda en fazla tezin 2022 yılında yayımlandığı görülmüştür. Üniversite türlerine göre en fazla tez devlet üniversitelerinde yayımlanmıştır. Devlet üniversiteleri adı altında yer alan Gazi Üniversitesi’nde en fazla yayım yapıldığı görülmüştür. Siber güvenlik alanında yapılan çalışmaların Fen Bilimleri Enstitüsü ve Bilgisayar Mühendisliği Ana Bilim Dalı’nda en fazla olduğu tespit edilmiştir. Bilgisayar Mühendisliği Bilimleri-Bilgisayar ve Kontrol konusunda en çok yayım bulunmaktadır ve araştırma yöntemi olarak ise en fazla nicel yöntem kullanılmıştır. Çalışma bu konuda gelecekte yazılacak lisansüstü tezleri desteklemeyi hedeflemektedir.

Anahtar Kelimeler— siber güvenlik, bibliyometrik analiz, lisansüstü tezler

Bibliometric Analysis On Cybersecurity Studies

Abstract—In today's world where competition is rapidly increasing for businesses, there are many infrastructure studies related to digital transformation, and it is known that cyber security is effective in businesses managing their data protection processes in a healthy way. In this study, a bibliometric analysis was presented under the keyword cyber security. Comparative data on 408 postgraduate theses in the archive of the Higher Education Council National Thesis Center (HECNTC) were analyzed and interpreted. As a result of the analysis conducted in the field of "Cyber Security", it was seen that the most theses were published in 2022. According to the types of universities, the most theses were published in state universities. It was seen that the most publications were made in Gazi University, which is under the name of state universities. It was determined that the studies conducted in the field of cyber security were the most in the Institute of Science and Technology and the Department of Computer Engineering. There are the most publications in the subject of Computer Engineering Sciences-Computer and Control, and the quantitative method was used the most as a research method. The study aims to support future postgraduate theses on this subject.

Keywords—cyber security, bibliometric analysis, graduate theses

1. GİRİŞ (INTRODUCTION)

Bilgi çağı olarak adlandırılan dönemde, bilginin değeri giderek artmaktadır. İnternet ve bilgisayar teknolojileri sürekli olarak gelişmekte olup, gereksinim duyduğumuz bilgiye daha hızlı ve kolay bir şekilde erişmemizi sağlamaktadır. İnsanlar interneti dijital dünyanın kısaltması olarak kullanırken, bu teknolojilerin arkasındaki insanlar ve toplumlarını nasıl etkilediğini içeren bir siber uzay bulunmaktadır[1]. Bilgi sistemleri teknolojilerindeki son ilerlemeler, işletmelerde çeşitli alanlarda otomasyonun benimsenmesine yol açmıştır. İşletmelerde verilerin giderek önemli bir kaynak haline gelmesi, verilere erişim, paylaşım, bilgi oluşturma ve kullanma ihtiyacını artırmıştır. Veri ve bilgi yönetimine olan talebin artmasıyla birlikte, veri tabanları, uygulamalar ve bilgi sistemlerinin güvenliğinin sağlanması son derece kritik hale gelmiştir. Verilerin ve bilgilerin yetkisiz erişimlere karşı korunması, yolsuzluklar kadar önemlidir. İnternetin yaygınlaşmasıyla, bu verilere birçok kişinin erişimi mümkün hale gelmiştir, bu nedenle verilerin ve uygulamaların korunması için etkili mekanizmalara ihtiyaç vardır[2]. Siber güvenlik, kontrol edilemeyen bir durumda endişe vericidir. Siber saldırılar ve siber suçlar potansiyel bir tehdit oluşturacaktır. Hükümetler arasında iç içe geçmiş ve birbirine bağımlı sistemlerden oluşan karmaşık ağlarda siber güvenlik saldırıları ve siber savaş tehdidi büyük maliyetler ve sorunlar getirebilir[3]. Siber saldırılar ve siber suçların artması, siber güvenlik araştırmalarına daha fazla odaklanmayı gerektirmekte ve bu konuda yapılan bilimsel çalışmaların kapsamı son yıllarda genişlemektedir. Bu çalışmanın amacı, siber güvenlik, siber suç, siber saldırılar, siber savaş araştırma eğilimlerinin bibliyometrik bir incelemesini sunmaktadır. Siber güvenlik alanındaki araştırmaların ve yayım performansının değerlendirilmesine ve alan hakkında farkındalık kazandırılmasına katkı sağlamak amaçlanmaktadır. Ayrıca siber güvenlik konusunda gelecekte hazırlanacak olan lisansüstü tezler için mevcut durumun ortaya konulması hedeflenmektedir. Böylelikle son dönemlerde giderek önemli hale gelen siber güvenlik konusunda yapılacak çalışmalara destek sağlanacağı düşünülmektedir. Yükseköğretim Kurumu Ulusal Tez Merkezi'nde (YÖKTEZ) bulunan tüm lisansüstü tezlerin siber güvenlik alanındaki çalışmalar Microsoft Excel tablolama aracı kullanılarak derlenmiştir.

2. KAVRAMSAL ÇERÇEVE (CONCEPTUAL FRAMEWORK)

1917'de Cole ve Eales, karşılaştırmalı anatomi tarihi üzerine yapılan çalışmaları 1550-1860 yılları arasında inceleyerek, araştırma alanının mevcut gelişimini bibliyometrik analiz ile ilk kez ortaya koyan araştırmacılar arasındadır[6]. "Bibliometric" terimi, Yunanca ve Latince kökenli olan "biblio" ve "metric" kelimelerinin

birleşiminden oluşur; "biblio" kitap veya belgelere, "metric" ise ölçüm anlamına gelir. Bu nedenle, "bibliometric" kelimesi kitap veya belgelere ilişkin bir ölçüm birimini ifade eder[7]. Bibliyometri, özünde sayım yapmaya dayalı bir inceleme alanı olarak belirtilmektedir. Bibliyometrik çalışmalar, belirli özelliklere sahip belgelerin veya yayımların analiz edilmesine dayanır ve bilimsel iletişimle ilgili çeşitli bulguların elde edilmesini sağlar[8]. Konu ile ilgili yapılan araştırmalarda; 2012 ve 2024 yıllarında yayımlanan çalışmalar yıllar bazında, üniversitelere, ana bilim dallarına ve araştırma yöntemi gibi bibliyometrik parametreler açısından incelenmiştir.

Çalışma sonucunda 2012 yılından itibaren "siber güvenlik" anahtar kelimesi altında yayımlanan tez sayılarının giderek arttığı görülmektedir. Bunun sebebi On Birinci Kalkınma Planında siber güvenlik konusunda farkındalık ve yetkinliklerin artırılacağından, ulusal siber güvenlik stratejisinin güncelleneceğinden, toplumsal yönelimin bu alana da odaklanması sağlanacağından, siber güvenlik ihtiyaçlarının planlanması için mekanizma geliştirileceğinden ve bu alanda yer alan firmalara teşvik verileceğinden bahsedilmiştir. Bu plan doğrultusunda 2019 yılında Siber Güvenlik ile ilgili yayımlanan tezlerde büyük bir artış olduğu görülmektedir[9]. On İkinci Kalkınma Planında ise ülkeler arası siber güvenlik güç dengelerinde belirleyici olması, siber güvenliğe ilişkin çalışmalara ağırlık verilmesi, siber güvenlik ürünlerinin kullanımının artırılacağı, bu konuda insanların bilinçlendirileceği beklenmektedir[10].

Bu çalışmada YÖKTEZ'de bulunan lisansüstü tezlerin, yıllara, üniversite türüne, üniversitelere, enstitülerine, ana bilim dallarına, konularına, araştırma yöntemlerine ve anahtar kelimelerine göre inceleme yapılmıştır. Kavramsal çerçeve incelendiğinde benzer bir çalışmaya rastlanmamıştır ve bu sebeple özgün bir içeriktir. Çalışma YÖKTEZ veri tabanında yer alan tezleri inceleyerek ileride yazılabilecek tez çalışmaları için yol gösterici olmayı hedeflemektedir. Siber güvenlik kapsamında farklı konularda çalışmalar yapılmış olsa da tezlerin mevcut durumunu inceleyen bibliyometrik bir analiz bulunmamaktadır. Bu bibliyometrik analiz mevcut durumu ortaya koyarak siber güvenlik konusunda gerçekleştirilecek lisansüstü tez çalışmalarına destek olmayı amaçlamaktadır.

3. YÖNTEM (METHOD)

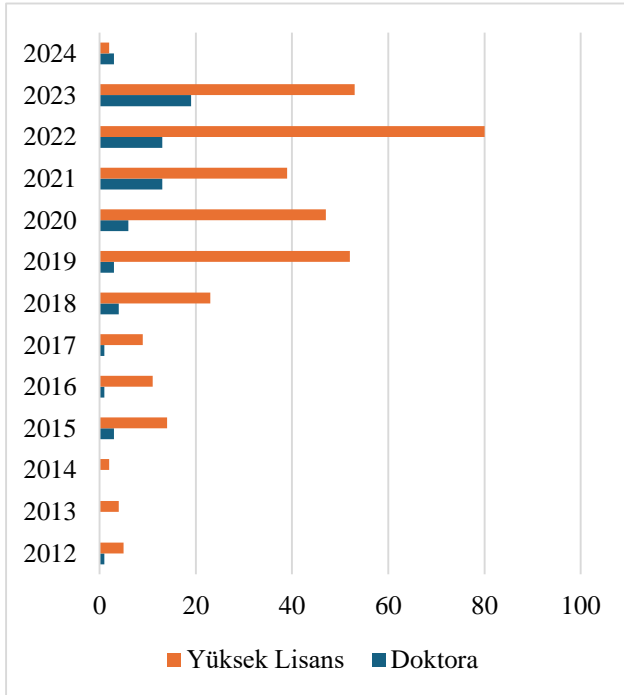
Bu çalışmada, tüm lisansüstü tezlerde "Siber Güvenlik" alanında hazırlananlar incelenmiştir. İncelemenin yapılabilmesi için YÖKTEZ'in tez tarama motorunda anahtar kelime bölümüne "Siber Güvenlik" yazılmıştır. Arama motorunda aranılacak alan olarak "özet" seçilmiştir. Tarama sırasında herhangi bir yıl sınırlaması yapılmamıştır. Siber güvenlik konusunda ilk çalışma 2012 yılında gerçekleştirilmiş olup 17.03.2024 tarihine kadar

olan çalışmalar analize dahil edilmiştir. Çalışma sadece YÖKTEZ veri tabanında bulunan ulusal tezleri kapsamaktadır. Gerçekleştirilen tarama sonucunda 341 yüksek lisans ve 67 doktora tezine ulaşılmıştır. Bu çalışma 17.03.2024 tarihinde sonlandırılmıştır. Tezlerin yıllara göre dağılımı, yayımlandığı üniversite, yayımlandığı enstitü, ilgili anabilim dalı, tezin araştırma yöntemi, tezin konusu, tez türü, tez yılı, tezin çözüm yöntemi, anahtar kelimeler olarak sıralanan kriterler belirlenmiştir. Derleme işleminde Microsoft Excel tablolama aracı kullanılmıştır.

4. BULGULAR (FINDINGS)

4.1. Tez Türlerinin Yıllara Göre Dağılımı (Distribution of Thesis Types by Years)

Siber güvenlik alanında gerçekleştirilen 341 yüksek lisans ve 67 doktora tezlerinin yıllara göre yayımlanan tez türü sayıları Şekil 1’de verilmiştir.



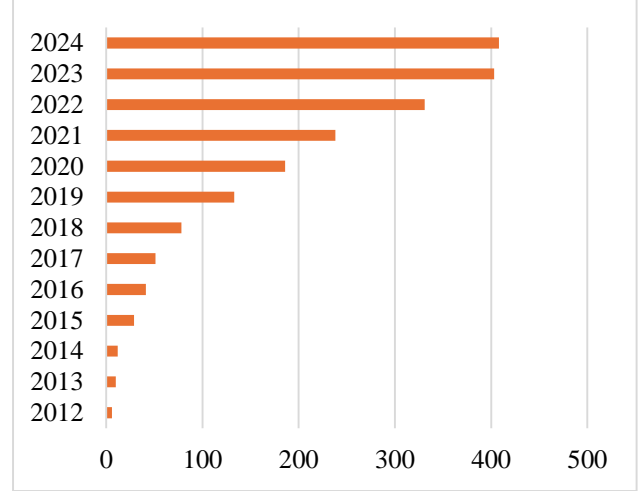
Şekil 1. Yıllara göre yayımlanan tez türü sayıları
(Figure 1. Number of published thesis types by years)

Çalışmaların özellikle 2019 yılında yüksek lisans tezi olarak artışa geçtiği görülmektedir. 2019 yılından 2024 yılına kadar üstel şekilde artış vardır. Benzer trend 2019-2024 doktora tezinde de görülmektedir. Ancak sayısal olarak doktora, yüksek lisans tezlerine göre daha azdır. Bu durum şöyle açıklanabilir. Akademik bir çalışmanın doktora seviyesine gelmesi için yüksek lisansın tamamlanması beklenmektedir. Doktora tezinin dört yıl sürdüğü düşünülürse 2024 yılına kadar artan bir trend olağandır. Doktora tezinin sayılarının azlığı genellikle akademik camiada doktora yapılması söylenebilir. Yüksek

lisans tezi gerek akademik gerekse sektörel olarak daha sık yapılmaktadır.

4.2. Lisansüstü Tezlerin Yıllara Göre Dağılımı (Distribution of Graduate Theses by Years)

Siber güvenlik alanında gerçekleştirilen 408 tezin yıllara göre yayımlanan tez sayıları Şekil 2’de verilmiştir.

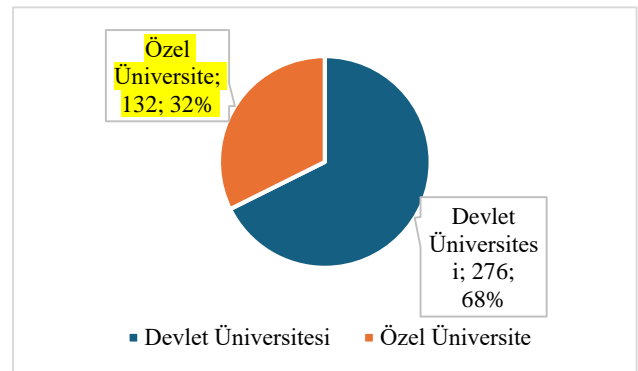


Şekil 2. Yıllara göre yayımlanan tez sayıları
(Figure 2. Number of theses published by year)

Lisansüstü tezlerin %1,22’si (5 tez) 2024 yılında, %17,64’ü (72 tez) 2023 yılında, %22,79’u (93 tez) 2022 yılında, %12,74’ü (52 tez) 2021 yılında, %12,99’u (53 tez) 2020 yılında, %13,48’i (55 tez) 2019 yılında yayımlanmıştır ve %19,14’ü (78 tez) diğer yıllarda yayımlanmıştır.

4.3. Üniversite Türüne Göre Dağılımı (Distribution by Type of University)

Siber güvenlik konulu lisansüstü tezlerin üniversite türlerine göre dağılımı Şekil 3’te verilmiştir.

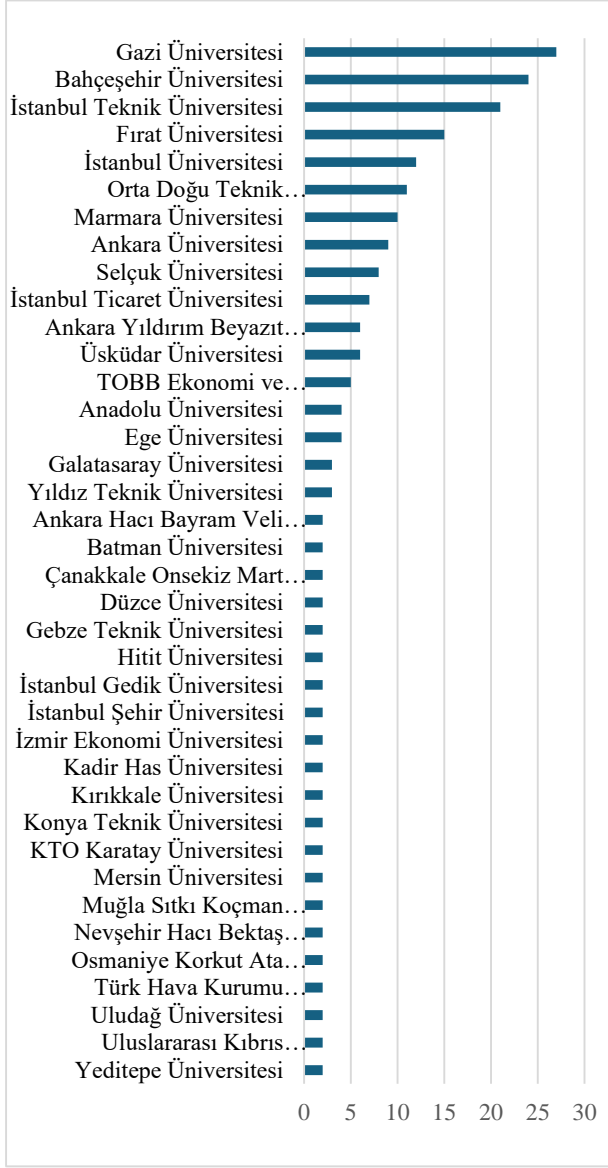


Şekil 3. Üniversite türü oranları
(Figure 3. Proportions of university type)

Yayımlanan tezlerin %68’si (276 tez) devlet üniversitesi ve %32’si (132 tez) vakıf üniversitesi olduğu gözlemlenmiştir.

4.4. Üniversitelere Göre Dağılımı (Distribution by Universities)

Siber güvenlik alanında gerçekleştirilen 408 tezin üniversitelerde yayımlanan tez sayıları Şekil 4'te verilmiştir.

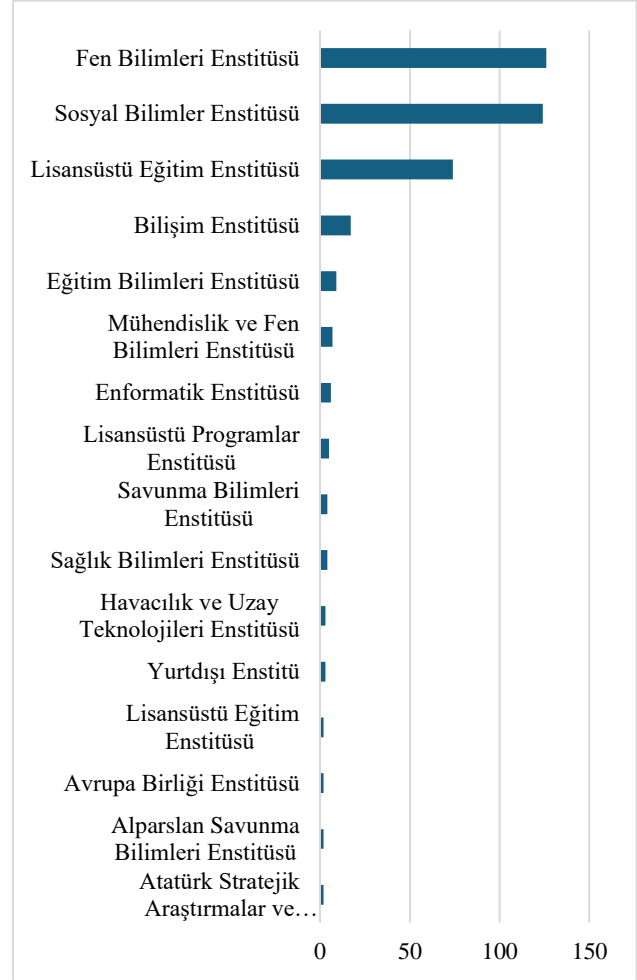


Şekil 4. Üniversitelerde yayımlanan tez sayıları
(Figure 4. Number of theses published in universities)

Lisansüstü tezlerin %6,61'i (27 tez) Gazi Üniversitesi'nde, %5,88'i (24 tez) Bahçeşehir Üniversitesi'nde, %5,14'ü (21 tez) İstanbul Teknik Üniversitesi'nde, %3,67'si (15 tez) Fırat Üniversitesi'nde, %2,94'ü (12 tez) İstanbul Üniversitesi'nde yayımlanırken %75,76'sı (309 tez) diğer üniversitelerde yayımlanmıştır.

4.5. Enstitülere Göre Dağılımı (Distribution by Institutes)

Siber güvenlik konusunda yayımlanan lisansüstü tezlerin enstitülere göre dağılımı Şekil 5'te verilmiştir.



Şekil 5. Enstitülere göre yayımlanan tez sayıları
(Figure 5. Number of published theses by institutes)

Lisansüstü tezlerin %30,88'i (126 tez) Fen Bilimleri Enstitüsü'nde, %30,39'u (124 tez) Sosyal Bilimler Enstitüsü'nde, %18,13'ü (74 tez) Lisansüstü Eğitim Enstitüsü'nde, %4,16'sı (17 tez) Bilişim Enstitüsü'nde, %2,20'si (9 tez) Eğitim Bilimleri Enstitüsü'nde ve %14,14'ü (58 tez) diğer enstitülerde yayımlanmıştır.

4.6. Ana Bilim Dallarına Göre Dağılımı (Distribution According to Main Departments)

Ana bilim dallarına göre siber güvenlik konusunda yayımlanan tez sayıları Şekil 6'da verilmiştir.



Şekil 6. Ana bilim dallarına göre yayımlanan tez sayıları
(Figure 6. Number of published theses according to main disciplines)

Lisansüstü tezlerin %16,17'si (66 tez) Bilgisayar Mühendisliği Ana Bilim Dalı'nda, %9,80'i (40 tez) Uluslararası Ana Bilim Dalı'nda %4,90'ı (20 tez) Siber Güvenlik Ana Bilim Dalı'nda, %3,92'si (16 tez) İşletme

Ana Bilim Dalı'nda, %3,67'si (15 tez) Yönetim Bilişim Sistemleri Ana Bilim Dalı'nda yayımlanırken %61,54'ü (251 tez) diğer ana bilim dallarında yayımlanmıştır.

4.7. Konulara Göre Dağılımı (Distribution by Subjects)

Siber güvenlik konusunda yayımlanan tezlerin konulara göre dağılımı Şekil 7'de gösterilmiştir.

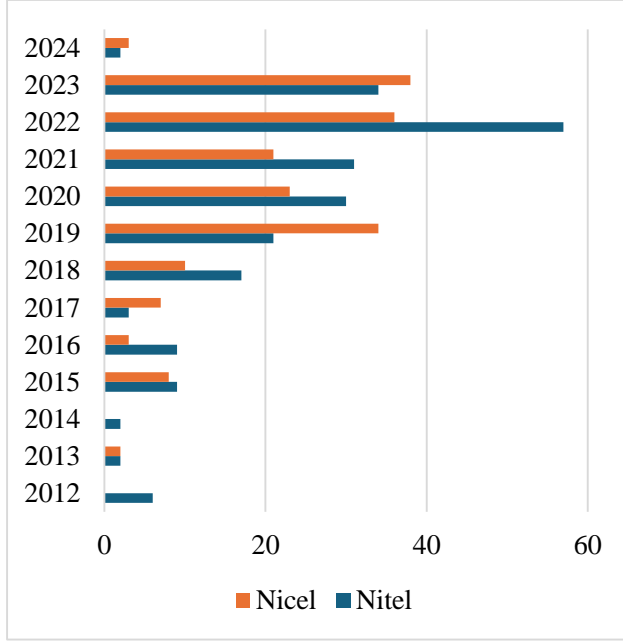


Şekil 7. Konulara göre yayımlanan tez sayıları
(Figure 7. Number of published theses by subject)

Lisansüstü tezlerin %40,44'ü (165 tez) bilgisayar mühendisliği bilimleri-bilgisayar ve kontrol, %9,31'i (38 tez) uluslararası ilişkiler, %6,37'si (26 tez) bilim ve teknoloji, %5,63'ü (23 tez) işletme, %3,92'si (16 tez) kamu yönetimi konularında yayımlanırken %34,33'ü (140 tez) diğer konularda yayımlanmıştır.

4.8. Lisansüstü Tezlerin Yıllara Göre Yöntem Dağılımı (Method Distribution of Graduate Theses by Years)

Lisansüstü tezlerin siber güvenlik alanında yıllara göre araştırma yöntemi dağılımı Şekil 8'de gösterilmiştir.



Şekil 8. Yıllara göre araştırma yöntemi sayıları
(Figure 8. Number of research methods by year)

Siber güvenlikle ilgili çalışmaların araştırma yöntemine göre dağılımı gösterilmiştir. 2023 yılındaki tezlerin %45,83'ü (33 tez) nitel, %54,17'si (39 tez) nicel araştırmadır. 2022 yılındaki tezlerin %45,16'sı (42 tez) nitel yayım iken, %54,84'ü (51 tez) nicel yayımdır. 2021 yılındaki tezlerin %40,38'i (21 tez) nitel, %59,62'i (31 tez) nicel araştırmadır. 2020 yılında %32,07'si (17 tez) nitel, %67,93'ü (36 tez) nicel araştırma olarak yayımlanmıştır. 2019 yılında %21,81'i (12 tez) nitel, %78,19'u (43 tez) nicel araştırma olduğu gözlemlenmiştir.

Siber güvenlik alanında yayımlanan tezlerde kullanılan yöntemler Tablo 1'de gösterilmiştir.

Tablo 1. Siber güvenlik alanında kullanılan yöntemler çizelgesi

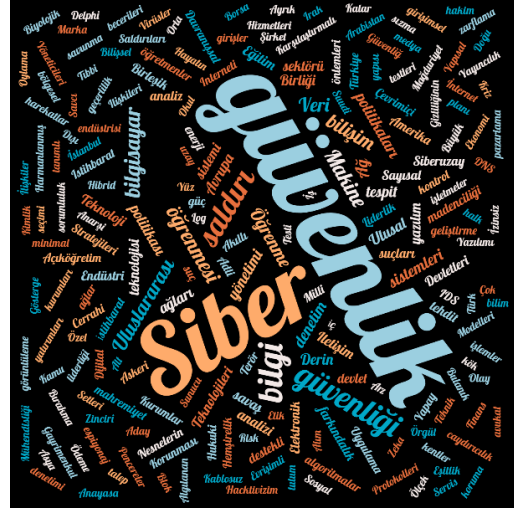
(Table 1. Chart of methods used in the field of cyber security)

Kullanılan Yöntemler	Miktar
Makine öğrenmesi yöntemleri	%25,49 (104 yöntem)
İstatistiki yöntemler	%14,71 (60 yöntem)
Derin öğrenme metotları	%7,11 (29 yöntem)
Matematiksel model algoritmaları	%3,19 (13 yöntem)
Karar verme yöntemleri	%2,70 (11 yöntem)
Nicel yöntemler	%46,81 (191 yöntem)

Tablo 1'de siber güvenlikle ilgili çalışmalarda kullanılan yöntemlerin miktarları ortak başlıklar altında verilmiştir. En fazla kullanılan nicel yöntem %48 (104 yöntem) ile makine öğrenmesi yöntemleri olmuştur.

4.9. Lisansüstü Tezlerde Kullanılan Anahtar Kelimeler (Keywords Used in Graduate Theses)

Lisansüstü tezlerde kullanılan anahtar kelimeler ile ilgili "Word Cloud Generator" üzerinden oluşturulan kelime bulutu Şekil 9'da verilmiştir.



Şekil 9. Tezlerde kullanılan anahtar kelimelerin bazıları
(Figure 9. Some of the keywords used in theses)

Şekil 9'da lisansüstü tezlerde kullanılan anahtar kelimelerin %9,38'inde (109 kelime) siber güvenlik, %3,87'sinde (45 kelime) siber saldırı, %2,15'inde (25 kelime) güvenlik, %1,63'ünde (19 kelime) bilişim teknolojisi, %1,37'sinde (16 kelime) bilgi güvenliği kelimeleri kullanılırken %81,6'sında (752 kelime) görülmektedir.

5. SONUÇ VE ÖNERİLER (CONCLUSIONS AND RECOMMENDATIONS)

Bibliyometrik analizlerde siber güvenlik ile ilgili yayımların danışmanları, tezlerin türleri, yayımlanma yılları, üniversite ve türleri, enstitüler, ana bilim dalları, araştırma yöntemleri, konuları ve anahtar kelimeleri yönünden analizleri yapılarak geleceğe yönelik yapılacak çalışmalara katkı sağlanması hedeflenmiştir. Özellikle mevcut durum ortaya koyularak lisansüstü tezler için gelecekte hazırlanacak çalışmalara destek vermek hedeflenmiştir.

Bu çalışma, lisansüstü tezlerin siber güvenlik alanındaki çalışmaları Microsoft Excel tablolama aracı kullanılarak derlemiştir. Yükseköğretim Kurulu Tez Merkezi veri tabanında 2012 yılından 2024 yılına kadar yer alan siber güvenlik anahtar kelimesi başlığı altında yayımlanan tezlerin bibliyometrik analizi yapılmıştır ve toplam 408 tez yayımı incelenmiştir.

Analiz sonucunda 2022 yılında en fazla tez yayımlandığı görülmüştür. Üniversite türlerine göre en fazla devlet üniversitelerinde tez yayımlanmıştır. Devlet üniversiteleri adı altında yer alan Gazi Üniversitesi'nde en fazla yayım yapıldığı görülmüştür. Siber güvenlik alanında yapılan çalışmaların Fen Bilimleri Enstitüsü ve Bilgisayar Mühendisliği Ana Bilim Dalında en fazla olduğu tespit edilmiştir. Bilgisayar Mühendisliği Bilimleri-Bilgisayar ve Kontrol konusunda en çok yayım bulunmaktadır ve araştırma yöntemi olarak ise en fazla nicel yöntem kullanılmıştır. İncelenen çalışmaların anahtar kelimeleri analiz edilmiş olup en çok kullanılan kelimenin siber güvenlik olduğu tespit edilmiştir.

Çalışmada elde edilen bulguların, gelecekte siber güvenlik ile ilgili gerçekleştirilecek çalışmalarda konu ile ilgili verilerin belirlenmesine katkı sağlayacağı düşünülmektedir. Analizde herhangi bir yıl kısıtlaması yapılmadan bütün tezler bibliyometrik analiz kapsamında değerlendirilmiştir. Elde edilen veriler literatür kapsamında bibliyometrik analiz için yeterli olduğu düşünülmektedir. Kapsamlı veri analizi için sistematik literatür çalışmaları önerilmektedir. Ayrıca siber güvenlik konusunda uluslararası tezleri kapsayan çalışmaların gerçekleştirilmesi önerilebilir.

KAYNAKLAR (REFERENCES)

- [1] W. P. inger, A. Friedman, "What everyone needs to know", Cybersecurity and cyberwar, Oxford University Press, 2014.
- [2] A. Hussain, N. Fatma, "Çin Kütüphaneciliği: Uluslararası Elektronik Dergisi (2006-2010) Kitabının Bibliyometrik Analizi", *Çin Kütüphaneciliği: Uluslararası Bir Elektronik Dergi*, 31, 1-14, 2011.
- [3] V. Kumar, J. Srivastava, A. Lazarevic, "Issues, Approaches, And Challenges. Springer", *Managing Cyber Threats* 2005.
- [4] U. Al,Türkiye'nin Bilimsel Yayım Politikası: **Atf Dizinlerine Dayalı Bibliyometrik Bir Yaklaşım**, Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, 2014.
- [5] E. Şahin, G. Akdağ, C. Çakıcı, O. Neslihan, "Gastronomi ve mutfak sanatları anabilim dallarında yayımlanan tezlerin bibliyometrik analizi", *Güncel Turizm Araştırmaları Dergisi*, 2(Ek1), 30-41, 2018.
- [6] M. McLennan, **The Global Risks Report 17th Edition**, İsviçre, 2020.
- [7] U. Al, R. Coştur, "Türk Psikoloji Dergisi'nin Bibliyometrik Profili", *Türk Kütüphaneciliği*, 21(2), 142-163, 2007.
- [8] M. N. Alabay, "Sosyal Medyada Tüketiciler ve Pazar Bölümleme Uygulamaları", *inet-tr*, 2011.
- [9] Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji Ve Bütçe Başkanlığı On Birinci Kalkınma Planı (2019-2023)https://www.sbb.gov.tr/wpcontent/uploads/2022/07/On_Birinci_Kalkinma_Planı-2019-2023.pdf, 07.2022.
- [10] Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji Ve Bütçe Başkanlığı On İkinci Kalkınma Planı (2024-2028)https://www.sbb.gov.tr/wpcontent/uploads/2023/12/On-Ikinci-Kalkinma-Planı_2024-2028_11122023.pdf, 12.2023.

A New Hybrid Model for Artificial Intelligence Assisted Tire Defect Detection: CTLDF+EnC

Araştırma Makalesi/Research Article

 Özcan ASKAR¹,  Ramazan TEKİN*²

¹Batman University, Graduate School of Education, Department of Information Technologies,
²Batman University, Faculty of Engineering and Architecture, Department of Computer Engineering,
ozcanaskar@gmail.com, ramazan.tekin@batman.edu.tr
(Geliş/Received:04.04.2024; Kabul/Accepted:08.07.2024)
DOI: 10.17671/gazibtd.1465294

Abstract— This paper focuses on an artificial intelligence based worn tire detection system proposed to detect cracks in the tires of vehicle drivers. Although drivers are generally aware of the importance of tire tread depth and air pressure, they are not aware of the risks associated with tire oxidation. However, tire oxidation and cracks can cause significant problems affecting driving safety. In this paper, we propose a new hybrid architecture for tire crack detection, CTLDF+EnC (Cascaded Transfer Learning Deep Features + Ensemble Classifiers), which uses deep features from pre-trained transfer learning methods in combination with ensemble learning methods. The proposed hybrid model utilizes features from nine transfer learning methods and classifiers including Stacking, Soft and Hard voting ensemble learning methods. Unlike X-Ray image-based applications for industrial use, the model proposed in this study can work with images obtained from any digital imaging device. Among the models proposed in the study, the highest test accuracy value was obtained as 76.92% with the CTLDF+EnC (Stacking) hybrid model. With CTLDF+EnC (Soft) and CTLDF+EnC (Solid) models, 74.15% and 72.92% accuracy values were obtained respectively. The results of the study show that the proposed hybrid models are effective in detecting tire problems. In addition, a low-cost and feasible structure is presented.

Keywords— transfer learning, deep features, ensemble learning, tire cracks.

Lastik Kusurlarının Tespiti için Yapay Zeka Destekli Yeni Bir Hibrit Model: CTLDF+EnC

Özet— Bu çalışma, araç sürücülerinin lastiklerindeki çatlakları tespit etmek için önerilen yapay zeka tabanlı bir aşınmış lastik tespit sistemine odaklanmaktadır. Sürücüler genellikle lastik diş derinliği ve hava basıncının öneminin farkında olsalar da, lastik oksidasyonu ile ilişkili risklerin farkında değildir. Ancak, lastik oksidasyonu ve çatlakları sürüş güvenliğini etkileyen önemli sorunlara neden olabilir. Bu makalede, lastik çatlağı tespiti için, önceden eğitilmiş transfer öğrenme yöntemlerinden elde edilen derin özellikleri topluluk öğrenme yöntemleriyle birleştirerek kullanan yeni bir hibrit mimari olan CTLDF+EnC (Basamaklandırılmış Transfer Öğrenme Derin Özellikler + Ensemble Sınıflandırıcılar) önerilmektedir. Önerilen hibrit model, dokuz transfer öğrenme yönteminden gelen özellikleri ve İstifleme, Yumuşak ve Katı oylama topluluk öğrenme yöntemlerini içeren sınıflandırıcıları kullanmaktadır. Endüstriyel kullanıma yönelik X-Ray görüntü tabanlı uygulamalardan farklı olarak bu çalışmada önerilen model herhangi bir dijital görüntüleme cihazından elde edilen görüntülerle çalışabilmektedir. Çalışmada önerilen modeller arasında en yüksek test doğruluk değeri %76,92 olarak CTLDF+EnC (İstifleme) hibrit modeli ile elde edilmiştir. CTLDF+EnC (Yumuşak) ve CTLDF+EnC (Katı) modelleri ile sırasıyla %74,15 ve %72,92 doğruluk değerleri elde edilmiştir. Çalışmanın sonuçları, önerilen hibrit modellerin lastik sorunlarını tespit etmede etkili olduğunu göstermektedir. Ayrıca, düşük maliyetli ve uygulanabilir bir yapı sunulmuştur.

Anahtar Kelimeler— transfer öğrenme, erin öznitelikler, topluluk öğrenme, lastik çatlakları.

1. INTRODUCTION

Tire problems are a difficult to detect issue that can cause serious risks when traveling at high speeds. The tire is one of the vehicle's key safety elements and even simple defects can affect safe driving and cause accidents. Therefore, tire quality is critical to the safe and smooth movement of the vehicle. Drivers often overlook problems such as tire aging, oxidation and cracking and do not carry out regular checks [1].

Contamination of tire components on the production line can lead to the production of defective tires, which can cause explosions at high speeds and risks to life. As traditional methods are not efficient, computer-aided systems are important to improve tire quality. In these systems, traditional visual detection [2,3] methods as well as deep learning [4,5] techniques can also be used. Tires may have cracks that can affect the safety of the vehicle while driving. Tire problems pose significant risks to drivers, especially at high speeds. However, while drivers are often aware of risks such as tire pressure and tread depth, they are not aware of problems related to tire life. Therefore, computer and AI-assisted applications using digital images are useful tools when manual checks are not safe [1].

Deep learning methods are used as a successful tool in many fields thanks to their ability to extract features [6–15]. However, they usually require large amounts of data and long training times. To overcome this problem, pre-trained Transfer Learning (TL) methods are used to speed up the training process of deep networks and provide effective learning with less data. In this study, the methods Xception [16], VGG16 [17], NASNet [18], ResNet50 [19], DenseNet [20], InceptionV3 [21] and MobileNet [22] advanced deep networks, such as the one used in this study.

In this study, a hybrid model is proposed to detect cracks on the tire surface and distinguish normal/cracked tires. The proposed model is a modified version of Özaydın and Tekin's DeepFeat-E model [9]. In this model, deep features from pre-trained TL models are reduced and then cascaded and applied to stacking [23] and voting [24] ensemble classifiers. This approach enables more stable and accurate detection of oxidation-related problems in tires. Thus, it can enable drivers to detect tire problems in advance and minimize risks by taking preventive measures. Furthermore, the proposed system has the potential to save time and labor and increase efficiency for organizations with a large number of vehicles, such as fleets, logistics companies and public transport providers.

2. LITERATURE REVIEW

Defective tire detection methods are generally divided into two main categories based on the type of defect detection algorithm: traditional visual detection and deep learning-based methods. Traditional visual detection

techniques can be classified as statistical, frequency and model-based methods. In the literature, there are several methods that use X-Ray images to detect tire defects during production. In a study by Zhao and Qin [3], tire texture images obtained by non-destructive X-Ray were used for the detection of internal structural tire defects on the production line. The researchers extracted the features of these images with the local inverse difference moment (LIDM) method, calculated the Hausdorff distance using the LIDM features to obtain a defect feature map (DFM) and proposed a pixel-level defect detection algorithm with high accuracy. Similarly, in another study by Guo et al. [2], an effective defect detection method that exploits the feature similarity of tire X-Ray images to detect tire surface abnormalities that may occur during production for the production of quality tires is proposed. In this method, the feature vectors of the tire images are obtained using a local kernel regression (LKR) descriptor to evaluate the feature dissimilarity of pixels and an abnormality map is generated. A simple thresholding process on the abnormality map was used to successfully locate the defects. It is reported that this method gives successful results for both cheek images and back images.

Another area of research is the studies in which convolutional neural network (CNN) models, which are deep learning methods, are included in the literature, as well as methods that directly use tire images. Wang et al. [5] successfully detected tire faults that occurred during production using X-Ray images of tires produced in an industrial area with a fully convolutional neural network (FCN) structure. The network architecture includes the first stage, where a traditional deep network method is applied to extract tire image features. In the second stage, a sampling layer is added to obtain outputs of the same size as the original images. In the last stage, the scaled feature map is added to the outputs and the correct defects are obtained as a result. In this study, the basic architecture of VGG16 was used using a total of 914 images and the results obtained were validated by comparing with AlexNet, VGG11, VGG13 and VGG16.

Another study by Zheng et al. [4] aimed to perform tire defect detection from X-Ray images to detect invisible defects in the internal structure of tires. In this research, a Concise Semantic Segmentation Network (Concise-SSN) model based on VGG16 is proposed for automatic tire visual inspection. This model performs purely pixel-based defect detection. The model combines the capabilities of an optimized semantic segmentation network and an integrated convolutional neural network. Experimental results show that the proposed model outperforms FCN, Mask R-CNN, Faster R-CNN, SegNet and U-net networks in segmentation and classification. The accuracy of the Concise-SSN model in detecting defective tires is reported to be 96.5%.

Another study by Zhang et al. [25] proposes a segmentation method based on wavelet transform using X-Ray images to detect tire surface defects. In this method, the edge information utilizes wavelet transform features that can be represented by larger coefficients in the high frequency band. The larger curve coefficients

corresponding to the edge information are selected by thresholding. Comparisons show that the proposed method outperforms traditional edge detection methods (such as Canny, Sobel and LoG).

The TireNet method developed by Li et al. [26] uses X-Ray images of tires to reduce the high rate of tire returns due to manufacturing defects. In this method, a Siamese network is used for image feature extraction and then the model is built by adding the structure of the Siamese network to the Fast R-CNN classifier. Using 120,000 labeled tire images, the TireNet model outperformed expectations by reducing the error rate to 0.17% compared to Faster R-CNN, SSD and YOLO models.

Finally, Lin [1] proposed a ShuffleNet model with a deep learning architecture for the detection of tire defects caused by oxidation. The proposed ShuffleNet model is able to effectively detect defects on normal tire images and outperforms the GoogLeNet, traditional ShuffleNet, VGGNet and ResNet models.

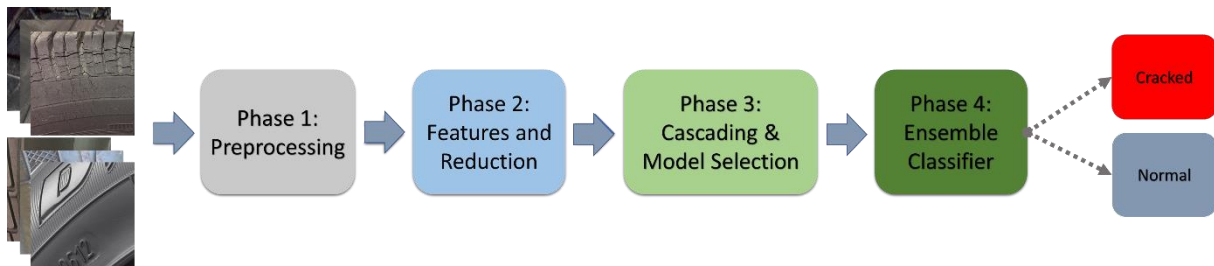


Figure 3.1. Stages of the proposed hybrid system.

Phase 1 (Preprocessing): This stage includes the data preprocessing step. All visual data are scaled to a size (224x224x3) suitable for the TL models. Then, image manipulation operations such as horizontal/vertical translation, size scaling, horizontal/vertical rotation, angular rotation, and brightness adjustment are applied for data enhancement. These operations are applied to make the model more robust to various conditions. Then, the dataset is divided into training and test subsets and analyzed, and the next step is taken.

Phase 2 (Features and Reduction): In this stage, the deep features obtained from each pre-trained TL model used are reduced. Each TL model is applied independently of the others to extract and then reduce the deep features of the training and test datasets.

Phase 3 (Cascading and Model Selection): In this stage, the deep features of the reduced training and test datasets obtained from each ML method are combined (cascading). Then, the five most successful ML methods are selected using the training set. The selection of the five best models is performed using 10 cross-validations and the AUC scale. The top five ML models are selected to be used in the next stage, the ensemble classifier stage.

Phase 4 (Ensemble Classifier): In this final stage, normal and cracked tires are predicted by stacking and voting based ensemble classifiers using the top five ML

3. MATERIAL and METHOD

In this paper, we propose a hybrid system, CTLDF+EnC, which aims to detect cracked and normal tires. This hybrid system combines deep features extracted from the outputs of a set of TL models, where traditional machine learning (ML) methods are evaluated in an ensemble classifier.

3.1. General Structure of the System

The hybrid architecture proposed in this study aims to detect cracked and normal tires. Figure 3.1 shows the general structure of the proposed system including four basic stages. The dataset used in the study contains 1028 images. These images include digital camera images of oxidized (cracked) and non-oxidized (normal) tires.

models identified in the previous stages and the resulting cascaded deep features.

All applications using the hybrid system proposed in this study and direct TL models are implemented in Python programming language using Tensorflow, Keras and scikit-learn libraries. The models and analyses were performed in a personal computer environment with an AMD Ryzen 7 (5800H) processor with 16 CPUs at 3.2GHz, 4GB GDDR6 memory / Nvidia GeForce RTX 3050 GPU at 1.5GHz and 16 GB RAM.

3.2. Data Set

In this study, we use a publicly available tire image dataset that includes cracked (oxidized) and normal (non-oxidized) tires. The dataset consists of a total of 1028 images, including tire sidewall and tread images [27]. The dataset used in this study has not undergone any augmentation or preprocessing, and the training and test sets are shared separately. The dataset is publicly available and can be downloaded from Harvard's dataverse webpage [27]. The current dataset contains 491 normal (non-oxidized) and 537 cracked (oxidized) tire images. Table 3.1 details the distribution of cracked and normal tire images between the training and test datasets.

Table 3.1. Data set image distributions.

	Normal	Cracked	Total
Train	376	327	703
Test	115	210	325
Total	491	537	1028

As seen in the table, since there are 703 samples in the training set and 325 samples in the test data set, the

training-test data set ratios are approximately 68% and 32% respectively. The resolutions of the images in the dataset are not standardized and are composed of images with different resolutions. For this reason, the images were adjusted to standard sizes in the preprocessing stage and data augmentation methods such as horizontal/vertical shifting, horizontal rotation, scaling, rotation, etc. were applied to prevent overlearning. Examples of normal and cracked images are shown in Figure 3.2.

**Figure 3.2.** Sample tire images, a) Cracked and b) Normal.

3.3. Proposed Hybrid System

In this study, two main applications were carried out. In the first application group, pre-trained TL models were directly used to predict tire conditions. In the second application group, the proposed hybrid model named CTLDF+EnC (Cascaded Transfer Learning Deep Features + Ensemble Classifiers) is used. This model has a structure in which deep features obtained from TL models are cascaded and applied to ensemble classifiers.

Figure 3.3 shows the block diagram of the CTLDF+EnC hybrid system proposed in this study. In the figure, the Deep Features Generator (DFG) block extracts the deep features of a given TL model and the Reduced Features Set (RFS) is generated by the feature reduction process. This process is performed separately for the 9 TL models used in this study. Since the highest performance values were obtained with the Extra Tree Classifier (ETC), ETC was used as the feature reduction method and IDs were created. The number of features before and after reduction for each TL is presented in Table 3.2.

Table 3.2. Number of features before and after reduction for each TL model.

TL Models	Before Reduction	After Reduction
Xception	2,048	527
NASNet	4,032	1,119
MobileNet	1,024	232
DenseNet169	1,664	360
DenseNet201	1,920	381
VGG16	512	139
InceptionV3	2,048	573
ResNet50V2	2,048	497
ResNet101V2	2,048	526
Total	35,714	4,354

In the Cascading and Ensemble Classification (CEC) block, the RFSs of each TL model are cascaded. The Model Selector evaluates 14 different ML methods after 10 cross-validations with these cascaded features, and the first k chooses the ML model ($k = 5$). Although the selected ML methods vary for each TL, RF (Random Forest), LDA (Linear Discriminant Analysis), LGBM (Light Gradient Boosting Machine), LR (Logistic Regression), ETC (Extra Trees Classifier), GB (Gradient

Boosting) and KNN (K Neighbors Classifier) ML models were generally used. When the features obtained from TL models were cascaded, LDA, ETC, LR, GB and LGBM were selected as the best $k = 5$ ML models (ranked according to the top 5 AUC metric). Table 3.3 presents the performance metrics of the ML Models ranked according to the AUC value.

Table 3.3. Best ML models ranked by AUC value.

Model	Accuracy	AUC	Recall	Prec.	F1
Linear Discriminant Analysis	0.9474	0.9863	0.976	0.9311	0.9522
Extra Trees Classifier	0.9361	0.9836	0.9841	0.9074	0.9433
Logistic Regression	0.9517	0.9828	0.9814	0.9337	0.9564
Gradient Boosting Classifier	0.9531	0.981	0.9894	0.9293	0.9579
Light Gradient Boosting Machine	0.9503	0.9805	0.9841	0.9288	0.9552
Random Forest Classifier	0.9375	0.9798	0.9841	0.9092	0.9444
K Neighbors Classifier	0.9488	0.9705	0.9706	0.9369	0.953
Ada Boost Classifier	0.9317	0.9684	0.9496	0.9261	0.9372
Naive Bayes	0.9261	0.9372	0.9733	0.8998	0.934
Decision Tree Classifier	0.8734	0.8727	0.8859	0.8791	0.882
Quadratic Discriminant Analysis	0.5065	0.5087	0.4814	0.5466	0.5086
Dummy Classifier	0.5348	0.5	1.0	0.5348	0.6969
SVM - Linear Kernel	0.9531	0.0	0.9788	0.9377	0.9574
Ridge Classifier	0.9559	0.0	0.9787	0.9428	0.9599

The best k ML models selected in the Ensemble Learning Model (EnsLM) block are used in the Ensemble Classifier sub-block. This sub-block predicts tire state (Cracked and Normal) using three different ensemble

classifiers, one based on Stacking and the other two based on voting, using selected ML models.

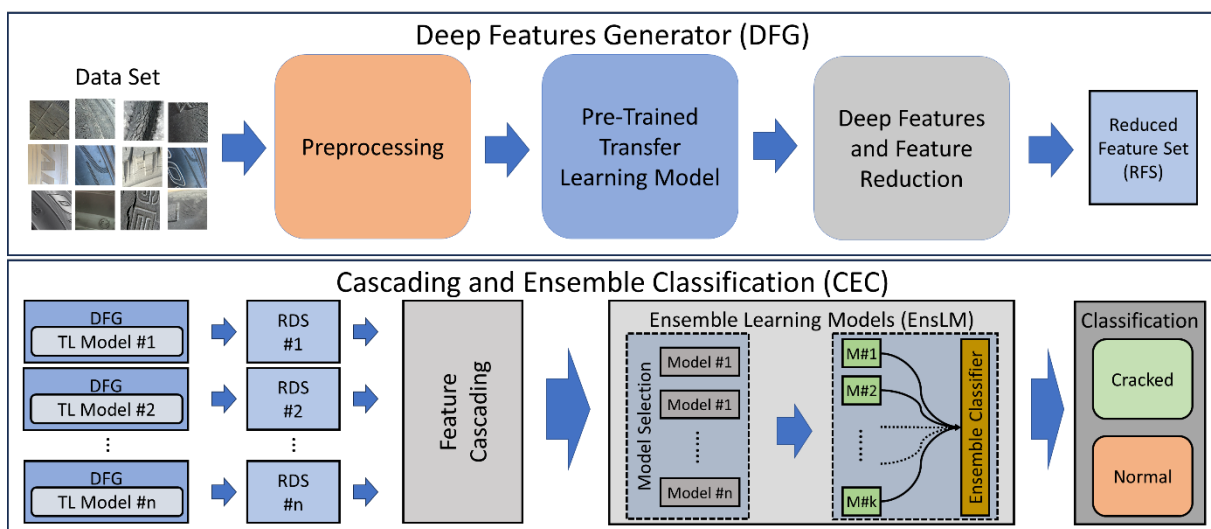


Figure 3.3. Schematic representation of the proposed hybrid system

3.4. Pre-Trained Transfer Learning Models

Transfer learning methods are a technique that enables the use of pre-trained models in a new problem or application and provides an effective solution in ML processes. This methodology is not considered as a different type of ML algorithm, but rather as a strategy or method used to train models. It involves applying the parameters and weights obtained from previous training to a new problem. The reused pre-trained model must have a high level of generality in order to be used in different problems. TL models can be applied to new or different problems without training, saving time and resources for training. Through reuse, problems such as resource shortages and long training times are minimized. With these advantages, TL offers more effective and efficient solutions in ML processes.

TL also offers a solution to the time-consuming process of accurately labeling large datasets [28]. This is a significant advantage, especially when considering the large datasets required to train a ML algorithm. Transfer methods are often preferred when large resources are required for the training phase in a system. Since the pre-trained structures of these models can be used directly, they can be used directly for similar problems related to the model. These aspects of transfer models make them general-purpose. TL allows for the development of more generalizable models rather than models being strictly bound to a training dataset. In this way, the models developed can be used under varying conditions and with different data sets.

3.5. Feature Selection with Extra Tree Classifier

Feature selection is an important step in ML models to improve prediction accuracy and reduce computational cost. For this purpose, the Extra Tree Classifier (ETC) method is used in this study to identify the most appropriate deep features obtained from ML models [29]. ETC is a decision-based method similar to the Random Forest classifier and provides a common framework between feature selection and classification. The method has the ability to generate many sub-trees and randomly select subsets. For feature selection, features are evaluated using the Gini metric and the most important features are identified. In this way, redundant or low-contributing features are eliminated and the complexity of the model is reduced, reducing computational cost and increasing success.

$$Gini = 1 - \sum_{i=0}^{c-1} p_i(t)^2 \quad (3.1)$$

where c is the number of unique classes at this node and $p_i(t)$ is the frequency of class i at node t .

3.6. Ensemble Classification Methods

Ensemble classification methods are approaches that use multiple decision makers instead of a single decision maker in order to reduce the number of incorrect predictions and to increase the achievement. In this study, the best five models are selected among traditional ML models and the best five models are selected among 14 models based on the AUC metric after 10 cross-validation procedures. In the applications, the features obtained from the pre-trained TL models were reduced and cascaded, and then applied to ensemble classifiers using these selected models. The common decision of the ensemble classifiers was determined by Stacking and Voting based ensemble learning methods. Voting strategies include two main approaches, Hard Voting and Soft Voting; Hard Voting is based on the majority decision, while Soft Voting uses the average of the prediction probabilities of the available classifiers. These strategies aim to achieve more efficient and reliable classification results.

3.6.1. Voting Based Ensemble Classifiers

Voting-based ensemble classifiers aim to obtain a common ensemble prediction by combining the classification predictions of various ML techniques. The most commonly used methods are called Hard Voting and Soft Voting. In Hard Voting, the prediction of the majority of classifiers in the ensemble is taken and this prediction is considered as the collective prediction of the ensemble. In Soft Voting, the ensemble prediction is formed by averaging the probabilistic weights of the classifiers' predictions. The contribution of each classifier is calculated with a specific weight and the final prediction is based on this weighted sum. These methods aim to achieve a more reliable and efficient classification result by combining the power of different classifiers in the ensemble. In this way, more accurate predictions can be made, avoiding the potential mispredictions that can arise from a single model. For example, these ensemble classifiers, when predicting the class of a sample x , allow to choose between k class $\{s_1, s_2, s_3, \dots, s_k\}$ predictions of n different classifiers $\{h_1, h_2, h_3, \dots, h_n\}$ according to certain criteria [24].

The Hard Voting method accepts the prediction of the majority of classifiers in the ensemble as the final prediction of the ensemble. In this method, the predictions of the classifiers in the ensemble are equally weighted and only the decision of the majority is taken into account. In this case, in the Hard Voting approach, the class of any instance x in the dataset is predicted as follows [24],

$$H(x) = s_{\sum_{i=1}^n h_i^j(x)} \quad (3.2)$$

The Soft Voting method makes class prediction by averaging the probabilistic weights of the predictions of the classifiers in the ensemble. This method aims to achieve a more accurate classification result by weighting the predictions of different classifiers in the ensemble according to their reliability. In this case, in the Soft Voting approach, the class of any instance x in the dataset is estimated as follows [24]:

$$H(x) = s_{\sum_{i=1}^n w_i h_i^j(x)} \quad (3.3)$$

The weight of the prediction of each classifier h_i among the other predictions in the ensemble is denoted by w_i .

3.6.2. Stacking Based Community Classifier

The Stacking Based Ensemble Classifier is a method proposed by Wolpert [23] and basically consists of two stages. In the first stage, the predictions of the different methods used in ensemble classifiers are obtained. Then, these predictions are processed by a meta-classifier to produce the final prediction of the ensemble. The aim of this approach is to improve the accuracy by balancing the incorrect predictions of a single classifier with the predictions of other classifiers. In this way, more reliable predictions are obtained through ensemble classifiers [23].

3.7. Performance Evaluation Metrics

In this study, a confusion matrix is used to analyze the performance of the proposed hybrid models and the TL methods used for comparison. The confusion matrix is presented in the form of a table as in Table 3.4, which shows the number of correct and incorrect classifications between the actual class labels of the examples in the dataset and the prediction labels of the models used.

Table 3.4. Confusion matrix

Actual/ Estimated	Cracked	Normal
Cracked	True Positive (TP)	False Positive (FP)
Normal	False Negative (FN)	True Negative (TN)

Where, TP (True Positive) refers to how many images with cracked true class labels are correctly predicted by the model used. TN (True Negative) refers to how many images with normal true class labels are correctly predicted by the model used. FP (False Positive) refers to how many images with real class labels cracked are incorrectly predicted as normal by the model used, while FN (False Negative) refers to how many images with real class labels normal are incorrectly predicted as cracked by the model used.

TP (True Positive): Refers to how many images with true class labels "Cracked" were correctly predicted by the model.

TN (True Negative): Refers to how many of the images with true class labels "Normal" were correctly predicted by the model.

FP (False Positive): Refers to how many of the images with true class labels "Normal" are predicted as "Cracked" by the model.

FN (False Negative): Refers to how many images with real class labels "Cracked" are predicted as "Normal" by the model.

The following four metrics were used to analyze and evaluate the performance of the models used in the study [30].

Accuracy: Refers to the ratio of correctly estimated samples to the total number of samples and is calculated by the formula below,

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3.4)$$

Precision: Refers to the probability that samples predicted as Cracked are actually Cracked and is calculated by the following formula,

$$Precision = \frac{TP}{TP + FP} \quad (3.5)$$

Recall: Refers to the proportion of samples whose true class is Cracked that are correctly predicted to be Cracked and is calculated by the formula below,

$$Recall = \frac{TP}{TP + FN} \quad (3.6)$$

f1-score: Provides a balanced measure of performance by taking the harmonic mean of precision and recall and is calculated by the following formula,

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (3.7)$$

4. EXPERIMENTAL ANALYSIS

In order to compare the performance of the hybrid model proposed in this study, firstly, a set of applications were performed in which each of the Transfer Learning (TL) models was used as a direct classifier separately. In another set of applications, within the proposed method, after the features of all TC models are obtained and reduced, they are combined together and Stacking, Soft and Hard Voting ensemble classifiers are used separately. In all the applications, the classifiers used were aimed to classify oxidized (Cracked) and non-oxidized (Normal) tires. Table 3.5 presents the values of the performance metrics obtained on the training and test datasets for the

first applications where the TL models were used directly. In these applications, the highest training and test accuracies were obtained with the ResNet50V2 TL model

with 91.89% and 70.77%, respectively. The lowest accuracy values were observed for the VGG16 TL model with 72.69% and 56.31%, respectively.

Table 3.5. Training and test success criterion values of the TL models.

Model Name	Train (%)				Test (%)			
	Acc.	Prec.	Rec.	f1-scr.	Acc.	Prec.	Rec.	f1-scr.
Xception	89.05	90.41	88.43	88.79	61.23	71.65	69.21	61.04
NASNet	91.18	91.4	90.94	91.09	66.15	71.79	72.04	66.15
MobileNet	87.2	88.41	86.58	86.91	57.54	67.1	65.18	57.32
DenseNet169	86.91	89.11	86.09	86.49	59.38	71.45	67.98	59.01
DenseNet201	90.9	91.2	90.61	90.8	68.92	74.24	74.77	68.91
VGG16	72.69	74.03	73.41	72.61	56.31	54.79	55.18	54.49
InceptionV3	88.19	88.39	87.93	88.08	65.54	71.48	71.56	65.54
ResNet50V2	91.89	92.11	91.66	91.82	70.77	74.22	75.61	70.68
ResNet101V2	90.9	91.8	90.41	90.73	63.08	71.07	70.05	63.03

Figure 3.4 shows the training and test confusion matrices of the most successful TL model, ResNet50V2. In the Confusion matrix for the training dataset in Figure 3.4.(a), out of 327 cracked tire images, 289 were correctly classified and 38 were incorrectly predicted. Of the normal tire images, 357 were correctly classified and 19 were incorrectly predicted. In total, there are 376 normal

tire images. In the complexity matrix for the test dataset in Figure 3.4.(b), 124 out of 210 cracked tire images are correctly classified and 86 are incorrectly predicted. Of the normal tire images, 106 were correctly classified and 9 were incorrectly predicted. In total, there are 115 normal tire images.

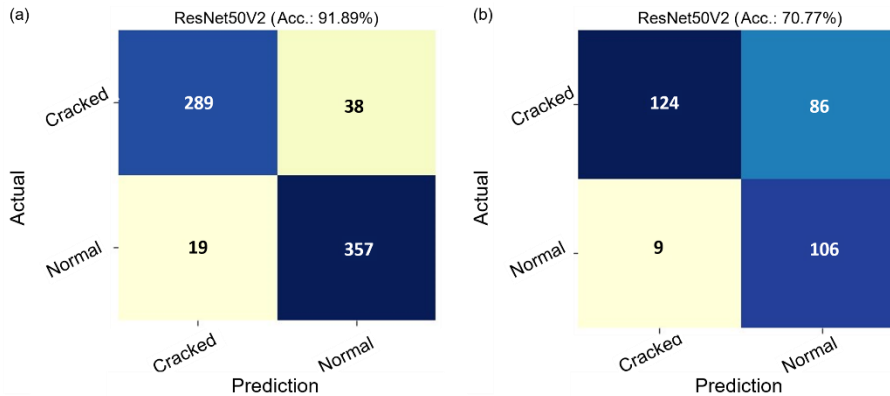


Figure 3.4 ResNet50V2 TL training and test confusion matrices.

Table 3.6 shows the performance of the CTLDF+EnC based ensemble classifiers on the test dataset. The table shows the performance of all TL models when deep features are applied to the ensemble classifiers. The highest success rate is obtained with the hybrid model with Stacking ensemble classifier (CTLDF+EnC(Stacking)), which has an accuracy of 76.92% and precision, sensitivity and f1-score of 79.10%, 81.36% and 76.75% respectively. The hybrid models of Soft CTLDF+EnC(Soft) and Hard CTLDF+EnC(Hard) voting methods have an accuracy of 74.15% and 72.92%

respectively. The precision, sensitivity and f1-score values of these models are given in Table 3.6. It was observed that all ensemble classifiers achieved 100% success rate on the training dataset. The reason for this difference between the accuracy values for the training and test datasets is considered to be overfitting. In the case of overfitting, while the model learns each sample in the training set very well, it loses its generalization capability and classifies the samples in the test data set, which it has never seen, with low accuracy.

Table 3.6. CTLDF+EnC based ensemble classifier performance values.

Model	Train (%)				Test (%)			
	Acc.	Prec.	Rec.	f1-scr.	Acc.	Prec.	Rec.	f1-scr.
CTLDF+EnC(Stacking)	100	100	100	100	76,92	79,1	81,36	76,75
CTLDF+EnC(Soft)	100	100	100	100	74,15	77,26	79,02	74,05
CTLDF+EnC(Solid)	100	100	100	100	72,92	76,6	78,06	72,85

Figure 3.5 shows the test data set confusion matrices for each ensemble classifier. Since the training dataset achievements are 100%, they are not presented separately. In Figure 3.5(a), the CTLDF+EnC (Stacking) model predicts 135 correct and 75 incorrect images of cracked tires. For normal tire images, 111 were correctly predicted and 4 were incorrectly predicted. In Figure 3.5(b) and Figure 3.5(c), the correct/incorrect prediction

values of the cracked tire images for the CTLDF+EnC(Soft) and CTLDF+EnC(Solid) models are 134/76 and 128/82, respectively. The true/false prediction values for normal tire images are 111/4 and 110/5, respectively. These matrices show the ability of each hybrid model to correctly and incorrectly predict cracked and normal tire images.

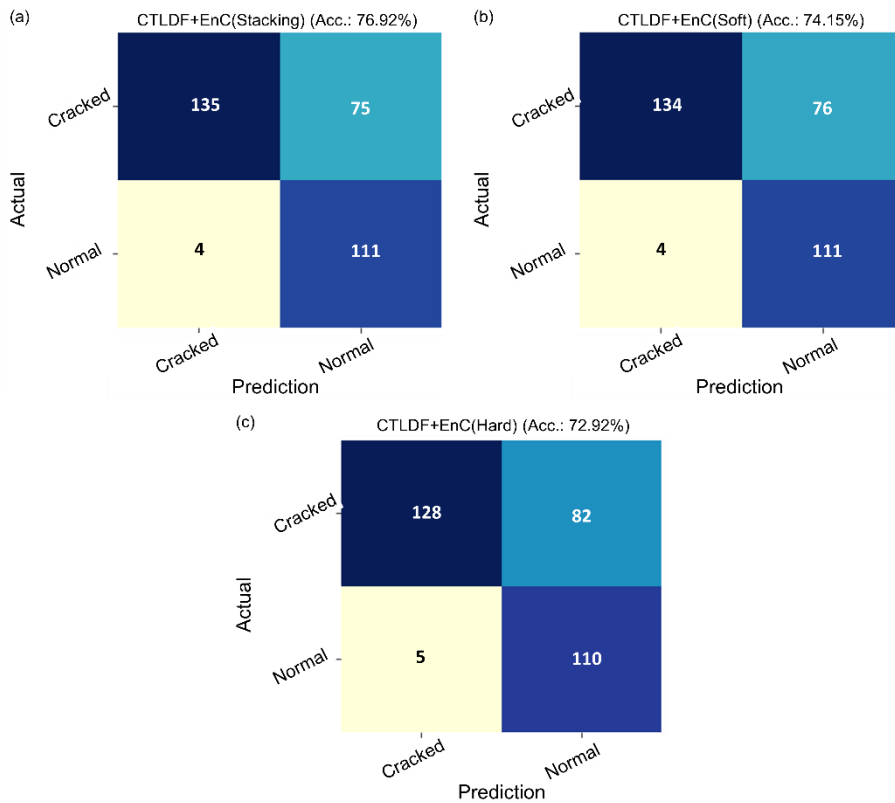


Figure 3.5. Ensemble classifier test confusion matrices

Within the scope of the study, Table 3.7 was created to determine the classification performance of hybrid models using the deep features of the proposed TL models and to more clearly demonstrate the advantages of deep feature-based models. In this table, the first column, TL Model (TLM), contains the success metrics obtained directly with the TL models, while the other columns show the test accuracies and improvement amounts obtained with ensemble classifiers such as Stacking Hybrid Model (SHM), Soft Voting Hybrid Model (SVHM) and Hard Voting Hybrid Model (HVHM), which are built using all the features obtained from the TL models. From Table 3.7, it can be seen that the proposed SHM, SVHM and HVHM ensemble

classifiers all exceed the performance measures of the TL models and provide more accurate predictions. This table clearly shows that the proposed deep feature-based hybrid models provide higher classification performance compared to the direct TL models.

Table 3.7. Comparison of the quantities of improvement in the test performance of the hybrid models.

Models	Imp. Dif. (%)			
	TLM	SHM	SVHM	HVHM
Xception	61.23	15.69	12.92	11.69
NASNet	66.15	10.77	8	6.77

MobileNet	57.54	19.38	16.61	15.38
DenseNet169	59.38	17.54	14.77	13.54
DenseNet201	68.92	8	5.23	4
VGG16	56.31	20.61	17.84	16.61
InceptionV3	65.54	11.38	8.61	7.38
ResNet50V2	70.77	6.15	3.38	2.15
ResNet101V2	63.08	13.84	11.07	9.84

In Table 3.7, it is seen that the highest accuracy values obtained with the SHM, SVHM and HVHM are 76.92%, 74.15% and 72.92%, respectively. When these values are compared with the accuracy values of the TL models, it is determined that the highest improvement differences are 20.61%, 17.84% and 16.61% for the VGG16 TL model, respectively. On the other hand, the lowest improvement differences were 6.15%, 3.38% and 2.15% for the ResNet50V2 TL model, respectively. It is expected that the improvement differences would be low considering that ResNet50V2 was the model with the highest success in the first group of applications. It is clear from Table 3.7 that the proposed hybrid models of SHM, SVHM and HVHM significantly improve the accuracy of ResNet50V2 in particular.

5. DISCUSSION

Vehicle drivers are often aware of the importance of tire tread depth and tire air pressure, but overlook the risks of tire oxidation. Tire oxidation and related cracks pose potential hazards that seriously affect driving safety. In this context, this study aims to detect cracks in tires based on the use of pre-trained TL methods and ensemble classifiers.

According to the World Health Organization (WHO), a large number of fatal traffic accidents occur every year and the majority of these accidents are caused by tire defects [31]. While manual detection of such defects can be difficult and inaccurate, AI-enabled systems have significant potential in this area. Therefore, we propose an AI-assisted model for easy and fast detection of tire defects by vehicle users. In the proposed model, three different hybrid models based on the five best classical ML algorithms are proposed by combining the deep features of nine pre-trained ML models obtained from tire images with ensemble classifiers: CTLDF+EnC(Stacking), CTLDF+EnC(Soft) and CTLDF+EnC(Solid). This approach provides an effective solution for tire defect detection.

The use of non-destructive testing techniques for the detection of tire defects is widespread. These techniques include laser shearing [32], ultrasonic methods [33] and electromagnetic pulse [34]. However, these methods are often expensive and difficult to implement and are not widely used. Moreover, most of the studies based on traditional visual detection [2,3] and deep learning [4,35]

methods use X-Ray imaging and these studies usually focus on the production line. In this study, we aim to detect tire defects that are worn or oxidized due to usage. Furthermore, a more efficient structure is proposed by replacing the expensive and complex X-Ray images with more cost-effective and easily available digital camera images. This approach provides a more accessible and practical solution for tire defect detection.

Although the use of digital images is almost non-existent in the literature, a similar approach was adopted in a study by Lin [1]. However, since the dataset of this study was not shared, the success of the hybrid models proposed here could not be tested on this dataset. This shows that each study has its own unique datasets and model success is dataset dependent. However, on the basis of the applications with TL models, it is seen that the hybrid models proposed in this study significantly increase the success. This shows that the study offers a new and effective approach.

In this study, the proposed CTLDF+EnC (Stacking) hybrid model has the highest test accuracy of 76.92%. The other hybrid models, CTLDF+EnC (Soft) and CTLDF+EnC (Hard) architectures have 74.15% and 72.92% accuracy respectively. According to these results, it is concluded that the proposed hybrid model performs at an acceptable level in general and can be an effective tool for tire defect detection.

When the performance of the proposed hybrid models is compared to the performance of the directly used TL models, it is seen that they are more successful. For example, in Table 3.5, the highest test accuracy rate is 70.77% with ResNet50V2 when the TL models are used directly. On the other hand, the test accuracy rate obtained with the proposed CTLDF+EnC (Stacking) method is 6.15% higher than ResNet50V2. As can be seen in Table 3.5 and Table 3.7, all of the ensemble TL models have lower accuracy values than the proposed hybrid classifiers. These results show that the proposed hybrid models are more effective than the TL models in detecting tire problems and have a reasonable level of success. In this context, the proposed hybrid models are considered to be a promising method for future studies in tire defect detection.

The original contributions of this work can be listed as follows: First, we propose a new architecture, CTLDF+EnC, which enables inspection from tire images. This architecture represents an important step forward in tire defect detection. Second, the architecture has the ability to combine deep features of TL models and classical ML methods in an ensemble classifier. This provides an efficient way to achieve more comprehensive and accurate results. Finally, the study emphasizes the use of images acquired with a digital camera, which is a cheaper and easy-to-use method to replace expensive and complex imaging techniques. This provides a more accessible and cost-effective solution for tire defect detection. Therefore, the proposed system can be considered as a useful tool to improve safety standards in the tire industry and solve tire-related problems more effectively.

6. CONCLUSION

This paper presents a novel approach, CTLDF+EnC, for tire defect detection. This architecture combines deep features derived from TL methods and classical ML methods to propose an effective solution for tire defect detection. The proposed architecture can work with cheap and easy-to-use digital images that can be acquired with regular digital cameras, which reduces the cost and increases the applicability. In addition, this work aims to detect problems that occur during the lifetime of tires, such as oxidation and aging, unlike the studies in the literature, which are usually aimed at detecting defects that occur on the production line. Nine different TL models were used to extract features, which were then combined and fed to ensemble classifiers containing classical ML models. Stacking, Soft and Hard voting methods are used as ensemble classifiers. The implementations show that the proposed architecture achieves satisfactory performance compared to other alternatives. The results show that the CTLDF+EnC architecture can successfully detect oxidized or worn tires and is advantageous in terms of cost and applicability. This study is considered to propose an effective system to improve safety standards in the tire industry and solve tire-related problems more effectively.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement

The data that support the findings of this study are openly available in “Oxidized and non-oxidized tire sidewall and tread images” at <https://doi.org/10.7910/DVN/Z3ZYLI>.

REFERENCES

- [1] S.-L. Lin, Research on tire crack detection using image deep learning method, *Sci Rep* 13 (2023) 8027.
- [2] Q. Guo, C. Zhang, H. Liu, X. Zhang, Defect detection in tire X-ray images using weighted texture dissimilarity, *J Sens* 2016 (2016).
- [3] G. Zhao, S. Qin, High-precision detection of defects of tire texture through X-ray imaging based on local inverse difference moment features, *Sensors* 18 (2018) 2524.
- [4] Z. Zheng, S. Zhang, B. Yu, Q. Li, Y. Zhang, Defect inspection in tire radiographic image using concise semantic segmentation, *IEEE Access* 8 (2020) 112674–112687.
- [5] R. Wang, Q. Guo, S. Lu, C. Zhang, Tire defect detection using fully convolutional network, *IEEE Access* 7 (2019) 43502–43510.
- [6] J. Huang, B. Kingsbury, Audio-visual deep learning for noise robust speech recognition, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013: pp. 7596–7599.
- [7] Y. Kaya, F. Kuncan, R. Tekin, A new approach for congestive heart failure and arrhythmia classification using angle transformation with LSTM, *Arab J Sci Eng* 47 (2022) 10497–10513.
- [8] J. Liu, D. Capurro, A. Nguyen, K. Verspoor, “Note Bloat” impacts deep learning-based NLP models for clinical prediction tasks, *J Biomed Inform* 133 (2022) 104149.
- [9] B. Özeydin, R. Tekin, *A Hybrid Model Based on Deep Features and Ensemble Learning for the Diagnosis of COVID-19: DeepFeat-E*, *Turkish Journal of Science and Technology* 18 (2023) 183–198.
- [10] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, *IEEE J Sel Top Signal Process* 13 (2019) 206–219.
- [11] B. Rim, N.-J. Sung, S. Min, M. Hong, Deep learning in physiological signal data: A survey, *Sensors* 20 (2020) 969.
- [12] H. Tung, R. Tekin, New Feature Extraction Approaches Based on Spatial Points for Visual-Only Lip-Reading., *Traitement Du Signal* 39 (2022).
- [13] D. Wang, J. Su, H. Yu, Feature extraction and analysis of natural language processing for deep learning English language, *IEEE Access* 8 (2020) 46335–46345.
- [14] S.-H. Wang, S. Xie, X. Chen, D.S. Guttery, C. Tang, J. Sun, Y.-D. Zhang, Alcoholism identification based on an AlexNet transfer learning model, *Front Psychiatry* 10 (2019) 205.
- [15] X. Yang, Y. Zhang, W. Lv, D. Wang, Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier, *Renew Energy* 163 (2021) 386–397.
- [16] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: pp. 1251–1258.
- [17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *ArXiv Preprint ArXiv:1409.1556* (2014).
- [18] B. Zoph, V. Vasudevan, J. Shlens, Q. V Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: pp. 8697–8710.
- [19] K. He, X. Zhang, S. Ren, Deep Residual Learning for Image Recognition, *Microsoft Research* 45 (2015) 1951–1954. <https://doi.org/10.1002/chin.200650130>.
- [20] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017) 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>.
- [21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: pp. 2818–2826.

- [22] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, (2017). <http://arxiv.org/abs/1704.04861>.
- [23] D.H. Wolpert, Stacked generalization, *Neural Networks* 5 (1992) 241–259.
- [24] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*, CRC press, 2012.
- [25] Y. Zhang, D. Lefebvre, Q. Li, Automatic detection of defects in tire radiographic images, *IEEE Transactions on Automation Science and Engineering* 14 (2015) 1378–1386.
- [26] Y. Li, B. Fan, W. Zhang, Z. Jiang, TireNet: A high recall rate method for practical application of tire defect type classification, *Future Generation Computer Systems* 125 (2021) 1–9.
- [27] J. Siegel, Oxidized and non-oxidized tire sidewall and tread images, Harvard Dataverse (2021). <https://doi.org/https://doi.org/10.7910/DVN/Z3ZYLI>.
- [28] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, G. Zhang, Transfer learning using computational intelligence: A survey, *Knowl Based Syst* 80 (2015) 14–23.
- [29] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach Learn* 63 (2006) 3–42.
- [30] M.K. Bohmrah, H. Kaur, Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model, *Global Transitions Proceedings* 2 (2021) 476–483.
- [31] S. Erdogan, Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey, *J Safety Res* 40 (2009) 341–351.
- [32] C. Cai, L. He, Improved Mach–Zehnder interferometer-based shearography, *Opt Lasers Eng* 50 (2012) 1699–1705.
- [33] H.L.M. Dos Reis, K.A. Warmann, Acousto-ultrasonic non-destructive evaluation of fatigue damage in steel-belted radial tires., *Int J Fatigue* 3 (1996) 216.
- [34] L.E. Roemer, N. Ida, Location of wire position in tyre belting using Bayesian analysis, *NDT & E International* 24 (1991) 95–97.
- [35] R. Wang, Q. Guo, S. Lu, C. Zhang, Tire defect detection using fully convolutional network, *IEEE Access* 7 (2019) 43502–43510.

Düzeltilmiş Makalesidir

Uyarlanabilir Gezinme Yapılarının E-Sağlık Uygulamalarına Uyumu

Araştırma Makalesi/Research Article

 Levent ÇETİNKAYA*¹,  Hafize KESER²

¹Bilgisayar ve Öğretim Teknolojileri Eğitimi, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, Türkiye

²Bilgisayar ve Öğretim Teknolojileri Eğitimi, Ankara Üniversitesi, Ankara, Türkiye

lcetinkaya@comu.edu.tr, k eser@ankara.edu.tr

(Geliş/Received:13.03.2023; Kabul/Accepted:12.01.2024)

DOI: 10.17671/gazibtd.1264449

Özet— Çevrimiçi uyarlanabilir gezinme ortamlarına yönelik ilkelerin e-sağlık uygulamalarına uyumu ve düzenlenmesi amacıyla gerçekleştirilen çalışmada, nitel ve nicel araştırma yaklaşımlarının işe koşulduğu yapılandırılmış çevrimiçi Delphi Tekniğinden faydalanılmıştır. Uzmanların çalışmaya katılımı ile ilgili onayları sonrasında başlatılan Delphi sürecinde tek bir soru kağıdında nitel ve nicel veriler eşzamanlı olarak toplanarak veri çeşitlenmesi yapılmıştır. Elde edilen veriler ve bu verilerin değerlendirilmesi ile elde edilen sonuçların uzmanlara sunumu şeklinde tekrarlanan süreç uzmanların uzlaşmaya ulaşmalarıyla son bulmuştur. Çalışmanın ilk turu sonucunda uzmanların gezinmenin uyarlanmasına yönelik ilkeleri yapısal olarak uygun bulduğu belirlenmiştir. Ancak, çevrimiçi öğrenme ortamlarına yönelik ilkelerin e-sağlık uygulamaları çerçevesinde yeniden yorumlanması gerekliliği uzmanlar tarafından bildirilmiştir. Uzman görüşleri doğrultusunda her turda yeniden şekillenen ilkeler üzerinde artık yeni fikirlerin üretilmediği, maddelere ilişkin tüm güçlü-zayıf yanların belirlendiği ve uzmanlardan herhangi bir görüş ya da önerinin gelmediği görülen üçüncü Delphi turunda çalışma sonlandırılmıştır. Çalışma sonucunda e-sağlık uygulamaları tasarımında gezinmenin uyarlanmasına yönelik tüm ilkeler üzerinde uzmanların uzlaşma sağladıkları ve edilen gözlemler arası uyumun (ICC=.930) mükemmel düzeyde olduğu belirlenmiştir.

Anahtar Kelimeler— e-sağlık, sağlık hizmetleri, uyarlanabilir sistemler, gezinmenin uyarlanması, uzlaşma, delphi tekniği.

Integration of Adaptive Navigation Structures Into E-health Applications

Abstract— The structured online Delphi Method, which included qualitative and quantitative research methods, was used in the study to integrate and regulate the principles for online adaptive navigation environments to e-health applications. Data triangulation was accomplished in the Delphi process by concurrently gathering qualitative and quantitative data in a single questionnaire, which was started following the experts' approval of their involvement in the study. The experts finally reached an agreement after repeatedly presenting the data acquired and the outcomes generated by the examination of this data to the experts. The first round of the investigation led to the conclusion that the experts deemed the principles for the adaptation of navigation were structurally adequate. Yet, experts reported that the principles for online learning environments should be reinterpreted within the framework of e-health applications. The study was terminated in the third Delphi round when it was observed that no new ideas were produced on the principles that were reshaped in each round in line with the expert opinions, all the strengths and weaknesses related to the items were determined, and no views or suggestions were received from the experts. As a result of the study, it was determined that experts reached a consensus on all the principles for the adaptation of navigation in the design of e-health applications. It was also concluded that the agreement between observers (ICC=.930) was excellent.

Keywords— e-health, health services, adaptive systems, adaptive navigation, agreement, delphi technique.

¹Bu makale, birinci yazarın ikinci yazar danışmanlığında tamamlanmış olduğu doktora tez çalışmasından türetilmiş olup elde edilen sonuçlar farklı bir alana uyarlanmıştır.

1. GİRİŞ (INTRODUCTION)

Sağlık sistemindeki dijitalleşme sürecinin hızlanmasıyla birlikte sağlık sistemini oluşturan paydaşlar ile vatandaşlar arasında dijital bir köprünün oluşum süreci de hız kazanmaya başlamıştır. Dijitalleşme ile birlikte sağlıkta paradigma kaymasının sonucu olarak vatandaşların sağlıkla ilgili süreçlerini yönetebilecekleri hizmetlerin de yaygınlaştığı görülmektedir. Bu noktada Avrupa Birliği ülkeleri ve ABD başta olmak üzere birçok ülkede dijital sağlık hizmetleri önemli ölçüde ivme kazanmaya başlamıştır. Bireyler ve toplumlar için insan sağlığı ile sağlık hizmetlerinin iyileştirilmesinde bilgi ve iletişim teknolojilerinin işe koşulması olarak tanımlanan dijital sağlık [1] ile sunulan hizmetler bireye kendi sağlığında aktif bir katılımcı olma fırsatı sağlamaktadır. Temelde sağlık hizmetlerini anlık olarak ulaşılabilir kılarak, tasarlanan bu sistemlerin insan yaşamının kalitesini doğrudan etkileme potansiyeline sahip olduğunu söylemek mümkün gözükmemektedir. Her ne kadar sağlıkta dijitalleşme sürecinin tarihi eskiye dayanıyor olsa da sağlık sistemi ile vatandaşları bir araya getirebilecek yapılar oldukça yenidir ve sayıları da hızla artmaktadır. Dijitalleşen sağlık sistemi teşhis, tedavi, önleme ve tedavi süreçleri için yeni fırsatlar sunmakla birlikte sağlık profesyonellerinin yanı sıra vatandaşların da sağlık verilerine kolaylıkla ulaşmalarına olanak sağlamaktadır. Bu yapılar aracılığı ile bireylerin mevcut içerikle etkileşim içinde olmaları ve sağlıkla ilgili bilgilerine erişimlerine olanak sağlanarak iletişim sürecinde de aktif olarak katılmalarının desteklenmesi amaçlanmaktadır [2].

Sağlık hizmetlerinden etkin bir şekilde faydalanılması noktasında e-sağlık hizmetlerinin verimli bir şekilde işe koşulması ve bireylerin bu hizmetlere erişebilmelerinin sağlanması oldukça önemlidir. Bu durumda e-sağlık hizmetlerinden faydalanabilecek kitlenin heterojen bir yapıya sahip olduğunun göz ardı edilmemesi gerekmektedir. Nitekim, bu kitlenin demografik özelliklerinin (yaş, cinsiyet, eğitim durumu, vb.) yanı sıra teknoloji ve sağlığa ilişkin okuryazarlık düzeyleri de birbirinden farklılık gösterebilmektedir [bkz., 3, 4, 5, 6, 7]. Yine bireylerin teknoloji ve sağlıkla ilgili yetkinliklerini ifade eden dijital sağlık okuryazarlığı günümüzde önemle üzerinde durulan yetkinliklerden biri olsa da tüm bireylerin bu yeterliliklere sahip olması beklenemez. Bireylerin bu becerilere sahip olmaları sağlıkla ilgili verileri anlamlandırabilmeleri ve süreçlerini yönetebilmeleri için ise hayati önem taşımaktadır. Ancak, bu bir süreç olup yetkinliklerin kazandırılabilmesi için eğitimin yanı sıra zamana da ihtiyaç vardır. Nihayetinde, dijital dönüşümler halk ve nüfus sağlığına fayda sağlamak için muazzam bir potansiyele sahip olsa da mevcut eşitsizlikleri şiddetlendirme konusunda da eşit derecede yeteneklidirler [7]. Bu noktada e-sağlık uygulaması ya da hizmeti geliştiricilerinin tüm kullanıcıların dijital sağlık okuryazarı olmadığı gerçeğini göz önünde bulundurmaları ve bu yapıları tüm paydaşlar açısından kullanılabilir nitelikte tasarlamaları gerekmektedir. Özellikle de istenilen bilgi ve bağlantılara, kullanıcıların kolaylıkla erişebilecekleri gezinme unsurlarını barındıran kullanılabilirliği yüksek

ortamların tasarlanması, temelde sağlık sistemi ile vatandaşları bir araya getirmeyi amaçlayan bu yapıların sürdürülebilirliği açısından da oldukça önemlidir.

Gelişen teknoloji, bilgiyi daha ulaşılabilir hale getirmiş ve sürekli artan bilgi miktarının karmaşaya yol açmadan bireyin ihtiyaçlarına göre sunulmasını gerekli kılmıştır. Bu gerekliliğin yanı sıra bilginin, bireyin özellikleriyle birlikte zamanla değişen ihtiyaçlar doğrultusunda yeniden şekillenen yapılar oluşturmak için uyarlanabilir sistemlerin geliştirilme çalışmaları hız kazanmaya başlamıştır. Bu duruma paralel olarak geleneksel ortamlara yönelik “biri hepsine uyar (one size fits all)” görüşü terk edilmeye [8, 9, 10] ve “biri hepsine uymaz (one size does not fit all)” görüşü [11] yaygın olarak benimsenmeye başlamıştır [12]. Uyarlanabilir çevrimiçi sistemler, yapay zeka, veri madenciliği gibi tekniklerin sürece dahil edilmesi ile birlikte hızlı bir şekilde gelişmeye devam ederek farklı kullanıcı gereksinimlerine ya da kullanıcının zamanla değişen gereksinimlerine uygun olarak değiştirilebilen yapılar haline gelmiştir. Genel olarak kullanıcı tercihlerine ve ihtiyaçlarına göre şekillenen ortamlar sunmayı amaçlayan uyarlanabilir yapıların [10] ön plana çıkmasındaki en önemli unsurların başında kişiselleştirilmiş servisleri sunmak için davranışların gözden geçirilmesi, ilgi ve tercihlerin göz önünde bulundurulması tasarımların gerçekleştirilebilmesi gelmektedir [13]. Sistemdeki mevcut bilgilerin yanı sıra kullanıcının sistem içindeki anlık davranışları dinamik bilgilerin anlamlı bir şekilde yeniden yapılandırılması kritik öneme sahiptir. Bu yapılandırma bireyin sistem içinde gezinme stratejileri ile birlikte istenilen bilgiye erişiminde kilit rol oynamaktadır. Özellikle çevrimiçi ortamlarda kullanıcının istediği bilgiye ulaşmada en kısa yolu bulmasını destekleyecek tasarımların gerçekleştirilmesi önemli bir unsur olarak karşımıza çıkmaktadır. Uyarlanabilir yapılar içerisinde gezinmenin uyarlanması önemli bir basamak olup, kullanıcının modelinin yapılandırılması ve uyarlamaların gerçekleştirilmesi noktasında oldukça önemlidir. Gezinmenin uyarlanması, kullanıcının ortamda izleyeceği en uygun yolu bulması için destek sunulmasıdır [14, 15] ve sistemin sağlıklı işleyebilmesinde kilit rol oynamaktadır. Nitekim çevrimiçi ortamlarda iyi düzenlenmemiş gezinme stratejilerinin başta kaybolma ve aşırı bilişsel yüklenme olmak üzere birçok olumsuzluğa yol açabileceği görülmektedir [bkz., 16, 17, 18, 19]. Bu tür ortamlarda aşırı bilişsel yüklenen ve buna bağlı olarak da kaybolan kullanıcıların performansları düşerek istedikleri bilgilere ulaşmada daha fazla zaman ve çaba harcamaları gerekecektir [20]. Bu noktada aşırı bilişsel yüklenmeyi engelleyeceği belirlenen tasarım ilkelerinin kullanıldığı araştırma sonuçları ise kullanıcıların başarı, hatırlama ve transfer puanlarının arttığını göstermektedir [16].

Genel olarak kullanıcı özelliklerinin çevrimiçi ortamda gezinmelerinde etkili olduğu ve düşük ön bilgi düzeyine sahip bireylerin gezinme ve bağlantı yapılarını anlama konusunda daha çok sorun yaşamaya yatkın oldukları görülmektedir [bkz., 16, 19, 21, 22, 23]. Özellikle yeni kullanıcıların sistem içinde hangi bilgiye nasıl

ulaşabileceklerini kestirebilmeleri çok daha güçtür ve sistemin kullanıcıyı destekleyici nitelikte tasarlanması gerekmektedir. Bu durum hızla yaygınlaşan ve ilk kullanıcı sayısı hızla artmaya devam eden e-sağlık uygulamaları için de geçerlidir. Yaşamın birçok alanına neredeyse durdurulamaz bir şekilde giren dijitalleşme, politikacılar ve sağlık sigortacıları tarafından, genellikle gelecekteki sorunların çözümünün en azından bir parçası olarak görülmektedir [24]. Ancak, diğer taraftan da teknik altyapı, hızla artan veri hacmi ile birlikte sistemle etkileşim, istenilen verilere etkili bir biçimde erişiminin sağlanması ve korunması gibi çözümlenmeyi bekleyen birçok sorun da beraberinde gelmektedir. Her ne kadar veri madenciliği ve yapay zeka tekniklerinden faydalanılarak makine temelli çözümler üretilmeye başlanmış olsa da özellikle son kullanıcıların kolaylıkla istenilen bilgiye erişim sağlayabilecekleri sistemlerin tasarlanması bir zorunluluktur. Yine bu doğrultuda kullanıcının sistem içerisinde gezinirken kaybolması ve aşırı bilişsel yüklenmesi gibi risklerin göz önünde bulundurularak iyi tasarlanmış uyarlanabilir gezinme yapılarının geliştirilmesi gerekmektedir. Bu çerçevede gerçekleştirilen çalışmada, Çetinkaya (2013) tarafından uyarlanabilir öğrenme ortamlarına yönelik geliştirilen tasarımı ilkelerinden [25] hareketle çevrimiçi e-sağlık uygulamalarının son kullanıcılarına yönelik uyarlanabilir gezinme ortamlarının düzenlenmesine ilişkin ilkelerin uzman görüşlerine dayalı olarak belirlenmesi amaçlanmıştır. Bu genel amaç doğrultusunda çalışmada;

- uzman görüşlerine göre Çetinkaya (2013) tarafından geliştirilen uyarlanabilir öğrenme ortamlarına yönelik tasarımı ilkeleri, çevrimiçi e-sağlık uygulamalarının son kullanıcılarına yönelik uyarlanabilir gezinme ortamlarının düzenlenmesinde de benimsenebilir mi?
- uzmanların söz konusu ilkelerin uygunluğuna katılım düzeyleri nedir? sorularına yanıt aranmıştır.

2. YÖNTEM (METHOD)

2.1. Araştırma modeli (Research model)

Çalışmada nitel ve nicel araştırma becerilerinin işe koşulduğu yapılandırılmış çevrimiçi Delphi tekniğinden faydalanılmıştır. Genel olarak bir grup uzmanın görüşü ile kontrollü geri bildirim sağlanarak, uzlaşma aranabilir bir anket elde etme süreci olarak tanımlanan Delphi tekniği [26] büyük ölçüde amaca hizmet edebilen, deneysel bilgi eksikliği nedeni ile sosyal süreçlerin anlaşılacağı durumlarda kullanılmaktadır [27].

2.2. Çalışma Gurubu (Study Group)

Delphi tekniği sürecinde katılımcıların, konular hakkında bilgi ve ilgi sahibi olan bireylerden oluşması gerektiğinden [28, 29], çalışma grubunun belirlenmesinde amaçlı örnekleme yaklaşımı benimsenmiştir. Amaçlı örnekleme yaklaşımında temel amaç, araştırmanın konusunu oluşturan kişi, olay ya da durum hakkında ve belirli bir amaç doğrultusunda derinlemesine bilgi toplamaktır [30]. Gerçekleştirilen bu çalışmada çevrimiçi Delphi uzman

grubu katılımcılarını belirlemek için amaçlı örnekleme kapsamında ölçüt örnekleme yoluna gidilmiştir. Bu çerçevede oluşturulan çalışma grubu doktora eğitimini tamamlamış olmak koşuluyla;

- uyarlanabilir ortamlar, yapay zeka ve öğretimsel uygulamalara yönelik araştırmaları olan (n=7),
- sağlık bilişimi, tıp bilişimi ve tıp eğitimi alanlarında ders vermiş ya da araştırmaları olan (n=8) uzman kişilerden oluşmuştur.

Delphi tekniğinin işe koşulduğu çalışmalarda katılımcı grubun uzmanlık ölçütlerine göre homojenliğinin söz konusu olduğu araştırmalarda 10-15 kişinin yeterli olduğu bildirilmiştir [26]. Bu doğrultuda belirtilen ölçütleri kapsayacak biçimde oluşturulan çalışma grubunu çevrimiçi ortamda ulaşılabilen 15 uzman oluşturmuş ve bu uzmanlar çalışmanın tüm aşamalarında katılım sağlamıştır. Çevrimiçi ortamda gerçekleştirilen Delphi çalışmasının güvenilirliğini sağlamak için veriler anlık olarak kayıt altına alınmıştır.

2.2. Verilerin Toplanması (Data Collection)

Yapılandırılmış çevrimiçi Delphi yönteminin benimsendiği çalışmada metodüçi karışım (Intramethod mixing) yaklaşımı işe koşulmuştur. Bu yaklaşım tek bir soru kağıdında eşzamanlı olarak veri toplamak için kullanılan bir karma yöntem stratejisidir [31]. Bu strateji doğrultusunda tekrarlanan turlar içeren yapılandırılmış çevrimiçi Delphi uygulama süreci aşamaları aşağıda belirtilmiştir.

1. İlkelerle yönelik görüş belirlenmesi amacıyla önceden belirlenen 22 uzmana çalışmanın amacını belirten bir metin ile birlikte davet mektubu gönderilmiştir. Davete olumlu yanıt veren 15 uzman ile süreç başlatılmıştır.
2. Çevrimiçi öğrenme ortamlarının tasarımında gezinmenin uyarlanması sürecinde Çetinkaya (2013) tarafından geliştirilen ilkeler [25] üzerinde değişiklik yapılmadan çevrimiçi ortama aktarılmış ve uzmanlara gönderilmiştir. İlkelerle ilişkin açıklama ve yönergelerin de yer aldığı çevrimiçi anket formu, her bir ilkenin altına ilkeye yönelik uzmanların varsa görüş ve önerilerini belirtmeleri amacıyla serbest metin alanı oluşturularak düzenlenmiştir. Ayrıca, ölçme aracının en sonunda eklemek istedikleri yeni bir ilke önerisi ya da ilkelerin geneline ilişkin varsa görüş ve önerilerini paylaşabilecekleri serbest metin alanı oluşturulmuştur.
3. Uzmanların Çetinkaya (2013) tarafından geliştirilen gezinmenin uyarlanmasına ilişkin ilkeleri [25], e-sağlık uygulamaları çerçevesinde irdeleyerek yaptıkları değerlendirme sonucunda gelen veriler araştırmacılar tarafından incelenmiştir. Uzmanların yorumları ile birlikte ilkelere yönelik değerlendirmeleri gözden geçirilip düzenlenmiştir. Düzenleme sırasında öğrenme ortamlarına yönelik geliştirilen ilkelerin e-sağlık uygulamalarına uyumu ve uygunluğu göz önünde bulundurulmuştur. Bu aşamada ilkelerin uygunluğunun belirlenmesi amacıyla 5'li (Tamamen uygun, Uygun, Kısmen uygun, Uygun değil, Hiç uygun değil) likert tipi çevrimiçi değerlendirme formu oluşturulmuştur.

Form uzmanların her fikri gözden geçirmesi, güçlü ve zayıf yanlarını belirtmesi ve yeni fikirlerini ekleyebilmelerine olanak sağlayabilecek şekilde düzenlenmiştir. Kolaylıkla kullanılabilir ve uzmanların her fikri serbestçe paylaşabilecekleri şekilde tasarlanan ikinci anket formu açıklayıcı yönergelerle birlikte uzmanlara gönderilmiştir.

- Gerçekleştirilen ikinci Delphi turu sonucunda uzmanlardan gelen değerlendirme sonuçları araştırmacılar tarafından incelenmiştir. Yapılan incelemeler sonucunda ilkelerin ifade edilmiş biçimine yönelik uzmanlar tarafından küçük düzeltme önerileri geldiği belirlenmiştir. İlkelerin uygunluğuna yönelik derecelendirme verileri doğrultusunda ise uzmanların uzlaşmaya ulaşma dereceleri hesaplanmış ve bir sonraki anket formu hazırlanmaya başlanmıştır. Üçüncü Delphi turu için hazırlanan anket formu; uzmanların ikinci Delphi turundaki önerileri doğrultusunda yeniden düzenlenen ilkelerin yanı sıra tüm uzmanlara ait yanıtların ortalama, standart sapma değerleri ile birlikte uzmanın maddeye kendi verdiği puanında bulunduğu bir formatta yeniden düzenlenerek gönderilmiştir.
- Uzmanlardan gelen üçüncü Delphi turu verilerinin incelenmesi sonucunda artık yeni fikirlerin üretilmediği, maddelerin tüm güçlü-zayıf yanlarının belirlendiği ve uzmanlardan ilkelere düzenlenmesine yönelik herhangi bir görüş ya da önerinin gelmediği görülmüştür. Uzmanların tüm maddeler üzerinde uzlaşmaya vardıkları belirlenen bu turdan sonra uzmanlara katılımları için teşekkür edilerek en son veriler doğrultusunda oluşturulan ilkeler yönergeleri ile birlikte gönderilmiştir.

2.3. Verilerin Analiz Edilmesi (Analysis of the Data)

Her turunda verilerin toplandığı Delphi çalışmasının ilk turunda, Çetinkaya (2013) tarafından gezinmenin uyarlanmasına yönelik geliştirilen ilkeler [25] gönderilmiştir. Ayrıca, bu turda gönderilen ilkelere yönelik ek görüş belirtebilecekleri bir serbest yazım alanını içeren veri formu düzenlenerek nitel veriler elde edilmiştir. İkinci ve üçüncü turda ise, hem nitel hem de nicel veriler elde edilirken sürecin son turu olan dördüncü turu sadece nicel verilerden oluşmuştur. Uzmanlardan gelen görüş ve önerilerden oluşan nitel veriler araştırmacılar tarafından incelenerek ilkeler üzerinde gerekli düzenlemeler yeniden uzman görüşüne sunulmak üzere gerçekleştirilmiştir. Uzmanların uzlaşmaya ulaşma derecelerinin belirlenmesi için ise; aritmetik ortalama, standart sapma ve ilk çeyrek kayma değerlerinden faydalanılmıştır. Her bir ilkeye ilişkin uzmanların uzlaşma düzeyleri tablo 1’de belirlendiği gibi dört kategoriden oluşmuş [32] ve her etki altında standart sapma ile ortalamaya göre belirlenmiştir.

Tablo 1. Kategori Ölçeklendirme Matrisi

Ortalama Değeri	Standart Sapma Değeri	
	Düşük	Yüksek
	I	II
Yüksek	Büyük Öne Sahip Yüksek Düzeyde Anlaşma	Yüksek Öne Sahip Düşük Düzeyde Anlaşma
	III	IV
Düşük	Alt Öne Sahip	Alt Öne Sahip

	Yüksek Düzeyde Anlaşma	Düşük Düzeyde Anlaşma
--	------------------------	-----------------------

Delphi yönteminin kullanıldığı çalışmalar incelendiğinde aritmetik ortalama, standart sapma ve medyan (ortanca) değerlerinin uzlaşma derecelerinin belirlenmesinde işe koşulduğu görülmektedir. Ancak, kategorik sınıflandırmalarda bu değerlerin aralıklarına yönelik belirli bir standart oluşturulmamıştır. Bu nedenle her bir ölçütün birlikte ele alınarak uzlaşma kategorilerinin oluşturulduğu ve Çetinkaya (2013) tarafından düzenlenen değer aralıkları [25] uzmanların uzlaşmaya ulaşma derecelerinin belirlenmesinde işe koşulmuştur (bkz. tablo 2).

Tablo 2. Uzlaşma Aralıkları

\bar{X}	Ss	Q_1	Yorum	
1.00-1.49	$Ss \geq .75$	$3 \leq Q_1$	Önemsiz uzlaşma olması	1
1.50-2.99			Orta derecede uzlaşma olması	2
3.00-3.74			Ekseriyetle uzlaşma olması	3
3.75-4.49	$.50 \leq Ss < .75$	$3 < Q_1 \leq 4$	Önemli derecede uzlaşma olması	4
4.50-5.00	$Ss < .50$	$4 < Q_1 \leq 5$	Neredeyse mükemmel uzlaşma olması	5

Delphi süreci sonunda belirlenen ilkelerin tamamına yönelik değerleyici güvenilirliğini belirlemek amacı ile ise Sınıf-içi Korelasyon Katsayısı (Intraclass Correlation Coefficient-ICC) hesaplanmıştır. Değerleyici güvenilirliği, değerleyiciler arasındaki uyuma derecesini ölçmek amacıyla işe koşulan Sınıf-içi Korelasyon Katsayısı [33] ile ilgili güvenilirlik aralığı tablo 3’de verilmiştir [34].

Tablo 3. Sınıf-içi Korelasyon Katsayısı Güvenirlik Aralığı

ICC	Yorum
$< .69$	Düşük düzeyde güvenilirlik
0.70— 0.79	Yeterli/Orta düzeyde güvenilirlik
0.80 — 0.89	Yüksek düzeyde güvenilirlik
0.90 — 1.00	Mükemmel düzeyde güvenilirlik

3. BULGULAR VE YORUMLAR (FINDINGS AND INTERPRETATIONS)

Gezinmenin uyarlanması, kullanıcının ortamda izleyebileceği en uygun yolu bulması için destek sunulmasıdır [14, 15]. Çetinkaya (2013) tarafından gerçekleştirilen çalışmada gezinmenin uyarlanmasına yönelik tüm ilkeler [25] üzerinde bağıl önemlilik ölçütlerine göre uzmanların mükemmel uzlaşma sağladığı sonucuna ulaşılmıştır. Yine araştırmacı bulguların sağlamlığı değerlendirme ölçütleri çerçevesinde alanyazına dayalı olarak elde edilen veriler doğrultusunda tüm ilkelerin güçlü araştırma desteğine sahip olduğunu belirlemiştir. Bu noktada alanyazından elde edilen bulgular ile birlikte uzman görüşlerine dayalı olarak ilkelerin geçerliliğinin sağlandığı ve farklı sistemler üzerinde de değerlendirilmesinin faydalı olabileceği belirtilmiştir. Bu bulgulardan hareketle gerçekleştirilen çalışmada hızla yaygınlaşan ve neredeyse tüm toplumun sağlığa ilişkin ihtiyaçlarını karşılamayı hedefleyen e-sağlık uygulamalarının tasarımında da uyarlanabilir gezinme ilkelerinin uyumunun değerlendirilmesi amaçlanmıştır. Bu doğrultuda elde edilen ilkelerle birlikte her bir ilkenin

açıklayıcı biçiminde düzenlenen yönergeler Tablo 4'te sunulmuştur.

Tablo 4. E-Sağlık Uygulamalarında Gezinmenin Uyarlanmasına Yönelik İlkeler

Gezinmenin Uyarlanması (GU.)		*GU	Bağl** Önem
GU.1	Kullanıcı modelindeki bilgiler doğrultusunda, bir sonraki bağlantıya ulaşmasını sağlayın.	5/5	5
Kullanıcı modelindeki verilere dayalı olarak kullanıcıların bir sonraki bağlantı ya da içeriğe ulaşmasını sağlanması gerekmektedir. Bu doğrultuda kullanıcılara uyarlanabilir bağlantı yapıları sunularak aşırı bilişsel yüklenme ve kaybolma problemleriyle karşılaşmalarını önlemek için gezinme unsurlarında gerekli düzenlemelerin yapılması gerekmektedir. Bu kapsamda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.93$, $Ss=0.26$, $Q1=5.00$).			
GU.2	İlgili bağlantılara açıklama/yönerge ekleyin.	5/5	5
Gezinme sırasında bağlantı açıklamaları kullanıcıların sistem içinde kaybolmalarını ve boğulmalarını engelleyerek gezinmelerini destekleme potansiyeli taşımaktadır. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.79$, $Ss=0.59$, $Q1=5.00$).			
GU.3	Kullanıcı modelindeki bilgiler doğrultusunda dinamik bağlantı yapıları sağlayın.	5/5	5
Kullanıcı modelindeki bilgiler doğrultusunda uygun dinamik bağlantı yapılarının sunulması ve bu doğrultuda ilgililik durumlarına göre bağlantıların gösterilmesi gerekmektedir. Böylece kullanıcılara dinamik bağlantı yapıları sunularak ortamda kaybolmalarını engellenebileceği gibi istenilen bilgi ya da bağlantıya ulaşım kolaylaştırılabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.80$, $Ss=0.41$, $Q1=5.00$).			
GU.4	Kullanıcının uygulamanın neresinde olduğunu görmesini sağlamak için kişiye özel geribildirim sağlayın.	5/5	5
Ortamda gezinme sürecinde kullanıcıların mevcut durumları bilgi almaları, buldukları konuma nasıl geldikleri ve bir sonraki ilgili bağlantıya görebilmeleri; kaybolmaları ve aşırı bilişsel yüklenmelerinin engellenmesinde olumlu katkı sağlayabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.93$, $Ss=0.26$, $Q1=5.00$).			
GU.5	Kullanıcılara yönelik uyarlanabilir içerik haritası oluşturun.	5/5	5
Kullanıcıların ortamın içeriğini oluşturan unsurların genel yapısını görmelerini sağlayan içerik haritalarının sunulması gezinmeye olumlu katkı sağlayabilir. Bu durum tüm içeriği ve ilgili bağlantıların bir noktada görülmesine olanak sağlayarak sistem içerisinde istenilen içerik ya da bağlantıya kolaylıkla ulaşabilmelerini destekleyebilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.78$, $Ss=0.36$, $Q1=5.00$).			
GU.6	Kullanıcıların daha önce girmiş oldukları bağlantı, içerik ve sayfalara ilişkin istatistiksel bilgileri görmelerine izin verin.	5/5	4
Kullanıcıların gezinme sürecindeki hareketlerine ilişkin istatistiksel verileri görmeleri, gezinmelerine olumlu katkı sağlayabilir. Böylelikle kullanıcının ortamda yaptığı işlem ve girilen bağlantıların görülebilir olması, geçmişe ait sürecini nicel olarak ve hızlı bir biçimde gözden geçirmesine olanak tanır. Bu durum ortam içerisinde kullanıcıların aşırı			

bilişsel yüklenmesi ve kaybolmaları gibi problemlerin de önüne geçilmesine katkı sağlayabilir.

Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, önemli derecede uzlaşma sağladığı belirlenmiştir ($X=4.33$, $Ss=0.72$, $Q1=4.00$).

GU.7	Kullanıcı modelindeki bilgiler göz önünde bulundurularak işlevsel arama sonuçları sağlayın.	5/5	5
Kullanıcının gereksinim duyduğu bilgiye kolaylıkla ulaşabilecekleri bir sistemin geliştirilmesi, ortam içerisinde gezinmelerine ve istenilen içeriğe ya da bağlantıya kolay ulaşabilmelerine olumlu katkı sağlayacaktır. Arama motorlarının tasarımında uyarlanabilir sistem özelliklerinin kullanılması kişiye özel en ilgili içerik ya da bağlantının sunulmasına olanak sağlayabilir. Bu noktada uyarlanabilir yapıların işlevsel sonuç çıkarma mekanizmalarının ortama entegre edilmesi istenilen kullanıcıların bilgi ya da bağlantıya tek bir yol ile ulaşarak sistem içerisinde kaybolmaları riski azaltılabilir. Yine böylelikle kullanıcının gereksiz içerik ya da bağlantılarla karşılaşmaları engellenerek sistem içinde boğulmalarının da önüne geçilebilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.87$, $Ss=0.35$, $Q1=5.00$).			
GU.8	Kullanıcının arama performansını geliştirmek için kişiye özel ipuçları sağlayın.	5/5	5
Arama motorları aracılığı ile kullanıcının gereksinim duyduğu bilgi ya da bağlantıya kolaylıkla ulaşabilecekleri bir sistemin geliştirilmesi, ortamda gezinme sürecine olumlu katkı sağlayacaktır. Ancak, arama motorları kullanılarak en uygun içerik ya da bağlantının bulunabilmesi noktasında kullanıcılara doğru arama stratejileri yapabilmeleri için ipuçları sağlanmalıdır. Uyarlanabilir sonuç çıkarma mekanizmalarından kullanıcıların etkili bir biçimde faydalanmaları sağlanmalı ve bu konuda kişiye özel destekleyici ipuçları verilmelidir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.80$, $Ss=0.41$, $Q1=5.00$).			
GU.9	Kullanıcı modeline bağlı kalınarak kullanıcının içerik tercihlerine göre, yeni eklenen içeriği kolaylıkla takip edebilmesini sağlayın.	5/5	5
Kullanıcıların ilgili içeriğe ya da bağlantıya kolaylıkla erişebilmesi için uygun gezinme araçlarının işe koşulması gerekmektedir. Özellikle e-sağlık uygulamaları ile beraberinde verilen hizmetlerin çeşitlenerek arttığı göz önünde bulundurulduğunda, kullanıcının yeni içerik ya da hizmetleri kolaylıkla görebilmeleri sağlanmalıdır. Bunu sağlayabilecek araçlardan biri olan RSS (Rich Site Summary) yapıları aracılığı ile yeni eklenen içerikler kolaylıkla takip edilebilir. Ayrıca, uyarlanabilir sistemler ile entegre çalışabilir olan bu yapılar aracılığıyla kullanıcı özelliklerine göre uyarlanabilen ve dinamik olarak güncellenebilen içerik haritaları da sunulabilir. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse mükemmel uzlaşma sağladığı belirlenmiştir ($X=4.87$, $Ss=0.35$, $Q1=5.00$).			
GU.10	Kullanıcının önemli gördüğü alanları (bağlantı, metin, konu, sayfa, vb.) işaretlemelerine izin verin.	5/5	5
Kullanıcının kendine göre önemli gördüğü konu, sayfa ya da bölümleri işaretleyebilmelerine olanak tanınması gezinmelerine olumlu katkı sağlayacaktır. Özellikle gezinme sırasında kolaylıkla işaretleme ya da not almalarına olanak sağlayan sistemlerin geliştirilmesi kullanıcının ortama bir sonraki girişinde geçmişte yapmış olduğu işlemler ile birlikte süreç içinde önemli gördükleri unsurların hatırlanmasına da olanak sağlanmış olacaktır. Bu doğrultuda gerçekleştirilen çalışma sonucunda belirlenen ilkenin e-sağlık uygulamaları için de geçerli olduğu ve uzmanların, neredeyse "mükemmel uzlaşma" sağladığı belirlenmiştir ($X=4.93$, $Ss=0.26$, $Q1=5.00$).			

*GU: Bağlı önemlilik/Bulguların Sağlamlığı: Gezinmenin uyarlanması [25]

**Bağlı Önem: Gezinmenin uyarlanması ilkelerinin e-sağlık uygulamalarına uyumu

Elde edilen sonuçlar Çetinkaya (2013) tarafından geliştirilen modelin uyarlamaların gerçekleştirilmesi başlığı altında yer alan uyarlanabilir çevrimiçi öğrenme ortamlarında gezinmenin uyarlanmasına yönelik 10 ilke üzerinde elde edilen uzlaşımın [25], e-sağlık uygulamalarına yönelik geliştirilen çevrimiçi ortamlar için de sağlandığını göstermektedir. Bu noktada belirlenen 10 ilkenin 9'u üzerinde bağlı önemlilik ölçütlerine göre tüm uzmanların uzlaşımına vardıkları ve "neredeyse mükemmel uzlaşım (5)" sağlandığı belirlenmiştir. Belirlenen ilkelere 1'in de ise uzmanların, "önemli derecede uzlaşım (4)" sağladığı belirlenmiştir.

Tablo 5. Delphi Uygulaması Anket Turları Uzlaşım Dereceleri

	Delphi Uygulaması Anket Turları		
	1. Tur	2. Tur	3. Tur
Uzmanlar arasındaki uzlaşım (ICC)	0.799	0.892	0.930
Ortalama değerleri	4.42	4.62	4.79

Gezinmenin uyarlanmasına yönelik ilkelerin e-sağlık uygulamalarına ilişkin uyumu noktasında değerlendiriciler arasındaki uyuma derecesini belirlemek amacı ile hesaplanan sınıf-içi korelasyon katsayısı (ICC= .930), ilkelere ilişkin uzmanlar arasında yüksek düzeyde güvenilirlik sağlandığını göstermektedir. Veriler üzerinde uzlaşımın sağlandığını gösteren ve verilerin uyumluluğunu analiz etmek için kullanılan sınıf-içi korelasyon katsayısı değerleri tablo 6'da sunulmuştur.

Tablo 6. İlkeler Üzerinde Delphi Panelistleri Arasındaki Uyuşma

	Sınıf-içi Korelasyon	Güven Aralığı (%95)		F	p
		Alt Sınır	Üst Sınır		
Örneklem Ortalaması	0.930	0.861	0.972	14.237	.000*

Bağıl önemlilik değerlendirme sonuçları, e-sağlık uygulamalarında gezinmenin uyarlanmasına ilişkin belirlenen ilkelere yönelik alan uzmanlarının ortak eğilimlerini göstermektedir.

4. TARTIŞMA, SONUÇ VE ÖNERİLER (CONCLUSION, RESULT AND RECOMMENDATIONS)

Yapılan analizler sonucunda elde edilen sonuçlar alanyazınla birlikte tartışılmış ve bu doğrultuda geliştirilen öneriler başlıklar altında sunulmuştur.

4.1. Tartışma ve Sonuçlar (Conclusion and Results)

Dijital sağlık hizmetlerinin, sağlık profesyonelleri ve vatandaşlar açısından faydalarından bahsetmek elbette mümkün, ancak bu hizmetlerin aynı zamanda kullanılabilirlik açısından da ele alınması oldukça önemlidir. Nitekim hızla artan veri miktarı ile birlikte sistemi kullanan paydaşların istenilen bilgiye kolay erişiminin sağlanması kullanılabilirliği yüksek e-sağlık uygulamalarının geliştirilmesiyle mümkün olabilir. Bu

noktada farklı sistemleri bir araya getirebilecek ve uyum içerisinde çalışabilmelerine olanak sağlayan işlevsel yapıların oluşturulması gerekmektedir. Uyarlanabilir sistemler bu bağlamda önemli bir alternatif olarak karşımıza çıkmaktadır. Uyarlanabilir sistemlerin geliştirilmesinin disiplinler arası araştırma yaklaşımı gerektiriyor olması; çoğu zaman tasarım, metod ve değerlendirmeye uyum sağlayamayan ortamların geliştirilmesine sebep olabilmektedir. Geliştirilen farklı sistemlerin birlikte işlerlik kazanabilmesi noktasında ortak bir dilin oluşturulması ve ortak bir paydaşa uzlaşımın sürdürülebilir verimli yapıların oluşturulması açısından oldukça önemlidir. Bu durum göz önünde bulundurularak gerçekleştirilen çalışmada gezinmenin uyarlanmasına yönelik önceden geliştirilen ilkelerin e-sağlık uygulamalarına uyumu ve düzenlenmesi hedeflenmiştir. Bu doğrultuda nitel ve nicel araştırma becerilerinin işe koşulduğu yapılandırılmış çevrimiçi Delphi tekniğinden faydalanılan çalışmaya 15 uzman dahil edilmiştir. Metodüçi karışım yaklaşımının benimsendiği veri toplama sürecinde tek bir soru kağıdında nitel ve nicel veriler eşzamanlı olarak toplanarak veri çeşitlemesi yapılmıştır. Uzmanlardan elde edilen veriler ve bu verilerin değerlendirilmesi ile elde edilen sonuçların uzmanlara sunumuyla tekrarlanan bir sürecin izlendiği çalışma, uzmanların uzlaşımına ulaşmalarıyla son bulmuştur.

Çalışmanın ilk turu sonucunda uzmanların gezinmenin uyarlanmasına yönelik ilkeleri yapısal olarak uygun olduğu belirlenmiştir. Ancak, çevrimiçi öğrenme ortamlarına yönelik ilkelerin e-sağlık uygulamaları çerçevesinde yeniden yorumlanması gerekliliği uzmanlar tarafından bildirilmiştir. Uzman görüşleri doğrultusunda her turda yeniden şekillenen ilkeler üzerinde artık yeni fikirlerin üretilmediği, maddelerin tüm güçlü-zayıf yanlarının belirlendiği ve uzmanlardan ilkelerin düzenlenmesine yönelik herhangi bir görüş ya da önerinin gelmediği görülen üçüncü Delphi turunda çalışma sonuçlandırılmıştır. Uzman görüşleri çerçevesinde uyarlanabilir yapılarda gezinmenin uyarlanmasına yönelik geliştirilen 10 ilke üzerinde tüm uzmanların uzlaşımına e-sağlık uygulamalarına yönelik uyumu sağlanmıştır (bkz., Tablo 4). Diğer taraftan belirlenen ilkelere ilişkin değerlendiriciler arası uyuma derecesini ölçmek amacı ile kullanılan sınıf-içi korelasyon katsayısı (ICC= .930), mükemmel düzeyde güvenilirlik sağlandığını göstermektedir. Bağıl önemlilik ölçütlerine göre uzlaşım sağlanan ilkelerin tamamı gezinmenin uyarlanmasında ilişkili olmakla birlikte sistemin diğer bileşenlerinin işleyişiyle de doğrudan ilişkilidir. Bu noktada gezinmenin uyarlanmasına yönelik tasarımın sistemin sağlıklı bir şekilde işleyebilmesi için kritik öneme sahip olduğunu söylemek mümkün gözükmektedir. Özellikle bilgi yoğunluğu ve etkileşimin yüksek düzeyde olduğu çevrimiçi ortamların tasarımında kullanıcıların kendi istekleri ve ihtiyaçlarına uygun şekilde sistem içinde gezinmelerine olanak sağlayabilecek şekilde düzenlenmesi gerekmektedir. Aksi takdirde aşırı bilişsel yüklenme ve kaybolma gibi birbirleriyle ilişkili sorunların ortaya çıkmasına neden olunabilir. Özellikle çok fazla bileşen, yol ve araca sahip ortamlardaki seçenekler nedeniyle,

kullanıcıların sistem üzerinde kaybolarak kafalarının karışması sonucu istenilen bilgilere ulaşmada daha fazla zaman ve çaba harcamaları söz konusu olacaktır [18, 20]. Bu durum özellikle sayıları gün geçtikçe hızla artan e-sağlık uygulamaları için de söz konusu olmakla birlikte gezinmeye yönelik unsurlara dikkat edilmemesi durumunda risklerin çeşitlenerek artabileceği de söylenebilir.

Dijital dönüşüm bireylerin hizmetlere erişim ve eşitliğin sağlanmasında muazzam bir potansiyele sahip olsa da tüm bireylerin bu hizmetlerden faydalanabilmelerine olanak sağlayan ortamların geliştirilmesi gerekmektedir. Özellikle de sağlık sisteminin hızla dijitalleşmesi ile birlikte sağlığa erişimde e-sağlık uygulamalarının kullanımı kritik öneme sahip olmaya başlamıştır. Bu noktada e-sağlık uygulamalarının yeni kullanıcıların istenilen bilgiye hızlı ve doğru bir şekilde ulaşmalarının desteklenmesi gerekmektedir. Bunun için ise uyarlanabilir gezinme yapılarıyla birlikte kullanıcıların doğru bir şekilde yönlendirilmeleri sağlanarak sistem içinde kaybolmaları ve boğulmaları engellenmelidir. Özellikle bilginin yapılandırılması ve anlamın şekillenmesinde doğru yolun izlenerek en kısa sürede istenilen içeriğe ulaşılmasının oldukça önemli olduğu göz önünde bulundurulduğunda gezinmeye ilişkin unsurların dinamik olarak uyarlanabilir bir yapıyla tasarlanması kritik rol oynamaktadır. Nitekim bireylerin bilgiye ulaşma stratejileri farklılık gösterebilmekte ve bu durum bilginin edinimi sürecinde bireysel farklılıkların gözetildiği uyarlanabilir ortamların geliştirilmesini zorunlu hale getirmektedir. Bunun yapılabilmesi için ise kullanıcının sistemle ve sistem ile ilişkili tüm bileşenlerle etkileşim sürecinin iyi tasarlanması gerekmektedir. Bir başka ifadeyle kullanıcının sistem üzerindeki tüm hareketleri gözlenerek bir sonraki adımlarında sistem üzerinde kaybolmalarını engelleyecek nitelikte uyarlanabilir gezinme yapıları sunulmalıdır. Aksi takdirde sağlık gibi kritik bir temel hakka erişimde eşitsizliklerin artmasına yönelik risklerin yanı sıra dijital sağlık hizmetlerinin sürdürülebilirliğinin sağlanması noktasında da sorunlarla karşılaşılması söz konusu olabilir.

4.2. Öneriler (Recommendations)

Araştırmanın amacı doğrultusunda elde edilen sonuçlara dayalı olarak geliştirilen öneriler aşağıda sunulmuştur.

1. Çalışma sonucunda gerek alanyazın gerekse alan uzmanlarının ortak eğilimleri ile elde edilen ilkelerin, e-sağlık uygulamalarında gezinme unsurlarının tasarımında işe koşulmasının faydalı olacağı söylenebilir.
2. Uyarlanabilir gezinmeye ilişkin ilkeler çevrimiçi ortamlara yönelik olarak geliştirilmiş olup gerçekleştirilen bu çalışmada ise e-sağlık uygulamalarına uyumu değerlendirilmiştir. Bu noktada çevrimiçi olmayan ya da farklı cihaz türlerinde (tablet, akıllı telefon, vb.) belirlenen ilkelerin geçerliliğine bakılabilir.

3. Hızla yaygınlaşan mobil e-sağlık uygulamaları için de benzer bir çalışmanın planlaması çok daha doğru bir yaklaşım olacaktır.
4. Çalışmada belirlenen ilkeler alanyazın doğrultusunda uzman görüşlerine dayalı olarak belirlenmiştir. Bu belirlenen ilkelerin deneysel etkililiğinin test edilmesine yönelik çalışmaların da yapılması faydalı olacaktır.
5. E-sağlık hizmetlerine yönelik tasarımların, çalışmada belirlenen ilkelere uygunluğunu değerlendirmeye yönelik çalışmalar yapılmalıdır.

KAYNAKLAR (REFERENCES)

- [1] P. Kostkova, "Grand challenges in digital health", *Frontiers in Public Health*, 3(134), 1-5, 2015.
- [2] U. H. Bittlingmayer, K. Dadaczynski, D. Sahrai, S. v. Broucke, O. Okan, "Digitale gesundheitskompetenz – konzeptionelle verortung, erfassung und förderung mit fokus auf kinder und jugendliche", *Bundesgesundheitsblatt- Gesundheitsforschung- Gesundheitsschutz*, 63, 176-184, 2020.
- [3] A. Scheerder, A. van Deursen, J. van Dijk, "Determinants of Internet skills, use and outcomes: A systematic review of the second-and third-level digital divide", *Telematics and informatics*, 34(8), 1607-1624, 2017.
- [4] K. Sørensen, J. M. Pelikan, F. Röthlin, K. Ganahl, Z. Slonska, G. Doyle, J. Fullam, B. Kondilis, D. Agrafiotis, E. Uiters, M. Falcon, M. Mensing, K. Tchamov, S. van den Broucke, H. Brand, HLS-EU Consortium, "Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU)", *European journal of public health*, 25(6), 1053-1058, 2015.
- [5] A. Odone, S. Buttigieg, W. Ricciardi, N. Azzopardi-Muscat, A. Staines, "Public health digitalization in Europe", *European journal of public health*, 29(Supplement_3), 28-35, 2019.
- [6] I. Kickbusch, D. Piselli, A. Agrawal, R. Balicer, O. Banner, M. Adelhardt, E. Capobianco, C. Fabian, A. Singh Gill, D. Lupton, R. P. Medhora, N. Ndili, A. Ryś, N. Sambuli, D. Settle, S. Swaminathan, J. V. Morales, M. Wolpert, A. W. Wyckoff, L. Xue, ... Secretariat of the Lancet and Financial Times Commission, "The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world", *Lancet (London, England)*, 398(10312), 1727-1776, 2021.
- [7] R. van Kessel, R. Hrzic, E. O'Nuallain, E. Weir, B.L.H. Wong, M. Anderson, S. Baron-Cohen, E. Mossialos, "Digital Health Paradox: International Policy Perspectives to Address Increased Health Inequalities for People Living With Disabilities", *J Med Internet Res*. 24(2):e33819, 2022.
- [8] P. Brusilovsky, "Adaptive and intelligent technologies for web-based education", *Künstliche Intelligenz*, 13, 19-25, 1999.
- [9] P. Brusilovsky, "Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation", *British Journal of Educational Technology*, 34(4), 487-497, 2003.
- [10] P. Brusilovsky, C. Peylo, "Adaptive and intelligent web-based educational systems", *International Journal of Artificial Intelligence in Education*, 13(2), 156-169, 2003.

- [11] C. M. Reigeluth, "A new paradigm of ISD?" *Educational Technology & Society*, 36(3), 13-20, 1996.
- [12] S. Somyürek, **Uyarlanabilir Eğitsel Web Ortamlarının Öğrencilerin Akademik Başarısına ve Gezinmesine Etkisi** (Yayınlanmamış Doktora tezi), Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara, 2008.
- [13] C.-M. Chen, "Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance", *Expert Systems with Applications*, 36, 8816–8829, 2009.
- [14] P. Brusilovsky, & L. Pesin, "An intelligent learning environment for CDS/ISIS users", In J.J Levonen and M.T Tukianinen. (Eds.), **Proceedings of the interdisciplinary workshop on complex learning in computer environments (CLCE94)** (29-33). Joensuu, Finland, May 16-19, 1994.
- [15] P. De Bra, "Adaptive Hypermedia on the Web: Methods, techniques and applications, **Proceedings of the AACE WebNet'98** (220-225), Orlando, Florida, USA, November 7-12, 1998.
- [16] E. K. Çakmak, "Çoklu Ortamlarda Dar Boğaz: Aşırı Bilişsel Yükleme", *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 27(2), 1-24, 2007.
- [17] A. E. Kılıç, Ö. G. Ş. Karadeniz, "Hiper Ortamlarda Öğrencilerin Bilişsel Yükleme ve Kaybolma Düzeylerinin Belirlenmesi" *Kuram ve Uygulamada Eğitim Yönetimi*, 40(40), 562-579, 2004.
- [18] T. Murray, "Characteristics and affordances of adaptive hyperbooks", **Proceedings of WebNet 2001**, Orlando, FL, USA, October 23-27, 2001.
- [19] J. Q. Young, J. V. Merrienboer, S. Durning, O. Ten Cate, "Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86", *Medical Teacher*, 36(5), 371-384, 2014.
- [20] S. McDonald, & R. J. Stevenson, "Disorientation in hypertext: the effects of three text structures on navigating performance", *Applied Ergonomics*, 27(1), 61-68, 1996.
- [21] E. Cuddihy, K. A. Mobernd, J. H. Spyridakis, "Web page previews: effect on comprehension, user perceptions, and site exploration", *Journal of Information Science*, 38(2), 103–117, 2012.
- [22] K. A. Lawless, P. G. Schrader, H. J. Mayall, "Acquisition of information online: Knowledge, navigation and learning outcomes", *Journal of Literacy Research*, 39(3), 289–306, 2007.
- [23] L. Salmerón, W. Kintsch, J. J. Cañas, "Reading strategies and prior knowledge in learning from hypertext", *Memory & Cognition*, 34(5), 1157–1171, 2006.
- [24] A. Hansen, M. Herrmann, J. P. Ehlers, T. Mondritzki, K. O. Hensel, H. Truebel, P. Boehme, "Perception of the Progressing Digitization and Transformation of the German Health Care System Among Experts and the Public: Mixed Methods Study", *JMIR public health and surveillance*, 5(4), e14689, 2019.
- [25] L. Çetinkaya, **Uyarlanabilir eğitsel içerikli web ortamlarının tasarım ilkeleri** (Yayınlanmamış doktora tezi), Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara, 2013.
- [26] A. L. Delbecq, A. H. Van de Ven, D. H. Gustafson, **Group techniques for program planning: A guide to nominal group and Delphi processes**, Scott, Foresman & Company, 1975.
- [27] M. Fry, G. Burr, "Using the Delphi technique to design a self reporting triage survey tool", *Accident and Emergency Nursing*, 9(4), 235- 241, 2001.
- [28] W. Alexander, R. Serfass, **Futuring tools for strategic quality planning in education**, Milwaukee, WI: American Society for Quality Press, 1999.
- [29] T. Hatcher, S. Colton, "Using the internet to improve HRD research: The case of the web-based Delphi research technique to achieve content validity of an HRD-oriented measurement", *Journal of European Industrial Training*, 31(7), 570-587, 2007.
- [30] J. A. Maxwell, **Qualitative research design: An interpretative approach**, Thousand Oaks, CA: Sage, 1996.
- [31] A. Tashakkori, C. B. Teddlie, **Handbook of mixed methods in social and behavioral research**, Thousand Oaks, CA: Sage, 2003.
- [32] H. A. Shah, & T. S. Tillman, "An international delphi study to build a foundation for an undergraduate level lean manufacturing curriculum", *International Transaction Journal of Engineering, Management & Applied Sciences & Technologies*, 9(3), 521-530, 2011.
- [33] K. L. Gwet, **Handbook of inter-rater reliability**, Gaithersburg: StatAxis Publishing, 2001.
- [34] J. W. Youdas, J. R. Carey, & T. R. Garrett, "Reliability of measurements of cervical spine range of motion—comparison of three methods", *Physical Therapy*, 71(2). 98-104, 1991.